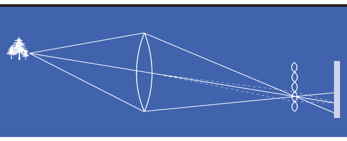
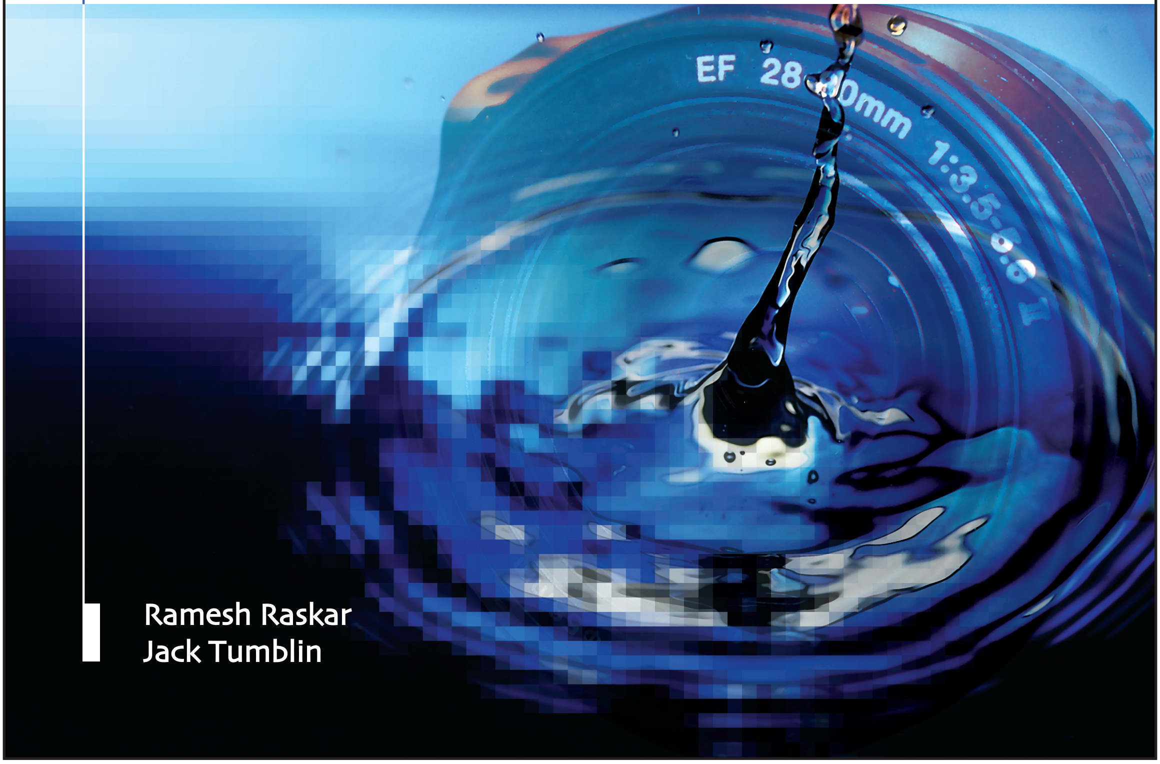


Computational Photography

Mastering New Techniques for Lenses, Lighting, and Sensors



Ramesh Raskar
Jack Tumblin

Preliminary Editorial Draft March 13, 2014

Computational Photography

Mastering New Techniques for Lenses, Lighting, and Sensors

Computational Photography

Mastering New Techniques for Lenses, Lighting, and Sensors

Ramesh Raskar
Jack Tumblin

preliminary editorial draft
March 13, 2014

This book is currently in active editorial development. Please send any comments, suggestions, and corrections to Randall Warniers at warniers@alum.mit.edu.

Contents

Chapter 1

Introduction

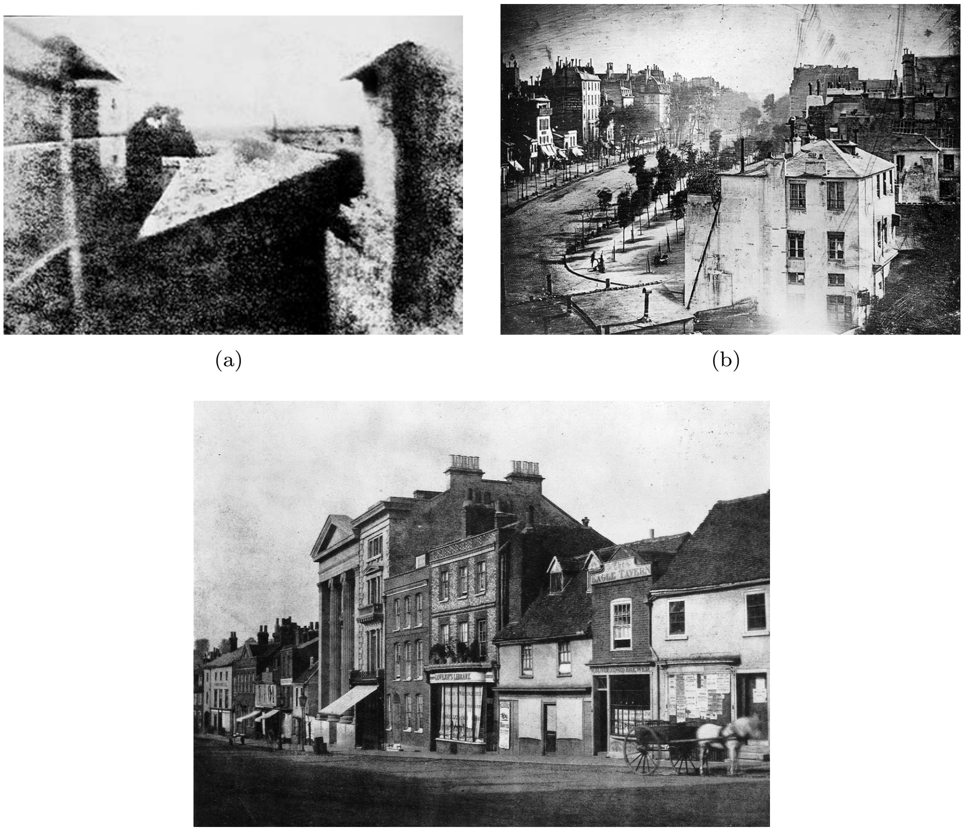
Photography—literally, 'drawing with light'—is the process of making pictures with a camera by recording the visually meaningful changes in the light reflected by a scene. This goal was envisioned and realized for plate and film photography over 150 years ago by pioneers Nicéphore Niépce,[1] Louis-Jacques-Mandé Daguerre, and William Henry Fox Talbot (Figure 1.1). Niépce is widely credited for making the first photograph in 1826. Daguerre, with limited materials, captured exquisitely detailed landscapes and portraits with great subtleties of expression. Fox Talbot made the important discovery that a latent photographic image on paper could be chemically developed into a negative, which was then used to produce multiple photographic images.[2]

The art and technology of traditional film-and-chemistry-based photography developed rapidly in the nineteenth century. Over time, cameras and lenses became better and lighter, and film became faster and easier to use. As a result photography became a powerful medium of visual communication throughout the twentieth century. The world of photography changed profoundly, however, when digital photography arrived at the end of the century. In less than a decade, the well-established techniques of traditional film photography and film processing became obsolete. Photography today is now an all-digital workflow from camera to computer to print.

Modern digital photography is truly revolutionary, but limited in many ways. For most purposes it is an electronically implemented duplication of the style and function of film photography, with an electronic sensor replacing the film. The imaging goals of a film camera—enabled and limited by chemistry, optics, and mechanical shutters—are highly similar to the imaging goals of a digital camera. Both cameras copy an image formed by a lens, without imposing judgment, understanding, or interpretive manipulations. In effect, film cameras and digital cameras are both faithful but mindless copiers of the scene in front of the lens.

[1] A description of the important role of Nicéphore Niépce in the history of photography and a summary of his many experiments that led to the first photograph in 1826 are found at the website www.hrc.utexas.edu/exhibitions/permanent/wfp/.

[2] William Henry Fox Talbot's 1844 book *The Pencil of Nature* is considered to be the first published book on photography.



(a)

(b)

(c)

Figure 1.1: (a) View from the Window at Le Gras, Saint-Loup de Varennes, France (1826) by Nicéphore Niépce. This camera obscura photograph, which took over eight hours to expose, is widely considered the first permanent photograph ever made. (b) Boulevard du Temple, Paris (1838) by Louis-Jacques-Mandé Daguerre. The sensitivity of photographic materials increased greatly in the decade following Niépce's photographs, allowing shorter exposure times and much more realistic detail in the imagery. (c) London Street, Reading (1845) by William Henry Fox Talbot. Fox Talbot introduced the photographic negative, which was used to make reproducible prints.

For the sake of simplicity and clarity, let's use the term "film-like" to describe photography accomplished with traditional film cameras and with current digital cameras. Both work by recording an image, as seen by a lens, on film or on a digital sensor. Both presume (and often require) artful human judgment, intervention, and interpretation at every stage to choose viewpoint, framing, timing, lenses, film properties, lighting, developing, printing, display, search, index, and labeling.

Film-like digital photography is convenient and powerful, but it currently ignores a great new promise inherent in the digital representation of visual information. Even though current digital methods eliminate or simplify much of the variability of traditional methods (e.g., processing chemistry), the shift to digital has far more profound consequences than merely streamlining the imaging process. Digital methods do not just replace negative emulsions with solid-state sensors, they replace one-of-a-kind fragile negatives with numerical grids and mathematical abstractions, and they replace chemical baths that act uniformly on the entire emulsion with user-defined instructions applied to each value in the numerical grid. The emerging methods of computational photography greatly expand the potential of digital photography. They allow us to create, measure, construct, and manipulate visual information, and then make that information amenable to any computable algorithm, interaction, or image-related task.

The ideal possibilities of photography—showing what we want to capture rather than what we can capture—were perhaps clearest to its pioneers, who were both artists and scientists. They saw that photography could be more than just a picture of a scene. For example, Edgar Degas wanted to capture the emotional atmosphere induced by closeness and lighting. "Daylight is too easy," Degas said in 1895. "What I want is difficult—the atmosphere of lamps and moonlight." Others expanded our visual capabilities as well. Edweard Muybridge studied fleeting motion phenomena, James Clerk Maxwell captured the full visual spectrum of light, and William Conrad Röentgen used X-rays to see bones in living flesh (Figure 1.2).

The possibilities of computational photography are not readily apparent at first. Those of us who grew up with film photography are like zoo animals released after years of confinement; we are wandering in a new world, astonished at the changes we see around us. By habit, we comply with old constraints. We tend to stay within the familiar limitations and tradeoffs we learned from our work in film, even though these constraints are now gone or easily surmountable. The vast new landscape of photography is difficult to comprehend at first, and we slowly familiarize ourselves with all that is different.

The transformation to digital photography comes with a cultural price. Digital methods nullified much photographic expertise that had accumulated since the 19th century. We lost exquisitely perfected varieties of thin-film emulsions with subtle mutable responses to light, including magnificent Kodachrome slide film, which finally disappeared from stores in 2009. We lost darkrooms and enlargers and archival printing skills. Vast collections of classic cameras with perfectly matched lenses, meters, and shutters became useless in the digital age. And the subtle processing skills and well-honed expert judgments of dedicated photographers and darkroom technicians are no longer important. So much magnificent engineering and so much film expertise is now marginalized or even gone forever. Despite our regrets for these losses, we lose even more if we are satisfied with digital photography

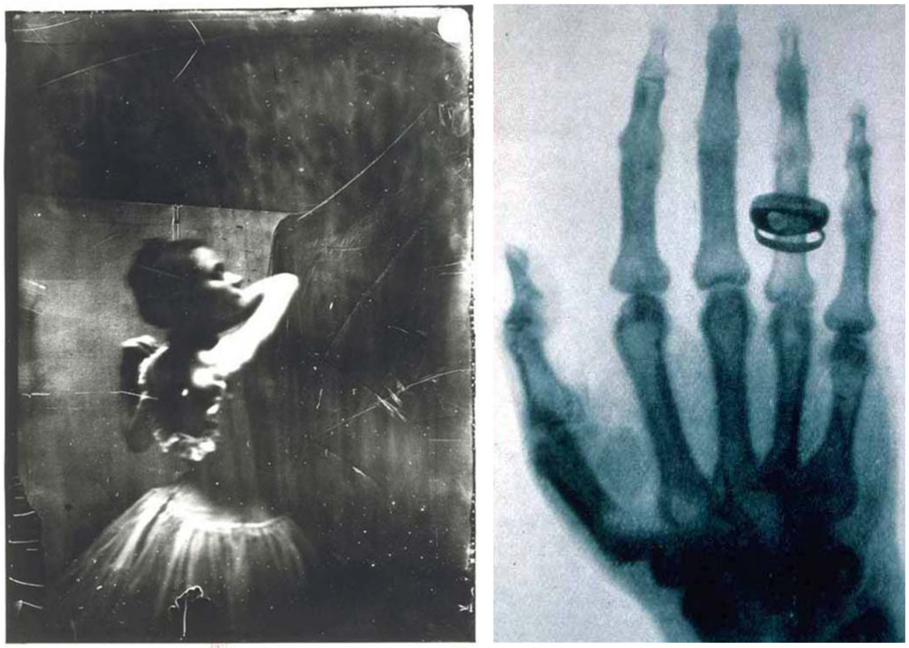


Figure 1.2: (a) Danseuse (1895), a photograph by French painter Edgar Degas. (b) X-ray photograph of the hand of Albert von Kölliker (1896), by Willhelm Conrad Röentgen.

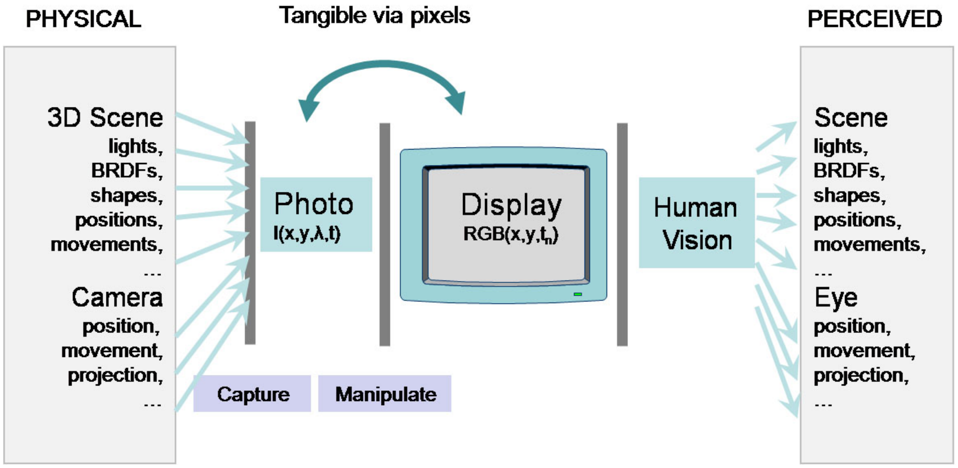


Figure 1.3: Why do we like photographs? Traditional two-dimensional photographs show us basic—but often partial—information about boundaries, shape, occlusion, lighting, shadows and texture in a scene. Even though these images are often limited by compromises on dynamic range, resolution, noise, and lighting, traditional photography can still provide—whether through film or pixels—a tangible connection between the physical world and perceived sensation. Research in computational photography has targeted novel capture and manipulation methods that can enhance this connection. [THIS FIGURE NEEDS TO BE "EXPLAINED" MORE THOROUGHLY, AND THE CAPTION EXPANDED. ALTERNATIVELY, THIS FIGURE AND AN EXPANDED CAPTION COULD BE FORMATTED AS AN INDEPENDENT SIDEBAR.]

as little more than the reconstruction of film-like photography.

The changeover to truly new and innovative digital cameras has come slowly. Many commercial digital cameras today still include tiny viewfinders and the comforting sounds of a mechanical shutter release, which helps photographers imitate the old ways of shooting. These film-like features keep our attention and our skills focused on the creative traditions of the past, far away from new and exciting opportunities inherent in digital imaging. [NOTE: THE INTRODUCTION COULD USE ONE OR MORE INTRODUCTORY EXAMPLES OF COMPUTATIONAL PHOTOGRAPHY WORK HERE.]

Where is the new horizon? Which directions promise the most breathtaking new ways to capture a better record of the visually important content of a scene? What can computational photography do to help us make strong images, retain fond memories, keep a personal record of our lives, and extend both the archival and the artistic possibilities of photography? Let's look away from film-like photography and march outward to explore the digital world. Much is there to discover (Figure 1.3). Let's survey new possibilities, and outline a more comprehensive technology that exploits plentiful low-cost computing with new kinds of digitally enabled sensors, optics, probes, smart lighting, and communication.

1.1 What is Computational Photography?

[Dan Raviv comment. ”I would write Section 1.1 entirely different. A. concepts: where information is not just pixels. B. practical: photography and computation, etc.”]

Computational photography captures a machine-readable representation of our world, allowing us to hyper-realistically synthesize the essence of our visual experience. It is an emerging field of research, with many paths, and we do not fully know where the paths will lead. By capturing a machine-readable representation of a scene, we can use modern computational photography techniques to go beyond the limitations of traditional cameras and two-dimensional imagery. These techniques employ unusual optics, modern sensors, programmable illumination and sophisticated processing to record the scene.

With this idea of what computational photography can do, we can broadly classify major new themes in digital capture and image synthesis. Three important themes in digital capture are summarized here.

(i) Computational photography exploits computing, memory, interaction and communications to overcome the inherent limitations of photographic film and camera mechanics that have persisted in film-like digital photography. These limitations include constraints on dynamic range, image resolution, depth of field, and field of view.

(ii) Computational photography attempts to record a rich multilayered visual experience by capturing more information than just a simple set of pixels. It delivers the recorded representation of the scene as an abstract data type in far more machine-accessible form, containing much more than a simple visual grid of pixel values. The new metadata may include scene depth or its precursors such as corresponding-feature candidates shared among multiple viewpoints, fused photo-video representations, or multispectral imagery. The representation may also include compensating information, generated by the camera itself, which might aid in post-processing, such as modulation patterns, global positioning system (GPS) position and orientation records, and specialized lighting information. This information can be gathered wirelessly from the scene, or from local photo-assistance devices, the user’s own external metering devices, or even other cooperating nearby cameras. [A FIGURE ILLUSTRATING AN EXAMPLE OF DIGITAL CAPTURE, SUCH AS AN HDR IMAGE, COULD BE INCLUDED HERE]

(iii) Computational photography enables the recording of new classes of visual signal and provides decomposition of the visual signal into perceptually critical components. New visual signals include the ‘moment,’ [86] shape boundaries for non-photorealistic depiction [330], estimates of three-dimensional structure [e.g., [430]][3] and portions of multiview light fields. Decomposable signals include foreground versus background mattes [84][4], direct versus indirect illumination [292], specular versus diffuse components[5] and reflection versus transmission layers [39]. With modest additional information, the list of additional photographic capabilities can grow dramatically. For example, why can’t our cameras remove unwanted reflec-

[3] graphics.stanford.edu/workshops/ibr98/Talks/Lance/silhouettes.html

[4] grail.cs.washington.edu/projects/digital-matting/image-matting/

[5] Debevec team [SPECIFIC DEBEVEC REFERENCES NEEDED HERE]

tions when we photograph through a window or a display case? [ADD OTHER EXAMPLES OF ADDITIONAL CAPABILITIES HERE]

In addition to improving the description of a scene beyond a two-dimensional set of pixels, the methods of computational photography might selectively emphasize visually important features. Such hyper-realistic synthesis departs from strict recording of the scene's light intensities, which imposes simplistic forms of interpretation on the contents of the recorded scene. By conflating measurement and rendering, the final display of the scene is visually accurate, but not necessarily photometrically accurate. The idea is to utilize additional, often subtle, pictorial elements to depict what cannot be seen by the human eye, such as a single viewpoint or a single lighting situation, and to create a plausible illusion of a reality that is most informative but might be physically unrealizeable. [THE IDEAS IN THIS PARAGRAPH WORK WELL HERE, BUT THEY ALSO COULD BE INCLUDED IN ITEM 2 ABOVE.]

Computational photography also gives us many important opportunities for meaningful image synthesis, as summarized in the following three broad categories.

(i) Computational photography enables unprecedented post-capture control for synthesis including relightable photos [254][6], and interactive displays that permit users to change lighting [295][7], viewpoint[8], and focus [300][9].

(ii) Computational photography enables synthesis of "impossible" photos that could not have been captured with a single exposure in a traditional camera. It also enables the synthesis of "plausible" but non-existent scenes. Such impossible photos include wrap-around views that use multiple-center-of-projection[10], fusion of time-lapsed events [330], the motion microscope (motion magnification [250]), and video textures and panoramas [36]. It supports seemly impossible camera movements such as the bullet-time sequences in the 1999 movie *The Matrix*, which were made with multiple cameras using staggered exposure times, and free-viewpoint television (FTV) recordings [e.g., [253]][11]. Plausible photographs also include images created from a mosaic of non-local scenes to replace or insert new scene parts[12] and for believable but generally unfilmable special effects.

(iii) Computational photography provides easy access to previously exotic forms of imaging and data-gathering techniques in astronomy[13], microscopy [239][14], tomography [403][15], and other scientific fields. These methods create images by inventive and previously impractical computation to reveal what cannot be seen by human observations. [A FIGURE OF COMPUTED IMAGERY COULD BE INCLUDED HERE, SUCH AS AN IMAGE FROM THE HUBBLE TELESCOPE.]

Computational photography is closely related to computer vision, machine learning, applied optics and visual aesthetics. Figure 1.4 illustrates how computational

[6] www.hpl.hp.com/research/ptm/

[7] portal.acm.org/citation.cfm?id=1027414

[8] www.cs.washington.edu/homes/seitz/vmorph/vmorph.htm;phototourism

[9] graphics.stanford.edu/papers/lfcamera/

[10] Unwrap Mosaics, Sig08; Rademacher and Bishop 1998

[11] graphics.tu-bs.de/people/magnor/publications/sig03.pdf

[12] scene completion, Hayes and Efros REFERENCE?

[13] www.paulcarlisle.net/old/codedaperture.html

[14] see Wikipedia entry on confocal microscopy; also graphics.stanford.edu/papers/confocal

[15] www.cs.ubc.ca/~heidrich/Projects/Tomography/

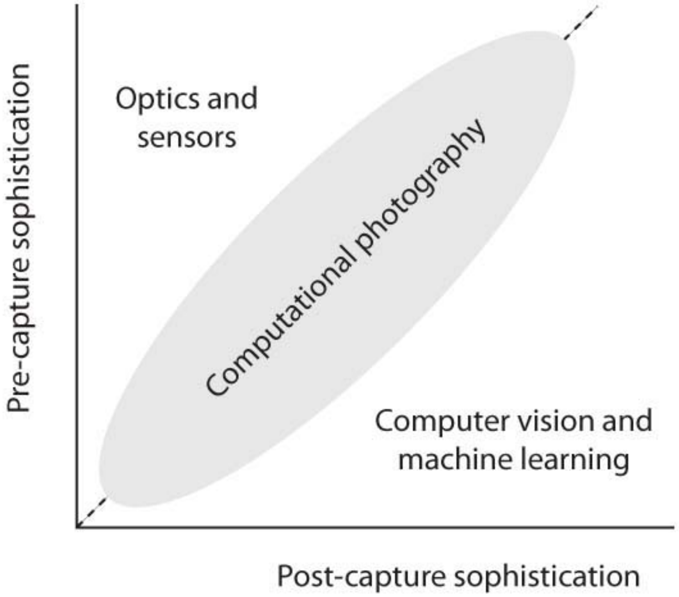

Optics and sensors
Computational photography
Computer vision and machine learning
Pre-capture sophistication
Post-capture sophistication

Figure 1.4: Computational photography extends the processing that can be done by optics and sensors before an image is captured, and it prepares imagery for further processing by techniques in computer vision and machine learning.

photography uses the pre-capture techniques of optics and sensors, and the post-capture techniques of computer vision and machine learning, but is distinct from these fields of research. For the purpose of this book we will mainly examine techniques of digital capture and image synthesis. Our goals are the synthesis of a visual experience rather than the judgment of a scene. Thus goals of scene measurement and analysis, such as counting people, determining range to a point, or planning collision avoidance, among others, will not be covered here.

1.2 Elements of Computational Photography

Traditional film-like digital photography involves (a) a lens, (b) a two-dimensional planar sensor and (c) a processor that converts sensed values into an image. In addition, such photography may include (d) external illumination from point sources (e.g., flash units) and area sources (e.g., studio lights).

In accordance with Shree Nayar's insightful overview of computational photography in 2005 [288], we can consolidate recent advances into four broad areas: generalized optics; generalized sensors; computational illumination methods; and generalized processing, reconstruction, and display. As Figure 1.5 illustrates, these categories encapsulate the entire scene-recording and scene-reproduction process.

Unlike camera-centered models that describe photography as a record of the two-dimensional image formed behind a lens, Nayar's model is dominated by rays. We regard rays (or more precisely, changes in bundles of rays) as the most sensible

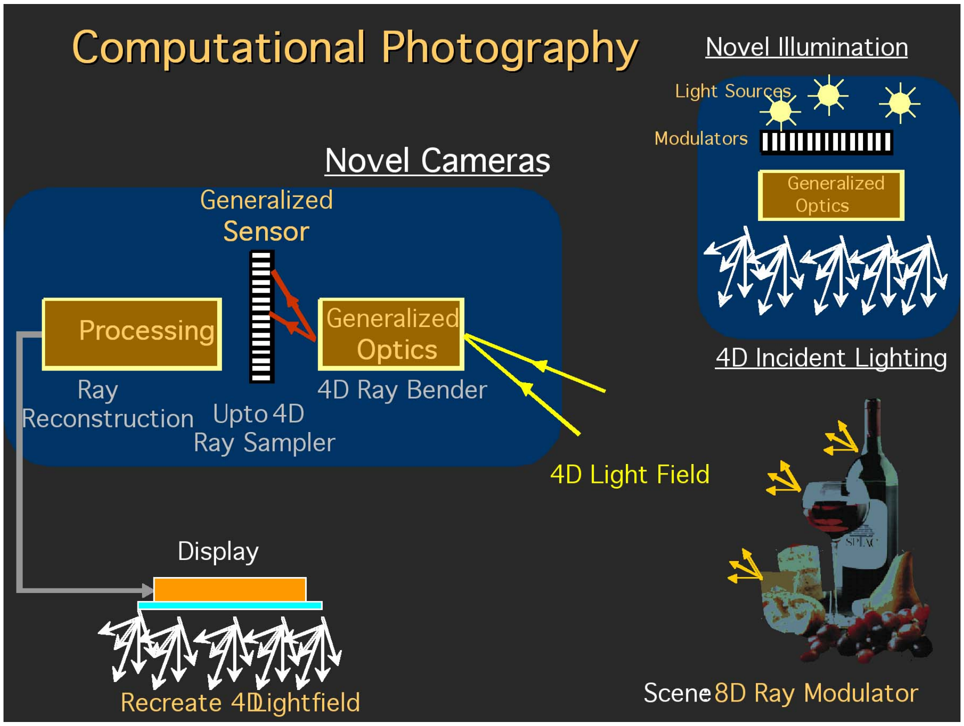


Figure 1.5: Elements of computational photography: generalized optics; generalized sensors; computational illumination; and generalized processing, reconstruction and display. [WE CAN DO A LOT TO CLARIFY AND VISUALLY IMPROVE THIS FIGURE.]

way to describe all the signal-producing mechanisms of photography. By using rays we can trace all the possible paths from any point on a light source to any point in the scene, all the paths from any point in the scene to any point on the camera's optics, all the paths through the optics to the sensor, all the possible ways an advanced sensor might measure the energy transferred by those rays, and finally all the possible ways to generate rays from the display to the eyes of the audience. This last step includes all the possible processing, storage, and transmission stages that can be used to deliver imagery to a viewer.

In traditional film-like digital photography, each camera image records one view of the scene as a two-dimensional array of pixels. Computational photography attempts to understand and analyze a higher-dimensional representation of the scene by using rays as the fundamental primitives. Thus the light transport via rays can be conceptualized at any plane in the optical path. The rays incident on this plane can be parameterized by using a four-dimensional position-angle representation: two dimensions for position on the plane and two dimensions for the incident angle. This 4D ray space is alternately called the *plenoptic function*, the *4D light field* or the *Lumigraph.* In this book we will use the term *light field.* In addition to the four geometric dimensions, we can include other dimensions such as wavelength (or color), time and polarization.

1.2.1 Generalized Optics

In computational photography, each optical element in the lens is treated as a 4D ray-transformer that modifies a light field between a scene and the sensor. The incident 4D light field for a given wavelength is transformed by the optical element into a new 4D light field. Generalized optics, which are discussed further in Chapter 4, can be classified in three general ways, based on where and how the optical element impacts the light propagation.

Each optical element spreads, rearranges and re-bins the incident rays. Depending on where the element changes the rays, the structure of rays emerging from a scene point can ultimately impact the sensor in different ways. When the optical element spreads rays from a scene element to a large section of the sensor, we consider the optics to be a *transform first element*. These can include camera arrays and other optical elements in place before the traditional lens optics. When the optical element affects most of the scene in a similar fashion, we call it a *transform middle element*. Examples include conventional thin lenses, coded amplitude masks, coded phase masks, and gratings in the aperture. When the optical element affects the sensor locally, close to the sensor surface, we call it a *transform last element*. These include lenslet arrays and Bayer mosaics.

Computational methods have already expanded the usable set of lenses in modern imaging systems. Traditional lenses form a single-point perspective by orderly refraction of incident rays, combined with suppression of all other changes. Generalized optics, on the other hand, can intentionally include refraction, reflection, attenuation, scattering and even intentional diffraction. Refractive elements such as those in traditional lenses are primarily ray-benders; they map one incoming direction to one outgoing direction. As a result, traditional lenses have a quadratic thickness profile to bend all rays from the same direction toward a single focused point on the sensor.

But no single refractive lens element can achieve the high optical standards of modern lens designs. Instead optics designers combine multiple materials and refracting elements with linear profiles (prisms), cubic profiles (wavefront coding), convex and concave profiles, aspheric and astigmatic profiles, and other polynomial profiles to form an adjustable optical assembly with superb image-forming abilities. In particular, prisms induce additional optical axis [154], and depth of field is extended computationally by wavefront coded optics [100].

The refractive index of any lens depends strongly on wavelength, making wide-spectrum traditional optics challenging or impossible to build. Computational methods can help by adding reflective elements to decode or rearrange rays and form more coherent images. In addition, reflection plays an important role in dichroic filters that selectively pass light of a small range of colors while reflecting other colors. Dichroic materials use the principle of interference to pass or reduce light at different frequencies, which leads to optical coatings that enhance or suppress controlled reflection. Similarly, mirrors [296] outside the camera can be used to adjust the linear combinations of ray bundles reaching the sensor pixel, and adapt the sensor to the imaged scene.

Apertures, aperture stops and coded masks block or attenuate light. In some imaging methods [445], and in coded-aperture imaging [192] used for gamma-ray and X-ray astronomy, the traditional lens is absent entirely. In modern coded aperture photography, lenses are combined with coded apertures [413] to achieve invertible blur in out-of-focus regions. Elements that diffuse and scatter light are rarely used because they induce a blur in ray-space, although naturally occurring scattering mediums sometimes form part of the optical path. Diffraction gratings and holograms with physical features comparable to the wavelength of light can achieve controllable scattering of light. Combinations of diffractive and refractive elements can create useful programmable spectrum cameras [279] or reduce chromatic aberration by using diffractive optics lenses [74]. Volume holograms can create images in which out-of-focus parts are not only blurred but also darkened [54].

1.2.2 Generalized Sensors

All light sensors can measure the projection of a 4D light field along with the additional dimensions of wavelength, time and polarization. Traditional sensors capture only a discrete two-dimensional projection of this light field by integrating along angle and time dimensions to create a two-dimensional photograph. Video cameras sample the time dimension as well. Computational photography attempts to capture a richer, higher-dimensional representation by using planar, non-planar or even volumetric sensor assemblies. Traditional sensors trade spatial resolution for wavelength measurement (i.e., color) by using a Bayer grid of red, green and blue filters on individual pixels. Another modern sensor design, from Foveon, determines photon wavelength by sensor penetration, permitting several spectral estimates at a single pixel location [138] (Figure 1.6).

Generalized sensors can extend dynamic range by using a gradient processor that measures differential intensity between neighboring pixels [406]. High-speed detectors achieve precise time-gating to measure time-of-flight of light transport for depth detection. The ZCam cameras from 3DV Systems were early commercial examples. In computational photography, such depth-sensing cameras are being used for powerful post-capture control [89]. Careful time-integration sequences can deal with motion blur by preserving information about high spatial frequencies [331]. They can compute sharp images of a fast moving object from a single image taken by a camera with a ‘fluttering’ shutter. Within a single exposure, controlled sensor motion in or out of plane can be used to preserve image spatial frequencies [235] or create shallower depth of field [278]. In the future, optics and sensing will fuse into hybrid elements. We already see this in modern wafer level cameras where the optical components are fabricated on glass wafers in a manner similar to that of fabricating integrated-circuit chips on silicon wafers. More detailed information about sensors is found in Chapter 6.

[Dan Raviv comment. ”There are many 3D equation methods. Why just focus on TOL?”]

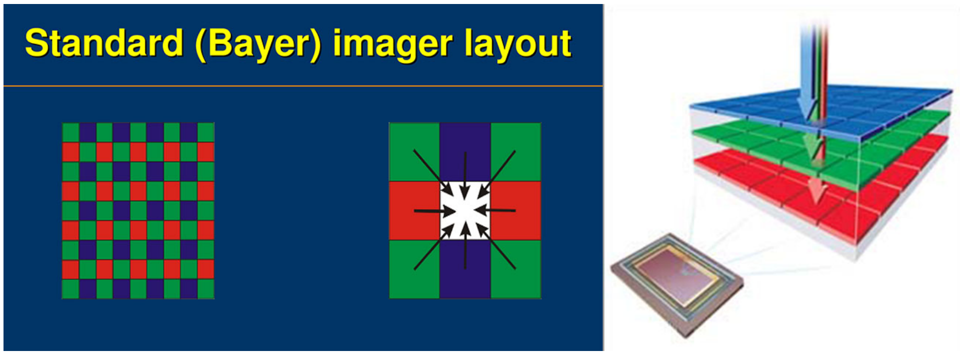


Figure 1.6: (a) The Bayer mosaic applies a pattern of red, green, and blue filters to adjoining sensor pixels. The RGB image pixel is formed by combining red, green, and blue intensities from neighboring sensor pixels. (b) The Foveon sensor uses stacked red, green, and blue layers to record the separate color frequencies for each sensor pixel. The RGB image pixel is formed by combining the red, green, and blue intensities from each corresponding sensor pixel in the stack.

1.2.3 Computational Illumination

Photographic lighting devices have changed considerably in the past century, but the nature of the light they produce has changed hardly at all. Increasingly sophisticated studio lighting continues to be an important tool for the professional photographer confined to a studio space. The introduction of flash bulbs in the 1930s and electronic flash units in the 1950s gave photographers much greater flexibility in adding light to photographs made outside the studio. These lights are limited, however, in the structure of the light they add to a scene. Today, with digital video projectors, servos, and device-to-device communication, we have sophisticated new opportunities for controlling the sources of light. We don't need to be confined to bursts of light at uniform intensities. We can ask the question: What spatio-temporal modulations of lighting might better reveal the visually important contents of a scene?

In his pioneering research work, Harold Edgerton showed that high-speed strobe lighting offers tremendous new capabilities for capturing appearance (Figure 1.7). How many new advantages can we gain by replacing dumb flash units, static spot lights and passive reflectors with actively controlled spatio-temporal modulators and optics? Much research on structured lighting has already been done. For example, we can capture occluding edges with multiple flashes [329], exchange cameras and projectors by Helmholz reciprocity [362], gather relightable actor's performances with light stages [416] and see through muddy water with coded-mask illumination [238]. In every case, better control of lighting during image capture allows for richer representations of photographed scenes. More information about computational illumination is found in Chapter 5.

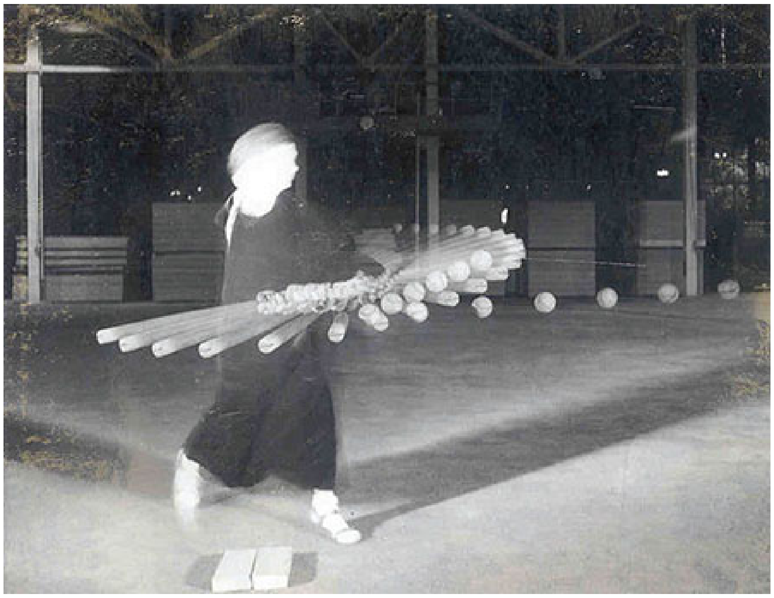


Figure 1.7: Harold Edgerton multiflash photograph. [WE COULD ADD A MORE RECENT EXAMPLE OF COMPUTATIONAL ILLUMINATION, SUCH AS RAMESH'S OCCLUDING EDGES WITH MULTIPLE FLASHES.]

1.2.4 Generalized Processing, Reconstruction and Display

Existing digital camera processors do a series of steps to overcome sensor limitations and improve the resulting photographs. For example, they perform demosaicing (i.e., interpolating the Bayer grid), remove fixed-pattern noise, and hide dead sensor pixels. While these steps improve a digital picture, recent work in computational photography shows that the conversion of raw sensor outputs into picture values can be much more sophisticated. The main approach is a co-design of optics and processing for optimal capture and post-capture resynthesis. This co-design a common theme in what is called coded photography.

In some cases, the outputs of current camera sensors can be manipulated by incorporating recent advances in image processing and computer vision. For example, modern filtering methods can reduce the impact of noise. Similar methods can also detect and recognize important image features, such as faces, as well as categorize and automatically assign higher-level labels. Recent advanced algorithms can remove blur due to defocus or motion by solving the ill-posed blind deconvolution problem and enforcing certain natural image statistics on the solution. In addition, advances in geometric operations on large sets of photographs allow anyone to explore their image content in 3D [376]. In such a data-rich environment, in which millions of photos on any object can be archived and retrieved at little cost, automatic processing must be a central feature of computational photography. These topics are discussed further in Chapter 7.

The statistical 'priors' exploit common observations that there are large gradi-

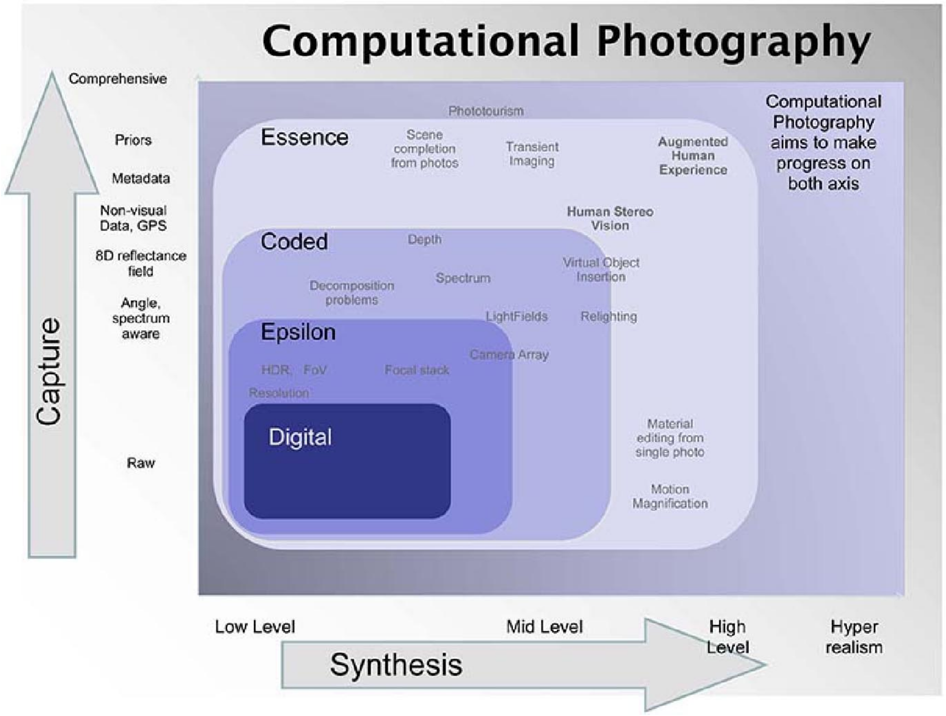


Figure 1.8: The progression of four different stages of computational photography, starting with the methods of conventional film-like digital photography, and extending onward to epsilon photography and then to coded photography, and finally stretching toward a horizon we call essence photography.

ents at sparse image locations or that the histogram of gradients of natural scenes is sharply peaked at zero. The even-increasing online photo collections are allowing rapid progress in data-driven, probabilistic and inferential methods. Cartoon rendering from photos is redefining what it means to be photorealistic. [THIS SHORT PARAGRAPH IS A FRAGMENT THAT DOESN'T CONNECT TO THE TWO PREVIOUS PARAGRAPHS. IT NEEDS TO BE REVISED OR DELETED.]

1.3 Sampling the Dimensions of Imaging

Computational photography is a multidisciplinary field emerging at the intersection of optics, signal processing, computer graphics, computer vision, electronic hardware, art, and online sharing in social networks. We can visualize the extent of computational photography along two axes: capture process versus synthesis goals (Figure 1.8). One axis represents the process of higher-dimensional capture and manipulation, which provides a greater degree of freedom in a machine-readable representation of the scene. The other axis represents the goal of synthesizing visual experience.

Here is a summary of the four stages of computational photography, as illustrated in Figure 1.8.

(i) *Film-like photography.* The first stage describes today's consumer-level digital cameras and basic processing techniques that mimic film photography. The capture is two-dimensional raw images [USE OF 'RAW' HERE IS CONFUSING, SINCE 'RAW' IS AN IMAGE OUTPUT FORMAT. DO YOU MEAN 'TRADITIONAL' OR 'UNPROCESSED' INSTEAD?].

(ii) *Epsilon Photography.* The second stage describes methods to expand camera capabilities by manipulating data from a conventional digital camera. It corresponds to low-level visual processing of pixels and localized scene features. The goal is enhanced performance in the traditional parameters of photography, such as dynamic range, field of view, or depth of field. Because of the tradeoffs inherent in existing cameras, the process typically involves sampling the scene via multiple photographs, each captured by a small, or epsilon, variation of the camera parameters.

(iii) *Coded Photography.* The third stage describes new tools that go beyond the capabilities of the highest-quality conventional digital camera. It corresponds to mid-level visual processing techniques, including segmentation, organization, and the inference of shapes, materials, and edges. These methods perform reversible encoding of information about the scene in a single photograph (or a very few photographs) so that the corresponding decoding can recover light fields, motion-deblurred images, global- and direct-illumination components, or distinctions between geometric versus material discontinuities.

(iv) *Essence Photography.* The fourth stage describes machine-assisted efforts to determine image semantics and the visual content of the scene. Essence photography goes beyond a simple record of the scene's radiometric quantities and challenges the notion that a camera should mimic a grid of light meters. It aims at understanding the contents of the scene as a cognitive aid to human vision. Instead of relying solely on recovering physical parameters, the process exploits non-visual metadata and other priors. The goal is to capture the visual essence of the scene and analyze the perceptually critical components; its results may loosely resemble depiction of the world after high-level visual processing. The goal is a hyper-realistic synthesis that holds promise for spawning new forms of artistic expression and communication.

These four stages of computational photography are described in greater detail in the subsections below.

1.3.1 Film-Like Digital Photography

Even though photographic equipment has undergone continual refinement in the past century, the basic approach to making photographs remains unchanged. A lens admits light into an otherwise dark box, and forms an image on a surface inside the box. (Figure 1.9) This early concept of the camera, known as *camera obscura*, has been understood and explored for over a thousand years,[16] but became known as photography only in the 19th century, when the imaging system was combined

[16] R. L. Verma (1969). "Al-Hazen: Father of Modern Optics."

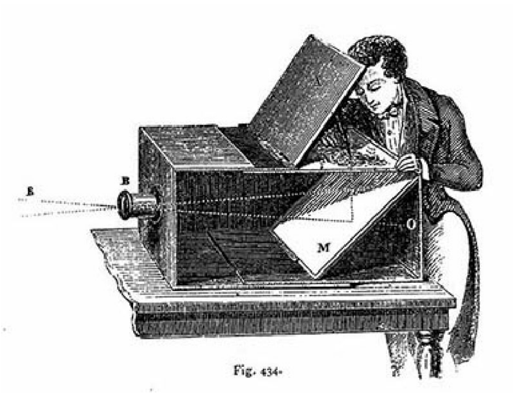


Figure 1.9: Lens-based camera obscura in 1526.

with light-sensitive materials to record the incident light for later reproduction.

Early lenses, boxes, and photosensitive materials were crude in every way. In 1826, with a simple lens, Niépce needed an eight-hour exposure to capture the image of a sunlit farmhouse onto chemically altered asphalt-like bitumen, resulting in a coarse and barely discernible image. Within two decades, other capture strategies based on the light-sensitive properties of various silver salts had reduced the typical exposure time to minutes. By the 1850s these various strategies were displaced by wet-plate collodion emulsions prepared on a glass plate just prior to exposure. (Figure 1.10)

Early photographic history is full of fiercely contested battles over patents and precedence between overlapping processes with different names. William Henry Fox Talbot published a detailed description of his calotype method in 1842, well after Daguerre's huge commercial success with less sophisticated methods that were closely held trade secrets. Talbot made his earliest successful photos in 1835 (but they weren't shown until 1839, in response to Daguerre's claims), and devised the first photographic negatives. Higher sensitivity, better tone reproduction, and the new ability to reproduce photos from negatives caused wet-plate methods to rapidly displace more limited non-reproducible methods based on Daguerre's work. Even though wet plates were messy, complex and noxious to prepare, they produced photos that were larger and more subtly shaded. The materials were also fast enough to record human portraits in shorter sitting intervals, which brought to an end the fixed stony stares that resulted from trying to remain perfectly motionless while squinting into direct sunlight.

George Eastman, annoyed with the expense and complexity of the wet-plate processes of the time, introduced a flexible gelatin film stock in 1874. By the late 1870s, pre-manufactured gelatin dry plates largely replaced the cumbersome collodion wet plates of Talbot and others. Unfortunately, these early film stocks were dangerous. They used a nitrate-cellulose base sheet that would burn ferociously if ignited. Plus, if the film stocks were improperly stored, the chemical byproducts of film degradation resembled dynamite and gun-cotton. In the world of cinema, early large-theatre movie projectors employed arc-lamp illumination, which some-

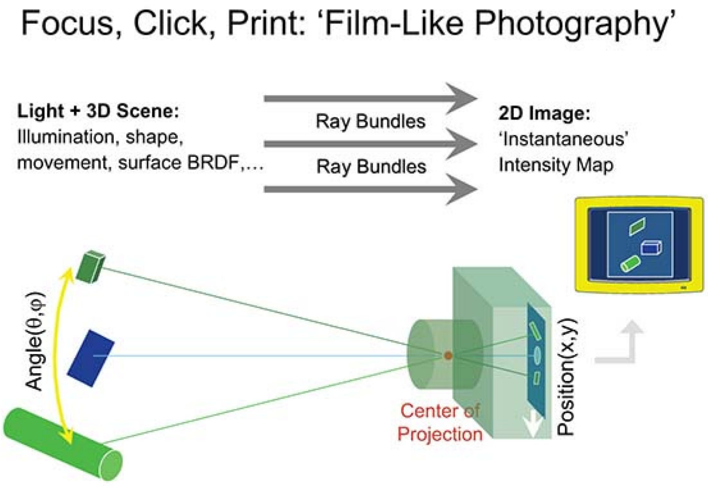


Figure 1.10: Film-like photography uses a lens to form and record an image on a surface with light-sensitive materials. Practical limits such as lens light-gathering efficiency, sensitivity, and exposure time necessitate tradeoffs. [THIS FIGURE, WHICH INCLUDES RAYS, BRDF, AND A MONITOR SCREEN, IS OUT OF PLACE HERE IN A SECTION ON EARLY METHODS OF PHOTOGRAPHY. WE SHOULD FIND A SUBSTITUTE IMAGE]

times ignited reels of nitrate film, causing many deadly movie-theatre fires. It wasn't until 1951 when acetate-based safety films finally replaced nitrocellulose for all photographic purposes.

The emulsions on these films continued to rely on sensitized silver-halide salts, but they advanced from crude single coatings to multilayer thin-film emulsions. The extra layers maximized light absorption, resolution, and sensitivity. Monopack color films, introduced in the 1930s, further complicated emulsions by inserting color filters between multiple light-sensitive layers.

Rapid advances in the design of cameras, lenses, and lighting complemented these advances in thin-film chemistry. Camera manufacturers developed high-speed shutter designs and accurate aperture mechanisms. Lens manufacturers produced complex multi-element lens systems. Portable lighting devices advanced from the crude ignition of magnesium flash powder to synchronized sets of Xenon flash units. Studio lights advanced from simple ceiling louvers that controlled sunlight, as in Thomas Edison's rotating Black Maria film studio, to multi-kilowatt color-balanced lamps with directional reflectors, focusing lenses, removable scrims, motorized gobos, and barn doors, all of which shaped the direction and intensity of light.[17]

With each set of technical improvements, photographers expanded the creative choices that affected the appearance of the captured image. We recall that the ear-

[17] For an authoritative technical review of early photography, see "The Theory of The Photographic Process," 4th edition (1977), edited by T.H. James, New York: Macmillan. [NOTE: THIS BOOK IS CURRENTLY NOT IN PRINT]

liest cameras had neither shutters nor aperture mechanisms. Photographers chose a lens (or the camera attached to it), adjusted focus on a ground-glass viewer, replaced the ground glass with a light-sensitive plate, uncapped the lens and waited for the camera to gather enough light to record the image. As light-sensing materials improved, exposure time dropped from minutes to seconds to milliseconds. Adjustable shutters made exposure times more precise, and adjustable lens apertures regulated the amount of light passing through the lens during exposure. By the 1880s, the basic camera settings used for photographic exposure were well defined. Digital cameras continue to use these basic settings, and have improved and extended them only slightly.

The basic components of digital photography and their associated settings are (1) the lens, which controls aperture, focusing distance, and focal length; (2) the shutter, which controls exposure time; (3) the sensor, which controls light sensitivity (i.e., the ISO value), latitude (i.e., tonal range or dynamic range), and color-sensing properties; (4) the camera, which controls location, orientation, and the moment of exposure; and (5) the auxilliary lighting, which controls the position, intensity, and timing of the illumination. Most film-like digital cameras can automatically determine the focus, aperture, shutter speed, sensor sensitivity, and lighting settings at the moment a picture is made. Once the user trips the shutter release, these settings are fixed, and the resultant image is one among many photographs.

At the instant the shutter clicks and an exposure is made, the following camera settings have also been determined:

Field of view. The focal length of the lens determines the angular extent of the picture. A short (wide) focal length gives a wide-angle perspective on the scene; a long (telephoto) focal length gives a narrow perspective. For a fixed sensor resolution, the field of view dictates the angular resolution of the scene. Although the image can be cropped to reduce perspective (at a corresponding loss of resolution), it cannot be made wider.

Exposure and dynamic range. The chosen lens aperture, exposure time, sensor ISO sensitivity, and sensor latitude contribute to how the light in the scene maps to individual pixel tonal values. Larger aperture settings, longer exposure times, and higher sensitivities map dimly lit scenes into acceptable pictures, while smaller apertures, shorter exposure times, and lower sensitivities map brightly lit scenes into acceptable pictures. Poor choices in these settings may result in the loss of visible image details in brightly lit areas or in dimly lit areas, or both. Within the sensitometric response curve of any sensor (the intensity ratio between the darkest and lightest details), the latitude of the dynamic range of the sensor is not usually adjustable, and falls typically between 200:1 to 1000:1.

Depth of field. The lens aperture, the lens focal length, and the size of the sensor together determine the range of distances that will appear acceptably in focus at any given focus distance. A small aperture and a short (wide) focal length gives the greatest depth of field, while a large aperture and a long (telephoto) focal length gives the smallest depth of field. Note that increased depth of field normally requires a smaller aperture, which may entail increased exposure time or higher sensor sensitivity (which in turn increases digital noise in the image).

Spatial resolution. For a well-focused image, the sensor itself sets the spatial resolution. The image can be artificially blurred to diminish resolution, but no sharpening can recover more detail than that already recorded by the sensor. Note that increased spatial resolution reduces depth of focus and often increases visible noise due to the reduced size of the sensor pixel. In addition, increased spatial resolution correspondingly increases the required data storage and bandwidth requirements.

Temporal resolution. A chosen exposure time interval determines how long the camera will collect light for each point in the image. If the exposure is too long, moving objects will appear blurred; if the exposure is too short, the camera may not gather enough light for a proper image.

Wavelength resolution. Color-balance and saturation settings on the camera set sensitivity to color. Current film-like digital cameras sense color by measuring three color primaries (usually in the RGB color space) with fixed, overlapping spectral response curves. Different sensors offer varying spectral curves, but none of these curves for a given sensor are adjustable.

Film-like digital photography forces us (or the camera) to choose the camera settings, to make tradeoffs among interdependent parameters, and to lock in those choices in a single photo at the moment we click the shutter. The fundamental nature of film-like photography, which all photographers understand, forces these tradeoffs. They are inescapable because of the hard limits of simple image formation and the measurement of light. But photographers would like more capabilities with fewer tradeoffs! We would like to capture any viewed scene, no matter how transient and fast-moving, in an infinitesimally short time period. We would like to have the ability to choose any aperture setting, even a very tiny one in dim light, to control the depth of focus. And we would like unbounded resolution that would allow capture of a very wide field of view.[18]

New methods of computational photography offer a steadily growing number of capabilities to escape the restrictions of these tradeoffs. Even though existing film-like digital cameras are already excellent imaging devices, and they offer a range of adjustment for each of these imaging parameters, we can be increasingly confident of finding computational strategies to expand these parameters. The next section, on epsilon photography, describes some of the strategies that have been discovered.

1.3.2 Epsilon Photography

We can think of cameras at their best as defining a box in the multidimensional space of imaging parameters such as dynamic range, field of view or depth of field. The first and most obvious thing we can do to improve digital cameras is to expand this box in every parameter dimension. The goal is to build a super camera with enhanced performance in each of the traditional parameters. We call this expanded performance *epsilon photography.* In general, a single scene is recorded by making multiple images that vary one or more of the camera parameters by some small amount (i.e., epsilon). This is similar to the concept of epsilon geometry [350], which computes an exact solution for any perturbed version of the input within a small epsilon neighborhood. The goal of epsilon photography is to make a robust

[18] Obviously, in the limit, an infinitesimally small aperture and zero exposure interval would gather no photons at all!

estimation of an image from samples in a epsilon neighborhood of its film-like photographic parameters.

For example, normalized pixel intensity (intensity value divided by exposure time) may be difficult to estimate because of saturation or underexposure. While the exposure time can vary from a few femtoseconds to a year,[19] in practice we can take a series of photographs by making minor changes around the exposure time within an epsilon interval, a technique widely known as *exposure bracketing.* From these few measurements, we can robustly estimate the normalized pixel intensity in the presence of serious sensor response non-linearities and clamping. Similarly, increasing an image field of view by creating a panorama involves changing the camera's pose, and altering two of its six position-setting coordinates [position (x, y, z) and orientation (roll, pitch, yaw)] within a modest neighborhood of values. In another example, successive images (or even neighboring pixels) may have different settings for parameters such as exposure, focus, aperture, view, illumination, or timing of the instant of capture. Each epsilon setting allows partial information about the scene to be recorded, and the final image is reconstructed by combining all the useful parts of these multiple observations.

Epsilon photography is thus the concatenation of many such boxes in parameter space, where multiple film-style photographs are computationally merged to make a more accurate photographic description. While the merged photograph is superior, each of the individual photographs is still useful and independently comprehensible. The merged photograph contains the best features from each photograph in the group. Thus epsilon photography corresponds to low-level vision; the process estimates and merges pixels and pixel features from multiple observations and selects those with the best signal-to-noise ratio.

Here are descriptions of how epsilon photography can improve the six parameters settings we described earlier in the section on film-like photography.

Field of view. A wide field-of-view panorama is achieved by stitching and mosaicing images taken by rotating a camera around a common center of projection or by translating a camera across a planar scene.

Exposure and dynamic range. A high dynamic range (HDR) image can be formed by accurately merging photographs taken at a selected series of exposure values.[20]

Depth of field. A photograph that is entirely in focus, foreground to background, is constructed from a series of images taken by successively changing the plane of focus [39].

Spatial resolution. Higher resolution is achieved by tiling multiple cameras and assembling a spatially varying mosaic from individual images [425], or by jittering the position of a single camera [225].

Temporal resolution. High-speed imaging is achieved by staggering the exposure time of multiple low-frame-rate cameras. The exposure durations of individual cameras can be overlapping [367] or non-overlapping [425].

Wavelength resolution. Conventional digital cameras sample only three basic

[19] Welsley long exposure photos

[20] Mann and Picard 1993, "Compositing Multiple Pictures of the Same Scene," by Steve Mann, in IS&T 46th Annual Conference, Cambridge, Massachusetts, May 9–14, 1993; Debevec and Malik 1997; Kang et al. 2003.

RGB color primaries. Multispectral imaging (using multiple color frequencies in the visible spectrum) or hyperspectral imaging (using wavelengths beyond the visible spectrum) can expand wavelength resolution by successively changing color filters in front of the camera during exposure, or by using tunable wavelength filters or diffraction gratings [279].[21]

Photographing multiple images while varying camera parameters can be done in several ways. Images can be taken with a single camera over time. Or images can be captured simultaneously by using the technique of assorted pixels, where each pixel is attuned to a different value for a given parameter [286]. Just as some early digital cameras captured a sequence of scan lines, including those that moved a single linear detector array across the image plane, we can imagine detectors that intentionally randomize each pixel's exposure time to form a tradeoff between motion blur and resolution. This technique was previously explored for interactive computer graphics rendering [90].[22] Simultaneous capture of multiple samples can also be achieved by using multiple cameras, each camera having different values for a given parameter. Two designs are currently being employed for multi-camera solutions: a camera array [425] and single-axis multiple parameter (co-axial) cameras [281].

The techniques of epsilon photography have evolved significantly, and the field remains an active area of research with rich potential. Some camera manufacturers have already implemented aspects of epsilon photography in their consumer products. Burst-mode features with optional change of parameters between successive photographs (e.g., Casio Exilim EX-F1 [THIS CAMERA IS 4 YEARS OLD NOW. WE NEED A BETTER EXAMPLE!]) are examples of how epsilon photography will make a significant impact. We believe that panoramas, mosaics, extended depth of field, superresolution and HDR capture methods are only the beginning of what can be achieved with epsilon photography (see Figure 1.12).[23] Many of the traditional photographic parameters and tradeoffs are enticing targets for clever new computational exploits. Table 1 shows some of these promising new directions. (A temporary jpeg version of Table 1 is shown here as Figure 1.11)

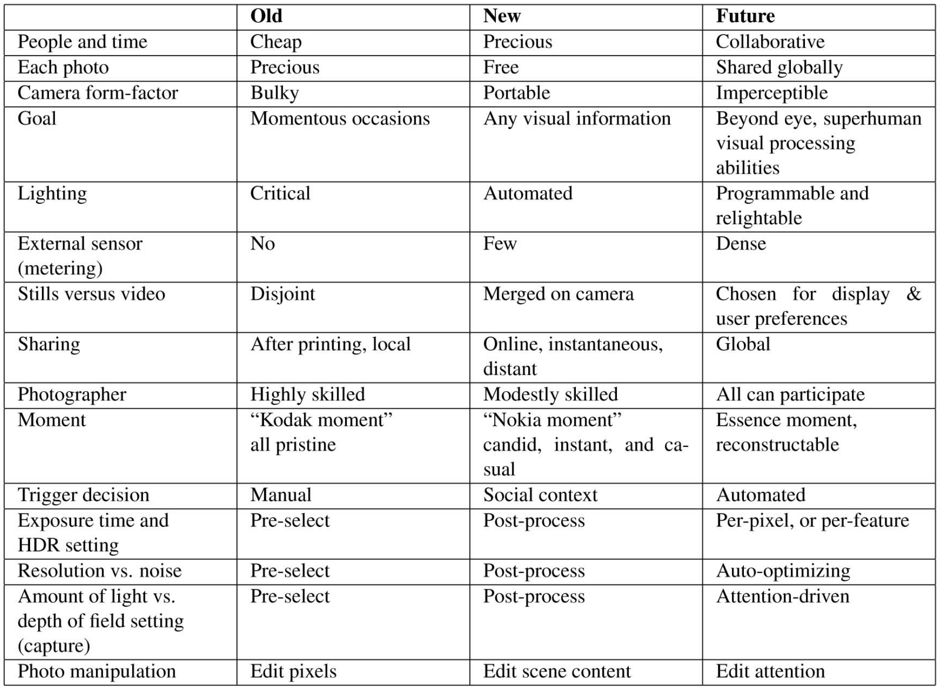
1.3.3 Coded Photography

By combining the best features of multiple conventional photographs we can extend the capabilities of any existing camera, and produce what could be called the best possible super camera. But we wish to go far beyond this idea. Instead of high-quality pixels, our goal is to capture and convey the mid-level cues, including the shapes, boundaries, materials and organization in a scene. In a traditional camera, the light incident at a pixel is integrated along angular, temporal and wavelength dimensions during the exposure interval to record a single intensity value. Distinctly different scenes may result in identical projections (images) and hence identical pixel values. Thus we are challenged to estimate scene properties that are not directly observable. These properties are critical for post-capture manipula-

[21] www.blackwell-synergy.com/doi/abs/10.1111/j.1467-8659.2008.01169.x

[22] www.cs.virginia.edu/~luebke/publications/pdf/afr.egsr.pdf and www.cs.virginia.edu/~luebke/

[23] scalarmotion.wordpress.com/2009/03/15/propeller-image-aliasing/



	Old	**New**	**Future**
People and time	Cheap	Precious	Collaborative
Each photo	Precious	Free	Shared globally
Camera form-factor	Bulky	Portable	Imperceptible
Goal	Momentous occasions	Any visual information	Beyond eye, superhuman visual processing abilities
Lighting	Critical	Automated	Programmable and relightable
External sensor (metering)	No	Few	Dense
Stills versus video	Disjoint	Merged on camera	Chosen for display & user preferences
Sharing	After printing, local	Online, instantaneous, distant	Global
Photographer	Highly skilled	Modestly skilled	All can participate
Moment	"Kodak moment" all pristine	"Nokia moment" candid, instant, and casual	Essence moment, reconstructable
Trigger decision	Manual	Social context	Automated
Exposure time and HDR setting	Pre-select	Post-process	Per-pixel, or per-feature
Resolution vs. noise	Pre-select	Post-process	Auto-optimizing
Amount of light vs. depth of field setting (capture)	Pre-select	Post-process	Attention-driven
Photo manipulation	Edit pixels	Edit scene content	Edit attention

Figure 1.11: This temporary figure is a jpeg version of Table 1.

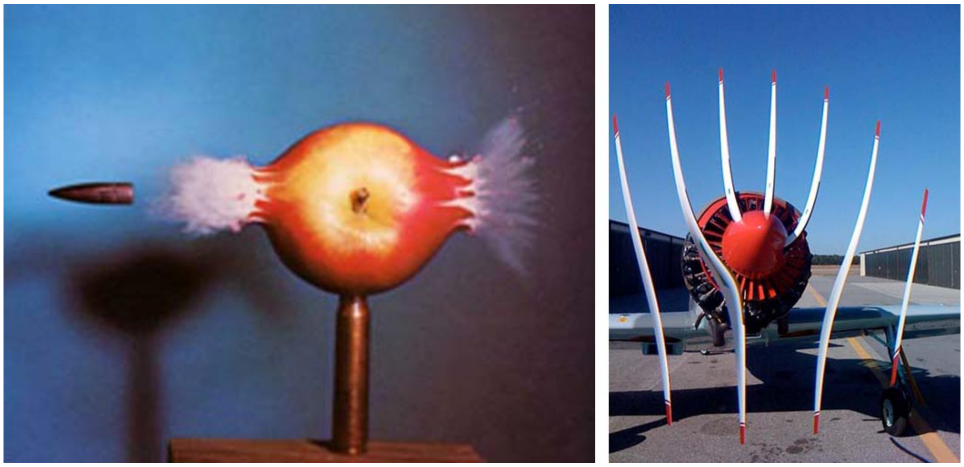


Figure 1.12: Photographs help us understand our world and how our world looks and works, usually in ways our eyes can never see. (a) The multiple flash pulses in stop-motion strobe photography, such as the classic work done by Harold Edgerton, is a form of epsilon photography with expanded illumination. (b) Surprising slit-shutter images of propellor motion taken by a simple cell phone camera, show us features of the world that are beyond our visual experience. [THESE PHOTOS WOULD WORK BETTER EARLIER IN THE CHAPTER, AS EXAMPLES OF HOW PHOTOGRAPHY CAN EXPAND OUR VISUAL UNDERSTANDING OF THE WORLD. NEITHER OF THESE PHOTOS DEMONSTRATE EPSILON PHOTOGRAPHY AS YOU DESCRIBE IT. DO YOU HAVE OTHER EXAMPLES USING CURRENT DIGITAL METHODS?]

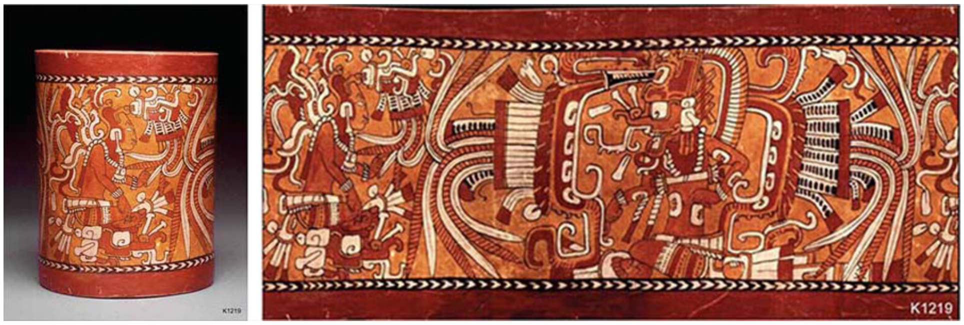


Figure 1.13: (a) What is a photograph? Does it always have to be an image of a scene from a single point of view? (b) A virtual camera can capture images of an object from multiple points of view, such as a ring of camera positions, and construct a single "photograph" that contains only a small number of pixels from each individual camera position. A coded image such as this can then be decoded to reproduce a representation of the original scene.

tion and synthesis. Coded photography reversibly encodes information about the scene in a single photograph (or a small number of photographs) so that the corresponding decoding allows powerful decomposition of the image into light fields, motion-resolved images, global/direct illumination components or distinctions between geometric versus material discontinuities.

Here's a simple example. Instead of increasing the field of view just by panning a camera, can we also create a wrap-around view of an object? Panning a camera allows us to concatenate and expand the box in the camera parameter space in the dimension of field of view. But a wrap-around view spans multiple disjoint pieces along this dimension. We can virtualize the notion of the camera itself if we consider it as a device for collecting bundles of rays leaving a viewed object in many directions, not just toward a single lens, and virtualize it further if we gather each ray with its own wavelength spectrum.

Coded photography is an out-of-the-box photographic method in which individual (ray) samples or data sets may not be comprehensible as images without further decoding, re-binning or reconstruction. For example, as shown in Figure 1.13, a wrap-around view might be constructed from multiple images taken from a ring or sphere of camera positions around the object, using only a few pixels from each input image for the final result.[24]

Can we find a better, less wasteful way to gather information beyond pixel intensities? We can start with a stereo pair of cameras that encode the depth of a scene, or a camera array that captures a light field in a scene for novel view synthesis. While estimating depth continues to be a challenging problem, we can estimate boundaries and regions more robustly. By using multiple flashes and analyzing the slivers of shadows created at depth discontinuities, we can distinguish between geometric versus reflectance boundaries [329].

[24] research.famsi.org/kerrmaya.html. Rollout photograph ©Justin Kerr. Suggested by Steve Seitz.

Capturing higher-dimensional signals on two-dimensional sensors requires some jugglery in optics and sensing. A new strategy involves careful manipulation and coding of the point spread function. In a coded exposure technique, the shutter of a camera can be rapidly fluttered open and closed in a carefully chosen binary sequence as it captures a single photograph. The fluttered shutter encodes the motion that conventionally appears blurred, and this reversible encoding then allows us to compute a moving but unblurred image [331]. Similarly, coded aperture techniques, inspired by work in astronomical imaging, preserve the high spatial frequencies of light that passes through the lens so that out-of-focus blurred images can be digitally refocused [413] or resolved in depth [231].

An important aspect for post-capture manipulation is the ability to decompose the scene into meaningful components. These decomposition problems are at the heart of many new coding techniques. By coding illumination, it is possible to decompose radiance in a scene into direct and global components [292]. We can also segment foreground from background by using various matting techniques [84, 383]. Other examples include confocal synthetic aperture imaging [239] that let us see through murky water, and techniques to recover glare [ED: do you mean "veil glare" or "recover imagery from glare"?] by capturing selected rays through a calibrated grid [389] or multiple lens sub-apertures [332]. Coding the sensor for differential measurement, such as with a gradient camera [406], provides higher dynamic range.

In the emerging field of coded photography, we continue to look for other surprising capabilities that occur when we combine computation with new combinations of sensing and scene appearance.

1.3.4 Essence Photography

The next phase of computational photography will go beyond radiometric quantities. It will challenge the notion that a synthesized photo should appear to come from a device that mimics the information-gathering and comprehension of the human eye and visual system. Instead of recovering physical parameters, the goal will be to capture the visual essence of a scene and scrutinize its perceptually critical components. This 'essence' photography may loosely resemble depiction of the world after high-level vision processing. (Figure 1.14)

Essence photography does not limit a camera to photon sensing and light measurements alone, nor does it rely on a single isolated recording device. A camera may measure geographical location coordinates, identify scene contents, and recognize gestures by communicating with other probes, actuators, light sources, or nearby wireless services. With sophisticated algorithms we will exploit priors based on natural image statistics [133] and online community photo collections to explore scenes in 3D [376] or to enhance our own photographs with information gathered from those collections [178]. We will challenge physical notions of time and space by capturing and synthesizing multilinear perspective [435], unwrapped mosaics [334] and video texture panoramas [36].

Non-photorealistic synthesis will become a new artistic tool, for example to exaggerate motion via a motion microscope [250] or to highlight subtle features via illumination [131]. Beautification of photographs by directly manipulating scene elements will become an option by using data-driven enhancement of visual features

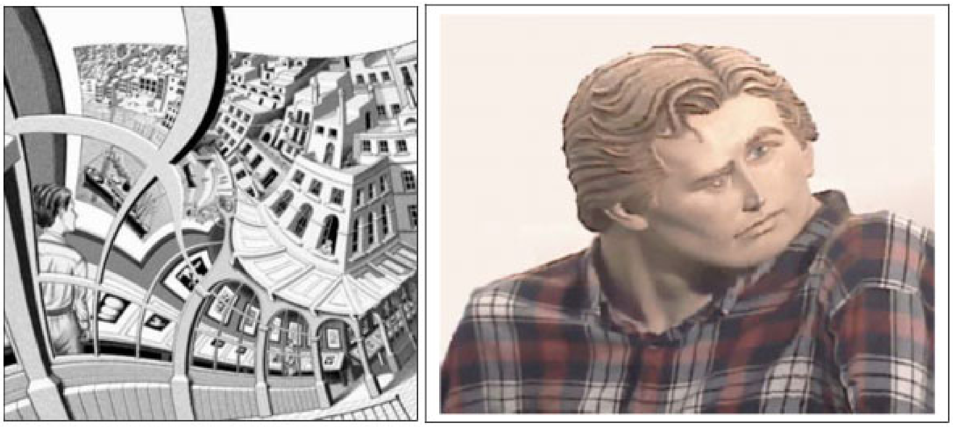


Figure 1.14: Essence photography recognizes that inference and perception are important, as are the intent of the photographer and the purpose of the photograph. Instead of a straightforward recording of light, perhaps what we really care about is a form that captures a meaningful subset of the visual and emotional experience.

such as facial attractiveness [242]. Fusion and synthesis of non-visual data not normally associated with imagery will be critical. Can we render the calories or aroma of food? Can we synthesize the wind on a roller coaster ride? Can we highlight irony? Essence photography, with a fresh field of discovery in front of it, will spawn new and unexpected forms of visual and artistic expression and communication. We are moving toward a much more capable box of parameters, in ways we can't yet fully recognize, with quite a bit of innovation yet to come!

1.4 Where Is Photography Going?

The first decade of this century has been an exciting period of dramatic change for digital photography. Ten years ago only a few expensive digital cameras could approach the quality of film, and software processing was limited. New consumer digital cameras were fun to play with, but they were not much more than diversions. Film was still king. Photographers knew how to work with film and they had a well-defined workflow to process and print images. Few people thought film would ever disappear.

Today everything has changed. Digital photography is now the dominant imaging technology throughout the world, and digital cameras are ubiquitous. People adapted to the new paradigm, and quickly became comfortable with new cameras and necessary computer skills. We've seen astounding developments in cameras and sensors, and the computational promise is enormous, as we've just begun to describe.

But we have many questions. What will the future of photography look like? How will the technology of photography continue to develop? What will a camera look like in ten years? In twenty years? In fifty years? What will photojournalism look like? Will we all be photojournalists in a networked online world? How will powerful new movie-making capabilities change the nature of photography? Will photography as we know it disappear into a soup of unlimited media possibilities? How will online photo collections transform visual social computing? How will a billion portable networked cameras change the social culture?

All these questions will be answered as researchers explore new imaging possibilities. Computational photography, in which photographs of the future will be computed rather than recorded, has already started to change the workflow of imaging and give us new and expanded opportunities for seeing. It will continue to transform the new world of digital photography just as dramatically as digital photography transformed the traditional world of film.

Chapter 2

Camera Fundamentals

Photographers who learned how to make images in the traditional world of film already have a good qualitative grasp of the capabilities and limitations of film-like digital photography. These well-known features have been understood and accepted as tradeoffs for decades. Here is a list of some of them.

- A longer exposure time, additional scene lighting (such as flash), or greater sensitivity makes brighter pictures. But a longer exposure time can increase blur, additional lighting can disrupt scene appearance, and greater sensitivity results in increased noise.

- Larger lens apertures gather more light, but reduce the depth of focus in the image.

- Wide-angle (i.e., wide field-of-view) lenses shrink scene features and exaggerate foreshortening (depth-dependent size). Narrow-angle (i.e., telephoto) lenses enlarge scene features and reduce foreshortening.

- Wide-angle lenses gather more light than narrow-angle lenses of the same aperture.

- Wide-angle lenses offer very short minimum focus distances and have very large depth of focus. Narrow-angle lenses require larger minimum focus distances and have shallower depth of focus.[1]

After more than a century of use and understanding, these and other so-called limitations may seem to be inescapable and fundamental laws of photography, but they're not. Each of these limitations is a direct consequence of laws of physics, image formation, and light transport, as applied to well-grounded assumptions of film-like photography. We will challenge and transcend each of these conditions in subsequent chapters.

[1] A tiny bit of dust on the front surface of extra-wide fisheye lenses at small aperture settings can easily cast fiber-shaped shadows on the image sensor.

To prepare, we first embark on a brief quantitative review of these limitations and tradeoffs. We review light measurement, lenses, apertures, and image formation, and we reveal the reasons behind these underlying principles so that we can address them with new computational photography techniques.

2.1 What is Light? Waves and Particles

Let's start with a bit of history. In the 4th and 5th centuries BCE, the ancient Greeks speculated extensively about the nature of light. Empedocles, Euclid, and Plato thought that light is projected from the eye. Lucretius believed that the sun sent out particles of light, while Pythagoras argued that objects emitted particles of light. Aristotle even proposed a theory of wave propagation of light. Some of these ideas had elements of truth, but they were not particularly useful. They did not advance knowledge because they were not based on a consistent theoretical framework, and they were not tested in experiments. Ultimately, these early ideas provided only a set of possibilities.

Some observations and insights in the ancient world carried over into the scientific study of light. For example, Euclid in the 4th and 3th centuries BCE noted that light appears to travel in a straight line, and Hero in the 1st century BCE concluded from the study of reflections that light follows the shortest path between two points. These early concepts were useful, but not fruitful. Modern thinking on the nature of light and optics didn't originate until the 17th century, when two ideas dominated theory and experiment: (1) light is a propagating wave and (2) light consists of streams of particles, or corpuscles.

Sir Isaac Newton (1642–1727), shown in Figure 2.1, was perhaps the most influential scientist of that period. He carried out exquisite optical experiments to determine the nature of color, and developed a valid foundational theory of color. He concluded that light rays consist of particles rather than waves because light travels in straight lines along ray-like paths. Anyone who has played with a laser pointer might easily come to the same conclusion, but today we know that this perception is incomplete. It does not include the deflection and decomposition that occurs when light interacts with sharp edges of opaque objects, narrow slits, or collections of narrow slits in gratings.

In 1663 James Gregory noticed that sunlight passing through a feather is diffracted into spots of different colors. He wrote, "I would gladly hear Mr. Newton's thoughts of it." Later, in 1665, Francesco Grimaldi published a book on the effects caused when beams of light passed through small apertures. These effects indicated light had wave-like properties. Apparently, Newton did not consider the possibility that light might be a wave. He continued to see light as a particle that traveled in straight lines, even though light when closely observed behaved as if it had a very small wavelength.

The particle theory of light was, in fact, continually undermined by observations of diffraction. In 1678 Christiaan Huygens presented a wave theory of light that explained such phenomena, although with certain unproven assumptions. According to wave theory, diffraction intensity patterns consisting of bright and dark lines result from interference effects, or the addition of the amplitudes of waves.

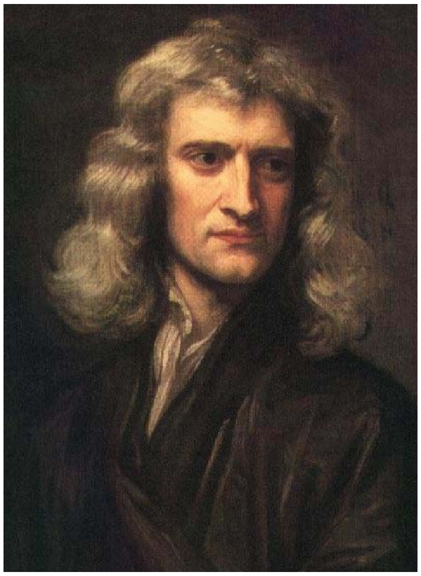


Figure 2.1: Sir Isaac Newton (1642–1727) dispersed light with a prism and developed a theory of color.

In constructive interference the positive and negative peaks add together to give intensity maxima; in destructive interference the positive and negative peaks cancel to produce dark lines, or intensity minima, as illustrated in Figure 2.2.

Several contemporany scientists confirmed the validity of the wave theory. Most notably, Thomas Young in 1802 demonstrated the interference pattern produced by diffraction by passing a light beam at a selected wavelength through two closely spaced slits in an opaque sheet, causing a fine pattern of bright and dark lines to form on a screen placed behind the slits. Joseph Fraunhofer and Augustine Fresnel also investigated diffraction effects by passing light through small holes of different shapes. They presented rigorous theories for calculating diffraction patterns, thus advancing the wave hypothesis.

By the early 19th century the wave properties of light were well understood. Scientists knew clearly how light rays behave and how perceived colors result from the mixing of light beams. The diffraction of light from the surface of a compact disc, with 625 tracks per millimeter, provides a beautiful illustration of this wave effect, as shown in Figure 2.3.

There still was, however, an open question. “What is light?” An experiment to measure the speed of light provided a solid line of evidence in the search for an answer. To casual observers the speed of light appeared to be infinite, but well-planned experiments showed that assumption was not accurate. In 1676 Olaf Romer used observations of eclipses of the moons of Jupiter to determine an early approximate value of the speed of light. In 1860 newer measurements of the speed of light in air had been made by studying light pulses (Armand Fizeau) and rotating mirrors (J.L. Foucault).

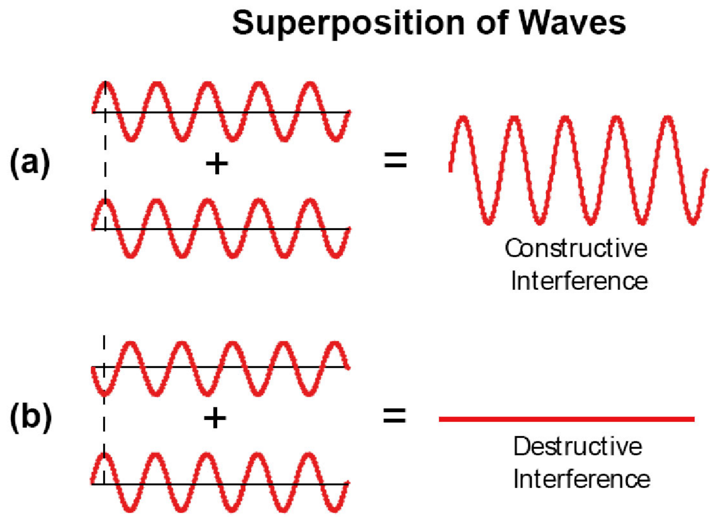


Figure 2.2: (a) Constructive and (b) destructive interference of waves. (Source: Charles Johnson book)

Figure 2.3: Diffraction of light from the surface of a compact disc. The pitch of the data tracks is 1.6μm, and each angle of reflection selects a specific spectral color. (Source: Charles Johnson book)

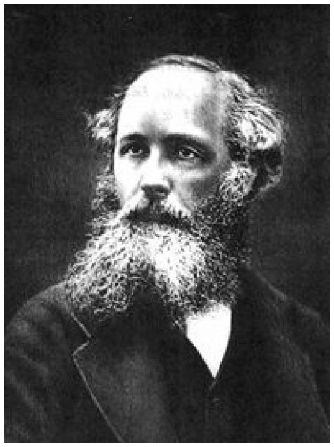


Figure 2.4: James Clerk Maxwell (1831–1879) identified light as electromagnetic radiation. (Higher resolution source image and permissions needed for this figure.)

James Clerk Maxwell's monumental achievements in the 1860s provided a striking conclusion to this part of the search. Maxwell developed a set of partial differential equations to describe the interrelationship of electric and magnetic fields. He concluded that oscillating electric fields are always accompanied by oscillating magnetic fields, and together they propagate through space as electromagnetic waves. In 1862 he computed the velocity of electromagnetic waves and discovered that they have the same velocity as light.

This revelation prompted Maxwell to write in an 1864 paper, "...we have reason to believe that light itself (including radiant heat and other radiations) is an electromagnetic disturbance in the form of waves propagated through the electromagnetic field according to electromagnetic laws." The simple question "What is light?" now had an answer—light is electromagnetic radiation.

But the story of light does not end with Maxwell's equations. In 1905 Albert Einstein showed that light is composed of quanta, or particles of energy, and every experiment since that time has confirmed his conclusion. He discovered that light quanta are necessary to explain the interaction of light with electrons in metals. For this amazing conceptual breakthrough Einstein was awarded the Nobel Prize in 1921.

Einstein showed that Newton was right about the particle nature of light, but he was right for the wrong reasons. We always detect light quanta (now called photons), but the wave theory permits us to compute the probability of finding a photon at a certain location. Each photon carries an amount of energy that depends on the frequency of the light, and the number of photons arriving each second determines the intensity of the light. A few photons are required to activate a nerve in the eye, and the creation of a latent image in photographic film requires the absorption of photons by atoms in silver halide grains. Today we have many instruments that can detect single photons.

So we have a beautiful and subtle story. Light is made of particles and light is

Figure 2.5: Albert Einstein (1879–1955) discovered that light is quantized. (Higher resolution source image and permissions needed for this figure.)

also a wave. Maxwell's equations are sufficient when there are many photons, and quantum theory is necessary when there are only a few photons. No wonder the ancient thinkers had a hard time getting the story straight!

2.2 Measuring Light with Rays

The behavior and measurement of light is both obvious and elusive at the same time, especially if we use rays to describe light, as we will do for most of this book. Intuitively, a ray is a line-like path for light. It is the two-way tracing through space that we imagine a single photon might leave behind as it flashes through a scene, reflecting, bending, or scattering at the surface boundaries between different materials. Figure 2.6 illustrates this concept, showing how rays travel from an illuminating source to different surfaces and onto the image plane of a camera.

Rays provide us with a simple way to describe digital photography. We can imagine that each pixel in a digital image copies the color of a single ray through a single point or pinhole called the *center of projection* (COP) somewhere inside the lens, as shown in Figure 2.6 [ADD THE COP TO THE FIGURE]. But rays are also powerful decomposition tools. They let us explore complex optical systems locally, just one path at a time, describing each coating and lens surface by how sets of rays are affected, and decomposing each set of rays into simple individual ray-bending events. Selective occlusion and bending in sets of rays allows us to illustrate the causes of soft shadows and caustics, explore the effects of multi-element lenses, and classify whole families of image formations beyond the familiar approach of a single-point perspective.

These ray paths are also exactly reversible. The light attenuation along the

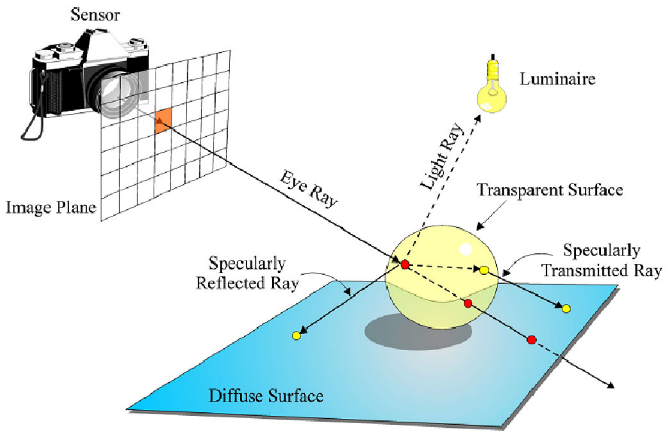


Figure 2.6: Each ray specifies a reversible line-like path that a photon might take as it travels through a scene, moving from an illuminating source to different surfaces and into the camera. (Figure by Ann McNamara, 2000)

path of a single ray from source to destination is exactly the same if we reverse the path, no matter how complicated that path might be. Ray-traced renderings routinely exploit this reversibility property (known as *Helmholtz reciprocity*) to create beautiful computer graphics imagery. The reciprocity of rays has helped computer vision researchers such as Zickler, Belhumeur, Kriegman and others neatly untangle shape and complex reflectance properties in novel forms of stereo photography [442, 443].

Quantitatively, however, one single isolated ray is problematic. It cannot exist in isolation and it does not carry any measurable amount of light. A light ray is not just a very narrow beam; it is infinitesimally narrow and it does not spread out with distance. Its angular extent is zero. A light ray doesn't emerge from (or arrive at) a small area. It comes from a single point with zero area. Thus a ray is doubly infinitesimal, which makes it impossible to measure individually, and which makes its units maddeningly obtuse. As a result, we need to define rays clearly and carefully.

First, let's briefly review some fundamental theories of how light interacts with matter. A century ago, physicists firmly established the astonishing duality of light as both propagating electromagnetic waves and as streams of particles or photons. This duality is also known as *wave-particle complementarity*. Both explanations are necessary to describe how light interacts with matter.

2.2.1 Light as Waves

In physics, all electromagnetic radiation qualifies as light, and all wavelengths are governed by Maxwell's equations, from radio signals with kilometer-long wavelengths to cosmic rays and gamma rays with wavelengths shorter than the diameter of a single atom. All these electromagnetic waves consist of coupled localized changes in electric and magnetic fields, and these changes propagate rapidly through space and time. If entirely unobstructed, all electromagnetic waves in a perfect vacuum propagate outward in all directions from their sources at the constant rate of $c = 299,792,458$ meters/second, thus linking its temporal frequency in Hertz (Hz, or cycles/second) to its wavelength λ by:

$$f = c/\lambda \tag{2.1}$$

Light that is visible to humans falls within a narrow span of wavelengths between about 380 to 770 nm (1 nanometer = 10^{-9} meters), with our eye's greatest sensitivity around 555 nm or about 540 THz (1 terahertz = 10^{12} Hz). Rescaling can help make these tiny distances a bit less abstract; exactly 2000 cycles of 500 nm waves, which look cyan/green, stretch across one millimeter, and we need about 50,000 cycles to span one inch.

At all visible wavelengths (and in many broad swaths of nonvisible wavelengths), the propagation of light through air closely matches propogation in a perfect vacuum, especially if the air is calm, clear, dry, unobstructed, and uniform in temperature. If we measure nearby wavelengths (λ), a margin of about 100λ around the path of propagation is usually sufficient. For example, as we drive a car underneath a railway or highway overpass bridge, we can see the roadway at all times, but music from the car's AM radio receiver may fade or fall silent briefly as we pass through the roadway opening into the steel-and-concrete structure of the overpass. Why? At roughly 20×20 meters, the roadway opening through the bridge is vast when measured in visible wavelengths—about 40 million λ in each direction (2000 per mm × 1000 (mm/m) × 40 meters)—and thus does not obstruct the light that passes through it. Broadcast AM radio signals, which are another form of light, flickers to match the pressure variations of the sound signals it carries, but the light has wavelengths much longer (about 1MHz or 300 meters) than visible light. The bridge's aperture is simply too small (approximately 0.07 λ) to admit much of that light, and only a faint evanescent residue propagates far enough into the opening to reach our radio receiver, reducing its output to nothing as we drive under the bridge. [2]

Electromagnetic waves, which are usually illustrated as pure, clean, and perfectly complementary orthogonal sine waves (Figure 2.7), are rarely so picture per-

[2] THIS FOOTNOTE ON THE EM SPECTRUM COULD BE REFORMATTED AS A SIDEBAR WITH FIGURE. The measurable electromagnetic spectrum is astonishingly vast, covering at least 20 factors of ten, or 66 octaves (for details, see [1]). Visible light covers less than one octave of that range. The electromagnetic spectrum begins with with magma disturbances, (2–8Hz) and ELF radio waves (40–80Hz) for submarine communication (sent from 28-mile-long antennas [21]) that can resonate by circling the entire earth. Wavelengths shrink from kilometers to meters as frequencies rise through broadcast radio and TV bands, up through centimeter lengths for microwaves with uses from home cooking to radar. As wavelengths shrink further to sub-millimeter lengths, diffraction by man-made structures and multipath radio behaviors mix with more familiar optical properties at larger scales. Continuing up the spectrum, even smaller millimeter-wave (or terahertz frequency) signals are now under active investigation for use in security screening and back-scatter imaging. As frequencies climb, only the tiniest structures diffract, from millimeters for infrared and visible wavelengths to molecular structures for soft and hard ultraviolet wavelengths. Then begins the vast span reaching toward soft X rays, hard X rays, and gamma rays, where wavelengths are less than 10 picometers (10^{-15} m), which is smaller than a single atom, with frequencies above 10^{+19} Hz.

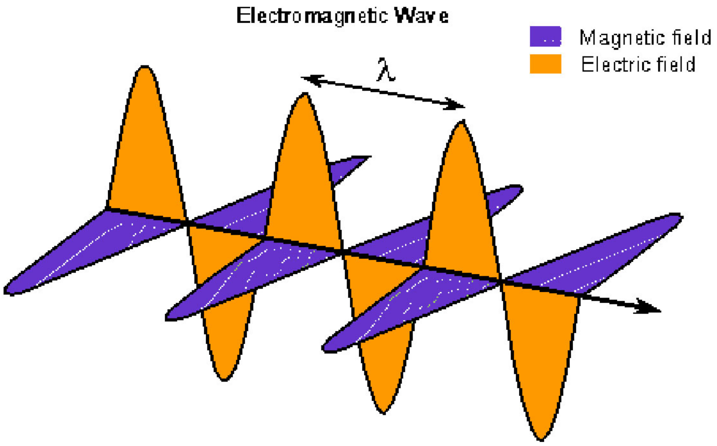


Figure 2.7: A linearly polarized, single-wavelength propagating electromagnetic wave, frozen at an instant in time. These sinusoids depict how electric and magnetic field strength and direction vary along a single ray-like propagation path through space, and how the orientation of these electric and magnetic field variations defines polarization. (Source: www.warren-wilson.edu/physics/physics2/Formal2000/sstephens/elecmag.gif. NOTE: an equally good source for a similar figure is jdsmith390.glogster.com/batescience-energy-and-waves.

fect or uniform, especially as they rattle and bounce their way through complex environments. Almost any electrical or magnetic disturbance will cause ragged and time-varying electromagnetic emanations, including flipping a light switch, chewing Wint-o-green flavored LifeSavers [**this example needs an explanation!**], firing a car's spark plugs, or moving a magnet, including a magnet spun in an electrical generator, tossed at a refrigerator door, or dropped on the floor.

Just as sounds can be decomposed into a weighted sum of pure sinusoidal tones, Fourier analysis permits us to decompose electromagnetic disturbances into a sum of individual sinusoidal waves such as those shown in Figure 2.7, each with its own distinct amplitude, frequency and phase. These decompositions let us describe the effects of materials on light in more detail than is possible by using rays alone. They form the essence of the methods used in Fourier optics to describe diffraction effects (see the classic text by Goodman [161]).

2.2.2 Light as Particles

Ray-based methods work well when they describe light propagation for large structures such as mirrors, lenses, apertures, or masks whose smallest features are larger than about 100λ. More complicated (and less intuitive) techniques of Fourier optics can provide accurate predictions of the interactions of light with much smaller structures, including holography, coherent laser light, interference effects produced by lens coatings and diffraction gratings, and structures whose size is similar to the wavelength of the light itself. Unlike ray-based models of lenses that suggest any lens could achieve unlimited resolution, Fourier-optics models of lenses accurately predict resolution limits caused by diffraction through lenses, coatings and apertures. These models also explain why larger lenses can achieve sharper images than smaller lenses, and why smaller apertures reduce resolution as they approach zero area.

Because the basic principles of computational photography are still being developed, current publications (including this book) use ray-based models to keep explorations mathematically straightforward. These models are occasionally augmented with simpler diffraction measures such as the Airy disk described in Section 2.3.6. We expect methods of Fourier optics to eventually supersede the ray-based models [305].

Photon and ray models of light propagation appear at first much simpler than electromagnetic waves. When entirely unobstructed, photons behave as individually distinguishable particles with curious properties. These fastest-possible particles have no rest mass, yet they transport energy as momentum in quantized packets, enabling sunlight to push solar sails on spacecraft [377] and lasers to nudge supercooled atoms together into clumps to form Bose-Einstein condensates [320, 321]. In the unobstructed vacuum of empty space or in uniform materials, photons follow straight-line paths and seem to pass through each other without any interaction. In confined spaces (e.g., pinhole apertures comparable to photon wavelengths) they diffract and behave as waves distributed across space, waves that interfere and combine to interact in ways impossible to describe by distinct particles. This wave phenomena was illustrated by the double-slit experiments of Huygens, G. I. Taylor, and others in 1909 [2].

In all cases the energy E transported by a photon is directly proportional to its wavelength in free space

$$E = hc/\lambda \tag{2.2}$$

where λ is the wavelength in meters, c is the speed of light in meters/sec, and h is Planck's constant (6.626×10^{-34} joules/sec, where 1 joule = 1 kilogram × meter^2/sec^2, or the work done by one watt of power applied for one second).

A single visible light photon transports a spectacularly tiny amount of energy: $(6.626 \times 34 \times 3.0 \times 10^8 / 500 \times 10^{-9})$ or about 0.4×10^{-18} joules. To get a better idea of just how tiny this is, and of the vast numbers of photons that stream outward from any visible object, consider the typical one-milliwatt output of an 'eye-safe' laser pointer used for video-projector presentations. If we spread that beam with a lens to cover an entire square meter on a diffuse-white sheet of poster board, the result is so dim that blinking the light on and off is barely discernible even in a dark room. But Planck's constant, one of the smallest of all known physical constants, reveals that this barely discernible illumination of 1 mW/m^2 consists of no less than (0.001 watt/0.4×10^{-18} joules) = 2.5×10^{15} photons/sec. Every millisecond, each square millimeter of that board (mm = inch/25.4) receives $2.5 \times 10^{15}/(10^3 \times 10^3 \times 10^3) =$ 2.5 million photons!

As we will see, this seemingly enormous number of photons is not enough to avoid photon-related noise and quantization errors in modern digital cameras, even though these errors are not apparent in human vision, even in the most extreme low-light conditions. As Hecht, Schlaer and Pirenne discovered in their pioneering experiments in 1942 [182], once our eyes are fully adapted to darkness we can detect a point-like flash of 510 nm (green) light from as few as nine or ten photons absorbed by rod cells. Our retina can almost count photons! Until recently, only photomultiplier tubes could match this performance, but as we will discuss in the chapter on sensors, even these quantum effects of light may be exploitable in future cameras by building on recent progress in solid-state single-photon detectors, such as the biologically inspired work by Memis and Mohseni et al. [269]

2.2.3 Rays of Visible Light

As light propagates into a camera or outward from a surface, its measurement is not as simple as we might suppose. Ray-based measurements of the strength of light that we can use for photographic purposes combine several factors: the amount of power transferred, the radiated directions, the area of real or imaginary surfaces, the wavelength of the light, and our ability to see those wavelengths. Let's examine each factor in turn.

Radiant Flux Φ Measures Power

The easiest-to-describe light source measurement is its radiant flux Φ, the total power carried away from a source by electromagnetic radiation. It is measured in SI units of watts, where 1 watt = 1 joule/sec. Radiant flux Φ includes all the light energy at all wavelengths leaving the source in all possible directions from all points on the surface of the light source. Don't assume radiant flux Φ describes the number of photons per second, because the energy carried by each photon is proportional to its frequency, as shown in Equation 2.2.

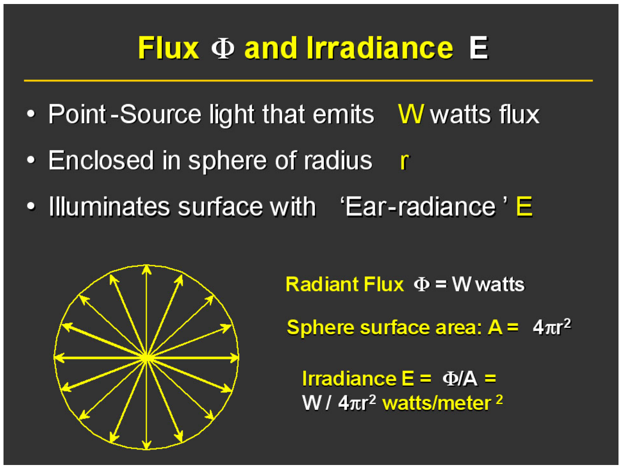
Definition 2.1. Radiant flux Φ measures the power transferred by light propagation, expressed in SI units of watts (1 watt = 1 joule/sec).

As most light sources are already small compared to the objects they illuminate, let's replace each one with a simple idealized point-source light, an infinitesimally small volume (a point) that radiates W watts of light power outward uniformly in all directions. To measure that light by using rays, we create an imaginary sphere of radius r meters centered at the source, and then apply a few intitive assertions, as shown in Figure 2.8:

- First, each ray from the point-source light leaves radially outward from the sphere's center, and arrives exactly perpendicular to the sphere's surface at the one point where that ray passes through the imaginary sphere.

- Second, the sphere's set of all surface points is interchangeable with the set of all rays: every possible ray defines a unique point on the sphere, and every point on the sphere defines a unique direction for a ray from the source.

- Third, the sphere's surface (or any fraction of it) contains an uncountable 2D infinity of points, and therefore the point-source light emits an uncountable 2D infinity of rays.



Flux Φ and Irradiance E

- Point-Source light that emits W watts flux
- Enclosed in sphere of radius r
- Illuminates surface with ‘Ear-radiance’ E

Radiant Flux Φ = W watts

Sphere surface area: A = $4\pi r^2$

Irradiance E = Φ/A =
W / $4\pi r^2$ watts/meter ^{2}

Figure 2.8: caption and source needed for this figure

- Fourth, as the set of all rays carries W watts, each individual ray carries only an infinitesimally small portion dW, which passes through the sphere at a point with infinitesimal area dA. Any measurable amount of power transferred from the point-source light must form a beam made by an uncountable 2D infinity of rays that pierce a measurable fraction of the sphere's surface area to form a measurable 2D angular extent.

Irradiance E Measures Power versus Area

Countable or not, the rays carry power from the point source to the sphere's surface, spreading the W watts of radiant flux from the point source uniformly across the sphere's entire surface area of $4\pi r^2$ meters^2. At the sphere's surface, the arriving light creates a spatial power density of $W/4\pi r^2$ watts/meter^2, a measurement of incident light called the *irradiance* E (note: we use the mnemonic “Ear-radiance” for E). Light leaving the imaginary sphere is measured in the same way, but is called the *radiant exitance* M of the surface.

Definition 2.2. Irradiance E measures the spatial power density at a point $d\Phi/dA$, defined as the radiant flux per unit surface area, and given in SI units as watts/meter^2.

Our sphere-and-point-source example is a special case that provides perfectly constant irradiance of $E = W/4\pi r^2$ watts/meter^2 at every point on the sphere surface. Clearly that irradiance will change by the square of distance: tripling r to $3r$ reduces E to $E/(3)^2 = E/9$. Thus point-source irradiance on a flat surface varies with position. As irradiance includes all the incident light on a surface from all sources and all directions, we also must add the contributions from any other light sources to find irradiance E for a surface.

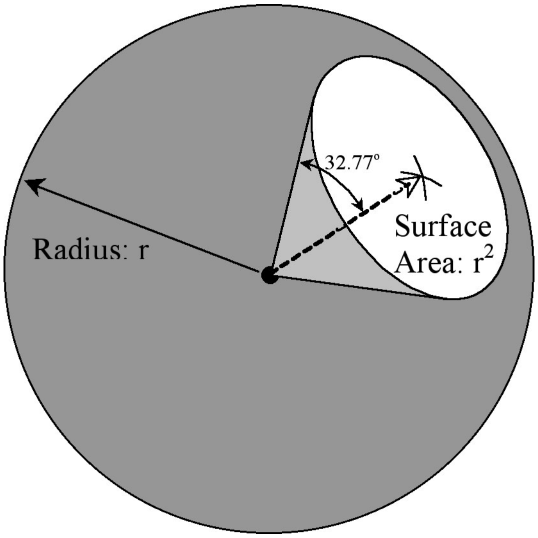


Figure 2.9: A cone-shaped beam of one steradian from a point-source light at the sphere's center illuminates an area of exactly r^2 meter^2 on a sphere of radius r meters. (Source: Jack Tumblin.)

Radiant Intensity *I* Measures Power versus Direction

Radiant intensity measures the angular power density of a light source, defined as radiant flux per unit solid angle. Let's suppose the point-source light at the center of our sphere is directional. Like a lighthouse, it emits all its radiant flux in a narrow beam of light that evenly illuminates a small circular spot on the enclosing imaginary sphere, a spot that covers (for example) just 1/100th of the sphere's surface area. Of course that spot is two dimensional, but simple angular measurements do not adequately describe its angular size. A searchlight beam that appears rectangular, measuring 20° horizontally and 30° vertically, doesn't cover exactly 600° squared because such 2D angle-products are nonuniform, especially for large beam angles. For example, what is the shape of a beam 240° wide and 260° high? The calculation of $240° \times 260°$ measures many directions more than once! In a similar example, what is the height and width of the nearly omnidirectional beam from an idealized light bulb that radiates light outward uniformly in all directions from a point at its center, but whose round, metal-capped base forms a cone-like shadow of about ±18°?

Measuring the area of a shadow or a beam on an enclosing sphere eliminates the inconsistencies and irregularities of these naive products of angles. Instead of measuring beam size by multiplying two 1D angles (e.g., 20° by 30°), we measure the area the beam illuminates on the surface of a sphere of radius $r = 1$, as shown in Figure 2.9. The sphere's curved surface complicates this calculation, but the result provides a simple and uniform measurement of any fraction of sphere coverage. This area measurement doesn't really depend on the size of the sphere or the units we use to measure its unit radius.

Figure 2.9 illustrates how a 1-meter^2 area on a 1-meter-radius sphere covers the same 2D span of angles as a 1-inch^2 area on a 1-inch-radius sphere. That coverage defines one *steradian* (or squared radian) of solid angle in derived-SI units. Lights that illuminate the entire sphere form the broadest possible beam, an omnidirectional beam that spans 4π steradians of solid angle (about 12.566 steradians). For the nominal one-steradian cone-shaped beam shown in Figure 2.9, the angle from the beam centerline to its edge is $\arccos(1 - 1/2\pi) = 32.77°$. Steradian measurements now allow us to describe the directional strength of the searchlight beam by its radiant intensity I.

Definition 2.3. Radiant intensity I measures the angular power density in one direction, $(d\Phi/d\omega)$, defined as the radiant flux per unit solid angle, and specified in SI units of watts/steradian.

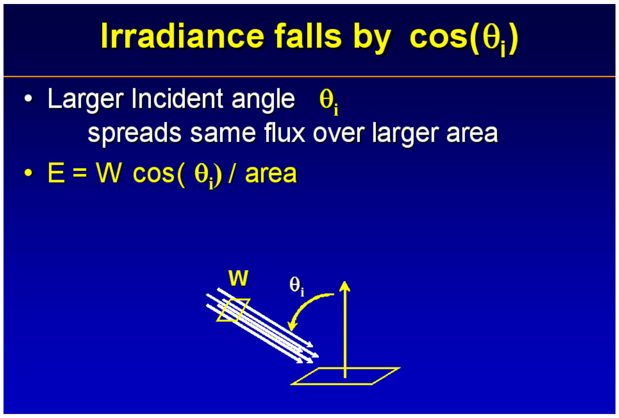
In our example of a narrow-beam searchlight, the W watts of radiant flux spread evenly over 1/100th of the whole sphere's 4π steradians yields a radiant intensity $I = W/(4\pi/100) = 25W/\pi = 7.96$ watts/steradian. Like irradiance E, the radiant intensity I measures power density and not power itself. Each and every ray in our idealized uniform searchlight beam has radiant intensity $I = 7.96$ watts/steradian, and zero watts/steradian for each and every other direction. (A more realistic beam would probably have radiant intensity that falls smoothly from the center ray to its outer edges, typically described by Gaussian beam profile functions.) The omnidirectional beam we use to define irradiance spreads the same W watts of radiant flux over the entire enclosing sphere, a full 4π steradians, trading beam intensity for angular coverage to supply $I = W/4\pi = 0.0796$ watts/steradian in all directions. As radiant intensity I measures only the directional strength of a light source, it does not vary with distance, unlike the irradiance E.

Illumination Angle Causes Falloff in Cosine Theta

The same point-source light and imaginary sphere we use to define irradiance E hides another measurable form of angular dependence. In addition to the sphere's constant distance r from the point source, the surface is exactly perpendicular to every point-source ray that pierces it. At every point on the sphere, the point-source light that illuminates it is always directly overhead.

In the same way that the sun heats the desert floor more vigorously at noon than at dusk, any irradiance on a flat surface will shink as the illuminator's angle of incidence grows. To illustrate, suppose we aim an idealized laser pointer at a flat surface, as shown in Figure 2.10. This fanciful idealized laser pointer emits exactly 1 mW of radiant flux as a bundle of perfectly uniform, perfectly parallel rays through a square aperture measuring exactly 5 mm on each side. For maximum irradiance E, we aim the laser perpendicular to a flat surface, where it will illuminate an area of 5 mm × 5 mm with spatial power density of $E = 10^{-3}$ watts / $(5 \times 10^{-3}$ meter$)^2$ = 40 watts/meter^2 at every illuminated surface point.

However, if we increase the laser pointer's angle of incidence from directly overhead ($\theta_i = 0$) to mid-afternoon or dusk-like angles, the same radiant flux illuminates a much larger surface area, namely 25 mm$^2/\cos\theta_i$. As that area grows, the irradiance decreases at each point within it. Irradiance E falls to 20 watts/meter^2 for



Irradiance falls by cos(θ_i)

- Larger Incident angle θ_i spreads same flux over larger area
- E = W cos(θ_i) / area

Figure 2.10: As incident illumination angle θ_i increases, irradiance E on surface area dA decreases by a factor of $1/\cos\theta_i$. (Source: Jack Tumblin.)

a late-afternoon angle of $\theta_i = 60°$, to 3.5 watts/meter^2 for the dusk-like angle of $\theta_i = 85°$, and to zero at the sunset angle of $\theta_i = 90°$ where all rays of light travel parallel to the surface and none irradiate it.

Equivalently, when the laser pointer is directly overhead, each small patch on the illuminated surface with area dA receives all the light that emerges from a dA-sized portion of the laser pointer's aperture. If we change the angle of incidence θ_i, the same illuminated surface area dA receives light from only a $\cos\theta_i \times dA$-sized portion of the laser pointer's aperture.

Equipped with this $\cos\theta_i$ angular falloff and radiant intensity I, we now have all we need to find the irradiance E at any point P from any combination of point-source lights. Any point source that emits W watts of light uniformly in all directions will have a radiant intensity of $I = W/4\pi$ watts/steradian in all directions, including the direction of point P. If the irradiance reaches point P with incidence angle θ_i, then that source adds an irradiance of $E = W\cos\theta_i/4\pi r^2$ watts/meter^2. This is written more simply as:

$$E = I\cos\theta_i/r^2. \tag{2.3}$$

Thus the radiant flux W is nice to know, but it often isn't necessary. All we really need is the radiant intensity I in the direction of point P along with its incidence angle θ_i, and from that we can find how much irradiance E the surface will get from a point source.

Radiance L: How We Measure Extended Sources

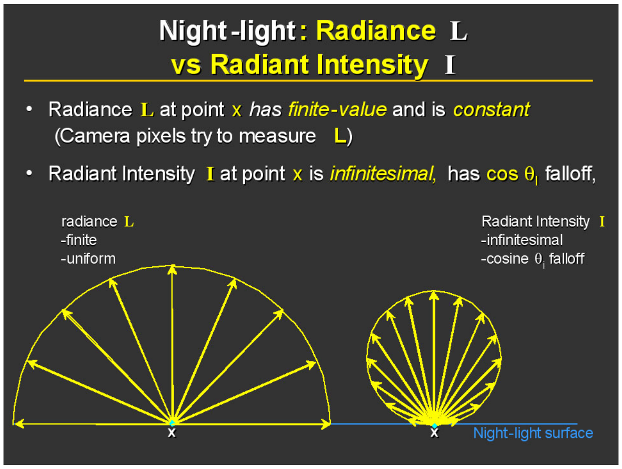
Almost all everyday light sources—from sunlight to campfires to long fluorescent tubes—are extended sources, with measurable shape and area, rather than point light sources. How do we measure the irradiance E of these extended light sources? If we wish, we could simplify the measurement of the emitted light by enclosing the extended source within an imaginary sphere, and then measure the light that passes through this sphere, rather than work with the complicated shape of the extended source. Unfortunately, with an extended source, the illumination of such a sphere is not simple. Instead of receiving a single ray of light, each surface point is pierced by its own varied bundle of rays that arrive from everywhere on the extended source's surface. Each bundle may have a different shape, and each ray in each bundle may have a different strength.

We need a new measurement to describe light from extended sources. The imaginary sphere's irradiance E at every point is not enough, because spatial power density E does not distinguish between differing amounts of light arriving from different directions. The radiant intensity I from the entire sphere is not enough, because angular power density I may be different at different points on the sphere. To describe those jointly directional and positional light strengths we must subdivide the spatial power density E by the angular power density I, or equivalently, we must subdivide the angular spatial power density by its positional power density. This power-density-density measurement is called *radiance* L, and it gives us a quantitative measurement of the strength of a single ray of light.

For example, suppose we wish to measure the output of a glowing blue-green night light, a simple 5-cm-diameter disk of electroluminescent materials. This uniform extended light source emits about 0.030 watts of radiant flux from its 0.001 meter^2 area [FOR A 5-CM-DIAMETER DISK THIS AREA SHOULD BE 0.00196 meter^2, OR ABOUT 0.002], and at every point on its surface the spatial power density (irradiance E, or equivalently the radiant exitance M) is $0.030/10^{-3} = 30$ watts/meter^2 [WITH THE CORRECT AREA THIS VALUE OF M WOULD BE ABOUT 15]. Its flat surface looks equally bright when viewed from any direction above the surface, a span of 2π steradians. When photographed with a digital camera, all pixel values in the night-light portion of the image are unchanged for any viewing direction and any viewing distance. [WE NEED TO FIX THE TINY MATH ERROR HERE TO BE CONSISTENT WITH NIGHT LIGHT COMPUTATIONS IN OTHER PARAGRAPHS]

Just how strong are the night-light's emissions from just one point and in just one direction? The radiant flux Φ must be infinitesimal (0.030 watts evenly distributed among a 2D uncountable infinity of surface points), but irradiance ($E = 0.030$ watts/0.001 meter^2) gives us a finite measure of the light strength sent out in all directions. If we could somehow measure the infinitesimal flux $d\Phi$ leaving just one point, wouldn't we find a uniform angular power density (radiant flux I, where $I = d\Phi/d\omega$) in all 2π directions? [DO YOU MEAN RADIANT INTENSITY I RATHER THAN RADIANT FLUX I ?]

Surprisingly, the answer is no! Suppose we view the night-light's emitting surface through a powerful telescope from 100 kilometers away. The telescope gathers light from every point on the surface as the light arrives from almost exactly the same direction. In the telescope's viewfinder we see an image of the night light, and if we ask an assistant to tilt the night light away from perpendicular to the telescope's viewing direction the night light will appear narrower due to foreshortening. Tilting the night light by $\theta_i = 60°$ does not obscure any of the night-light's surface points, but instead packs those points together into an area that decreases by $\cos\theta_i = 0.5$. If each of the night-light's surface points contributes the same amount of light to the telescope image (constant angular density I), then the tilted



Night-light: Radiance L vs Radiant Intensity I

- Radiance L at point x *has finite-value* and is *constant* (Camera pixels try to measure L)
- Radiant Intensity I at point x is *infinitesimal,* has cos θ_i falloff,

Figure 2.11: A glowing blue-green night light that appears uniformly bright in every direction does not have uniform angular power density I. Radiant intensity I leaving each point on the light's surface has a $\cos\theta_i$ falloff. The night-light's visually uniform quantity is not radiant intensity I, but radiance L, which is uniform in all directions. Radiance L describes the strength of the light ray, and radiant intensity I does not.

night light packs the same energy into half the area in the viewfinder image, and it must appear twice as bright, but obviously it does not.

Thus the radiant intensity I for each point on this extended source cannot be constant. It must fall by $\cos\theta_i$, as shown in Figure 2.11, to prevent brightening in the telescope image. Clearly the angular power density I alone does not describe what we see or what a camera captures from this surface! For that we need radiance L, which is constant in all directions from all points on the night-light's surface.

Radiance L Measures One Ray at One Surface

Radiance defines precisely the "ray strength" we think we already know intuitively, as it describes how much light leaves one point and arrives at another point along a straight line, and does not fade with distance. More importantly, radiance describes the physical quantity we try to measure with pixel values in an ideal camera with perfect lenses in perfect focus. It also describes the physical quantity supplied to us by a perfect pixel in a perfect digital display, whether printed on paper, viewed on a back-lit flat panel, or projected on a wall.

As a single ray carries only infinitesimal power,[3] radiance must instead measure

[3] We can illustrate every ray's infinitesimal power (radiant flux $d\Phi$) from either its destination or its source. **Destination:** Even the tiniest illuminated area A contains an uncountable 2D infinity of surface points. Distributing any finite Φ watts of radiant flux among those points ensures each point receives only an infinitesimal amount of power. As each point acts as the end of at least one ray from the light source, each ray can carry only an infinitesimal amount of power. **Source:** Even the narrowest beam from the tiniest light source (e.g., a laser) will spread apart with distance to span a finite solid angle. For example, most laser pointer beams are roughly

the ray's power *density* to provide nonzero values, and must measure that density in two different ways at once, combining angular power density (radiant intensity I in watts/steradian) with spatial power density (irradiance E in watts/meter^2).

The spatial power density component causes trouble, because we cannot measure it for a ray without defining a surface where we can measure area. Any surface will do, either real or imaginary, but we must know the ray's incidence angle θ_i at the surface. The simplest choice, a surface perpendicular to the ray, with $\theta_i = 0$, gives the simplest definition of radiance: the angular power density of the ray ($I = d\Phi/d\omega$), in watts/steradian, divided by the infinitesimal surface area dA where the ray delivers that radiant flux, or $L = d^2\Phi/(d\omega \cdot dA_\perp)$.

If we tilt the surface dA by ω_i, the same ray will cause a smaller spatial power density ($d\Phi/dA$) at the surface, as shown in Figure 2.10, even though the ray strength itself did not change. To compensate for this unwanted change, the definition of radiance contains a division by $\cos\theta_i$ to remove the surface-angle dependence. This inverse cosine term increases as we tilt the surface. It is the inverse of the cosine falloff used to compute irradiance E on the surface, so that radiance measures the ray's total power-carrying capacity, and not any tilt-induced spreading across the surface where the ray arrives. With this inverse term, the radiance from every point on our night light is constant in all directions, while radiant intensity I is not, as shown in Figure 2.11.

Definition 2.4. Radiance L measures ray strength. It measures combined angular and spatial power density delivered along a ray to a perpendicular surface ($d^2\Phi/dA_\perp \cdot d\omega$), with SI units of watts/(meter^2 steradian). Radiance for a ray that meets a non-perpendicular surface does not change with surface incidence angle θ_i because we cancel its reduction to spatial power density ($d\Phi/dA$) by using ($d^2\Phi/dA_\perp \cdot d\omega \cdot \cos\theta_i$).

For example, giant advertising searchlights that sweep the night sky create narrow rays of extremely high radiance. Larger units such as the USA Searchlights Inc. Astro 7000 [3] can emit 7000 watts of radiant flux Φ perpendicularly through an 18-inch diameter aperture (0.164 meter^2), and achieve a searing average irradiance E (or radiant exitance M) of about ($7000/0.164 = 42,600$ watts/meter^2) through the aperture's front surface. Its narrow cone-shaped beam (about 1° half-angle) spans only about 10^{-3} steradians, for an average radiant intensity I of $7000/10^{-3}$ or 7 million watts/steradian. The radiance for a single ray at the center of a uniform beam is $L = 7000$ watts/(0.164 meter$^2 \cdot 10^{-3}$ steradians), or about 42.6 million watts/(meter^2 steradian).

If we choose an imaginary measurement surface tilted across the beam at an angle $\theta_i = 60°$, we would spread the same 7000-watt beam across an area twice as large as the searchlight's 0.164 meter^2 aperture, but the $\cos\theta_i$ term in our definition of radiance would cancel this doubling, resulting in the same 42.6 million watts/(meter^2 steradian) radiance value for the same ray, even though we measured the beam with a different imaginary surface.

cone-shaped with about 1 milliradian (0.06°) half-angles. Even the smallest solid angle ω contains an uncountable 2D infinity of ray directions. Evenly distributing any finite Φ watts among these directions ensures that each ray carries only an infinitesimal amount of power.

Finding radiance for our earlier example of a night light is only slightly more complicated, but requires us to use this useful SI-units identity for any uniform diffuse emitter or reflector [102]:

$$\Phi = MA = LA\pi \tag{2.4}$$

The first portion is trivial. If we multiply irradiance E or radiant exitance M (watts/meter^2, or irradiance in reverse) by the area A we will, of course, find radiant flux Φ (watts). The identity's second portion solves the problem. Diffuse sources such as a night light supply constant radiance L at all points in all directions. Integrate L over the hemisphere of directions ω to show that its value is just M/π. For the night light, $\Phi/(A\pi) = 0.030$ watts/(0.001 meter^2 steradian)$\pi =$ 9.55 watts/meter^2 steradian. [THE PREVIOUS TINY MATH ERROR IN NIGHT LIGHT AREA AFFECTS THIS CALCULATION]

For a simpler example, let's find the sun's radiance L along a ray of light from the surface of the earth to the sun at noon. Near the earth's equator with the sun directly overhead on a clear day, the sun supplies approximately $E = 1000$ watts/meter^2 irradiance. The sun forms a disk in the sky, and each point on the ground receives its sunlight in a narrow cone-shaped beam with a half-angle of about 0.00465 radians (or 0.266°). Assume the sun disk is uniform, and the beam covers $\pi(0.0465)^2 = 6.78 \times 10^{-5}$ steradians [SHOULD THE AREA HERE BE 6.78 $\times 10^{-3}$ STERADIANS?]. Noonday sun radiance is then about $L = 10^3$ watts/$(6.78 \times 10^{-5}$ meter^2 steradian$)$ = 14.75 million watts/(meter^2 steradian).

Now suppose clouds arrive and merge to form a solid uniform gray overcast that hides the sun and looks the same in all directions, causing a noon-time irradiance of $E = 200$ watts/meter^2. What is the radiance for a vertical ray, one exactly perpendicular to the ground? For a ray with an incidence angle of 60°? For a ray parallel to the ground?

The answer is trickier than you might expect, because these problems encourage us to confuse the sky's uniform radiance L with radiant intensity I for points on the ground. As the sky looks equally bright in all directions, the assumption of constant angular power density leads us to answers that are simple, straightforward, and wrong. As a sanity check, remember eyes and cameras estimate radiance L well and angular power density poorly or not at all. Just as every point on the night light emits finite uniform radiance L in all directions, but an infinitesimal cosine-weighted radiant intensity I, the radiant intensity I received by each point on the ground is also infinitesimal and nonuniform. Both exhibit the $\cos\theta_i$ falloff illustrated in Figure 2.11.

Why? At first, you might choose to believe that a single point, one that is isolated and entirely unobstructed, would receive uniform (yet infinitesimal) power from all sky directions. Much like a point-source light run in reverse (absorbing rather than emitting light), this single point receives infinitesimal flux, but with seemingly uniform angular power density I. You might then surmise that points within a surface must suffer from partial occlusion by their neighbors to reduce the angular power density to the characteristic $\cos\theta_i$, but rigorous proof would force you to grapple with several ugly infinitesimal conundrums at once. Just how much occlusion does one point impose on another? How can we combine occlusions of multiple points? How can we extend that to an uncountable 2D infinity of points?

Ray measurements with radiance give us a better explanation of the cosine falloff of the radiant intensity I at a surface. Instead of examining dubious cases of infinitesimal occlusion at a point, the uniformity of the overcast sky should convince you that each and every ray it sends to the ground around us has the same strength, i.e., the same radiance value L. If that ray arrives perpendicular to a surface, it delivers its infinitesimal amount of radiant flux $d\Phi$ concentrated within the infinitesimal area dA. If it arrives at incidence angle θ_i that same flux gets spread across a larger area $dA/\cos\theta_i$. Now consider each point within that larger (but still infinitesimal) area. The flux falls by $\cos\theta_i$ for this ray. By induction, this same cosine falloff applies to all other rays that supply the point with flux from the sky. Despite constant radiance L for each ray, the received flux of the point varies with angle for those rays, imposing a cosine falloff on the radiant intensity $I = d\Phi_{sky}\cos\theta_i/d\omega$. Thus constant radiance from all directions causes cosine-weighted radiant intensity at each point on a surface.

But how can we find the constant radiance value L? We could integrate those radiant flux contributions across the entire hemisphere of the sky to determine how much radiance L is necessary to cause the given irradiance of $E = 200$ watts/meter^2, but we already have an easier way. As an overcast sky illuminates the ground with constant radiance L over the entire hemisphere of directions, it is a reversed version of our night light—a diffuse receiver instead of a diffuse emitter. The handy diffuse identity in Equation 2.4 side steps the messy integration: $EA = 200$ watts/meter$^2 \cdot A = LA\pi$, and thus $L = 200/\pi = 63.7$ watts/(meter^2 steradian) for sky rays from overhead, from $\theta_i = 60°$, and from any other direction, even parallel to the ground.

Extended light sources can dramatically reduce radiance without reducing the power they supply to illuminated surfaces (irradiance E). Without the dazzling sun disk, overcast skies reduce spatial power density (irradiance) on the ground by a modest factor of only 5, but the radiance from the sun's direction shrinks from 14.75 million watt/(meter^2 steradian) to 63.7—a factor of about $231,000 : 1$! This change isn't a simple result of the increased angular extent of our light source either. The light source grows from a tiny 6.78×10^{-5} steradian [AGAIN, SHOULD THIS BE 6.78 $\times 10^{-3}$ STERADIAN?] sun disk to the entire sky covering 4π [THIS NUMBER SHOULD BE 2π FOR A HEMISPHERE] steradian, a growth of about 185,000:1. [SHOULD THIS COMPUTATION BE 2π / 6.78 $\times 10^{-3}$, OR ABOUT 1000:1 ?] Both the strength and the direction of arriving light affects the power received at a surface, and their aggregate requires us to integrate radiance values.[4]

[4] Angular power density measurements I also reveal why the idealized 1 mW laser pointer of Figure 2.10 cannot exist. If each one of its perfectly parallel rays arrives perpendicular to a surface to create a uniform 5 mm × 5 mm illuminated spot, what is the radiant intensity I arriving at each illuminated point? What is the radiance for each ray? Answer: Every surface point receives just one ray, a beam of light that covers zero steradians. Thus the angular power density ($I = d\Phi/d\omega$) and the radiance ($L = d\Phi/(dA\, d\omega)$) of an idealized laser pointer both reach infinity! Real laser beams actually spread slightly with distance. Suppose we replace each parallel ray arriving at the 5 mm × 5 mm illuminated surface spot with an extremely narrow, uniform, cone-shaped beam of rays, with a half-angle of 0.8 milliradians (0.8×10^{-3} radians, a typical beam divergence). What is the radiant intensity I and the radiance L for the rays at the surface? Answer: each cone-shaped beam covers $\pi\cdot$ $(0.8\ \times 10^{-3})^2$ steradians, so $L = 0.5$ watts/ $(25 \times 10^{-3}$ meter$^2 \cdot \pi \cdot 0.64 \times 10^{-6}$

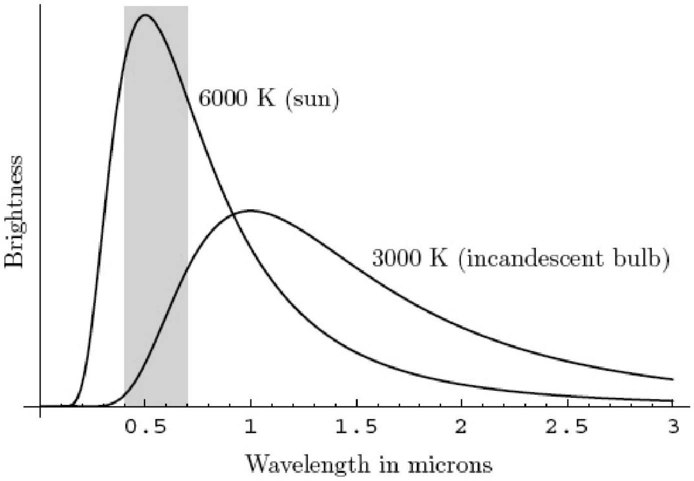


Figure 2.12: Visible light wavelengths straddle the peak of the mid-day sun's spectral distribution, but typical household incandescent bulbs emit most of their power at invisible infrared wavelengths around 1000 nm. Less than 2% of a bulb's light output is visible to humans. [THIS 2% FACT SHOULD BE EXPLAINED AND JUSTIFIED IN THE TEXT AS WELL AS GIVEN IN THE CAPTION. ALSO, THE VERTICAL AXIS LABEL IN THE FIGURE NEEDS TO BE CHANGED TO SPECTRAL RADIANT FLUX.]

2.2.4 Light Bulb Ratings and Photometry

Now that we have these four descriptors of light sources and destinations (radiant flux Φ, radiant intensity I, irradiance E and radiance L), you might think that at last we are fully equipped to describe the behavior of an ordinary light bulb. Surprisingly, even simple radiant flux measurements are often difficult to find because most lamps are labeled by their electrical power consumption or other ratings, and not by these four descriptors.

For example, an ordinary household 60-watt incandescent light bulb expends about 15 watts on heat conducted through its screw-socket base, and emits most of its 45-watt radiant flux at near-infrared wavelengths. Only about 2 watts leave the bulb at visible wavelengths. Nearly all incandescent bulbs, even halogens, emit light from their electrically heated tungsten filaments with a spectrum close to a black-body radiator between 2600 and 2800°. Wein's displacement law [4] shows that emissions peak at invisible infrared wavelengths around 1000 nm. A plot of relative power versus wavelength for an incandescent bulb shows a steeply tilted spectrum that expends most of its power in the longer (reddish) wavelengths (as shown in Figure 2.12). It is well known that the sensitivity of the human eye to light varies with wavelength, as shown in Figure 2.13, so we cannot see light in these longer wavelengths. If this same 2-watt [DO YOU MEAN 45-WATT?] radiant flux

steradian) = 0.5 watts/ $(25\pi \cdot 0.64 \times 10^{-9}$ meter^2 steradian) = 9.94×10^6 watts/(meter^2 steradian) radiance—a giant power density from a 1 mW light source!)

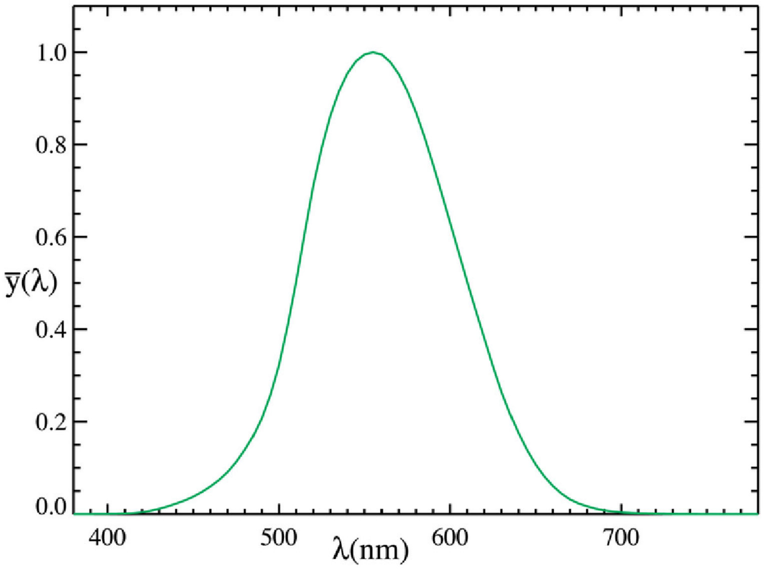


Figure 2.13: The sensitivity of the human eye to light varies with wavelength. The 1924 CIE photopic luminous efficiency curve (included in the 1931 CIE color standards) approximates how the light-sensing abilities of the eye vary with wavelength. Metering devices multiply spectral radiant flux by this curve (often achieved by an optical filter atop a silicon detector) and integrate the result to find the photometric intensity.

was uniformly distributed across all visible wavelengths, the light would appear much brighter.

Spectral Distributions

Spectral curves such as Figure 2.12 intuitively make good sense, but their measurement units can be confusing or misleading. Formally, the vertical axis shows spectral radiant flux, given by $d\Phi/d\lambda$, versus wavelength λ on the horizontal axis. Ian Ashdown described spectral radiant flux units concisely as "radiant flux per unit wavelength interval, at wavelength λ" [45]. Figure 2.12 does indeed show how the 45-watt radiant flux of a light bulb is spread smoothly across a wide band of wavelengths, with large amounts of power at wavelengths around 1000 nm and less power at shorter or longer wavelengths. We can find the radiant flux emitted within any chosen range of wavelengths by finding the area under the curve in that range. This integration confirms that the spectral radiant flux is the derivative of radiant flux with respect to wavelength, just as E and I are derivatives of radiant flux with respect to area and solid angle, respectively.

Do not confuse the height of this spectral curve at a single wavelength λ with the amount of power transmitted at that wavelength; that power is infinitesimal. Similarly, the SI units for spectral radiant flux are $d\Phi/d\lambda = d(\text{watts})/d(\text{meter})$, and not watts/meter. Like many other spectral plots, Figure 2.12 shows no absolute units, but only the correctly shaped curve. We can find the scale factor that converts this plot to units of absolute spectral radiant flux (e.g., differential units of watts/nm) for our light bulb by dividing 45 watts of radiant flux by the area under the curve.

Wavelength and Photometry

Radiometric measurements such as radiant flux Φ, radiant intensity I, and irradiance E are still not enough to measure the apparent visual strength of light; we must resort to *photometric methods* instead. If we measure the radiant flux only within the narrow 380–780 nm span of wavelengths that includes all visible light, we will still not get a reliable estimate of visual strength, because human sensitivity to light varies strongly with wavelength, as Figure 2.13 shows for wavelengths from 400 to 700 nm. This figure clearly describes how human perception of light at different wavelengths varies dramatically across the visible spectrum. Accordingly, to compensate, photometric measurements must apply a wavelength-dependent weighting function to radiant flux in order to create a more accurate estimate of the perceived visual impact of light measurements.

The curve shown in Figure 2.13 was originally determined in 1924, and was later incorporated into the 1931 color-measurement standard developed by the Commission Internationale d'Eclairage (CIE), or International Committee on Illumination. As the simplest and most widely used weighting function, the CIE standard statistically summarizes a set of experiments on more than a hundred (very patient) test subjects, who were asked to match the perceived brightness of light at different wavelengths. The resulting daylight-vision luminous efficiency curve, or *luminosity curve*, assigns a peak weight of 1.0 at the chosen standard for the most sensitive light frequency, which is 540×10^{12} Hz (540 THz), or a wavelength of 555.555 nm. At other wavelengths, the curve assigns lesser weights that fall toward zero in a Gaussian-like curve fitted to assessments by subjects in the tests. Later measurements have produced more refined results, since human spectral sensitivity varies considerably between night vision (scotopic) and daylight vision (photopic) [5]. The original 1931 CIE luminosity curve is still widely used to define photometric units and calibrate metering devices.

Photometric quantities differ dramatically from radiometric quantities, even when limited to visible light wavelengths. For example, a 1 mW green laser pointer (532 nm) will create a spot on a white sheet of paper that appears much brighter than that of an otherwise identical 1 mW red laser pointer (660 nm), despite their identical radiant flux. The weights on the CIE luminous efficiency curve [6] indicate that we must boost the output power (radiant flux) of the red laser pointer by a factor of $0.883/0.061 = 14.48$ to match the visual strength of the green laser pointer.

SI units provide a simple and orderly scheme for both radiometric and photometric measurements, but these names and units evolved through a lengthy and contentious historical muddle. [207] Since radiant flux Φ, in units of watts, is no longer appropriate when we apply the photometric efficiency function of Figure 2.13, we can measure these wavelength-weighted power units in lumens instead of modified existing power units (e.g.,"visible watts"). The SI standard-making body retained lumen and candela units in part to honor diverse historic candle- and gas-lantern-based light measurement systems[5]

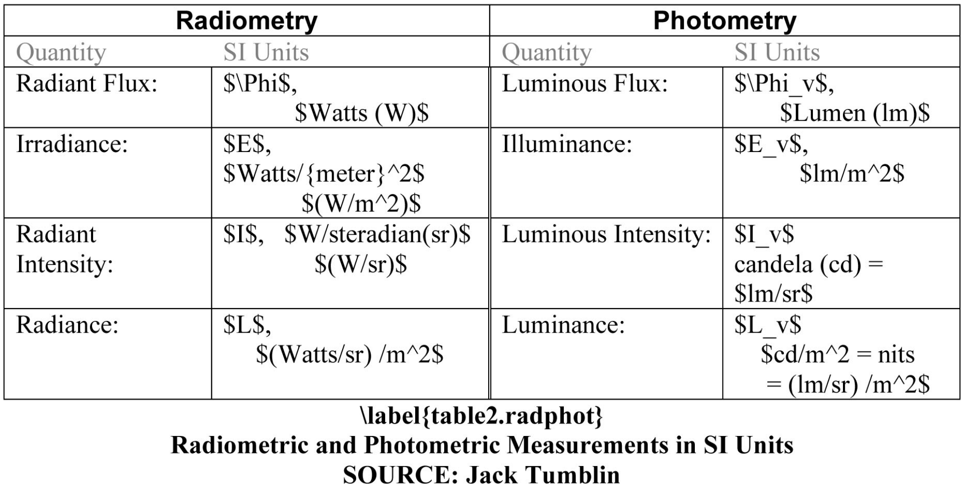
A single-wavelength laser emits exactly one lumen of visible light if it provides 1/683 watts (1.464 mW) of radiant flux at the wavelength of perfect (100%) photopic luminous efficiency, or 540×10^{12} Hertz ($\lambda = 555.55$ nm). To measure light power in lumens, we weight the radiant flux in watts by the luminous efficiency curve (Figure 2.13) and then multiply by 683. From our examples above, the 1 mW red laser pointer (660 nm) emits (0.001 watt radiant flux · 0.061 weight · 683 = 0.0417 lumens), while the green laser pointer (532 nm) emits 0.603 lumens. A typical 60-watt incandescent bulb is advertised to emit 900 lumens, which indicates that its radiant flux after weighting is only (900/683)= 1.32 "visible watts."

The SI system of units neatly generalizes our four radiometric measurements into four similarly named photometric measurements. In each measurement, it replaces radiant flux (watts) with luminous flux (lumens), but otherwise keeps the measurements unchanged, and assigns similar names as shown in Table 2.1.

Some published light-source specifications use either the term "efficiency" or the term "efficacy" to summarize the spectral weighting necessary for conversion from radiometric to photopic [PHOTOMETRIC ?] units. Values for luminous efficiency are unitless, between 0% and 100%. They specify the fraction of radiant flux included in the photometric flux, or luminous flux, measurement. Only a perfect 540 THz (555.555 nm) green source emits light with 100% luminous efficiency, and we get reduced values of 88.3% and 6.1% for our 532 nm green and 660 nm red laser pointers, respectively.albedo of

The most efficient light sources appear strongly colored. Light that appears

[5] SI units use candelas (1 cd = 1 lumen/steradian), instead of lumens, as one of its seven base units, but we explain lumens first for simplicity.



Radiometry		Photometry	
Quantity	SI Units	Quantity	SI Units
Radiant Flux:	Φ, $Watts (W)$	Luminous Flux:	Φ_v, $Lumen (lm)$
Irradiance:	E, $Watts/{meter}^2$ (W/m^2)	Illuminance:	E_v, lm/m^2
Radiant Intensity:	I, $W/steradian(sr)$ (W/sr)	Luminous Intensity:	I_v candela (cd) = lm/sr
Radiance:	L, $(Watts/sr) /m^2$	Luminance:	L_v $cd/m^2 = nits = (lm/sr) /m^2$

\label{table2.radphot}
Radiometric and Photometric Measurements in SI Units
SOURCE: Jack Tumblin

Table 2.1: Table of radiometry and photometry measurements (temporary version).

white is less efficient because it must include emissions at other less-efficient wavelengths within the 380–700 nm visible spectrum. The most efficient broadband white-light sources reach just 32% (the white-light spectrum resembles the visible portion of sunlight in the middle of the day). Artificial sources fall below that; up to 22% for the best LEDs, about 10% for compact fluorescent bulbs, and only 1.8% to 2.6% for incandescent bulbs.

The *luminous efficacy* gives the direct ratio between radiant flux (watts) and luminous flux (lumens). Only a perfect 540.0 THz laser light source can reach the absolute maximum luminous efficacy of 683 lumens/watt. Our green and red laser pointers at 603 and 41.7 lumens/watt show the strong dependence of luminous efficacy on wavelength. A white light whose spectrum exactly matches the luminous efficiency curve has an efficacy of about 240 lumens/watt, while a high-output 60-watt incandescent bulb achieves just 900/60 = 15 lumens/watt.

For an accessible but more thorough summary of light measurement, see Ian Ashdown's excellent online tutorial [7], adapted from his early book on radiosity rendering [44] and his useful resource webpage [8]. You can also find a particularly good summary of light measurement units, written by Steven Palmer, which is posted online [9].

2.2.5 Surface Transmission and Reflectance: Albedo, BRDF and Beyond

Now that we can measure the amounts of light leaving from a surface and arriving at a surface, their ratios can define surface reflectance and transmission. The simplest reflectance ratio, known as albedo ρ, has at least three definitions. Astronomers use the ratio of reflected to incident radiant flux ($1 \geq \Phi_r/\Phi_i \geq 0$) as *Bond Albedo* to describe planetary bodies. Some examples are brilliant foggy Venus at 0.75, midrange earth at 0.38, and our dusty-charcoal-like moon at 0.123. The more refined term *geometric albedo* describes Φ_r/Φ_i for light traveling perpendicular to the measured surface, or by incidence angles specified as *phase.* The *diffuse albedo* in physics and computer graphics sometimes describes scattering phenomena found in fog and smoke, but more commonly measures only the diffuse reflectance of a surface, defined as its uniform non-directional component.

A purely diffuse surface reflects uniform radiance L in all directions for any illumination, and the radiance is directly proportional to total irradiance E. Modifying the diffuse rule of Equation 2.4 to include albedo ρ reveals how albedo connects irradiance values E arriving at a surface to the radiance values L sent outward.

$$\Phi_r = E_i A\rho = \pi L_r A\rho \tag{2.5}$$

As you might expect, the unitless albedo ρ is a fraction between 0 (perfect absorber) and 1 (lossless diffuse reflector). Converting incoming irradiance E_i to outgoing radiance L_r includes the value π to account for distribution over the entire hemisphere of outgoing directions.

$$L_r = \rho E_i/\pi \tag{2.6}$$

in watts/(meter^2 steradian). For visible light measurements, we simply replace these radiometric units with photometric units. Albedo ρ remains unitless, radiance L_r in watts/(meter^2 steradian) becomes luminance L_{rv} in lumens/(meter^2 steradian) = candela/meter^2, and irradiance E_i in watts/meter^2 becomes illuminance E_{iv} in lumens/meter^2.

Albedo and its dependence on wavelength λ is a good first approximation for the appearance of many surfaces, but it lacks precision; few materials are truly diffuse and opaque. Commonplace materials such as paper, matte-finish plastics, non-glossy paint, soil, cloth and skin are approximately diffuse reflectors, but their reflectance changes for different illumination directions and different viewing directions. To measure reflectance for a single point x on a surface made of any material, we must specify what fraction of every possible incoming ray will get reflected in the direction of each and every possible outgoing ray. We cannot restrict ourselves to the hemisphere of ray directions above the surface, because some fraction of the incoming ray may emerge below the surface in a new direction, as in a lens (this is Snell's law, shown in Figure 2.16). For completeness, this new directional function must specify the coupling fraction of outgoing versus incoming light for the entire sphere of possible directions.

Definition 2.5. The Bidirectional Reflectance Distribution Function (BRDF), measured at a single point x on a surface, is the ratio of outgoing radiance L_i to incoming irradiance E_i, where irradiance is received from just one direction θ_i, ϕ_i. The BRDF is outgoing radiance L_r from surface point x toward a given direction θ_r, ϕ_r, divided by incoming irradiance E_i at surface point x from a given direction θ_i, ϕ_i,

$$BRDF = f_r(x, \lambda, \theta_i, \phi_i, \theta_r, \phi_r) = dL_r/dE_i \tag{2.7}$$

Photometric BRDF measurements are identical except for weighting by the luminosity function and replacement of watts by lumens. It is the ratio of outgoing luminance L_{iv} (cd/m^2) to incoming illuminance E_{iv} received from just one direction θ_i, ϕ_i.

BRDF, which describes what portion of the incident light from one direction will leave in another given direction, holds a few surprises among its important properties. First, incoming and outgoing directions (θ_i, ϕ_i) and (θ_r, ϕ_r) cover all directions both above and below the surface, thus enabling BRDF to describe reflection and transmission combined. Second, the BRDF function is symmetric for any and all materials. Its value is identical if we swap the incoming and outgoing directions. This long-confirmed property is known as Helmholtz Reciprocity.

Third, the BRDF f_r is a ratio of differential quantities that complicate intuition. It is not simply the ratio of incoming and outgoing radiance for two chosen rays, a ratio that would always fall between 0 and 1. It's the ratio of the outgoing ray's radiance L_r to the *irradiance E from only one direction.* If that surface is diffuse with albedo ρ, it spreads an incoming irradiance E from one direction (or any direction!) uniformly over a 2D infinity of exitance directions, producing an infinitesimal amount of radiant flux Φ along each ray. But radiance L_r describes the *density* of that flux, for a (one-sided, opaque surface) radiance of $L = \rho E \pi$. Thus a perfect diffuse reflector has a BRDF value of $1/\pi$ and not the value of 1.0 you might initially expect. Similarly, the BRDF of a perfect mirror irradiated from just one direction (e.g., a point-source light infinitely far away) is zero for all directions except for the mirror-reflection direction, where its value is infinite! Typical light absorption by real-world mirrors (where 80% to 99% of light is reflected) won't change this infinite BRDF, but their tiny surface imperfections (and diffraction limits) spread reflected rays into narrow beams of nonzero solid-angle, which produces BRDFs of finite values for even the best mirrors.

Fourth, the five input dimensions of the BRDF function (wavelength λ, illumination direction (θ_i, ϕ_i), and outgoing direction (θ_r, ϕ_r)) make empirical measurements vast and tedious to collect. Fortunately some high-quality BRDF databases are available online. To avoid the time and effort of gathering such measurements, computer graphics renderings resort to simpler directional models such as Phong shading (ambient, diffuse, and specular). They can also use a wealth of parameterized BRDF functions; some of which mathematically model material physics, such as He et al. [180] or LaFortune et al. [223], and others derived from or fitted to BRDF measurements such as Matusik [263].[6]

Scattering and BSSRDF

Current BRDF-based models used in computer graphics can create convincing renditions of metallic, transparent, and opaque materials, but they don't work well for partially translucent surfaces and materials such as carved marble, milk, and human skin. These materials fare poorly because the BRDF describes ray changes only at the surface of the material and not within the material itself. The BRDF model finds ray strengths leaving a point solely from the ray strengths that arrived

[6] For a brief survey of BRDF models, see [10].

there, and nowhere else, as if the material were either opaque or transparent. However, nearly all non-metallic materials admit some light below the surface. For each incident ray, some fraction of the light passes through, and chaotic scattering and absorption spread the light throughout a local volume. At the surface, some of the scattered and redirected light escapes from a small, glowing neighborhood around the point x where the light ray entered. You can see this scattering and absorption yourself with a laser pointer. In a dark room, aim the pointer at your hand; you will see a bright spot surrounded by a broad, dimmer reddish region caused by light scattering underneath the skin surface.

The aggregate of these ray paths largely determines the visual appearance of any material, but this appearance is quite difficult to describe without statistical models. Early models such as Kubelka-Munk's 1931 mathematical model (see [190] for details) provide convenient and predictive analysis tools for paints, lacquers and thin coatings. A new class of dipole models introduced by Jensen et al. in 2001 [202] describes subsurface scattering for semi-translucent materials such as marble and human skin. The representation of this scattering is called the Bidirectional Sub-Surface Reflectance Distribution Function (BSSRDF). While BRDF models of skin appear chalky and rough, BSSRDF renderings show the spatial blending characteristic of the translucency of skin.[7] These models and their approximations have greatly improved the accuracy and visual appearance of computer graphics renderings, and have been widely adopted for human characters in motion pictures and video games. The contributions of Jensen and his team to the accurate rendering of skin were formally recognized when they were awarded a 2004 Technical Academy Award.[8]

Rays, whether synthetically generated by computer graphics or arriving from a real scene with real lighting, need lenses to bend them into bundles and form them into a focused image. The next section presents an adapted excerpt on lenses from Charles S. Johnson's book *Science for the Curious Photographer*, which describes the fundamentals, limitations, and tradeoffs inherent in lenses. The complete book, with more detail and historical context, can be found online.[9]

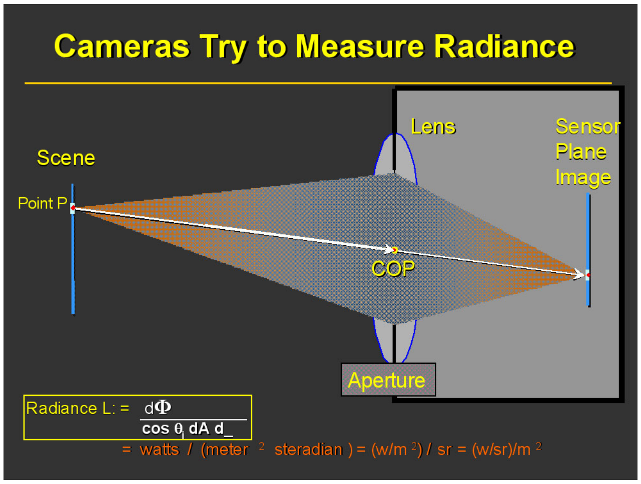
2.3 Ray Bending: Lenses, Apertures and Aberrations

Lenses bend rays. In the broadest sense, any ray-bending device qualifies as a lens, including mirrors, running water, and atmospheric anomalies. We can accurately

[7] Scattering may also help explain the surprisingly low contrasts of light reflected only from diffuse materials in a scene. These contrasts rarely exceed 50:1, with maximum diffuse albedo ρ no higher than about 0.92 for clean new snow and minimum diffuse albedo no lower than about 0.02 for finest-nap black velvet. The high-dynamic range captured in HDR photographs are the result of visible light sources, specular highlights, or lighting differences, and not albedo. These contrast limits changed recently when loosely packed synthetic carbon nanotube forests achieved a "super-black" albedo of 0.045 [11].

[8] A Technical Achievement Award (Academy Award) by the Academy of Motion Picture Arts and Sciences was given to Henrik Wann Jensen, Stephen R. Marschner, and Pat Hanrahan for "their pioneering research in simulating subsurface scattering of light in translucent materials."

[9] www.akpeters.com/product.asp?ProdCode=5817.



Cameras Try to Measure Radiance

Radiance L: = $\frac{d\Phi}{\cos\theta_i \, dA \, d\_}$

= watts / (meter ^{2} steradian) = (w/m ^{2}) / sr = (w/sr)/m ^{2}

Figure 2.14: Image formation. A camera lens bends rays from a scene into converging bundles that form a focused image on the camera's sensor plane.

describe nearly all optical devices and effects solely by the different ways they bend, split, or scatter rays. As BRDF and BSSRDF demonstrate, a good map that tells us how incoming rays get coupled to outgoing rays can describe the optical behavior of nearly anything, from lens coatings, diffraction gratings, optical fibers and LEDs to lasers, light sensors, meta-materials and even super-lattice or quantum-dot semiconductors. Rays describe *what* different materials and devices do to light, but explaining exactly *how* they do it may require anything from electromagnetic wave propagation to quantum electrodynamics.

Rays are sufficient for this book. We want to simplify and expand the definition of a lens to mean any device that bends incoming rays into useful patterns of outgoing rays. Later chapters in this book show how to combine computing with modified or non-traditional lenses for different or better ways to capture the appearance of a viewed scene.

To prepare for discussions to come, this section of the chapter briefly reviews conventional lenses that form images on a plane for film-like photography. Assembled from carefully shaped transparent discs mounted inside opaque tubes, these lenses bend rays from a scene into converging bundles that form a focused image on the camera's sensor plane (or *focal plane*) behind the lens, as shown in Figure 2.14. Rays reveal why the task of image formation is inherently difficult; the set of all possible lenses is uncountable and astoundingly vast. Plus we have only two methods to bend rays. We can change the refractive index of the lens or change the shape of the lens.

Rays define a dauntingly vast set of possible lenses, a continuum of at least nine dimensions with a unique lens behavior at every point. Each point on the 2D entry surface of the lens accepts rays from a 2D span (solid angle) of incoming directions, and each point on the 2D exit surface of the lens emits rays in a 2D span of outgoing directions. Plus the input-to-output coupling for each ray may (and usually does) vary with wavelength. As we learned from BRDF, the input-to-output transfer function of any lens is linear and bidirectional. Each input ray couples some fixed fraction of its power to each output ray, and that fraction does not change if we swap the names for the input and output rays. Like BRDF, a 'bidirectional lens distribution function' would describe the coupling between irradiance E and exiting radiance L, but it would measure coupling from each ray in the set of 4D incident rays to each ray in the set of 4D exiting rays. With wavelength included, this complete lens-describing function is 9-dimensional, and the set of all such functions fills a 9-dimensional space. Each point in this space defines one lens behavior, regardless how the lens was made.

Any change in the lens changes its describing function. It moves us to a different point within the 9-dimensional space, whether that change is caused by adjusting focus and zoom, a lens-surface scratch, a bent lens mount, or a water droplet from condensation. Inside this exceedingly vast set of behaviors, the set of lenses that form accurate, well-focused images all fall on a 3D manifold, as shown by Levin et al. [229] Most of these ideal lens designs cannot be manufactured because conventional lens-making techniques require many tradeoffs. What we can choose is lens size, the radius of curvature for each surface (almost all lenses have spherical surfaces), the materials and coatings for the lens, and the number, spacing, and sequence of lens elements.

Instead of measuring errors in a 9-dimensional space of ray mapping functions, traditional methods assess the image-forming abilities of a lens more directly—from the images they form. Common assessments of lens quality determine at least five competing image-formation errors, or aberrations, by measuring specific distances in the scene, the lens, and the image it forms. We summarize these terms below, and briefly explore how they guide high-quality multi-element lens designs for existing cameras.

2.3.1 Ray Bending: Snell's Law

All the basic rules for ray bending arise from general laws of physics. These rules are consistent with Maxwell's theory of electromagnetism as well as the quantum theory of photons and electrons. Richard Feynman's beautifully concise book *QED* [135] [391] [170] provides a wholly intuitive introduction to these deeper topics for interested readers.

Light rays bend at the boundaries between materials. As unobstructed light enters transparent materials such as air, water or glass, we approximate its behavior by claiming its propagation slows by the refractive index η, from c to c/η.[10] [11]

[10] More precisely, it is the phase velocity that changes by η, not the wave's propagation velocity. For some materials at X-ray wavelengths $\eta < 1.0$ but still do not violate relativity's 'speed limit' of c.

[11] According to the Theory of Relativity the speed of light is a constant in the universe, so how can light have a reduced speed in matter such as water or glass? It turns out that the index of refraction can be less than one and even less than zero. In Richard Feynman's interpretation, the speed of light always has the value c, but when traveling through matter its electric field isn't isolated. It disturbs all the electrons and protons in all the atoms, and each of them contributes to the electric field of the propagating light. The field we measure is the aggregate of all these uncountably numerous light-speed interactions. Their constructive and destructive combination

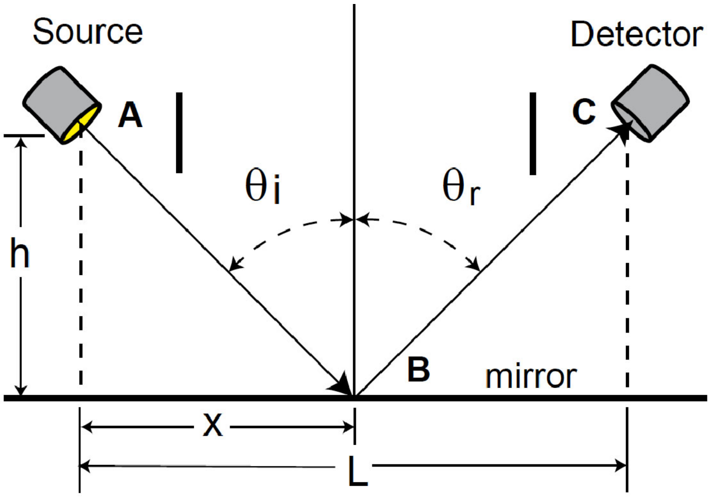


Figure 2.15: The angle of incidence equals the angle of reflection.

This apparent slowing causes propagating electromagnetic waves to change their local aggregate phase and direction at smooth boundaries, where some of their energy is reflected and some of their energy is transmitted into the new material in a new direction. In materials with smoothly varying refractive indices, light propagation can even follow curved paths. For example, practical graded-index materials have recently found widespread use in optical fibers and in seam-free replacements for bifocal eyeglasses. Such changes in direction are easy to describe by using rays. These descriptions are quite accurate for smooth surfaces, where features and variations are much larger than the light's wavelengths.

The refractive index of a material is a unit-free scale factor, defined as $\eta = 1.0$ for a perfect vacuum. It grows larger for denser materials that, in effect, slow down the light. At visible light wavelengths η is about 1.00029 for air. This value varies slightly with atmospheric temperature and pressure.[12] It's about 1.3 for water, 1.4 to 1.6 for various kinds of glass, 2.4 for diamond, and 4.0 for silicon, the raw material of CMOS detectors. The refractive index isn't constant either; it usually depends strongly on wavelength, temperature, pressure and other changes in material properties. For example, water at 25°C has a refractive index of 1.33, which slows the apparent speed of light in water to $c/\eta = 2.253 \times 10^8$ meter/sec. Refractive index measurements for almost any optically interesting material are readily available online. These mathematical models or measurement tables describe how η varies with wavelength, temperature, pressure and nearly any kind of disturbance.

matches well with the phase of waves described by seemingly 'slower' or 'faster' light given by c/η. The approximation is a good one for optics, because only the phase of these waves affects lens design, and not the time it takes for light to pass through the glass.

[12] The changes in refractive index in air and water vapor explain mirages, rainbows, and a wide range of atmospheric optical wonders. For a fascinating in-depth exploration see Marcel Minnaert's classic book entitled *Light and Color in the Outdoors* (1974, translated, republished in 1993).

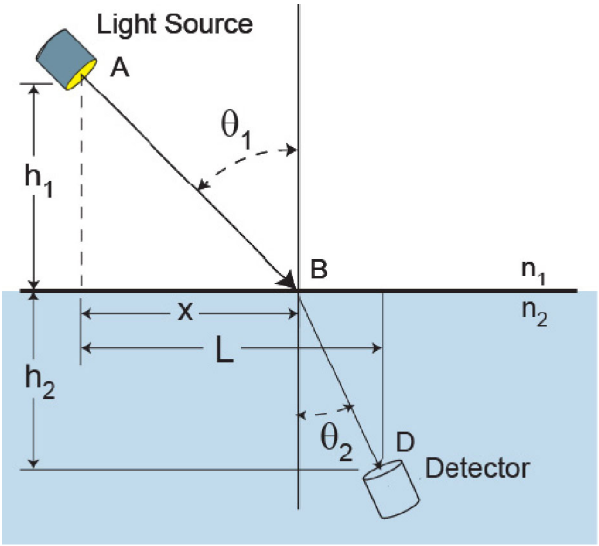


Figure 2.16: Snell's Law.

The value of the refractive index depends greatly on materials and conditions, while the rules for ray optics arise from fundamental laws of physics. Four of these laws are listed here.

1. In a material with a uniform index of refraction η, light rays travel in a straight line.

2. When rays reflect from a surface, the angle of incidence θ_i equals the angle of reflection θ_r, as shown in Figure 2.15.

3. Rays bend as they pass through a smooth boundary surface between two different materials, such as air and glass. Snell's Law, formulated in 1621, describes how much bending occurs at every point on every surface of a lens.[13] Light rays that arrive perpendicular to a surface enter the new material without bending; light rays that arrive at angles away from perpendicular bend as they enter the material. Higher-index materials weaken the angle and lower-index materials exaggerate the angle. If this exaggeration exceeds 90°, no light enters the new material. Instead, total internal reflection at the surface bounces away the incident light ray like a mirror. For lesser angles, Snell's law links together ray direction θ_1 through the first material with index η_1 to ray direction θ_2 through the second material with index η_2, as illustrated in Figure 2.16:

$$\eta_1 \sin\theta_1 = \eta_2 \sin\theta_2 \tag{2.8}$$

Rays and refraction governed by Snell's law give us a sufficient description of light transport to characterize ordinary lenses and traditional cameras that are

[13] Snell (or Snellius, born Willebrord Snel van Royen) was Professor of Mathematics at the University of Leiden when he described the law of refraction in 1621. For some reason he never published the law, and it didn't become widely known until 1703 when Christiaan Huygens revealed the equation in his Dioptrica.

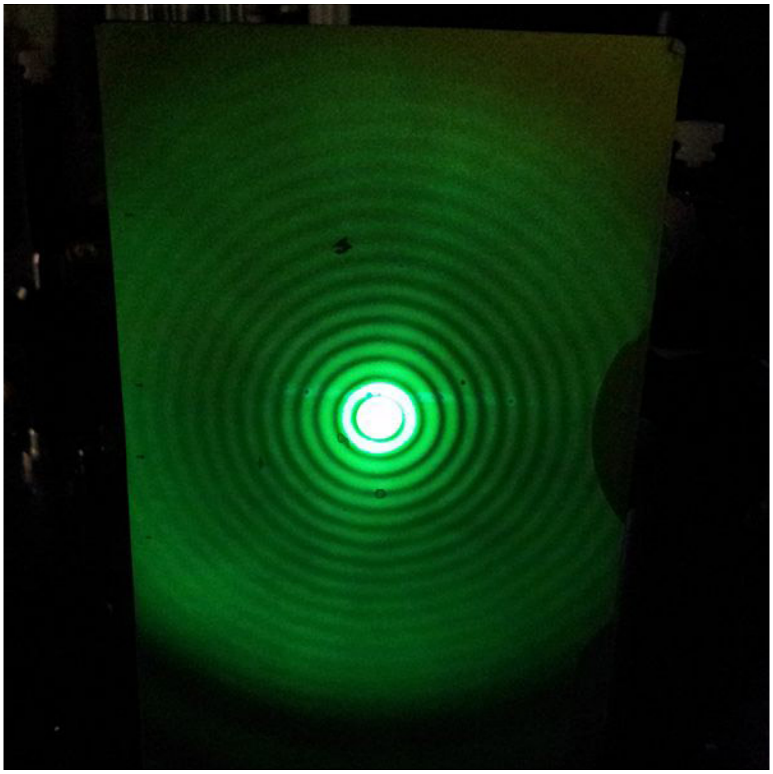


Figure 2.17: The image resolution of a pinhole camera decreases as the pinhole size approaches zero because of periodic fringing artifacts caused by the diffraction limits of light.

not diffraction limited—where resolution is not appreciably affected by the wave nature of light.

4. Nearby obstructions can also bend light rays. These diffraction effects limit the resolution and accuracy of any ray-based lens design. Diffraction modifies the paths of rays that pass through openings small enough to measure in single wavelengths (typically $< 10\lambda$). This is not really a rule, but rather a warning that rays cannot give the whole story of light propagation. Pinhole cameras demonstrate that light travels in a straight line, but as the pinhole size shrinks toward zero, the image resolution of the pinhole camera falls, degraded by periodic fringing artifacts, as shown in Figure 2.17

2.3.2 Ray Bending to Form Images

In an ideal perspective camera, as shown in Figure 2.14, a well-focused lens copies light from points in the scene (point P in the figure) to points on the sensor (point Q in the figure) to form an image [NOTE: POINT Q ON THE SENSOR PLANE IS MISSING FROM THE FIGURE]. The ray from each scene point P to each image point Q forms a straight line through a single point within the lens. This point in the lens is defined as the *center of projection* (COP). If we replace the ideal lens with an ideal (infinitesimal) pinhole at the COP, the camera captures the same image, but with the infinitesimal irradiance of only a single ray through the pinhole to the sensor plane.[14]

Any sensible lens will collect far more light and form far sharper pictures than any pinhole. All the rays that leave scene point P and pass through the camera's aperture define a cone-shaped beam or bundle of rays that spans a measurable solid angle. The lens bends each ray in the cone-shaped beam differently, causing all the rays to converge and to deliver all their light power to the sensor plane at image point Q. But just how much light do we get? Recall that a scene point P emits only infinitesimal radiant flux Φ, and point Q receives a small but finite fraction of that flux. Image points adjacent to Q receive similar fractions from scene points adjacent to P, each through their own bundle of rays. Through these ray bundles, some finite fraction of the radiant exitance M from the scene at point P becomes the irradiance E on the sensor at point Q. If the scene surface is diffuse (for example, if we photograph a blue-green electroluminescent night light), all those ray-bundles are uniform, and each ray in each bundle has the same radiance L. With a few size measurements for the ray bundles on either side of the lens, we could compute the sensor irradiance E that results from a (diffuse) surface that sends radiance L toward the camera.

Measuring the light in ray bundles gathered by lenses reveals why photographed objects (i.e., resolved objects—those with an image area that depends on distance, and not those that appear as a single point, such as a star) do not get brighter or dimmer with distance. If we double the distance from camera to object, each of these cone-shaped ray bundles gets narrower. Their angular radius falls by half, their solid angle (measured from scene point P) falls to 1/4 of their original value in steradians, and thus they deliver only 1/4 of the radiant flux to sensor point P. However, the ray bundles leaving each of scene point P's neighbors now converge at different locations around sensor point Q. Their distance to Q on the sensor plane is also cut in half, both vertically and horizontally. The irradiance E around sensor point Q increases by a factor of four, exactly cancelling the 1/4 reduction in the radiant flux from each scene point.

Put another way, as the camera moves away from an object, the camera gathers a narrower, less-powerful ray bundle from each point on that object. But inside the camera, these bundles converge to form an image that gets smaller and smaller, and these points pack together more tightly to keep the sensor irradiance constant.

[14] Pinhole cameras impose severely unfavorable tradeoffs. In a hobbyist's pinhole camera [12, 13] or a digital pinhole camera [14], the small size of the pinhole, the sensitivity of the paper, and the brightness of the outdoor scene are all necessary to raise the irradiance of the sensor plane high enough to form a usable image. Enlarging the pinhole even slightly to admit more light produces a disastrous boost in blurring, especially for nearby objects. From each scene point, the enlarged pinhole admits a larger cone-shaped bundle of rays, which blurs the image by spreading the light across the circular area of the sensor, instead of focusing the light onto a point.

2.3.3 Ideal Thin-Lens Parameters

Instead of integrating cones of rays, we can use easier and more conventional methods to describe simple image-forming lenses like the one shown in Figure 2.14. Plus we can tie these methods directly to our ray-based measurements of light. As illustrated in the figure, every ray gets bent only once, as it passes through the lens plane at the center of projection. Even though the diagram ignores the thickness of the lens and the front and back surfaces that actually perform the ray bending, for many lenses the image-forming results are the same. This simple approximation of the lens as a single ray-bending plane, known as the *paraxial* or *thin lens* model, is the actual design goal and the ideal result for many complex multi-element lenses. Three parameters describe an ideal thin lens completely: the focal length f (in millimeters), the aperture diameter δ (in millimeters), and the lens speed (unitless). Here are the definitions of focal length and aperture diameter.

Definition 2.6. The focal length f of a thin lens is the distance behind the lens where parallel incoming rays meet, forming an image focused at infinity. The inverse of focal length in meters ($1/f$) describes the focusing power of the lens in "diopters" (1 diopter $= 1/f$ meters).

Definition 2.7. The aperture diameter δ measures the distance across each cone-shaped ray bundle that passes through the thin lens. It is the diameter of the circular aperture shown in Figure 2.14.

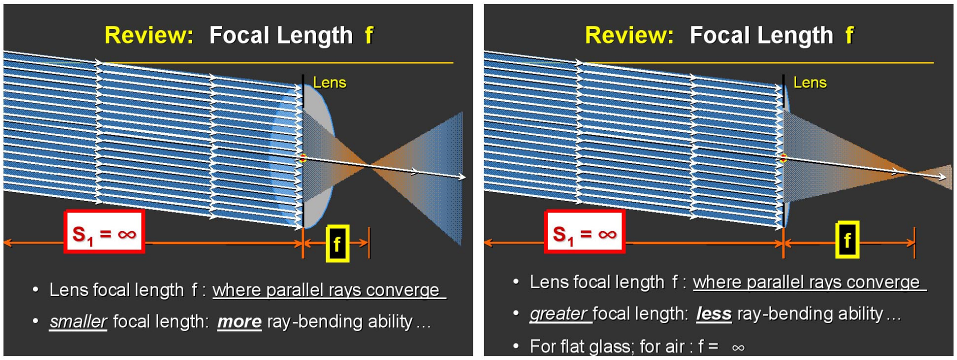
Thin lens focusing is sensible and intuitive. To focus at infinity, cameras move the COP of the lens to a distance f from the sensor plane. To focus on closer objects, lenses move the COP farther away, as shown in Figure 2.18. For an object at distance S_1 in front of the lens, and for a lens at a distance S_2 from the film plane, we have the thin lens formula:

$$\frac{1}{S_1} + \frac{1}{S_2} = \frac{1}{f} \tag{2.9}$$

The focal length f of the lens also determines the size of the image on the sensor plane. Therefore, larger cameras require lenses with greater focal lengths, and a so-called normal lens for a camera has a focal length that is approximately equal to the diagonal dimension of the image sensor. For example, popular 35-mm film cameras captured images in frames 24 mm tall × 36 mm wide; their 43.3 mm diagonal closely matches the normal 50 mm focal length lens supplied with new cameras. A normal lens assures that prints viewed from a convenient distance will match the field of view of the camera that took the picture.

The *speed* of a lens is a unitless ratio that describes the ability of that lens to transmit light; it is not a velocity of any kind. This lens parameter is a measure of the exposure time T required for the lens to collect a good image on film (hence the reason for the term). Fast lenses, as they are called, transmit enough light to capture a good image in a short time, while slower lenses take longer to capture the same light.

More formally, *exposure* in photography is the product of sensor irradiance E and the amount of time T that the camera shutter permits light to impinge on



Review: Focal Length f

- Lens focal length f : where parallel rays converge
- *smaller* focal length: ***more*** ray-bending ability …

Review: Focal Length f

- Lens focal length f : where parallel rays converge
- *greater* focal length: ***less*** ray-bending ability …
- For flat glass; for air : f = ∞

Figure 2.18: Focal length.

SIDEBAR ON LENS DESIGN

The theory of lens design was developed before the advent of photography, and a few well-designed lenses were available well before Nicéphore Niépce's first camera experiments in the 1820s. The Wollaston Landscape lens of 1812, probably the earliest well-designed lens, was a successful lens for many decades and was still being mass-produced for low-cost cameras in the middle of the 20th century. Improvements in camera lenses were slow in the decades after 1840 because opticians typically worked by trial and error rather than by the systematic application of optical principles.

Lens design is actually a difficult process that involves many compromises. Light rays must be bent and directed to exactly the correct point on the image, and rays of all desirable wavelengths need to focus at the same point. Current high-quality lens designs correct for seven independent aberrations, and include features to minimize reflections between the glass surfaces of the lens elements and the mounting barrels. Computer software now greatly enhances efficient lens design by quickly tracing the possible paths of light rays through combinations of mounted lenses. Acceptable results can then be scaled up or down in size. Of course, the actual construction of high-quality lenses is still a demanding task, and the required materials such as low dispersion glass (glass whose focusing abilities show only small variations with wavelength λ) can be very expensive.

The University of Central Florida's OSE 6265 course "Optical Systems Design," by James Harvey, offers excellent overviews and online course materials on the topic of lens design.

the sensor plane. In SI units, we measure exposure in Joules/meter^2 on the film. The unitless lens speed describes only the rate of delivery of that light energy. For example, if we focus the camera on a diffuse surface with radiant exitance M (or luminous exitance M_v), the speed of the lens describes the fraction of surface exitance M that arrives as sensor irradiance E. A faster lens on the same camera delivers a larger irradiance fraction E/M, and requires less time to capture the same image.

However, photographic conventions long ago settled on a different unitless ratio to describe lens speed. This number uses the focal length f and the size of the lens aperture δ to define speed. It has several different common names, including the f-number, the f-stop, or N.

Definition 2.8. The speed of a lens N is defined by the unitless ratio of focal length f divided by the aperture diameter δ, both usually given in millimeters (mm):

$$N = f/\delta \tag{2.10}$$

A lens speed of $N = 8$ is often denoted as an f-number of $f/8$ instead, or described as an f/8 stop. It is important to note that the lens speed N *grows larger* as the lens aperture becomes smaller and admits less light. In other words, faster lenses have smaller values of N.

In the world of general commercial photography, the *speed* of an adjustable-aperture lens typically represents the *maximum* aperture of that lens. (e.g., a lens with aperture settings from f/2.8 to f/16 has a speed of 2.8). For example, if the focal length f is 20 mm and the maximum aperture δ is 5 mm, then the lens speed N is equal to (20 mm)/(5 mm) = 4.0 . We would say this is an f/4.0 lens, and the lens barrel might be labeled (20 mm 1:4.0). Notice that the larger the maximum aperture, the smaller the value of N for any specific lens. Dividing by aperture diameter δ ensures that the f-number does not vary as we change lenses of different focal lengths. By matching the f-numbers among lenses we ensure the same radiant exitance M from a scene will cause the same irradiance E on the sensor. [15]

Lens speeds expressed as f-numbers represent the traditional (and rather cryptic) labeling of aperture settings found on most adjustable lenses. These aperture settings are typically the values 2.8, 4, 5.6, 8, 11, 16, and 22. The exposure (or sensor irradiance E) that reaches the sensor is proportional to the *area* of the lens aperture (where area $= \pi(\delta/2)^2$), instead of the aperture diameter δ. Hence, the aperture adjustment that doubles the admitted light increases the aperture diameter by a factor of $\sqrt{2} = 1.414$. This is the reason why the typical f-numbers listed on lenses increase by successive factors of 1.414, but cause exposure to change by a factor of two, with each doubling step called a *stop.* Fast lenses with small values of N have larger apertures, tend to be more difficult to make, and are usually more expensive (and heavier because of the additional glass needed). The larger apertures enable proper exposures with higher shutter speeds (shorter shutter open times), but introduce other tradeoffs.

[15] In the special case of closeup photography, the speed or brightness of the lens is diminished because the image plane is farther from the lens than one focal length. Effective speed N for these lenses must compensate for image magnification.

To explore lens, ray, and sensor behaviors interactively, we suggest the online optical-bench simulator [19]. This simulator is part of a growing family of *Physlets*, which are applets written in Java and designed to help teach physics.

2.3.4 Thin Lens Tradeoffs

As we described previously, three parameters are sufficient to specify thin lens performance for an ideal lens: the focal length f, lens speed N and aperture diameter δ. Different parameter choices will create different images in different cameras. Some choices yield photographic images that seem to have a natural appearance while other choices yield images that seem distorted. Nearby objects may appear too large, or distant objects may look flattened. Buildings may stand tall and straight, or their sides may tilt inwards as we aim our camera to capture their topmost floors. Why? Each effect is a consequence of the perspective projection used to capture 3D scenes as 2D images. These visual effects can result from pivoting the lens about its center of projection, sliding the sensor away from its centered position over the sensor plane, pivoting the sensor plane about the lens centerline, or some combination of these moves. In addition to these desirable planar-perspective effects, unwanted aberrations in lenses can impose non-linear distortions that map straight lines in the scene to curved lines in the image, as discussed in later sections.

Focal-length adjustments on a camera lens determine which portion of a scene we will capture in a single photograph, but they do not change how that 2D image was constructed from the 3D scene. As Figure 2.19 shows, the longer 135 mm focal-length lens captures a smaller but more detailed portion of exactly the same image captured by the 40 mm, 20 mm, and 10 mm focal-length lenses, taken from the same camera position. If we shrink the full-size 135 mm image enough, it exactly matches the central portion of the 10 mm image. Note that the ratio of focal lengths (135 mm/10 mm) doesn't exactly match the image scaling ratio, but it is often a reasonable approximation for scaling ratios between long focal lengths. No lens adjustments can change image contents; only moving the camera will change the relative sizes of viewed objects, the way one object occludes another, or the image perspective.

A normal or natural perspective depends on the angle of view of a photograph relative to that of the original scene. According to Kingslake's analysis,[16] the region of sharp vision of the human eye has an angular width of about 20° or a half angle of 10°, and motion is detected over a range of about 50°. To obtain a 'normal' view of a printed image, for example, a viewer should look at the print from a distance that causes the printed image to span the same field of view captured by the camera. In that way the angles between objects in the print match the angles between objects in the scene from the camera's viewpoint.[17]

Suppose, for example, that a 35 mm film camera with a 50 mm (approximately 2 inch) focal length lens produces a film negative (1 in × 1.5 in) that is used to

[16] Rudolf Kingslake, *Lenses in Photography* (Case-Hoyt, Corp., Rochester, N.Y., 1951). An updated version of this classic book on lens design is *Lens Design Fundamentals*, 2nd edition, by Rudolf Kingslake and R. Barry Johnson (Academic Press, 2009).

[17] The handy online calculator at www.canon.com/bctv/calculator/index.html finds the angular field of view versus focal length for most cameras.

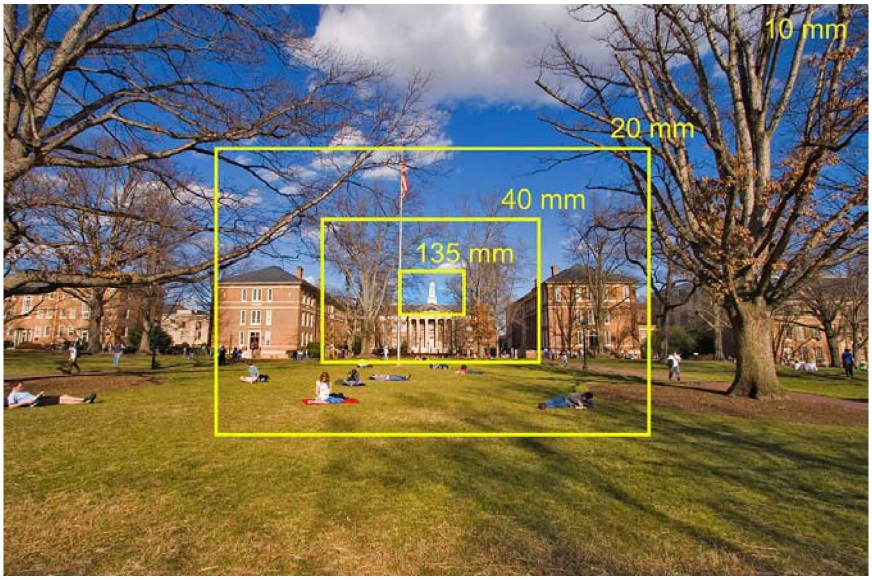


Figure 2.19: The effect created by changing the lens focal length while maintaining the same visual perspective. The Canon XTi digital camera used to make this image has a APS-C 22.2 mm by 14.8 mm sensor. (Source: Charles Johnson book.)

make an 8 in × 12 in printed photograph. The enlargement factor from negative to print is approximately 8 in/1 in = 8, so the proper viewing distance is 2 in × 8 = 16 in for 'normal' viewing. But for a wide-angle 17 mm (0.67 inch) focal length lens on the same camera, the 'normal' viewing distance for an 8 in × 12 in print shrinks to just 0.67 in × 8.0 = 5.36 in . Unfortunately, most 8 in × 12 in prints are viewed from approximately the same distance regardless of the point of perspective, and thus the apparent perspective suffers.

So what is a normal lens? By definition, a normal lens has a focal length approximately equal to the diagonal dimension of the film or the digital sensor. This choice insures that the camera's angle of view forms a good approximate match to the angle of view of the human eye. For a 35 mm film camera, the normal lens has a focal length of about 43 mm, and a simple computation shows that the angular (diagonal) field of view has a half-angle of $26.5°(\theta = \arctan(0.5))$, or a full angle of 53°. This is somewhat larger than the eye can view at a glance, but it is an acceptable compromise, as most photographic images are viewed from well behind their perspective point.

Wide angle lenses, of course, have shorter focal lengths than normal lenses. A typical wide angle lens has a diagonal field of view of about 75° or a focal length of 28 mm (full frame 35 mm equivalent). Ultra wide angle lenses have fields of view from about 90° to 100° and focal lengths of about 17 mm to 20 mm (full frame 35 mm equivalent). These lenses are often used for visually strange and distorted-looking photographs because any suitable normal viewing distance is inconveniently close to the image. This effect can be alleviated by making larger prints or murals.

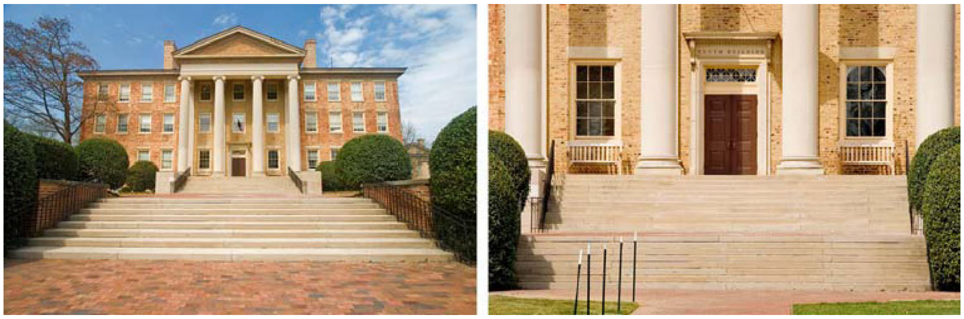


Figure 2.20: A set of stairs photographed with a 10 mm wide angle lens (left) and a 190 mm telephoto lens (right). The position of the camera was close to the stairs in the wide angle image and far from the stairs in the telephoto image. (Source: Charles Johnson book.)

Figure 2.20 illustrates the change in perspective when the same set of stairs was photographed at close range with a wide angle lens and from a considerable distance with a telephoto lens.

With a digital camera and panorama-making software such as the free Hugin Panotools project [15], anyone can stitch together images taken in many directions from a single viewpoint to obtain a single image with an extremely large field of view. Each image in the resulting panorama shares the same center of projection, and users can create the appearance of a "normal" image by printing or projecting a large picture for viewing.

Our final point about perspective involves camera tilt. It is common knowledge that when a camera is tilted up, the image will be distorted so that vertical lines converge and buildings appear distorted, as if they are leaning backward. This effect is often considered undesirable, and methods are available to correct for it. Holding the camera perfectly level removes the distortion, but the perspective can clip the tops of tall objects and may miss points of interest. Professional view cameras get around the tilt problem by allowing the photographer to raise the lens while keeping the lens plane parallel to the film plane. In 35 mm film photography or digital photography, special-purpose tilt/shift lenses are used to minimize this type of distortion.

Tilt corrections for existing film negatives can be made in a darkroom by tilting the paper holder under the photographic enlarger, or these negatives can be scanned into digital files and corrected with powerful software tools. Image processing software such as Adobe Photoshop can transform (distort) an image to correct for the convergence of vertical lines, as well as correct for lens aberrations and distortions, or even convert images taken with fisheye lenses to rectilinear form.

Curiously, these perspective effects are not always considered problems. Extreme camera tilt, for example, is often considered pleasing. Also, sideways tilt that makes horizontal lines converge at the horizon is not considered a defect. This effect is so common and accepted that artists often use geometric projections to build this kind of perspective into their images. In this and all the other cases mentioned, a natural view is obtained by using one eye at the perspective point. The use of two eyes (binocular vision) gives an unnatural appearance for small images. We don't notice this problem with natural scenes, and we can avoid it by making large images. One last note about perspective; our perspective point is determined when we select a seat in a movie theater, but that position is seldom the perspective point of the movie photographer [378].

2.3.5 The Simple Thin Lens and What It Does

The task in this section is to show how points in the object space (in front of a lens) are related to points in the image space (behind a lens) for ideal lenses. With modern equipment it is relatively easy to grind and polish the surface of a circular glass plate into a spherical shape; that is to say, having the shape of a portion of the surface of a sphere. In a plane that contains the optical axis (axis of symmetry), a spherical surface appears as a segment of a circle. To this day most camera lens elements are spherical in shape. We will show that aspheric surfaces are sometime desirable, but they are much more difficult to fabricate. The next step is to show how a lens can be shaped so that it will have a certain focal length. This is, in fact, the starting point for understanding the design of all lenses, even the compound, multi-element types.

Optical surfaces

No matter how complicated, lenses still consist of a collection of smooth boundaries between transparent materials with different indices of refraction. In principle, all we need is Snell's law of refraction to be able to trace the path of a ray of light through a lens, as shown in Figure 2.21. Here a ray of light (the red line in the figure) from an object arrives at a spherical surface at a distance h from the optical axis. This convex surface exhibits a positive radius of curvature R, with its center of curvature on the right side. (Note that this sign convention must be applied carefully and consistently; a negative radius R would form a concave surface with its center of curvature on the left.) The effect of the surface is to change the direction of a ray if the refractive indices on the two sides (η_1 and η_2) are different. The calculation of the change in angle from θ_1 to θ_2 relative to the surface normal requires the application of Snell's law.

Snell's law tells us that $\eta_1 \sin\theta_1 = \eta_2 \sin\theta_2$. From this it is easy to compute the angle of refraction for a single ray passing through the surface. The calculation must, of course, be repeated for each surface of a simple lens and many surfaces in a compound lens. Furthermore, we must compute its effects for cone-shaped bundles of rays from points on the object side in order to characterize the lens completely. This tedious task, nowadays assigned to a computer, can predict the performance of a particular lens design, but does not give much guidance about how to design a new lens. For that we need simple equations and rules of thumb to get us started.

Here, as in many areas of science, simplifying our assumptions can lead to useful approximate equations. For example, suppose that the rays remain close to the optical axis and that all of the angles are small. This means small enough that the $\sin\theta$ and $\tan\theta$ functions can be replaced with θ itself, where of course the

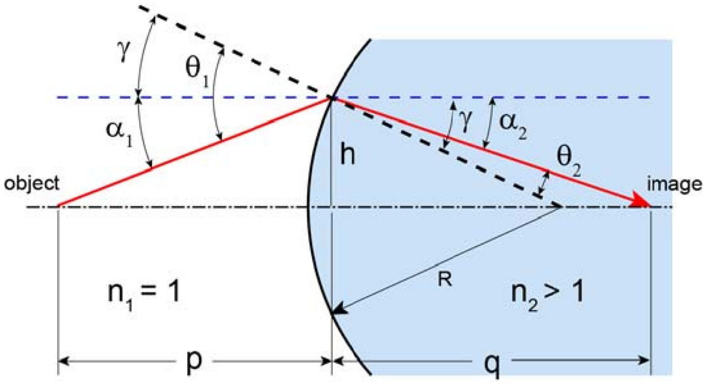


Figure 2.21: A spherical surface with radius R. The dotted line from the center of curvature is normal to the surface. (Source: Charles Johnson book, p. 37.)

SIDEBAR ON THE DEVELOPMENT OF LENSES

The earliest lenses were made of polished crystals, and it was not until the Middle Ages that glass lenses were produced. By the 13th century glass lenses were good enough to be used in spectacles for the correction of presbyopia, or farsightedness. Curiously, the development of lenses is credited to craftsmen, not scientists. Early philosophers dismissed lenses since their operation was not consistent with existing (European) theories of vision based on the ray emissions from the eye. They concluded that spectacles could only be a disturbing factor, and that one should not trust things seen through lenses. In fact, the emission theory of vision had been disproved by Ibn al-Haytham in the 11th century, by considering afterimages in the eye and the pain that results from looking at the sun.

More complicated optical instruments such as telescopes required lenses of sufficient uniformity to be used in pairs to obtain magnification. Hans Lippershey, the master spectacle maker of Middelburg, improved lens grinding techniques and in 1608 was able to present a telescope to Prince Maurits in The Hague. Others claimed priority in inventing the telescope, but the first record we have of an actual instrument appears in Lippershey's patent application. Everything changed in 1610 when Galileo made a telescope for himself and turned it on the heavens. He quickly discovered the moons of Jupiter and published his results. That same year Kepler used a telescope made by Galileo and worked out a new theory of optics based on two ideas: (1) light rays travel in all directions from every point in an object, and (2) a cone of light rays enters the pupil of the eye and is focused onto the retina. The revolution of 1610 opened a new world for scientific observations of objects both near (by the microscope) and far (by the telescope). Ironically, Lippershey never got his patent, but he was well paid for his services, and he now has a lunar crater and a minor planet named for him.

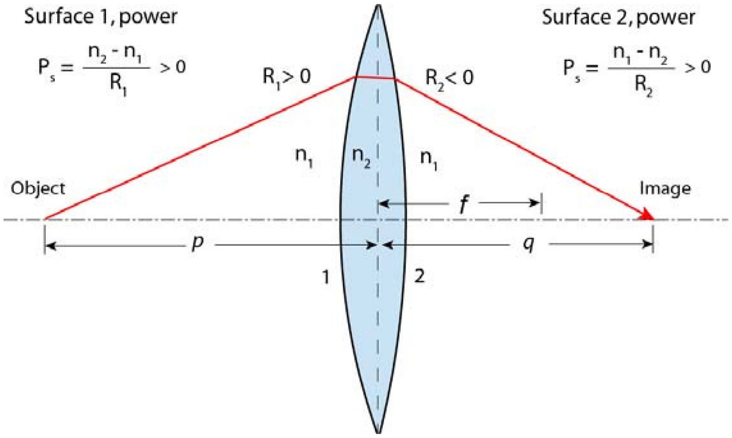


Figure 2.22: A simple convex-convex spherical lens.

angles are measured in radians (2π radians equals 360°). This is called the *paraxial approximation*, and Snell's law becomes $\eta_1\theta_1 = \eta_2\theta_2$.

With the paraxial approximation we can easily derive equations for a thin lens. In Figure 2.22, the radius R_1 for the first (entrance) surface is greater than zero, as it is in Figure 2.21, because the center of curvature is on the extreme right hand side, and the radius R_2 of the second (exit) surface is less than zero because its center of curvature is on the extreme left hand side. We first apply Snell's law to derive an equation for the focusing power of surface 1. From Figure 2.22 we find that $\theta_1 = \alpha_1 + \gamma$ and $\theta_2 = \gamma - \alpha_2$, where γ, α_1, and α_2 are shown in Figure 2.21. The paraxial approximation permits us to use the following expressions for the angles: $\alpha_1 = h/p_1$, $\alpha_2 = h/q_1$, and $\lambda = h/R_1$. When these quantities are substituted into Snell's law, we obtain the following equation for the focusing power of surface 1:

$$P_s = \frac{\eta_2 - \eta_1}{R_1} = \frac{\eta_1}{p_1} + \frac{\eta_2}{q_1} \tag{2.11}$$

Equations that describe the thin lens require that the powers of the two surfaces be combined [205]. These derivations yield

$$\left(\frac{\eta_2}{\eta_1} - 1\right)\left(\frac{1}{R_1} - \frac{1}{R_2}\right) = \frac{1}{p} + \frac{1}{q} \tag{2.12}$$

If the space before the lens is occupied by air, so that η_1 is close to unity, the focusing power depends only the refractive index of the lens material and the radii of curvature. If the entrance surface of the lens is in contact with water, which has a refractive index of approximately 1.3, then the effective power will be reduced. It can, of course, be compensated by decreasing the radius of curvature. For our purposes, we assume refractive indices in the object space and image space are both unity.

Now suppose that the object is moved very far to the left so that $1/p$ approaches zero. In this limit q becomes equal to the focal length f, and Equation 2.12 becomes the well-known Lens Maker's Equation:

$$\frac{1}{f} = \left(\frac{\eta_2}{\eta_1} - 1\right)\left(\frac{1}{R_1} - \frac{1}{R_2}\right) \tag{2.13}$$

Another useful equation is obtained by noting that the left hand side of (2.12) is equal to the right hand side of (2.13). Thus:

$$\frac{1}{f} = \frac{1}{p} + \frac{1}{q} \tag{2.14}$$

This is called the *conjugate equation* since it relates the conjugate distances p and q to the focal length f. Equation 2.13 is wonderful for lens designers, but is usually not of much interest to photographers. On the other hand, the conjugate equation (2.14) is very useful for understanding the way our lenses work. Right away we see that a simple lens can be focused by adjusting the lens-to-sensor distance q to accommodate the object-to-lens distance p. Furthermore, the magnification, that is the size of the image of an object relative to the actual size of the object, is equal to the ratio q/p and is, therefore, determined by the focal length. When the object distance p is much larger than the focal length f, the magnification is approximately equal to f/p. Note that when $p = q$, the magnification m is equal to one and that both p and q are equal to $2f$.

Caveat: In the paraxial limit, the thin lens is ideal in the sense that a cone of rays from a point in an object will be focused into a circular cone of rays that converges to a point (or to a diffraction limited disk) on the sensor plane. As the diameter of the lens (or the limiting aperture) increases as the focal length f shortens, or if we gather rays from scene points far askance from the lens axis, the paraxial assumption will no longer apply. Focused bundles of rays will distort, their shape will deviate from circular cones and take on a variety of patterns on the sensor plane. The patterns are characteristic of spherical lenses and are associated with well-known aberrations.

Compound Lenses: Paraxial Rules

The focusing power of the lens, taking into account both surfaces, is characterized by the inverse focal length $1/f$. Now suppose we have two thin lenses in contact as in Figure 2.23. What is the power of the combination? This is an easy question that can be solved by combining the conjugate equations for the two lenses. First we introduce the labels 1 and 2 to obtain:

$$\frac{1}{p_1} + \frac{1}{q_1} = \frac{1}{f_1} \qquad \frac{1}{p_2} + \frac{1}{q_2} = \frac{1}{f_2} \tag{2.15}$$

The image for the first lens falls to the right of the second lens and becomes its subject. Therefore, we can substitute $-q_1$ for p_2 in the second equation where the minus sign indicates that the subject is behind (to the right of) the lens. The equations can then be added to obtain:

$$\frac{1}{p_1} + \frac{1}{q_2} = \frac{1}{f_1} + \frac{1}{f_2} \tag{2.16}$$

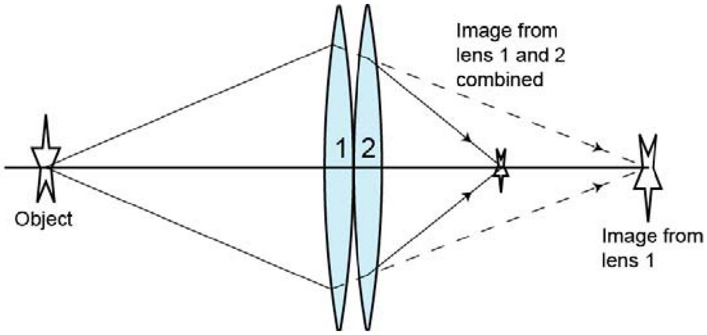


Figure 2.23: A pair of thin lenses with exaggerated thicknesses. (Source: Charles Johnson book, pg. 39.)

The subscripts on the left hand side can be dropped because the combined set of thin lenses only has one subject (object) distance and one image distance. The conclusion is that the focusing power of a set of lenses can be computed by simply adding their separate powers. We know that the quantity $1/f$ is defined as the power in *diopters* when f is measured in meters, and the rule is to add diopters when lenses are used in combination. If the two lenses are separated by the distance d, the equation becomes:

$$\frac{1}{p_1} + \frac{1}{q_2} = \frac{1}{f_1} + \frac{1}{f_2} - \frac{d}{f_1 f_2} \tag{2.17}$$

In conclusion, the simple thin lens provides a useful model for understanding all lenses. The conjugate equation is the starting point for estimating the behavior of lenses, and it even applies to compound lenses under certain conditions. For further information, consult *Optics: the Science of Vision* by Vasco Ronchi [344] or *Introduction to Modern Optics* by Grant R. Fowles [139] [16].

2.3.6 Resolution Limits and the Airy Disk

To understand the resolution limits of ideal thin lenses, we return to the pinhole camera and examine it more closely. A pinhole camera is a light-tight box with a pinhole in the center of one side of the box and a photographic film plate or digital sensor attached to the opposite side. It's a small version of the *camera obscura*, which played an important role in the history of art and the early development of photography. As illustrated in Figure 2.24, the formation of an image can be easily explained with ray optics. We assume that light beams move in a straight line, and we perform ray tracing. A ray of light from any point on an object in a scene passes through the pinhole and strikes the photographic plate or digital sensor on the opposite side of the box.

It is obvious from the illustration that a point on the object will become a small circle in the image because the pinhole aperture must have some diameter to admit the light. This situation creates a photographic dilemma. If we make the pinhole smaller, the light reaching the image will be diminished and the exposure will take

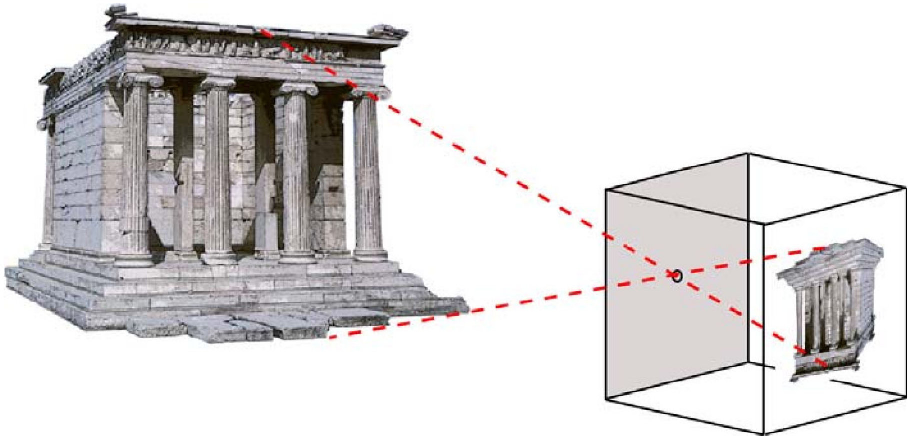


Figure 2.24: Illustration of a pinhole camera pointed at the Temple of Wingless Victory in Athens.

more time. On the other hand, if we enlarge the pinhole, the image resolution will suffer because of blurring caused by the larger pinhole size. At first there appears to be no optimum size for the pinhole, and all we can do is search for a convenient trade off between image resolution and the time required to expose the image. Unfortunately, that conclusion is not accurate. As we use smaller and smaller pinholes, at some point we find that the resulting image resolution actually decreases, contrary to what we might expect. The reason for this unfortunate development is the ubiquitous phenomenon of diffraction.

In the limit of very small apertures, the image of a distant point source of light, say a star, is not imaged as a point but as a circular disk with faint rings around it. The bright central region of this image, called the *Airy disk*, is illustrated on the right side of Figure 2.25. The diameter of the disk measured to the first dark ring is equal to $2.44\lambda N$ where λ is the wavelength of the light and N is the ratio of focal length to aperture, f/δ, where f is the f number described earlier and δ is the aperture diameter. The spot size for a distant point without diffraction is equal to the aperture diameter δ. Therefore, the spot size will increase for both large and small pinhole sizes.

The optimum pinhole size occurs when the contributions of these two effects are equal. Therefore,

$$\delta = 2.44\lambda f/\delta \tag{2.18}$$

The optimum aperture size is therefore equal to

$$\delta_{opt} = \sqrt{2.44\,\lambda f} \tag{2.19}$$

Since λ is approximately 500 nm in the middle of the visible spectrum, we find that

$$\delta_{opt} = 0.035\sqrt{f} \tag{2.20}$$

where δ_{opt} and f are measured in millimeters. When $f = 100$ mm, the optimum pinhole has the diameter $\delta_{opt} = 0.35$ mm.

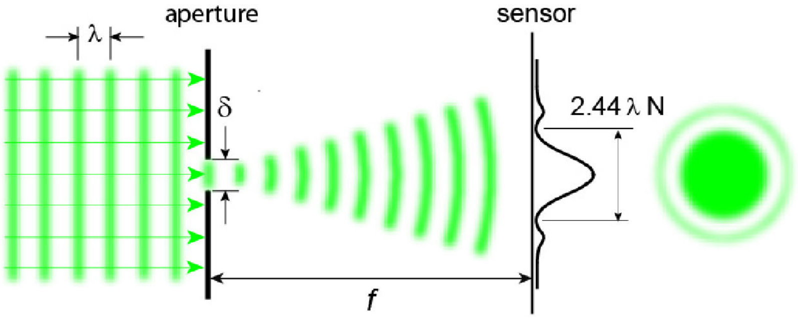


Figure 2.25: Incident light on a very small aperture results in diffraction effects in an image. The vertical scale is greatly exaggerated to reveal the intensity distribution in the diffraction spot.

This derivation is only approximate. It fails badly for nearby objects, where light rays from a bright point near the pinhole will diverge through the pinhole to make a spot that is larger than the pinhole. This results in an additional loss of resolution for nearby objects, indicating that the optimum size for the pinhole also depends on the distance of the object from the pinhole [17].

Photography with a pinhole camera requires an exposure time that depends on the f-number N of the aperture. From the definition we find that the f-number $N = 100$ mm/0.35 mm or $f/286$ for this optimum pinhole camera. How does this f-number compare with that on a lens-based camera? A typical aperture setting for outdoor photography with a lens-based camera is $f/8$. We know that every time N is doubled, the intensity of light in the camera is reduced by a factor of four. We must double an $f/8$ setting five times to reach $f/256$ (sufficiently close to $f/286$), which indicates there is about $4^5 = 1024$ times as much light intensity on the image plane in a conventional lens-based camera as there is in an optimum pinhole camera. A typical recommended exposure time for ISO 100 film or for a digital ISO of 100 in hazy sunlight is 1/125 of a second at $f/8$. The corresponding exposure with the optimum pinhole camera at $f/286$ would be about eight seconds.

One other point is worth noting. The ray tracing approach described above assumes that rays are undeviated (the path is unaffected) by the pinhole other than the spreading effect of diffraction. This works if the same medium, such as air, fills the interior of the camera and the outside space. Suppose instead that the space between the pinhole and the film is filled with water. A ray passing from air to water is subject to refraction and the path is bent toward the optical axis (a line through the pinhole and normal to the film plane). The unexpected result is a fish-eye view of the world. An air-filled pinhole camera with a glass window underwater would be the reverse situation, and should give a telephoto effect.

Lenses are obviously much better for imaging than pinhole cameras. They allow us to make sharp high-resolution photographs at large lens apertures, which means that adequate exposures can be obtained with fast shutter speeds (short exposures). In addition, the effects of diffraction can be reduced and the region of sharp focus (depth of field) can be controlled.

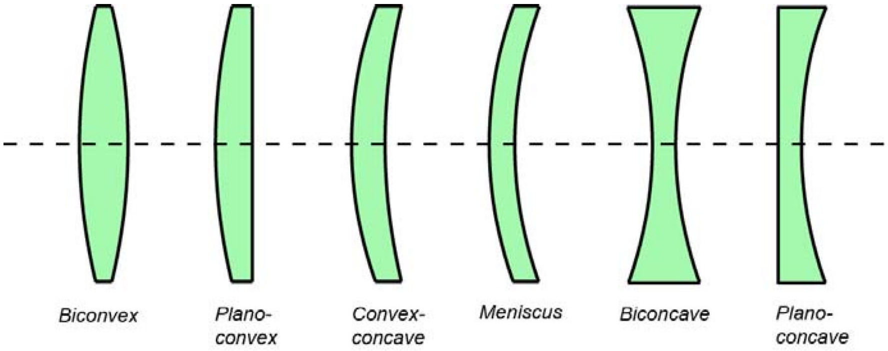


Figure 2.26: The different types of simple spherical lenses. (Image credit to be determined.)

2.3.7 How to Make Lenses That Are Good Enough for Photography

In this section we consider why multi-element lenses are necessary in order to minimize aberrations while providing large apertures. We will also show that multi-element lenses give unacceptable light transmission, ghosts, and flare in the absence of antireflective multi-coating.

Figure 2.26 illustrates the possible types of spherical lenses. Convex lenses (or positive lenses) are thicker in the center and cause light rays to focus, while concave lenses (or negative lenses) are thinner in the center and cause light rays to diverge. Suppose we select a single biconvex lens for our camera. The resulting image will look something like Figure 2.27. The image is sharp and in focus in the center of the scene but it deteriorates and is out of focus away from the center. The primary reason for this problem is that the image projected by the biconvex lens is not flat, but other aberrations exist as well.

The English scientist W.H. Wollaston discovered in 1812 that a meniscus-shaped lens (which is actually convex-concave), with its concave side to the front, can produce a much flatter image and, therefore, a much sharper photograph overall. Unfortunately, Wollaston's simple lens is not suitable for photography because it exhibits extreme chromatic aberration, focusing blue light at a different place on the optic axis than yellow light.

These limitations of biconvex and convex-concave lenses demonstrate the problems that must be overcome in attempting to improve a photographic image. It is relatively easy to produce lenses with spherical surfaces by grinding and polishing, but spherical lenses are not optimal. Monochromatic aberrations such as spherical aberration, coma, astigmatism, and distortion result from the inability of spherical lenses to produce perfect images. To use spherical lenses for photography we must correct aberrations by using a compound lens with multiple elements.

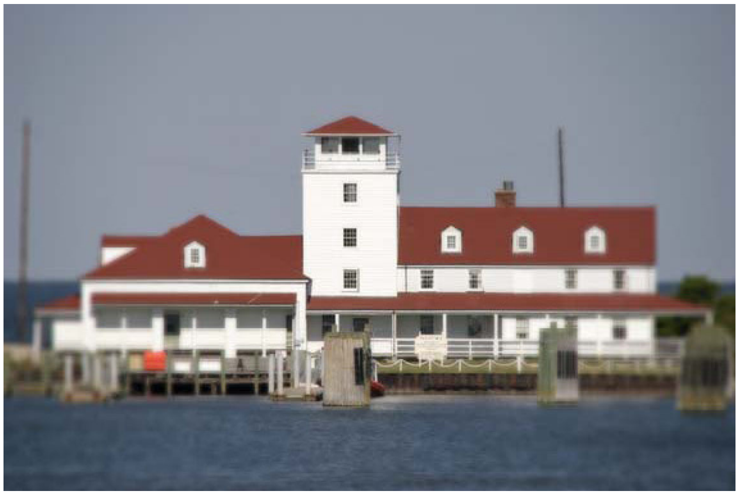


Figure 2.27: A simulated illustration of the effect of image curvature caused by a single-element lens. (Photograph by Charles Johnson.)

Aberrations and Coatings

The term ‘aberration’ appears frequently in the discussion of lenses so it is worthwhile to define it carefully[18]. Recall that Snell’s law describes the refraction of light rays at each optical surface. For rays that are close to the optical axis and make small angles with it, Snell’s law can be simplified by replacing the trigonometric function $\sin\theta$ with θ itself. This substitution represents paraxial or first-order theory and describes the perfect lens where $\sin\theta = \theta$ is satisfied. With practical lenses, however, this approximation is not justified and we require something better.

In 1857 Ludwig von Seidel used a Taylor Series expansion to find the next best approximation of $\sin\theta$ when θ is still small, namely $\sin\theta \simeq \theta - \theta^3/6$. He worked out the third-order optics to see what changes result from the extra terms. His calculation, still an approximation, revealed five independent deviations from perfect lens behavior for monochromatic light. These deviations are known as *Seidel aberrations*. More extreme effects that are not really independent appear at larger angles, but third-order optics provides a vocabulary and a good starting point for lens design. Seidel aberrations include spherical aberration, coma, astigmatism, field curvature and geometric distortion. [19]

If these aberrations result from the use of spherical lenses, then why not use aspheric lenses for photography? After all, aspheric lenses with appropriate shapes can eliminate spherical aberrations and reduce the number of elements required in compound lenses. As you might suspect, cost is a factor. Low-quality molded plastic and glass aspheric lenses are easy to produce, but high-quality glass aspheric lenses are much more expensive. Depending on the manufacturer, fabrication tech-

[18] 2. Klein, M. V., Optics, (Wiley, N.Y., 1970) chap. 4; 3. Photographic optics with illustrations of Seidel aberrations, www.vanwalree.com/optics.html.

[19] For descriptions and diagrams of these aberrations see the Melles-Griot web sites.

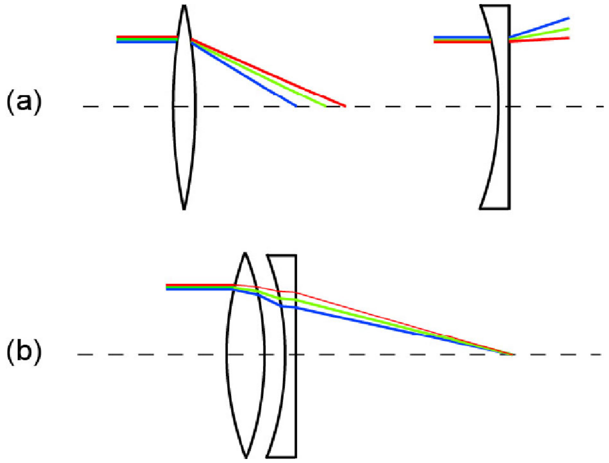


Figure 2.28: (a) The refraction of light by positive and negative lenses. Different frequencies of light are refracted at different angles, which causes the focal length of the lens to be frequency-dependent. (b) The combination of a positive lens with a negative lens having a higher index of refraction produces an achromatic compound lens.

niques can include shaping by diamond turning with a computer-controlled lathe, or applying optical resin that can be precisely shaped to the surface of a lens. In spite of high costs, many modern lenses in photography use one or more aspheric elements. For example. the Canon 17–40 mm and 10–22 mm wide-angle zoom lenses each have three aspheric lens elements.[20]

The problem of color fringes or chromatic aberration requires a discussion of optical materials. The refractive index of glass or any transparent material depends on the wavelength of light. This property is known as *dispersion*. Figure 2.28 illustrates how dispersion affects light rays as they pass through lens elements. The refractive index for blue light is higher than the refractive index for red light, so blue rays are bent through larger angles, as shown in Figure 2.28(a). This property causes the focal length of the lens to be different at different wavelengths.

A way to compensate for dispersion is to combine lens elements that have been carefully selected to cancel chromatic aberration at two well-separated wavelengths, as shown in Figure 2.28(b). The positive lens element is often crown glass (made from alkali-lime silicates) with relatively low refractive index, and the negative element is flint glass (silica containing titanium dioxide or zirconium dioxide additives) with a relatively high refractive index. These elements can be cemented together to make a single acromatic lens that has much less focal length variation over the selected wavelength range than is possible with a single glass element.

An *achromat* is a corrected lens with the same focal length at two wavelengths of visible light, usually in the blue and red regions of the spectrum. An *apochromat* is a corrected lens with three elements, giving identical focal lengths at three wavelengths and providing some spherical correction as well. A *superachromat*, the

[20] Canon, Inc., EF Lens Work III, The Eyes of Canon EOS (2006).

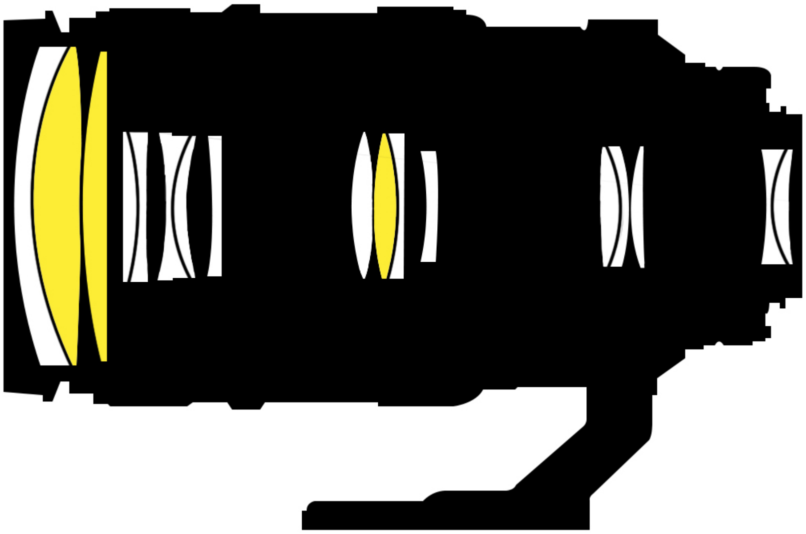


Figure 2.29: (a) Modern camera lenses have multiple lens elements in several groups.

best color-corrected lens, gives identical focal lengths at four wavelengths and often into the infrared region as well. This near-perfect correction requires the use of expensive low-dispersion optical materials.

Anti-Reflective Coatings

Current lens technology incorporates many lens elements to correct for various aberrations. Lens catalogs show sample lenses with large numbers of elements, including low-dispersion glass and aspheric elements in crucial positions to improve performance. Some examples are:

- Canon EF 100–400mm f/4.5-5.6IS USM, 17 elements in 14 groups (fluorite and Super UD-glass elements)
- Sigma APO 80–400mm f/4.5-5.6 ex OS, 20 elements in 14 groups (2 SLD elements)
- Nikon 80–400mm f/4.5-5.6 ED AF VR Nikkor, 17 elements in 11 groups (3 ED elements)

The modern multielement lens benefits from an important 19th-century technological breakthrough. At that time, lenses with more than four air/glass surfaces suffered from low contrast, lenses with six glass/air surfaces were sometimes considered acceptable, but lenses with eight air/glass surfaces, which corrected for aberrations, yielded film negatives with very low contrast. The problem was the reflection of light at the surfaces. When light passes from a region with a refractive index η_1 to a region with a refractive index η_2, some fraction R of the light is

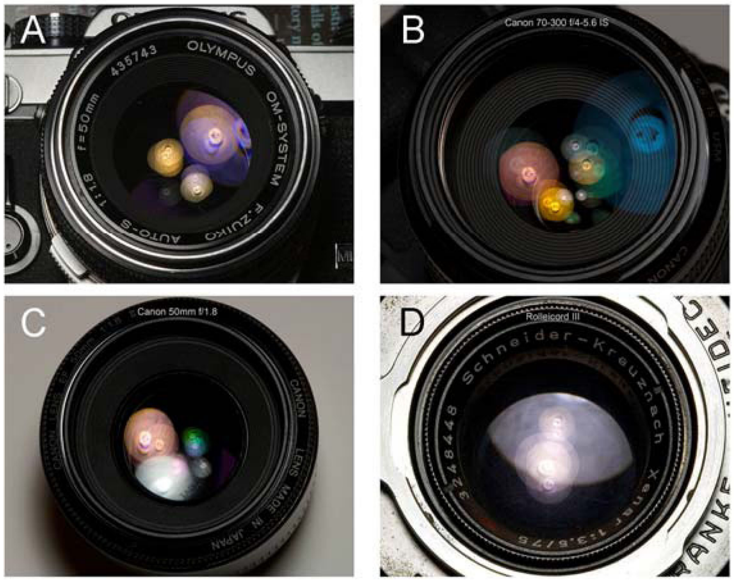


Figure 2.30: Reflections of a 5000K lamp from various lenses, showing evidence of the different coatings on the lens elements. (a) Olympus 50 mm f/1.8 (1972), (b) Canon 70–300 mm f/4-5.6 (2006), (c) Canon 50 mm f/1.8 (1987), (d) Schneider-Kreuznach 75 mm f/3.5 (1953). (Source: Charles Johnson book, pg. 48.)

reflected back from the new surface. Equations for R developed in the early 19th century by Augustin-Jean Fresnel showed that in the limit where the incident light is normal to the surface

$$R = \left[\frac{\eta_2 - \eta_1}{\eta_2 + \eta_1}\right], \qquad R = r^2 \tag{2.21}$$

where r and R are associated with the amplitude and the intensity of the reflected light, respectively. At an air/glass interface with $\eta_1 = 1.0$ and $\eta_2 = 1.5$, for example, the fraction of reflected intensity is $R = 0.04$ or about 4%. This fraction is lost at every air/glass surface. In a two-element lens with four surfaces about 15% is lost and only about 85% of the incident intensity is transmitted. This loss was, indeed, a serious problem for lens technology. Light was not only lost, but the reflected light bounced around inside the lens, producing ghosts, flare, and a general loss of contrast in the focused image.

The answer to the problem of light loss lies in sophisticated anti-reflection coatings. Everyone with a camera has observed, at the front of a lens, the beautiful violet and pastel shades that are reflected by the lens elements. Figure 2.30 shows reflections from lenses manufactured at different times. The colors reveal the presence of coatings that are themselves colorless. These coatings increase the transmission of a broad range of frequencies in the visible spectrum while reflecting a small amount of light that has not been completely cancelled by interference.

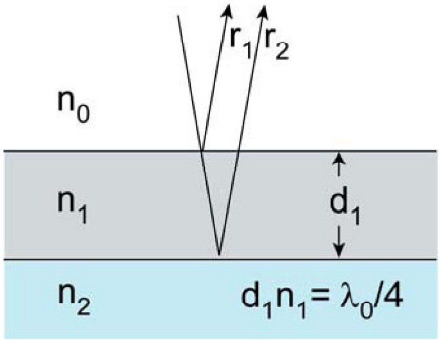


Figure 2.31: Single layer anti-reflective coating with exaggerated angle of incidence. Note: the n_x values in the figure should be redrawn as η_x. (Source: Charles Johnson book, pg. 49.)

How do anti-reflection coatings work? [181] Suppose we apply a thin transparent coat with a refractive index of $\eta_1 = 1.25$ to a lens made of glass with refractive index $\eta_2 = 1.5$. According to Eq. 2.21, the fraction of intensity reflected at the air/coat surface is 1.2%, and the fraction reflected at the coat/glass surface is 0.83%. The important question is how do we obtain the total reflection. Unfortunately, the numbers we calculated are not helpful at all! Experiments show that the total reflection depends critically on the thickness of the layer. Transmission and reflection from thin layers are interference phenomena, and amplitudes of the electric fields must be combined before squaring. The amplitude of a reflected wave, which should be thought of as the length of an arrow (vector), is given by r (not R) in Eq. 2.21 and each arrow has an orientation. [204] A single layer anti-reflective coating is illustrated in Figure 2.31. (Note: the arrows in this figure indicate the directions of light rays; the associated amplitudes denoted by the r values are different quantities.)

Suppose that $\eta_0 = 1.0$ (air) and $\eta_2 = 1.5$ (glass) in Figure 2.31. The coating layer must have a refractive index less than 1.5 (for there to be the same change in phase angle or turn of the amplitude arrow at the two surfaces). Plus, for effective cancellation of reflection, it is necessary that $r_1 = r_2$. This condition can be combined with the equation for r to show that the optimum refractive index for the coating layer is $\eta_1 = \sqrt{\eta_0 \eta_2}$, or in this case 1.225. We have a situation in which there are two ways for a photon to get from the source to the observer, and each of the paths can be represented by an arrow with a length and an angle. [204] These arrows must be determined, added together, and then squared to determine the intensity of the reflected light.

The lengths of the arrows are essentially equal because of the choice of η_1, and the difference in orientation (angles) depends only on the difference in the optical path lengths. The orientation of the arrow associated with r_1(path 1) is arbitrarily set at 12:00 o'clock, and the orientation of the arrow for path 2 is different only because of the time required for the photon to traverse the layer of thickness d_1 twice with the effective speed c/η_1 . A complete rotation of the arrow through 360°

occurs each time the path length increases by one wavelength; and with our choice of layer thickness, the arrow for path 2 is oriented at 6:00 o'clock or 180° out of phase with the arrow for path 1. The two arrows, when placed head to tail, add to zero, and there is complete cancellation of the reflected light at the wavelength λ_0. Since the sum of the intensities of the reflected and transmitted light must equal the intensity of the incident light, this coating permits 100% of light at λ_0 to be transmitted. We have simplified the example here by assuming the angle of incidence is close to zero and the amplitudes of reflection are very small, and by neglecting multiple reflections in the layer. Corrections for these simplifications are well known and do not change the qualitative picture.

Of course, things are not perfect because the single layer is only efficient at cancelling reflection at one wavelength. With a two-layer coating it is possible to zero out reflection at two wavelengths, a three-layer coating can be designed to cancel reflection at three wavelengths, and so on. Approximate analyses of these situations are quite easy with the vector method illustrated here.

From this discussion it is easy to see why modern lenses are so complicated. It should also be noted that the discussion thus far applies to fairly simple prime or non-zoom lenses. Also, most modern lenses contain autofocus mechanisms and in some cases vibration isolation systems that move a lens element to counteract low frequency camera vibrations. These important technological refinements are beyond the scope of this book, but some of the optical characteristics of compound lenses (e.g., primary surfaces and nodal points) will be considered in later chapters.

2.4 Cameras, Rays and Radiance

To gather any measurable amount of light power, a pixel detector (or photoreceptor in our eye), must gather a four-dimensional bundle of rays—rays that cover a non-zero solid angle for all points in a non-zero area. Accordingly, when we focus the lens in a camera (or our own eyes) on a screen point P, we gather together a sizeable cone of outgoing rays from P, and make them converge at point Q on the detector. This cone-shaped bundle provides nonzero irradiance, but only at Q. For measurable amounts of light power, a point is not enough: we need an irradiated area. Accordingly, a single-pixel detector (or single photoreceptor in our eye) measures power delivered to a small neighborhood area around Q from a small neighborhood area around P in the scene, and each point in this 2D neighborhood delivers an infinitesimal amount of power through its own 2D cone-shaped bundle of rays collected through the lens.

The seemingly erroneous inclusion of cosine falloff makes good sense if we reexamine imaging for these small neighborhoods around P and Q. We begin with our camera placed directly above the display screen, and assume the lens performs no magnification, so that the neighborhoods around both P and Q have the same size and shape. Additionally, we assume that the incident angle θ_i is zero or nearly zero for all the rays it gathers. Because a camera's lens aperture covers only 0.005 steradians out of the span of 2π directions for light leaving the screen in the neighborhood of point P, then only 1/1256.6 of this light reaches the detector at point Q. More formally, the lens maps the screen's uniform radiant exitance $W/2$ watts/meter^2 to a detector irradiance of $W/2513$ watts/meter^2.

Now suppose we move the camera to the side, so that its rays to point P form incident angles at or near 60°. From this viewpoint, the radiance for each ray has fallen to half its previous value—yet the detector irradiance at Q stays constant! To see why, trace rays from the corner points of the neighborhood of Q through the lens and back onto the screen. As the screen plane tilts, more and more of its surface area falls within the frustum viewed by the one-pixel area around Q, exactly counteracting the cosine falloff of radiance. Thus the cosine term in radiance measurements ensures that uniformly emitting surfaces such as our display screen produce the same pixel values from any viewing angle.

Suppose we want to measure the light that a back-lit photo sends to our eyes, or perhaps the light from a perfect, artifact-free CRT or LCD screen. Like many computer graphics and imaging publications, we assume that a digital image consists of a grid of pixels at integer locations in a plane, and each pixel describes one ray of light from the pixel's location in the image plane through the center of projection of a projector or camera. A pixel is not "a little square" [372] but an infinitesimal point on a continuous image, and the pixel's RGB value describes the color value of the image at that infinitesimal point. We don't really know the image values between the points, but displays construct a continuous image by using pixel values as estimates of neighborhood values.

Instead of imagining a digital image as a grid of little squares, think of the pixels as equally spaced point-source lights covered by a frosted-glass plate. Each pixel's RGB value sets the radiant flux of one point source, and the frosted glass disperses that light over a small neighborhood, forming a continuous image as a sum of blobs of light formed by the frosted-glass scattering function, also known as its *point-spread function.* While some digital micromirror device (DMD) projectors do use little squares as their point-spread functions, the highest quality and most visually uniform point-spread functions are blob-like with slightly negative side-lobes. [274]

Further text after this point to be written and included

Chapter 3

Epsilon Photography

In this chapter we continue to think of photographs, whether captured digitally or on film, as fixed and static records of a viewed scene, and straightforward copies of the 2D image formed on a plane behind a lens. How might we improve photographs from traditional cameras if we apply unlimited computing, storage, and communication to them? The past few years have yielded a wealth of new cameras and enhanced imaging opportunities, and we have started to see new and exciting results. How can these new imaging and computing opportunities continue to improve conventional forms of photography?

Currently, adjustments and trade-offs dominate film-like photography, and most camera decisions are locked in once we press the camera's shutter release. Poor choices lead to poor photos, and an excellent photo may be possible only for a narrow combination of settings taken with a shutter-click at just the right moment. Can we elude these adjustments and trade-offs? Can we defer choosing the camera's settings somehow, or change our minds and alter the settings later? Can we compute new images that expand the range of settings, such as a month-long exposure time? What new flexibilities might allow us to take a better picture now, and also keep our choices open to create an even better one later?

We need to broaden our thinking about photography to avoid missing opportunities. So many of the limitations and trade-offs of traditional photography have been with us for so long that we tend to assume they are inescapable, a direct consequence of the laws of physics, image formation and light transport. We are misled by our strong beliefs in how photography is done. Surely every photo-making process has to employ a high-quality optical system for high-quality results. Surely any good camera must require accurate focusing, an appropriate focal length, a good point of view, and the best framing of the subject scene. To achieve the results we aspire to, surely we must choose our exposure settings carefully, seek out the optimal trade-offs among ISO sensitivity, digital noise, and the length of exposure needed to capture a good image. Surely we must keep the camera stable as we aim it at our subject. Surely we must match the color balance of our sensor (or film) to the color spectrum of our light sources, and later match it to the color spectrum of our display device. Surely we must choose appropriate lighting and pose the subject for the most flattering appearance (and say ”cheese!”). Only then are we ready to click the shutter. Right?

Well, no, not necessarily, not any longer. We can break each of these conventions with computational methods. The technical constraints change radically for each of these conventions if we're allowed to combine results from multiple photographs and/or multiple cameras. This chapter points out some of those assumptions, describes a few current alternatives, and encourages you to look for more.

A few inescapable limits, though, do remain:

- We cannot measure infinitesimal amounts of light, such as the strength of a single ray, but instead must measure a bundle of rays; a group that impinges on a non-zero area and whose directions span a non-zero solid angle.
- We cannot completely eliminate noise from any real-world sensor that measures a continuum of values (such as the intensity of light on a surface).
- We cannot create information about the scene not recorded by at least one camera.

Beyond these basic irreducible limits, we can combine multiple photographs to substantially expand nearly all the capabilities of film-like photography.

3.1 Epsilon Photography

This is a single-strategy chapter. Because existing digital cameras are already extremely capable and inexpensive, here we will explore different ways to construct combined results from multiple cameras and/or multiple images. By digitally combining the information from more than one image, we can compute a picture superior to what any single camera could produce. We can also create interactive display applications that let users adjust and explore settings that were fixed in film-like photography.[1]

This strategy is a generalization of *bracketing*, an exposure technique that is already familiar to most photographers. Bracketing lets photographers avoid uncertainty about critical camera settings such as focus, exposure, or white balance. Instead of taking just one photo at what we think are the correct exposure settings, we make additional exposures at several higher and lower settings that bracket the chosen one. If our first best-guess setting was not the best choice, the bracketed set of photos almost always contains a better one. Perhaps the most common use for bracketing is in situations where the camera's in-built metering might get confused. In this mode the camera automatically captures additional photos that

[1] HDRShop from Paul Debevec's research group at USC-ICT (projects.ict.usc.edu/graphics/HDRShop) helps users construct high-dynamic-range images from bracketed-exposure image sets, then lets users interactively adjust exposure settings to reveal details in brilliant highlights or the darkest shadows; Autostitch from David Lowe's group at UBC (www.cs.ubc.ca/~mbrown/autostitch/autostitch.html) and AutoPano-SIFT (user.cs.tu-berlin.de/~nowozin/autopano-sift/) let users construct cylindrical or spherical panoramas from overlapped images; and HD View (research.microsoft.com/ivm/hdview.htm) from Microsoft Research allows users an extreme form of zoom to explore high-resolution panoramas, varying smoothly from spherical projections for very wide-angle views (e.g., > 180 degrees) to planar projections for very narrow, telescopic views (< 1 degree).

Application	Epsilon over **sensors**	Epsilon over **time**	Epsilon over **pixels**
	Parameter Major	Time Major	Position Major
	Cost/space trade-off	Time trade-off	Sensor resolution trade-off
	Space uniform; parameter varying	Time uniform; parameter varying	Space varying
HDR	SAMP (different ND filter)	Bracketing	Assorted Pixels
Color	3 CCD	Color wheel in the aperture	Bayer mosaic; Foveon
Field of View	Camera Array with telephoto lens	Unstructured panoramas	
Resolution		Shift based superresolution	
Depth of Field	SAMP (focused at different depths)	Focal Stacks	
Frame Rate	Camera Array [Irani]		Mosaic (virtual exposure)
Noise Reduction	SAMP (same camera settings)	Multiple photos	

Table 3.1: Epsilon photography: applications and approaches. [THIS TABLE COULD HAVE AN ADDITIONAL COLUMN FOR EPSILON OVER MULTIPLE AXES]

straddle the camera's exposure estimate. A similar technique is used for assembling high dynamic range images.

The methods in this chapter are analogous but often use a larger set of photos. This is because multiple settings may be changed and we may digitally merge desirable features from multiple images in the set rather than simply select just one single best photo. The process of capturing the scene multiple times with one or more parameters, slightly modified each time, and creating a composite image that contains information from all individual images, is called *epsilon photography.* While bracketing is usually done by changing a single parameter and capturing multiple photos over time, epsilon photography encompasses changing parameters by epsilon (i.e., a small amount) over (a) multiple sensors, (b) time, (c) pixels, or (d) multiple axes. Most applications of epsilon photography involve capturing low-level scene information (i.e., pixel intensities). Table 3.1 shows several common applications of epsilon photography produced by changing parameters over the three axes—sensors, time, and pixels. While it would seem as though any application can be solved by using either of the three approaches, some entries are empty either because they do not make much sense, or because no feasible solution yet exists. Trade-offs in epsilon photography improve one measurable aspect of a photograph at the expense of another. We discuss these trade-offs in greater detail for each of the three axes over which parameters are varied.

3.1.1 Epsilon over Sensors

Epsilon over sensors is perhaps the most brute-force approach to capturing images with varying parameters. It simply involves using multiple image sensors with different parameters to capture the same scene at the same time. This is achieved

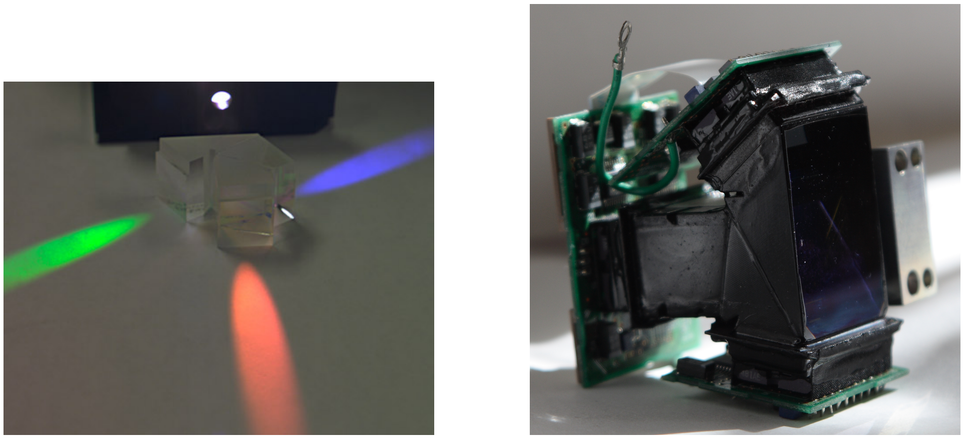

(a) (b)

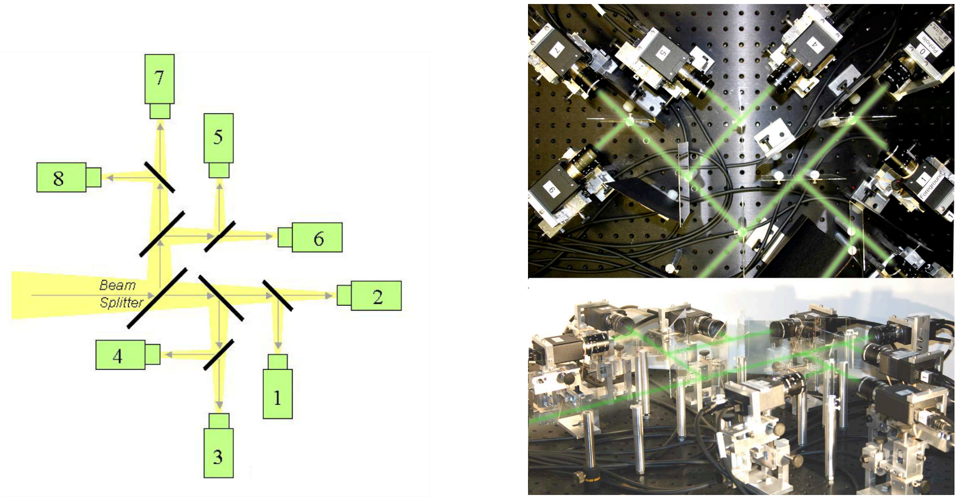
Figure 3.1: Trichroic beam splitter prism used to split incoming light into its color components for imaging with multiple sensors in a typical three-CCD setup. (Figures from wikipedia (en.wikipedia.org/wiki/Three-CCD).

by using clever optics to effectively co-locate the sensors, or by using adjacent synchronized cameras as in a camera array, or by using some other arbitrary camera arrangement.

Three-chip cameras (more common for high-quality video applications than for still photos) use a dichroic prism assembly behind the lens to split the image from the lens into three wavelength bands for three separate image sensors (see Figure 3.1). Three images are captured simultaneously and combined together to create a single color image.

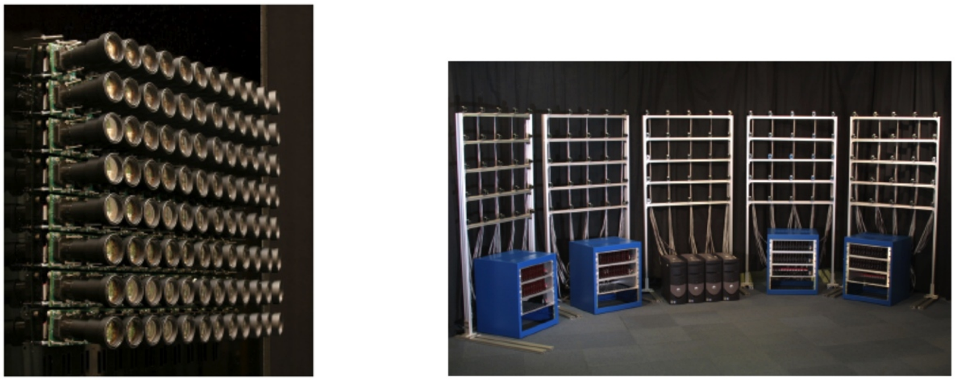
The three-CCD concept can be generalized and extended to a larger number of camera sensors, each with its unqiue parameter settings. The single-axis multiparameter (SAMP) camera [265] uses an optical splitting tree as shown in Figure 3.2; it is perhaps the most generalized version of epsilon over sensors. This arrangement uses a series of beamsplitters and cameras that can simultaneously capture pixel-aligned images for various epsilon photography applications such as HDR capture, focal stacks, and high frame-rate capture with staggered exposures. Changing parameters over various sensors is accomplished by placing color or neutral density filters in their optical path.

A closely related approach is the use of an array of cameras where each camera uses a different set of parameters to capture the same scene (see Figure 3.3). This approach is different from SAMP because each camera's spatial location (and possibly also view direction) is unique. This can be thought of as yet another parameter that is varied among the sensors. Wilburn et al. [427] built an array of one hundred relatively inexpensive video cameras, and demonstrated epsilon photography applications such as enhanced spatial resolution, high dynamic range, and high frame rate. They also used the array to achieve synthetic apertures and scene refocusing. In a related commercial application, ViewPLUS Inc. manufactures a 5 × 5 camera



(a) (b)

Figure 3.2: Single-axis multi-parameter (SAMP) camera captures multiple photos with different parameters at the same time. (Figures from McGuire et al. [265].)



(a) (b)

Figure 3.3: Camera array setups from Stanford University can capture multiple photos simultaneously from slightly different viewpoints for epsilon changes in parameters over sensors. (Figures from Wilburn et al. [427].)

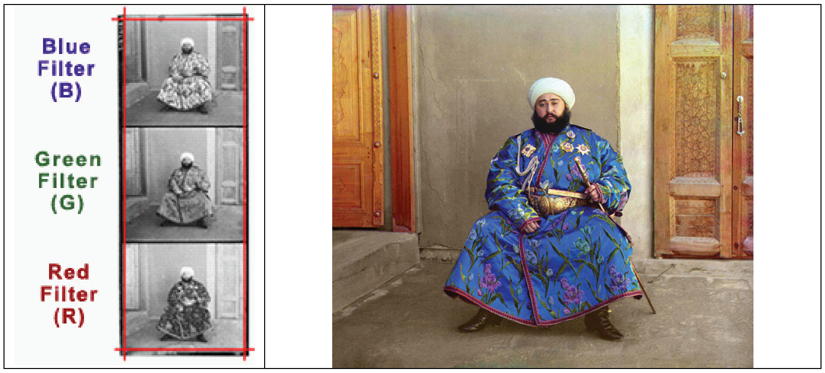


Figure 3.4: Russian photographer Prokudin-Gorskii captured scenes of Tsarist Russia with a custom-built, sort-first, time-multiplexed camera that captured three color-filtered images in rapid succession on a tall, single-plate negative [326].

array called the ProFUSION 25, which captures images at a resolution of 640 × 480 at 25 frames per second.

While this epsilon-over-sensors approach is the easiest to understand, the resulting cameras and arrays are usually quite cumbersome. This approach also typically costs more than others because of the additional sensors and cameras. In spite of these limitations, camera arrays are popular for capturing light fields and for synthetic aperture applications, as discussed in Chapter 4.

3.1.2 Epsilon over Time

In the epsilon-over-time approach, we capture a sequence of photographs with one camera, with different parameters settings for each. Each photo forms one complete image, taken with just one complement of camera settings. Each image is ready to use as output, and we need no further sorting of the image contents to construct a viewable output image (though we may still merge several photos to make the output even better). Bracketing of any kind is a good example of sort-first photography.

For example, in the early 1900s, commissioned and equipped by Tsar Nicholas II, Sergei Mikhailovich Prokudin-Gorskii (1863–1944) surveyed the Russian Empire in a set of beautiful color photographs gathered by his own method for color photography. His single-lens customized view camera took a rapid sequence of three separate photographs, each through a different color filter in front of the lens. In 2003, the U.S. Library of Congress digitized a large set of these negatives and merged them to construct conventional color photographs (see Figure 3.4 and website www.loc.gov/exhibits/empire/gorskii.html).

Another example of epsilon over time is capturing photographs for assembly into a panoramic image showing a 360-degree view from a single viewpoint. We mount a single camera on a tripod, use a lens with a field of view of D degrees, and take a time-multiplexed sequence by rotating the camera D degrees or less between

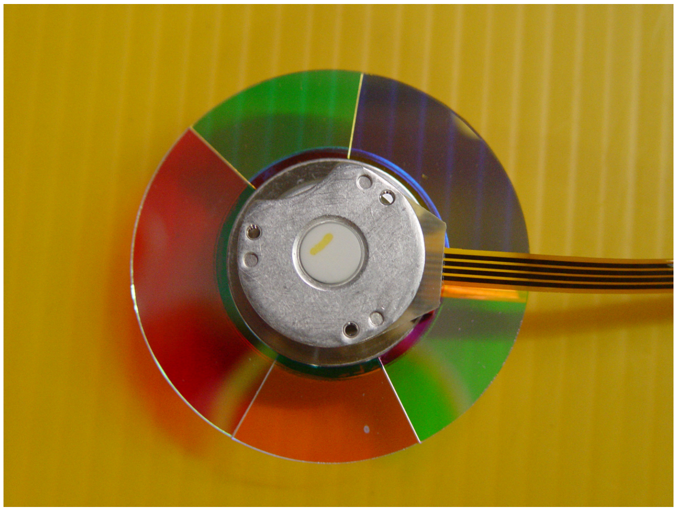


Figure 3.5: A color wheel with red, green, blue and clear portions, as used in a color DLP projector.

each exposure. With an unchanging scene and a camera with little or no radial distortion, we can gather a set of photographs that match each other perfectly in their overlapped regions, Any conventional panorama-making software will produce a good single image from this sequence. However, any movement or lighting changes within the scene during this process will introduce inconsistencies that are much more difficult to resolve. Clouds in the first photograph might not align at all with clouds in the last photograph, but alignment is not impossible. Tools such as Adobe Photoshop are suitable for manually resolving modest mismatches.

A similar approach is used in projectors that are based on Texas Instruments' DLP technology. These projectors use a color filter in front of a digital micromirror device (DMD). The DMD consists of an array of tiny controllable mirrors. Each mirror can be flipped in two distinct orientations, one of which corresponds to an "on" pixel, and the other to an "off" pixel. A synchronized color filter (see Figure 3.5) is used to project the red, green, and blue channels of the image sequentially. At any given time, an image of only one of the three color channels is projected. The phenomenon of persistence of vision in the human visual system aids in the perception of a natural-looking full-color image. Here's an experiment to try in a DLP movie theater. Take a photo with a digital SLR camera of the image projected by a DLP projector. If the shutter is fast enough, the camera will capture an image that has a single color (red, green, or blue), or even two colors for different parts of the image during a color wheel transition.

Varying parameters over time and capturing multiple photos is relatively easy and is possible with minimal changes to existing cameras. However, this technique assumes that both the camera and object are perfectly still while the multiple photos are captured. The illuminating light in the scene should also not change. Suppose we attempt to photograph a scene as clouds cover or reveal the sun during sort-

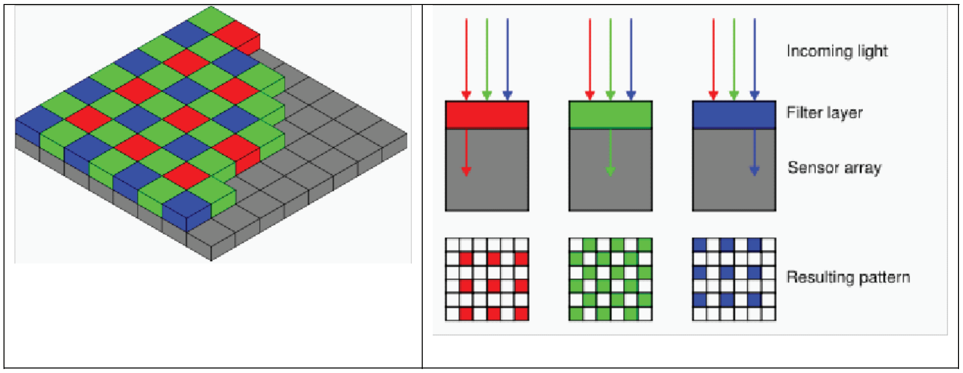


Figure 3.6: The Bayer color mosaic pattern in many modern digital cameras employs sort-last color sensing. Demosaicing techniques employ edge-following, estimation, and interpolation methods to approximate a full-resolution color image from these measurements. Alternatively, three-chip video cameras follow the sort-first method, and sense three complete, independent color images simultaneously. (Figure from wikipedia: en.wikipedia.org/wiki/Bayer_filter.)

first exposure bracketing. Our first high-exposure photo, taken before the sun goes behind clouds, appears overly bright, but our subsequent mid- and low-exposure photos are darker than they should be, due to falling light levels (the clouds will be moving as well). A situation like this yields no usable photos at all. In general, this approach requires a camera on a tripod, and non-moving or relatively slow changing scenes. Some recent applications of this technique allow for moderate camera and scene motion, and compensate for it using feature tracking and motion compensation.

3.1.3 Epsilon over Pixels

The epsilon-over-pixels approach mixes several different parameter settings within the pixels of a single photo. After capturing the photo, we must sort the contents of the photos, and rearrange and recombine them somehow to construct a suitable output image.

The Bayer color mosaic pattern found on nearly all single-sensor digital cameras is perhaps the most widely used example of the epsilon-over-pixels approach. Figure 3.6 illustrates how Individual, pixel-sized color filters cover adjacent pixels on the imaging sensor, forming a red, green, and blue filter pattern. Even though the sensor loses spatial resolution because of this multiplexing, we can measure all three colors at once and interpolate sensible values for every pixel location (a process called *demosaicing*) to give the impression of a full-resolution image with all colors measured for every pixel.

Unlike epsilon over time, the epsilon-over-pixels technique requires modifications to the image sensor, and may be harder to implement. Semiconductor manufacturing techniques can impose additional restrictions on the pixel parameters that may

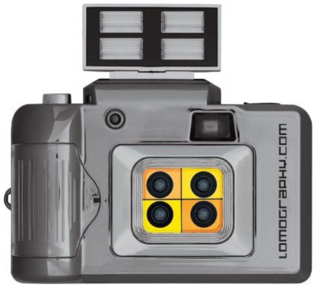


Figure 3.7: The Lomography Actionsampler Flash 35mm film camera uses multiple flashes and four lenses to capture multiple photos in quick succession. Each photo is taken from a slightly different point of view, and at a slightly different time.

be modified by using this technique. Additionally, the effective pixel resolution is reduced by a factor of the number of parameter settings used. On the other hand, multiple simultaneous measurements in a single photo make this method less susceptible to scene variations over time, reducing the chance that a transient scene value will escape successful measurement.

3.1.4 Epsilon over Multiple Axes

The classification on the basis of the axis along which parameters are changed is not rigid. Hybrid systems of video cameras or still cameras enable capture of each step of a complicated event over time in order to understand the event better, whether captured as a rapid sequence of photos from one camera (a motion picture), a cascade of single photos taken by a set of cameras, or something in between. The Lomography Actionsampler Flash 35mm film camera shown in Figure 3.7 is essentially a 2×2 camera array where each camera takes a photo sequentially over time, thus effectively combining epsilon over sensors and epsilon over time.

Schechner and Nayar [356] rigidly attached a spatially varying mask some distance in front of the camera lens (as shown in Figure 3.8). Since the mask is not the limiting aperture of the optical system, different scene points are attenuated differently by the mask. Moving the camera-mask setup yields multiple measurements for each scene point under different optical settings, resulting in image mosaics with additional scene information such as extended dynamic range and multispectral data. This technique is called *generalized mosaicing.* The registration algorithm is non-trivial because of spatially varying effects of the filter. A vision-based algorithm [358] synchronizes a changing mask in the optical system to the corresponding acquired image, thus allowing for uncontrolled modulation of the imaging system.

3.2 Improving Dynamic Range

The sensors in digital cameras have a limited input range. They cannot record image details in bright highlights and dark shadows at the same time. Too much

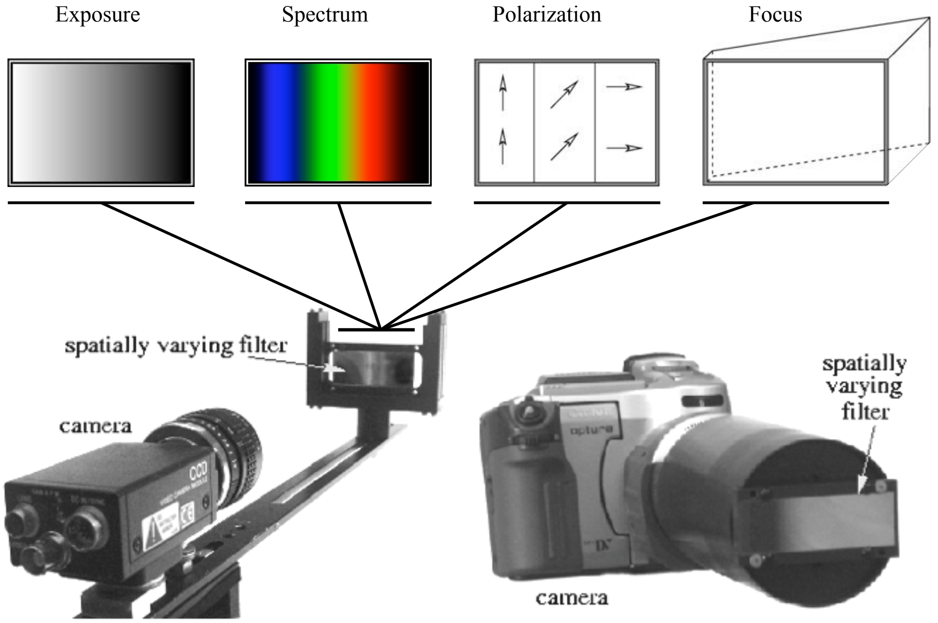


Figure 3.8: Setup for generalized mosaicing [356], which is a simple and effective method for extracting additional information at each scene point. As a camera moves, it senses each scene point multiple times. An optical filter with spatially varying properties is attached in front of the camera lens so that each scene point is measured multiple times but under different optical settings. (Figure from Schechner and Nayar [356])

light can overwhelm the sensor, ruining the image with a featureless white glare, while too little light can leave image features indistinguishably lost in shadowy darkness. Most film-like cameras provide automatic metering and exposure options to match the camera's overall light sensitivity to the amount of light in a viewed scene. These options include adjusting lens aperture to control the light admitted through the lens (which affects the depth of field), adjusting exposure time (which can result in motion blur), placing neutral density filters in front of the lens (which results in longer exposure times), or adjusting the ISO sensitivity of the sensor itself (which can add digital noise to an image). Despite such trade-offs, these options combine to give modern cameras an astoundingly wide range of sensitivity to light, rivaling or exceeding that of the human eye (which adapts to over 16 decades of light intensity from the absolute threshold of vision at about 10^{-6}cd/m^2 up to the threshold of light-induced eye damage near 10^{8}cd/m^2).

Camera adjustments for light sensitivity still aren't enough, however, to match the ability of the eye to sense the wide variations in light intensity of a high-contrast scene. Scenes with bright highlight regions and dark shadow regions are difficult to photograph well. The intensity ratio between the brightest and darkest regions overwhelms the ability of the camera sensor to record these intensities accurately. The sensor cannot capture details of the darkest blacks and the brightest whites simultaneously. Troublesome high-contrast scenes often include large visible light sources aimed directly at the camera, scenes with strong backlighting and deep shadows (for example, a dimly lit indoor scene with a background window showing a brightly lit exterior scene), or scenes with reflections and specular highlights, such as shown in Figure 3.9. Film-like photography offers us little recourse other than to add light to the shadowy regions with flash or other fill lighting. Rather than adjust the camera to suit the scene, we adjust the scene to suit the camera!

Unlike a camera's adjustable sensitivity to light, a camera's maximum contrast ratio, known as its *dynamic range*, is not adjustable. Formally, dynamic range is the ratio between the brightest and darkest light intensities a camera sensor can capture in a single image without losing image detail. In other words, this is the maximum intensity ratio between the darkest detailed shadows and the brightest textured brilliance, as shown in Figure 3.9. No one single sensitivity setting (or *exposure* value) will suffice to capture a high dynamic range (HDR) scene that exceeds the camera's contrast-sensing ability.

The lens and the sensor together limit the camera's dynamic range. In a high-contrast scene, unwanted scattering of light within complex lens structures causes glare and flare effects. The degree of these effects depends on the image itself, but traces of light from bright parts of the scene typically leak into dark image areas, washing out shadow details and limiting the maximum image contrast the lens can form on the sensor. The maximum contrast is usually between 100,000:1 to 10 million:1 [264, 390]. The dynamic range of the sensor itself (typically $< 1000 : 1$) imposes further limits. Device electronics (e.g., charge transfer rates) set the upper bound on the amount of sensed light, and the least amount of light distinguishable from darkness is set both by the sensitivity of the sensor and by its *noise floor*, which is the combined effect of all the camera's noise sources (quantization, fixed-pattern, thermal, EMI/RFI, and photon arrival noise).

The range of visible intensities in many scenes often exceeds the ability of cam-

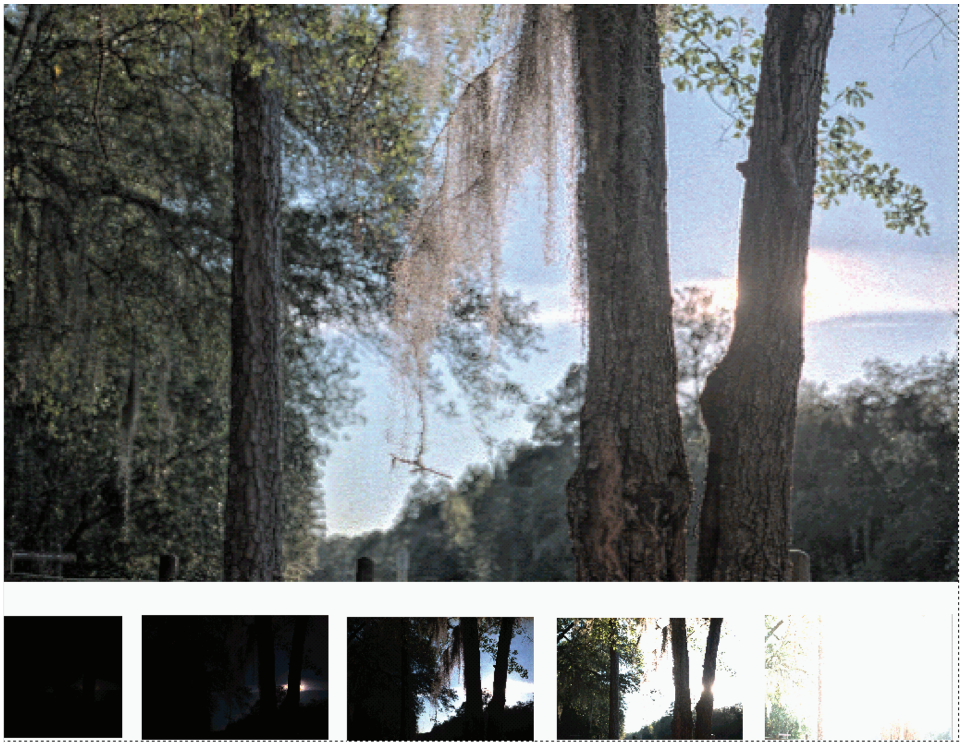


Figure 3.9: Tone-mapped high dynamic range (HDR) image from [82]. The high contrast of outdoor scenes such as this can easily exceed the dynamic range of most digital cameras. The bottom row shows the original scene intensities scaled by progressive factors of ten. Scene highlight intensities in the clouds are approximately 10,000 times brighter than shadow details in the forest. This contrast range is well beyond the typical 1000:1 dynamic range of conventional CMOS or CCD camera sensors.

eras to record the scene and the ability of displays to represent the scene. When plotted on a logarithmic scale (where distance depicts ratios, and each tic mark represents a factor-of-10 change), the range of human vision spans about sixteen decades, while typical film-like cameras can capture no more than five or six decades and displays can represent no more than two or three decades. For the daylight-to-dusk (photopic) intensities (upper two-thirds of scale), humans can detect contrast differences as small as 1–2% (1.02:1, which divides a decade into 116 levels $(1/log_{10}1.02)$). Accordingly, 8-bit image quantization with a maximum of 256 levels is barely adequate for cameras and displays whose dynamic range may exceed two decades (100:1). Many cameras and displays use 10, 12, or 14-bit internal image representations to avoid visible contouring artifacts.

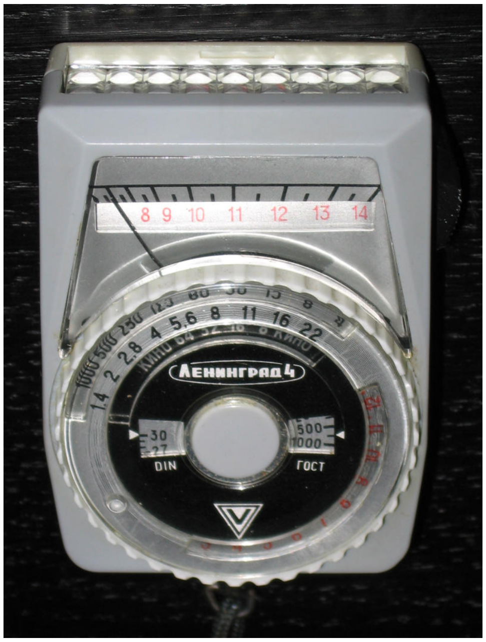


Figure 3.10: Analog Incident Light Meter (source: wikipedia)

3.2.1 Exposure Metering

Photographers have long recognized the importance of correct exposure. Experienced photographers often use the $f/16$ rule (also called the sunny-16 rule) to estimate exposure without using a light meter. The rule is as follows; on a sunny day, set the lens aperture to $f/16$ and the shutter speed to 1/ISO, where the ISO value is the exposure index of the film or the digital sensor. This reliable rule, together with luck and experience, has produced many iconic photographs, particularly in street photography and photojournalism. The rule results in acceptable exposures for black-and-white film and some color print films with wide exposure latitude. Slide film and digital sensors, however, are less forgiving of exposure errors, so the $f/16$ rule is not ideal. Fortunately, nearly all current cameras rely on an in-camera light meter to estimate the light values in a scene, allowing the camera to automatically determine the shutter speed and aperture setting for appropriate exposure.

There are two general types of light meters: reflected-light meters and incident-light meters. Reflected-light meters measure the light reflected by objects the scene. All in-camera meters are reflected-light meters. A spot meter measures the reflectance from a very small part of the scene (a subtended angle of 1° or less at the light meter), while a center-weighted meter takes the average of reflectances from a larger central portion of the entire field of view. In either case, the meter is calibrated to show the appropriate exposure for an average scene (typically an 18% gray scene). A scene with higher reflectance, such as a scene with snow, sand, or specular highlights, would affect the average reading of a light meter and lead to underexposure in the image. An incident-light meter that integrates the light arriving at a scene point can measure scene light values more accurately, and is less likely to lead to incorrect exposures, but it is inconvenient since it requires placing the meter at the scene point prior to capturing the photo.

The American photographers Ansel Adams and Fred Archer developed the Zone System as a means of determining optimal exposure with the use of a handheld or in-camera light meter [32]. The Zone System is essentially an enhanced version of the sunny-16 rule that relies on the photographer's experience to estimate exposure. In the Zone System, measurements are made of individual scene elements, and exposure is adjusted based on the photographer's knowledge of what is being metered (a photographer knows the difference between freshly fallen snow and a black horse, while a meter does not). Many books have been written on the Zone System, but the concept is simple—render light subjects as light, and dark subjects as dark, according to the photographer's visualization. The Zone System assigns numbers from 0 through 10 to different brightness values, with 0 representing black, 5 middle gray, and 10 pure white; these values are known as zones. To make zones easily distinguishable from other quantities, Adams and Archer used Roman rather than Arabic numerals. Strictly speaking, zones refer to exposure, with a Zone V exposure (the meter indication) resulting in a mid-tone rendering in the final image. Each zone differs from the preceding or following zone by a factor of two, so that a Zone I exposure is twice that of Zone 0, and so forth. A one-zone change is equal to a one-stop difference in exposure, corresponding more closely to standard aperture and shutter controls on a camera. Evaluating a scene is particularly easy with a light meter that indicates exposure value (EV), because a change of one EV is equal to a change of one zone.

In 1983, Nikon introduced *matrix metering* in the Nikon FA camera. This technology was perhaps the first commercial implementation of multizone metering. Since then all camera manufacturers have implemented a similar metering technique, but call it by different names (such as evaluative metering, honeycomb metering, and multisegment metering). The camera measures reflected light intensity at several distinct points in the scene, and intelligently combines the results to find optimal exposure settings. The number of points or zones varies from five to several thousand, depending on the camera. Newer cameras measure light in different color channels, and also incorporate information about focusing point, focus distance, scene mode, and the presence or absence of backlight to estimate the exposure. Current multizone metering algorithms are sophisticated, and produce nearly perfect exposure in most circumstances, freeing the photographer to better explore more creative aspects of the image-making process.

3.2.2 Capturing High Dynamic Range

Film-like photography is frustrating for high-contrast scenes because even the most careful attention to exposure settings will not allow us to capture the visible contents of a whole scene in a single picture. Exposure sensitivity that is set to reveal shadow details will cause severe overexposure in brightest parts of the scene; exposure sensitivity that is set to capture the brightest highlight details will cause severe underexposure in the darkest parts of the scene. Fortunately, several practical methods are available that allow a photographer to capture all the scene contents in a usable way.

High dynamic range (HDR) imaging is a rich and active area of research. We refer the reader to existing books (such as Reinhard et al. [336]) and recent surveys [53] for a more complete treatment of the subject. [WE COULD ADD A FEW MORE REFERENCES HERE, INCLUDING A REFERENCE TO THE UPDATED 2010 VERSION OF THE REINHARD BOOK]

HDR by Multiple Exposures.

The sort-first approach is highly suitable for capturing HDR images. [EXPLAIN WHAT YOU MEAN BY "SORT-FIRST"] To capture the widely varying intensities of light in a high dynamic range scene, we stabilize the camera on a tripod and take multiple images at different exposure settings, and then merge these images in imaging software. In principle, the merge is simple; we divide the pixel value of each pixel by the light sensitivity of the camera as it took that picture, and combine the best estimates of scene radiance at that pixel for all the pictures we took, ignoring badly overexposed and underexposed images.

This form of image merging, which is quick to compute, found widespread early use as *exposure bracketing* [282, 72, 252, 404]. Many of these methods assumed the linear camera response curves typically found on instrumentation cameras. However, most digital cameras intended for photography introduce intentional nonlinearities in their response to light, often mimicking the S-shaped log-log film response curves called Hurter-Driffield or H-D curves. These H-D curves enable digital cameras to capture a wider usable range of intensities, which provides a visually pleasing representation of HDR scenes, even at the extremes of overexposure and underexposure. Some authors have proposed acquiring images at different exposures to estimate the radiometric response function of an imaging device and then use the estimated response function to process the images before merging them [255, 91, 275]. This approach has proven robust and is now widely available in commercial software tools (Adobe Photoshop, CinePaint) and open-source projects (HDRShop (www.hdrshop.com), PFStools (www.mpi-inf.mpg.de/resources/pfstools), and others).

HDR by Exotic Image Sensors

While easy and popular for static scenes, exposure bracketing methods are not the only option available for capturing HDR scenes. They are particularly unsuitable for scenes that vary rapidly over time. In later chapters we will explore exotic image sensor designs that can sense higher dynamic range in a single exposure. They include logarithmic sensors, pixels with assorted attenuation [287], multiple sensor designs with beam splitters, and gradient-measuring sensors [406]. In addition, we will explore techniques for dealing with high dynamic range scenes with video cameras [109] or for capturing panoramas with panning cameras via attenuating ramp filters [107, 111].

3.2.3 Tone Mapping

Tone mapping is used to display the contents of a high dynamic range (HDR) image on a device or material of limited dynamic range, such as a computer monitor, an LCD display, or paper. By finding an optimal tone mapping, all or most of the HDR image information can be represented on the limited device. Conversely, when low dynamic range (LDR) images are to be displayed on HDR devices, we need to reversely convert the LDR content to HDR content in a visually convincing way. In both cases, human perception is a key factor to be incorporated.

Dynamic range compression

Certain tone mapping techniques, used primarily in photography, compress the extended high dynamic range of a real scene into the relatively limited contrast range of a device or medium. Some tone mapping techniques are global and act equally on all the pixels in the image, while other tone mapping techniques are local and manage tonal values and contrast in selected portions of the image.

Global Manipulation of Dynamic Range

Global operations to compress high dynamic range are spatially uniform non-linear functions that are based on the luminance and other global variables of the image. These variables are used to estimate an optimal transfer function for the image, such that every pixel in the image is mapped by this transfer function into new values, independent of the values of surrounding pixels. These techniques are simple, quick, and easily implemented, but they often result in a decrease in image contrast. Examples of common global tone mapping methods are brightness adjustment, contrast reduction, and color inversion. The well-known gamma transfer function, or gamma correction, is a typical global operator.

Local Manipulation of Dynamic Range

Local operations to compress high dynamic range are spatially varying non-linear functions that are determined at each pixel, according to local features extracted from the image parameters of surrounding pixels. Local tone mapping algorithms are more complicated to implement than global techniques. They often result in image artifacts such as ringing, which produces a thin bright line (the halo effect) at well-defined transitions in contrast. The output from local operations can look unrealistic, but they often provide the best overall image correction, since human vision is mainly sensitive to local contrast.

Local manipulation methods are content aware, and thus more sophisticated than global methods, but they typically require more computation. Some examples of local manipulation methods are gradient domain manipulation, bilateral filtering, and constraint propagation approaches.

Gradient domain manipulation methods, such as those developed by Fattal et al. [130] and Mantiuk et al. [256], concentrate on preserving the contrast between neighboring regions rather than adjusting the absolute magnitude of tonal value. This approach is motivated by the fact that human perception is more sensitive to

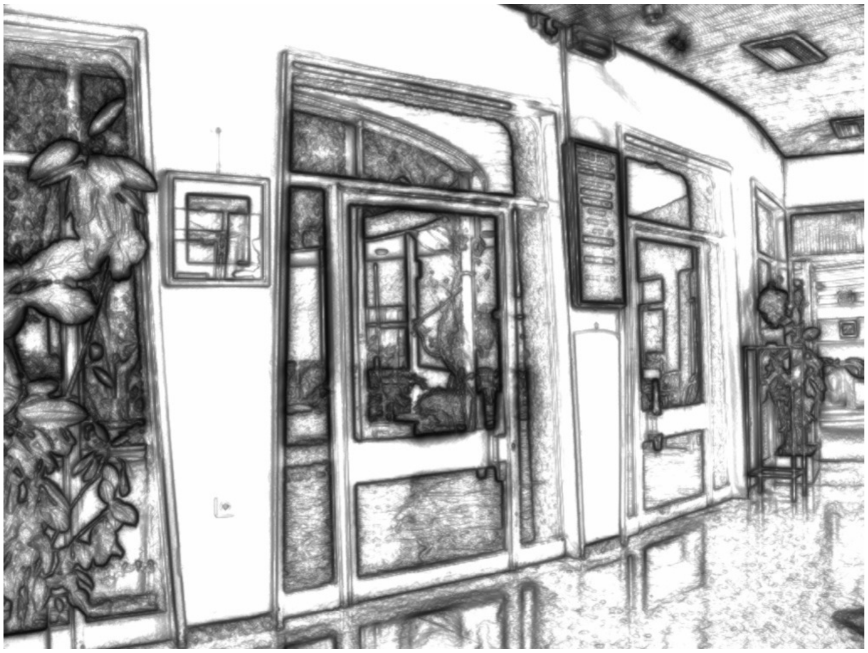


Figure 3.11: Gradient attenuation factors are computed locally for the Belgium House HDR radiance map [130] and are used to compress the tonal values of the HDR image. Darker shades indicate smaller scale factors, which leads to stronger tonal attenuation.[THE BEFORE AND AFTER IMAGES COULD BE HERE AS WELL]

changes in image contrast than to changes in absolute intensities. Fattal's method examines the gradient field of the luminance image and attenuates the magnitudes of large gradients. A new low dynamic range image is obtained by processing the modified gradient field of the high dynamic range image. The method compresses dynamic range (often drastically), while preserving fine image details and avoiding common artifacts such as halos, gradient reversals, or loss of local contrast. The method also enhances ordinary images by bringing out detail in dark regions. Figure 3.11 shows an example of an attenuator map for dynamic compression determined by this technique, made from the Belgium House HDR radiance map [130]. The darker the shade in this map, the stronger the attenuation in the high dynamic range image. [DO YOU HAVE THE BEFORE AND AFTER IMAGES FOR THIS ATTENUATOR MAP?]

Bilateral filtering methods proposed by Durand et al. [101] also reduce contrast in HDR images while preserving image details. Bilateral filtering decomposes the image into two layers—a base layer and a detail layer. The base layer is adjusted to reduce contrast, while the detail layer remains unaltered to preserve image details. Durand et al. then implemented a real-time bilateral filtering–based tone-mapping framework [311], as well as applied it to interesting photographic manipulations [50].

Constraint propagation approaches proposed by Lischinski et al. [249] can perform interactive tone mapping with an edge preserving property, which means the adjustments will be performed only on a local region, without leaking outside of the edge.

Mantiuk et al. [110] proposed an adaptive-tone-mapping operator for different display devices. The operator weights contrast distortions according to their visibility, as predicted by the model of the human visual system. [THIS SENTENCE IS TAKEN FROM THE PAPER ABSTRACT]]

Inverse—from Low Dynamic Range to High Dynamic Range

Inverse tone mapping techniques are needed now that HDR displays are available in the consumer market and when the HDR visual content does not have great dynamic range. This dilemma is similar to 3D displays and the problem of 2D-to-3D conversion. Rempel et al. [337] proposed a robust algorithm for converting legacy LDR video and photographs to HDR versions in real time, which can be played on HDR displays.

Reverse tone-mapping is an under-constrained problem, as a result, a reasonable evaluation system is needed for designing algorithms and improving results. Masia et al. [259] proposed a method for evaluating the reverse tone mapping algorithms on the basis of varying exposure conditions.

To be included?

Tumblin and Rushmeier [407]

LCIS by Tumblin and Turk [408]

Gradient domain HDR compression by Fattal et al. [129]

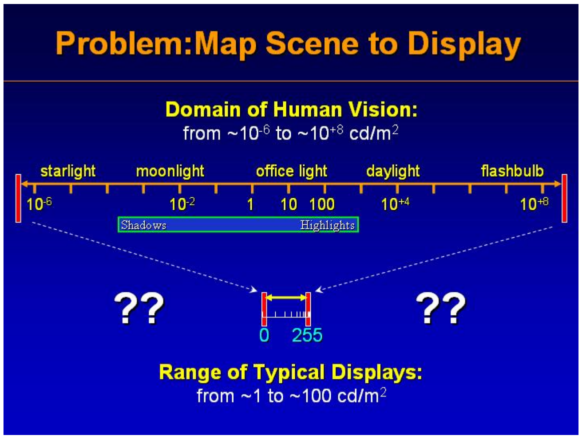
Bilateral filter; Trilateral filter

Ledda et al. [228] compared various tone mapping operators using a HDR display.

3.2.4 Compression and Display

An HDR image covers a dynamic range that is much wider than conventional eight-bit or ten-bit image file formats can express. A limited number of bits of information per color channel per pixel is inadequate to depict the full range of HDR light intensities in a camera sensor or display device (see Figure 3.12). Early file formats, using extravagant amounts of memory to represent light intensities, employed simple grids of floating-point pixel values. One popular solution used 8-8-8-8 bit pixels that featured a shared exponent E and 8-bit mantissas in a compact, easy-to-read RGBE format. This representation was devised by Greg Ward [421], and used in his photometrically accurate 3D renderer RADIANCE [420]. A psychophysically well-motivated extension to the format was proposed for the TIFF 6.0 image standard [226], which formed the basis for the slightly simpler format used by HDRShop. Later, the openEXR format developed in 2003 by Industrial Light and Magic and independent partners provided a simpler storage format, combining flexible bit-depth, compression capabilities, backwards compatibility, suitability for motion-picture workflows, computing platform independence, and open-source licensing. This format has gained widespread acceptance. [2].

[2] See www.openexr.com



Problem:Map Scene to Display

Domain of Human Vision:
from ~10^{-6} to ~10^{+8} cd/m^2

Range of Typical Displays:
from ~1 to ~100 cd/m^2

Figure 3.12: Visual dynamic range mismatch between a real life scene and a typical display or capture device.

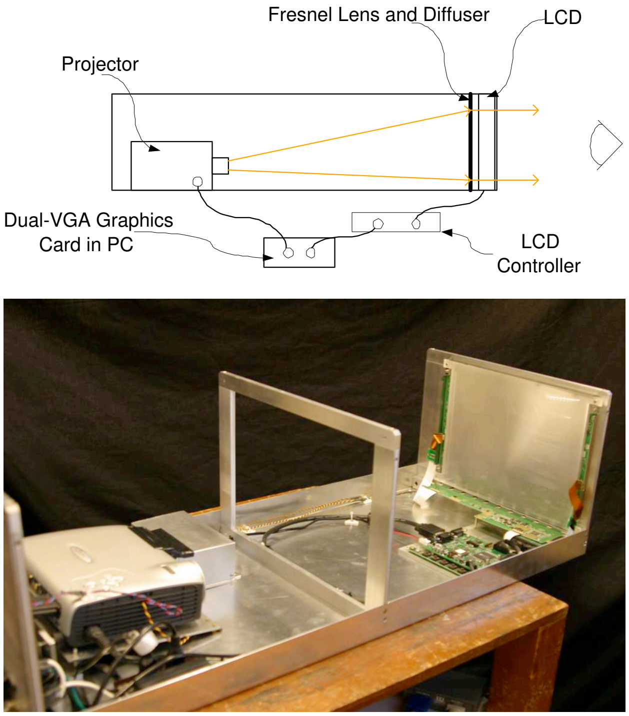


Figure 3.13: High dynamic range display demonstrated by Seetzen et al. [360] uses two LCDs placed one after the other to create higher contrast ratios than possible with a single LCD. (Figure from [360]).

Seetzen et al. [360] developed display devices capable of directly displaying HDR images without the use of tone mapping. They used two displays—a low resolution display behind a standard LCD panel. The overall contrast of the combined device is given by $(c_1 \cdot c_2) : 1$, where $c_1 : 1$ and $c_2 : 1$ are the contrast ratios of the two displays. In one prototype they used a low resolution array of high power LEDs behind a standard high resolution LCD to achieve a dynamic range beyond $50,000 : 1$. (See Figure 3.13).

3.3 Beyond Tricolor Sensing

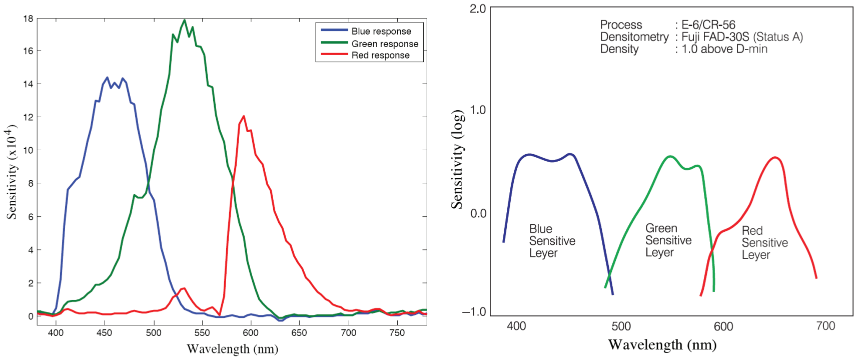
Existing photographic capture and reproduction methods mimic the well-understood trichromatic response of human vision. Three fixed color primaries—red, green, and blue (RGB)—are used to represent any color in the color gamut of the device. Unfortunately, fixed-spectrum photography limits our ability to detect or depict several visually useful spectral differences. In the common phenomenon of metamerism, for example, the spectrum of available lighting used to view or photograph objects can cause materials with notably different reflectance spectra to have the same apparent color because they evoke equal responses from the broad, fixed color primaries in our eyes or the camera. Metamers are commonly observed in fabric dyes where two pieces of fabric might appear to have the same color under one light source, and a very different color under another.

Fixed color primaries also impose a hard limit on the gamut of colors that a device can accurately capture or reproduce. As the well-known CIE 1931 color space chromaticity diagram illustrates [WE CAN ADD THE CIE CHROMATICITY DIAGRAM HERE AS A FIGURE], each set of fixed color primaries defines a convex hull of perceived colors within the space of all humanly perceptible colors. The device can reliably and accurately reproduce only the colors inside the convex hull defined by its color primaries. In most digital cameras, the fixed, passive Bayer RGB filter grid overlaid on pixel detectors sets the color primaries. Current digital micro-mirror device (DMD) projectors use broadband light sources passed through a spinning wheel that holds similar passive RGB filters. These filters compromise between narrow spectra that provide a large color gamut and broad spectra that provide greatest on-screen brightness.

At first glance an increase in the spectral resolution of camera, lights, and projectors might not seem to offer any significant advantages in photography. [PARAGRAPH IS INCOMPLETE. LIST THE ADVANTAGES HERE.]

3.3.1 Metamers and Contrast Enhancement

Black-and-white photographers often use yellow, orange, red, or green lens filters for specific visual effects in their images. Without filters, white clouds and blue sky are often rendered at roughly the same intensity in a black and white photograph. A yellow, orange, or red filter on the lens makes the sky appear progressively darker than the clouds, thus rendering sky and clouds as different tones on black and white film. A red filter essentially attenuates the wavelength corresponding to blue and green colors in the scene, thus creating strong tonal differences between the



(a) (b)

Figure 3.14: Comparison of the spectral response of a typical color film and digital camera sensor. (a) Spectral response of the Nikon D70 sensor [276]. (b) Spectral response of Fujichrome Velvia for Professionals color slide film [145]. [THE NIKON D70 IS A NINE-YEAR-OLD CAMERA. WE COULD FIND A NEWER SENSOR EXAMPLE.]

clouds and the sky in the resulting photograph. This is a classic case of effectively modifying the illumination to distinguish between metamers.

Unfortunately, photographers can carry only a limited number of filters with them. These filters are often broadband and useful for only standard applications. A camera that allows arbitrary and instantaneous attenuation of specific wavelength ranges in a scene would give a photographer increased flexibility. The camera could iteratively and quickly work out the best effective filter to achieve a metamer-free high contrast photograph for a given scene. Similarly, with an "agile" light source guided by our camera, we might change the illumination spectra enough to disrupt the metameric match. Or, we might interactively adjust and adapt the illuminant spectrum to maximize contrasts of a scene, both for human viewing and for capture by a camera.

[ARE CAMERAS WITH SUCH CAPABILITIES AVAILABLE COMMERCIALLY? IT SEEMS LIKE A PROPOSED CAPABILITY, RATHER THAN AN EXISTING EPSILON CAPABILITY. THIS SECTION ON TRICOLOR SENSING DOESN'T GO INTO ANY DETAIL ABOUT WHAT IS "BEYOND" TRICOLOR SENSING. THIS SECTION NEEDS MUCH MORE ATTENTION TO MAKE IT FIT BETTER IN THIS CHAPTER ON EPSILON PHOTOGRAPHY.]

3.4 Wider Field of View

Human vision provides us with seemingly endless visual richness and detail; the more we look, the more we see. We are almost never conscious of angular extent or the spatial resolution limits of our eyes, nor are we overly concerned with where we stand as we look at something interesting, such as an ancient artifact behind glass

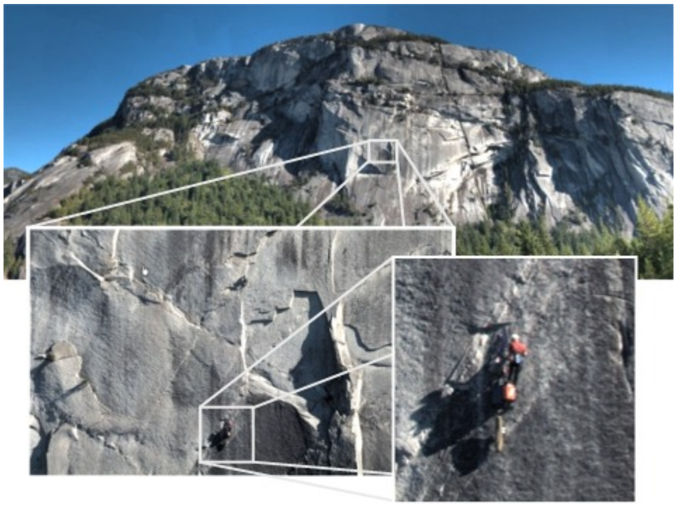


Figure 3.15: HDview—Capture and viewing of gigapixel panoramic images [220]. [CAPTION NEEDS MORE DETAIL ABOUT THIS IMAGE]

in a display case. Our visual impressions of our surroundings appear seamless, enveloping and filled with unlimited detail apparent to us with just the faintest bit of attention. Even at night, when rod-dominated scotopic vision limits spatial resolution, and the world looks dim, gray, and soft, we do not confuse a tree trunk with a telephone pole, or the distant grassy field beyond it.

Like any optical system, our imperfect lens and photoreceptor array offer little or no resolving ability beyond 60–100 cycles per degree, yet we have the experience clarity and visual detail at a much higher level. We infer that the edge of a knife blade is sharp and discontinuous. It appears disjoint from its background, and is not optically mixed with with surrounding details on even the most minuscule scale. Our impressions of our surroundings seldom lack subtlety and richness; we seek out mountaintops, ocean vistas, spectacular sunsets, and dramatic weather effects in part because the more we look around, the more variety we see in these visually rich scenes. With close attention, we almost never exhaust our eye's abilities to discover interesting visual details, including the fine vein structure of a leaf, the slow boiling formation of a thunderstorm, and the magnificently complex composition of luxurious fur on an animal. Of course we cannot see behind our heads, but we rarely have any sense of our limited field of view.

By comparison, camera placement, resolution, and framing are key governing attributes in nearly all photographs. How might we move past these current photographic requirements and limitations to achieve a more free-form visual record computationally? How might we construct a photograph to better achieve the impression of an unlimited and unbounded field of view, along with the visual richness revealed with little more effort than an intent gaze?

A panorama is a first-step solution to an unbounded field of view. To capture a panorama the photographer must point the camera at interesting parts of a larger scene while ensuring adequate overlap between adjacent captured photos. The overlapped photos are then processed, or stitched, by a software application into a nearly seamless larger image. The stitching process works best for relatively static scenes far from the camera, such as landscapes, large crowds, or city scenes. Panoramas are popular because of the visual appeal of a "wide" shot, and because the composite stitched image has a higher resolution—and more visual detail—than a single image from a camera's digital sensor. Additionally, the photographer can select exactly the parts of a scene to photograph and the parts to ignore. The resulting panorama might not have a traditional rectangular shape, but it will contain all the visual information the photographer finds appealing.

The main disadvantages of panoramas are that they require careful capture of multiple photos, so lens choice, camera stability, image overlap, and proper exposure are critical, and the stitching software can be less than perfect and produce visible seams in the final image. In addition, as camera resolution and field of view increase, panoramic image files become larger in size and more awkward to store, transfer, and display. Nevertheless, panoramic photography is more popular than ever. Many current digital cameras and smartphones now have built-in capabilities to create panoramas, without additional software processing. These are fun to use, even though they sacrifice ultimate image quality for convenience and speed.

The great advantage of carefully made panoramas is a pleasingly wide perspective and heightened image detail. Recently, several efforts have led to progress in capturing gigapixel resolution images via panoramic stitching, and viewing those gigapixel images using novel interfaces, such as HDview [220]. The HDview system cleverly selects the image-browsing parameters by continuously varying the blend between spherical and planar projections. Figure 3.15 shows an example of an HDview image. [MORE DETAIL ABOUT THIS GIGAPIXEL-IMAGE EFFORT AND OTHER PANORAMA CREATION EFFORTS COULD BE SUMMARIZED HERE.]

3.5 Improving Resolution

The spatial resolution in film-based cameras is set by the film emulsion. The *point spread function (PSF)* describes the distribution of light recorded by the film for an in-focus point source, due to internal scattering and film grain [153]. The effective resolution often depends on the film sensitivity, which is determined by the ISO exposure index of the film. Film with lower ISO has finer film grain and higher resolution, while film with higher ISO has coarser film grain and lower resolution. Photographers usually prefer using a lower ISO film for most portraits, landscapes, and object photography. For example, film such as Ilford PAN F Plus at ISO 50 has extremely fine grain and is widely used for high resolution photography in scientific, technical and copying applications. On the other hand, film such as Kodak Tri-X at ISO 400 is often used in low light conditions, in action scenes that require fast shutter speed, and for creative purposes where the higher noise and grain more effectively capture the essence of the scene.

The sensor in a digital camera replaces and mimics the film in a traditional camera. Digital sensors have a finite number of discrete light-sensing pixels that

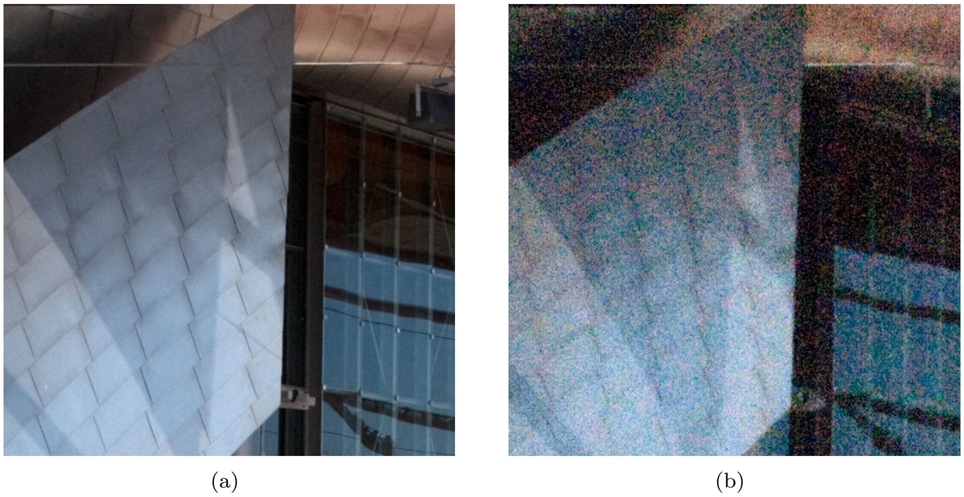


Figure 3.16: Image resolution decreases as ISO sensitivity of the digital sensor increases. The amount of digital noise produced at high ISO values depends on the camera sensor, and is especially noticeable in shadow areas of an image. (a) Sensor sensitivity at ISO 100 produces an image with fine details. (b) Sensor sensitivity at ISO 1600 produces a noisy image that obscures fine details.

set the maximum spatial resolution of a captured photograph. Sensor resolution is fixed at the factory, and replacing the sensor on a commercial camera is not possible (whereas a roll of film is easily replaced). Most digital sensors allow the user to modify the ISO sensitivity within some reasonable range by changing the gain on the A/D converter. However, higher ISO sensitivity results in increased digital noise in the captured photograph because the pixel size is fixed. [OTHER FACTORS IN SENSOR FABRICATION AND PROCESSING ALSO AFFECT NOISE—THOSE FACTORS COULD BE LISTED HERE.] This is different from film where the effective "pixel size" of the film grain changes according to the ISO sensitivity of the film. Because of increased digital noise the usable resolution of a digital sensor actually decreases as we increase the ISO sensitivity, as shown in Figure 3.16. [I ADDED A FIGURE THAT ILLUSTRATES LOSS OF RESOLUTION WITH INCREASED NOISE.]

Researchers have developed many techniques to improve sensor resolution. One technique, called *superresolution*, improves image resolution by combining multiple low-resolution images to recover higher spatial frequency components lost to undersampling. Numerous algorithms and techniques have been proposed [405, 213, 193, 194, 215, 106] that first estimate the relative motion between the camera and the scene, and then register all images to a reference frame. The images are then fused, usually by interleaving filtered pixels, to obtain a high resolution image. Keren et al. [212] and Vandewalle et al. [411] use randomized or 'jittered' sensor positions that they estimate by using sub-pixel image registration. Komatsu et al. [219] integrate images taken by multiple cameras with different pixel apertures to get a high resolution image. Joshi et al. [209] merge images taken at different zoom levels, and Rajan et al. [327] investigate the use of blur, shading, and defocus for achieving superresolution of an image and its depth map. Most authors also applied modest forms of deconvolution to boost the image's high spatial frequency components that were reduced by the box filter. Park et al. [314], Borman and Stevenson [64], and the book by Chaudhuri [80] provide a unified survey and explanation of many current methods. Lin and Shum [246], and Baker and Kanade [52] analyze the limits on the actual level of enhancement that superresolution techniques provide, and place hard bounds on it.

[NOTE IN MS: talk about Ankit's superres by moving mask in lens aperture]

[NOTE IN MS: to do: should this paragraph below go to the optics chapter instead (coded projections)?]

Agrawal and Raskar [40] propose a technique for obtaining superresolution from a single photograph. They use the effect of motion blur to increase the resolution of a moving object. A larger blur size gives greater resolution enhancement with a corresponding increase in reconstruction noise.

3.6 Extending the Depth of Field

Chapter 2 describes how the depth of field of an imaging system is related to the size of the lens aperture. The larger the aperture, the shallower the depth of field. Conversely, the smaller the aperture, the greater the depth of field. Unfortunately, a small aperture size can create other problems, such as higher noise (because less light enters the camera) and more pronounced diffraction artifacts. Several techniques have been proposed to extend the effective depth of field while still using a relatively large aperture size.

Extending depth of field has been a long-standing problem in the field of microscopy. Microscopes typically have an extremely shallow depth of field—a few microns—at close-up distances, and they are almost always light deprived (thus the need for a large aperture). A technique often used to extend the depth of field is to capture a series of images with the lens focused at different sample depths. These images form a *focal stack*, which contains sharply focused information at varying depths in the separate images. The focal stack can be processed into a single optimized image that has the appearance of much greater depth of field. Because the images in a focal stack are made over a period of time, this technique is limited to static subjects rather than live moving organisms. A similar processing technique known as confocal microscopy is discussed in Chapter 4.

Agarwala et al. [37] developed the photomontage processing technique (discussed in detail in Chapter 7) to combine the various images of a focal stack and produce a single image with extended depth of field. They find high contrast regions in the stack of images, segment these regions by using graph cuts, and fuse segments from different images by using gradient domain fusion. Figure 3.17 shows that the resulting synthesized image has a much greater depth of field than possible with an image made by a large physical aperture. Recently, Hasinoff and Kutulakos [176] analyzed the use of focal stacks to achieve desired depth of field (Figure 3.18). They show that by optimally selecting aperture, exposure time, and focus setting on each

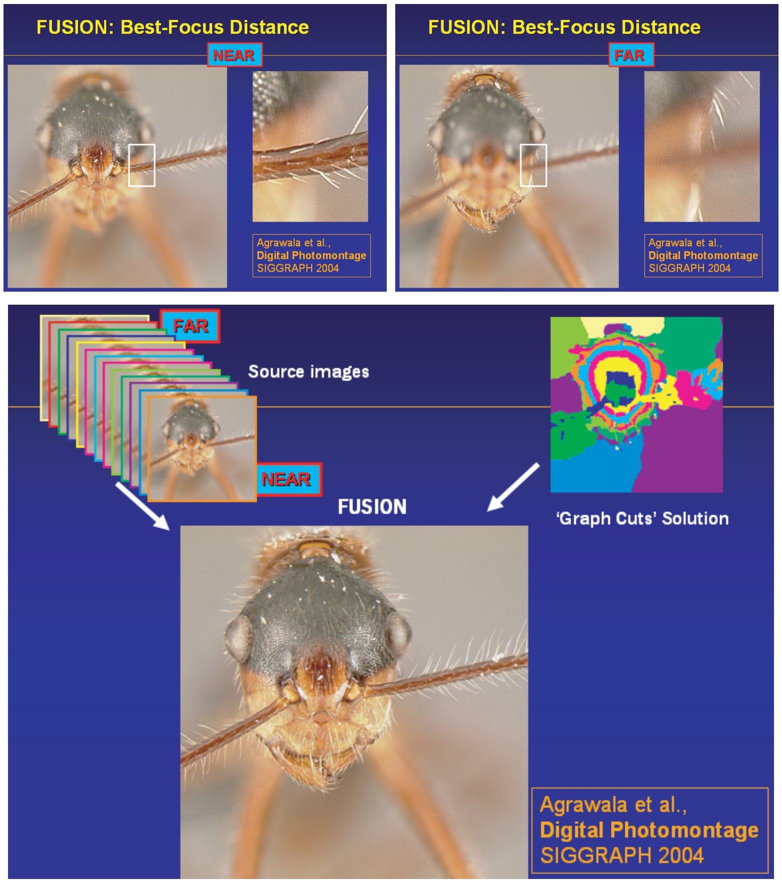


Figure 3.17: No single lens can achieve enough depth of field at close-up distances in microscopy to cover the full length of an object such as an insect. Because of extremely shallow depth of field at small distances, the lens can focus only on antennae or thorax or parts in between, as shown in the upper figure. By taking a series of focus-bracketed images, however, we can create a focal stack. With processing and optimization of the focused portions of each image, we can assemble a single image with much greater depth of field, as shown in the bottom figure. (Source Agarwala et al. [37].) [NOTE IN MS: replace with nicer figure.]

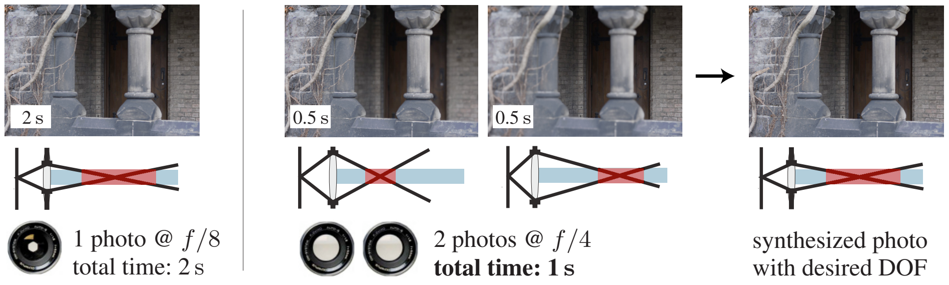


Figure 3.18: Light efficient photography. Capturing and merging multiple images taken at large aperture sizes is more efficient than capturing a single image taken at a smaller aperture size. (source Hasinoff and Kutulakos [176]).

photo, they can achieve a given depth of field with a given exposure level in less time than it takes to capture a single photo with the same depth of field. They derive a closed-form solution for a globally optimal capture sequence that may be used as an alternative to a single-shot narrow-aperture photograph. They use a technique similar to photomontage to combine the multiple images. [NOTE IN MS: to do: add more details for this?]

Section 4.4.3 discusses additional techniques that can achieve extended depth of field. A light field camera such as those described by Wilburn et al. [427], Ng et al. [303], and Veeraraghavan et al. [414] essentially capture the focal stack of a scene in a single shot. The focal stack is then processed to obtain a single extended depth of field image. Veeraraghavan et al. [414] and Levin et al. [231] use a mask at the aperture and a phase plate in place of a lens [100, 76], allowing them to use point spread function engineering to extend the depth of field.

3.7 Capturing Fast Phenomena

In the 1870s, well before the era of motion pictures, English photographer Eadweard Muybridge developed a method for capturing motion by constructing an elaborate multicamera system of wet-plate (collodion) cameras to take single short-exposure-time photographs in rapid-fire sequences. He devised a clever electromagnetic shutter-release mechanism triggered by trip-threads to capture action photos of people and animals in motion, such as the galloping horse in Figure 3.19. He refined the system by using electromagnetic shutter-release mechanisms triggered by pressure switches or elapsed time, which allowed him to record walking human figures, dancers, and acrobatic performances [3]). His sequences of short-exposure freeze-frame images provided the first careful examination of the subtleties of human and animal motion, which are typically too fleeting or complex for our eyes to absorb as they are happening. Going beyond the visual information in one perfect instant in a single traditional photograph, Muybridge's event-triggered image sequences contain valuable visual information that stretches across time and across a

[3] www.kingston.gov.uk/browse/leisure/museum

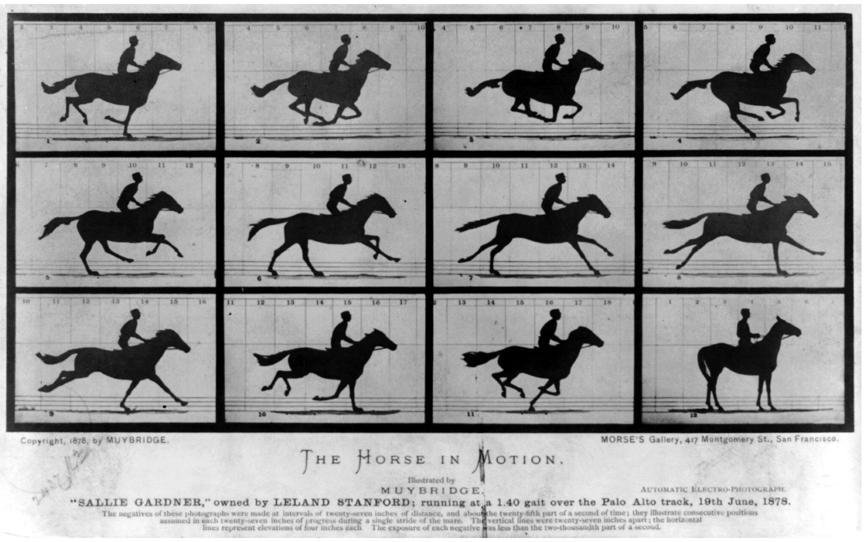

Copyright, 1878, by MUYBRIDGE.
MORSE'S Gallery, 417 Montgomery St., San Francisco.
THE HORSE IN MOTION.
Illustrated by
MUYBRIDGE.
AUTOMATIC ELECTRO-PHOTOGRAPH.
"SALLIE GARDNER," owned by LELAND STANFORD; running at a 1.40 gait over the Palo Alto track, 19th June, 1878.
The negatives of these photographs were made at intervals of twenty-seven inches of distance, and about the twenty-fifth part of a second of time; they illustrate consecutive positions assumed in each twenty-seven inches of progress during a single stride of the mare. The vertical lines were twenty-seven inches apart; the horizontal lines represent elevations of four inches each. The exposure of each negative was less than the two-thousandth part of a second.

Figure 3.19: Eadweard Muybridge sequence of a galloping horse, made in the 1870s with multiple cameras and a shutter release system timed to movement. (Source: en.wikipedia.org/wiki/Eadweard_Muybridge).

sequence of camera positions. Image sequences such as these are suitable for several different kinds of computational merging.

In some of Muybridge's pioneering efforts, two or more cameras were triggered at the same time to capture multiple views simultaneously. Modern work by Bregler and others on motion capture from video merged these early multiview image sequences computationally to infer the 3D shapes and the movements that caused them. Bregler et al. [68] found image regions undergoing movements consistent with rigid jointed 3D shapes in each image set, and computed detailed estimates of the 3D position of each body segment in each frame. They then re-rendered the image sets as short movies at any frame rate viewed from any desired viewpoint.

High speed video has gained wide popularity because of professional video cameras such as the Vision Research Phantom HD Gold and Phantom Flex [338] and consumer cameras such as the Casio Exilim Pro EX-F1 [125]. The professional cameras record impressive high-definition video at high frame rates of 1000 frames per second or more, and the consumer cameras give good results at slower frame rates and lower resolutions. These cameras provide a valuable perspective on temporal detail and motion, which our limited human vision doesn't allow us to see otherwise.

Several computational approaches to achieving higher frame rate videos were recently proposed. Shechtman et al. [368] extended superresolution to the spatio-temporal domain. They captured a dynamic scene with multiple cameras, each with relatively low resolution and low frame rate. They then exploit the sub-pixel spatial and sub-frame temporal misalignment between the cameras to combine the captured information with superresolution techniques. The multiple cameras offer

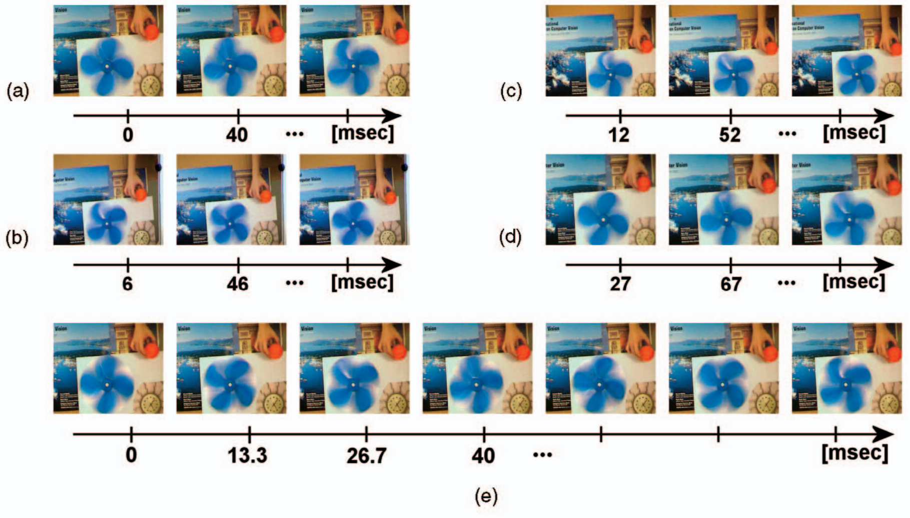


Figure 3.20: Shechtman et al. [368] proposed spatio-temporal superresolution. They combined images from multiple low-resolution, low-frame-rate cameras to produce a video with a higher spatio-temporal resolution. (Figure from Shechtman et al. [368].)

staggered exposures that are then combined. Figure 3.20 shows how their technique recovers rapid dynamic events that are not visible in any of the input sequences. They also analyze trade-offs between spatial and temporal superresolution. For example, motion blur (a spatial artifact) is resolved by increasing the temporal resolution. [WE COULD DESCRIBE ADDITIONAL COMPUTATIONAL APPROACHES TO HIGHER FRAME RATES.]

More than a hundred years after Muybridge's work, Marc Levoy and colleagues at Stanford University constructed an adaptable and reconfigurable array of 128 individual film-like digital video cameras that simultaneously perform both time-multiplexed and space-multiplexed image capture [426]. The reconfigurable camera array enabled a wide range of computational photography experiments. Using the valuable lessons learned from earlier arrays [211, 433, 262, 439], developers of the Stanford array added interchangeable lenses, custom control hardware, and a refined mounting system, all of which permitted adjustment of camera optics, positioning, aiming, and spacing between cameras. One configuration kept the cameras packed together, a inch apart, and staggered the triggering times for each camera within the normal 1/30 second video frame interval. The video cameras all viewed the same scene from almost the same viewpoint, but each viewed the scene during different overlapped time periods. By assembling the differences between overlapped video frames from different cameras, the team was able to compute the output of a virtual high-speed camera running at multiples of the individual camera frame rates and as high as 3,000 frames per second.

However, at high frame rates these differences were quite small, causing noisy results we wouldn't find acceptable in a conventional high-speed video camera. In-

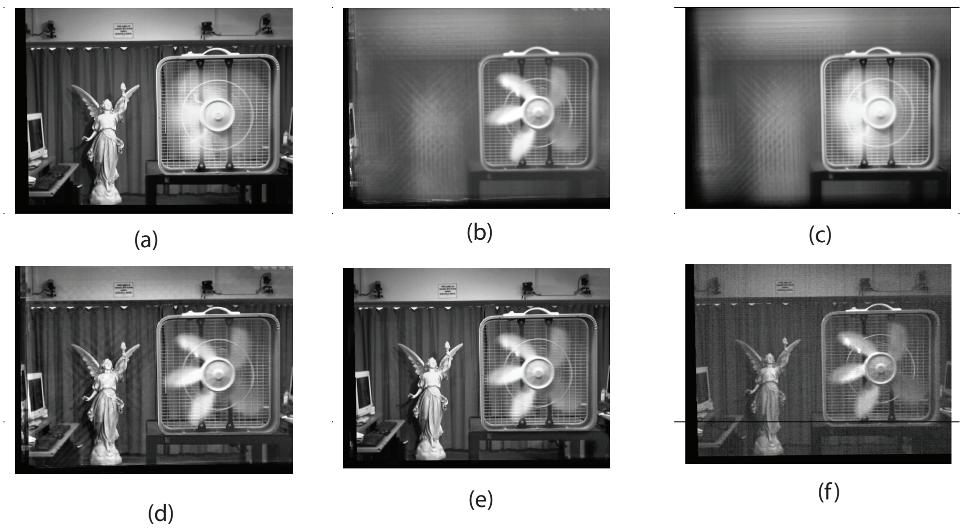
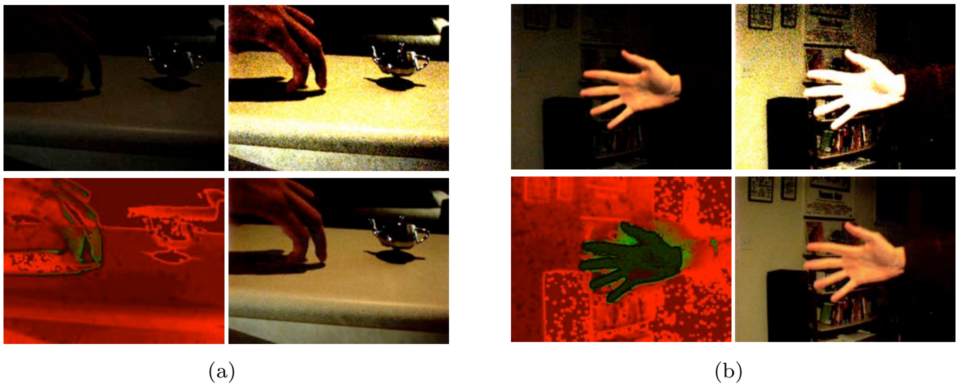


Figure 3.21: Staggered video frame times permit construction of a virtual high-speed video signal with a much higher frame rate via hybrid synthetic aperture photography [426]. Images of a scene are captured simultaneously through three different apertures: (a) a single camera with a long exposure time, (b) a large synthetic aperture with short exposure time , and (c) a large synthetic aperture with a long exposure time. Computing $(a + b - c)$ yields image d, which has aliasing artifacts because the synthetic apertures are sampled sparsely from slightly different locations. Masking pixels not in focus in the synthetic aperture images before computing the difference $(a + b - c)$ removes the aliasing in image e. For comparison, image f shows the image taken with an aperture that is narrow in both space and time. The entire scene is in focus and the fan motion is frozen, but the image is much noisier.



(a)

(b)

Figure 3.22: Bennett and McMillan [58] propose the concept of virtual exposure, where the effective exposure on each pixel is adaptively and independently varied, depending on the scene.

stead, the team simultaneously computed three low-noise video streams with different trade-offs by using synthetic-aperture techniques [238]. They made a spatially sharp but temporally blurry video I_s by averaging together multiple staggered video streams, providing high-quality results for stationary items but excessive motion blur for moving objects. For a temporally sharp video I_t, they averaged together spatial neighborhoods within each video frame to eliminate motion blur, but this induced excessive blur in stationary objects. They also computed a temporally and spatially blurred video stream I_w, to hold the joint low-frequency terms, so that the combined streams $I_s + I_t ? I_w$ [MS HAS ? HERE FOR AN OPERATOR—WHAT SHOULD ? BE?] exhibited reduced noise, sharp stationary features, and modest motion blur, as shown in Figure 3.21.

[NOTE IN MS: TO DO: Bullet time; sing bing kang]

3.8 Noise Reduction

Martinec [257] explains various noise sources in digital sensors and analyzes noise for modern digital SLR cameras.

Bennett and McMillan [58] introduce the concept of virtual exposure, where the exposure on each pixel is adaptively and independently varied in post-processing, based on the temporal and spatial neighborhood of the pixel. They estimate each pixel's exposure setting by using spatially uniform tone mapping of each frame. They then recreate the corresponding gain at each pixel by combining several temporal samples for static regions and spatial samples for dynamic regions of the scene. The resulting video has not only lower noise, but also a higher (tone mapped) dynamic range. The basic algorithm is extended to work with moving cameras by tracking feature points and compensating for camera motion. The results look impressive for static parts of the scene (see Figure 3.22), but detail may be lost in dynamic parts where spatial filtering is applied. [THIS PARAGRAPH READS

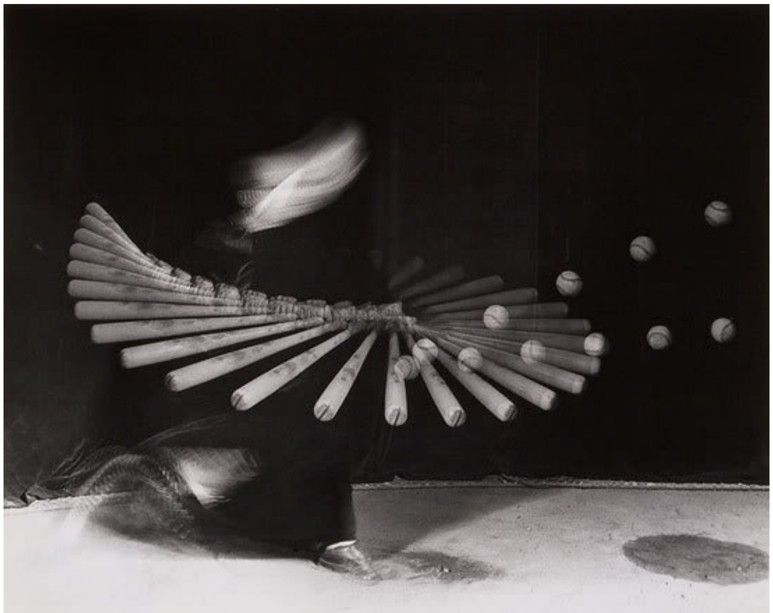


Figure 3.23: The elastic collision dynamics of a baseball. (Photo by Harold Edgerton.) [THIS FIGURE ALSO APPEARS AS FIGURE 1.7 IN CHAPTER 1. TEXT FOR THIS FIGURE WAS REMOVED BY RW.]

LIKE IT WAS TAKEN WORD FOR WORD FROM A PAPER ABSTRACT.]

Summary

The goal of film-like photography is to copy an image formed by a lens. We shift our attention in this chapter to the broader topic of capturing visual experience by looking for broader sets of solutions for gathering visual information. Just as biological vision systems include many different designs for different purposes, we believe that computational photography devices can achieve similarly broad diversity and novelty in design.

Chapter 4

Optics

The simplest optics for image formation is a pinhole placed at a distance from a sensor or film. This is the technique used in the well-known *camera obscura*. Because of its small size, however, a pinhole blocks nearly all the light needed to form an image, and the pinhole size creates diffraction effects that reduce image quality. Instead of pinholes, lenses are commonly used because they capture more light and avoid the diffraction effects created by pinholes. In principle a simple convex lens is sufficient for image formation, but in practice a compound lens with several optical lens elements is required to correct (as much as possible) the many optical aberrations that lenses produce.

New methods of imaging can form a coded intermediate image on the sensor. An intermediate image such as this may not be suitable for human observation, but software techniques can decode it and recover a richer representation of the imaged scene. In addition, many other techniques such color manipulators, patterned masks and arrangements that create a non-standard view and perspective are used to form an image.

4.1 Animal Eyes

Although the variety of eyes in the animal kingdom seems astonishing, physical laws have constrained solutions for collecting and focusing light to just eight types of eye optics. Of around 33 animal phyla, about one-third have no specialized organ for detecting light, one-third have light-sensitive organs, and the rest are animals with what we would consider eyes. Simple photon detectors aggregate incident light. But animal eyes have organs that also compare light from different directions. Biological pinholes, lenses, or mirrors are used to focus an image on photoreceptors.

As earliest evolution occurred in water, which transmits only a limited range of wavelengths, the mechanisms for photon response converged on biochemical solutions that set the course for subsequent evolution (3). The evolution of eyes very likely proceeded in stages. First were simple eyespots (early Cambrian period, 570 to 500 million years ago), with a small number of receptors in an open cup of screening pigment. Eyespots would distinguish light from dark but could not

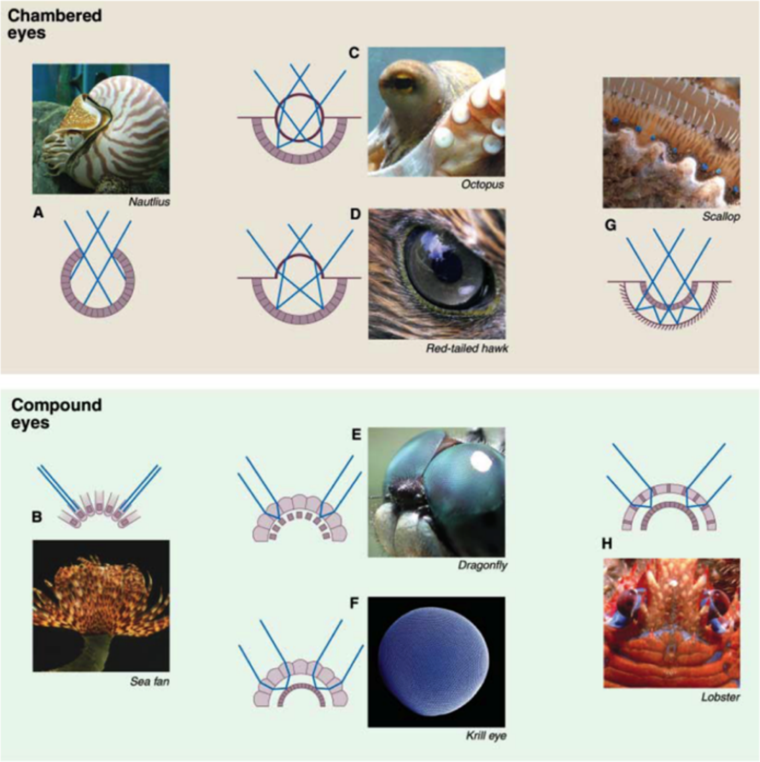


Figure 4.1: Categorization of eyes found in animals. They are broadly classified as chambered and compound eyes. (Figure from Fernald [134]).

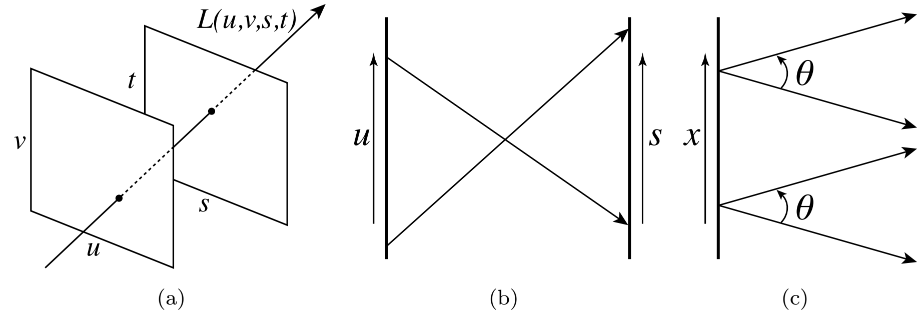


Figure 4.2: (a) Two-plane parameterization for light fields, (b) two-plane parameterization $L(s, u)$ in the flatland case, and (c) spatio-angular parameterization $L(x, \theta)$ in the flatland case.

represent complex light patterns. Invagination of this eyespot into a pit would add the capacity to detect the direction of incident light. Addition of receptors may then have led to a chambered eye, whereas duplication of an existing pit may have led to a compound eye (2). Adding an optical system that could increase light collection and produce an image would later dramatically increase the usefulness of an eye. Whereas primitive eyes can provide information about light intensity and direction, advanced eyes deliver more sophisticated information about wavelength, contrast, and polarization of light.

4.1.1 Traditional Optics in Cameras

(Depending on what we add in Chapter 3 about optics in Film-like photography, we may have to add more text here.)

Traditionally we see a (i) cascade of lenses (ii) aperture stop or iris (iii) filters such as polarization, color (iv) coatings and (v) additional optics for operations like range-finder, view-finder. [THIS SUBSECTION SHOULD BE A TRANSITION FROM THE INTRODUCTORY TEXT ON ANIMAL VISION TO A SUMMARY OF CAMERA OPTICS TO THE FOLLOWING SECTION ON LIGHT FIELDS.]

4.2 Light Fields and Ray Space Analysis

The light field is a 4D quantity that completely characterizes light transport in free space. While the concept of 4D ray-space has been used in computer graphics and vision for some time, it was first formalized and generalized in the computer graphics literature by Levoy and Hanrahan [240], and Gortler et al. [162]. The light field describes the set of all possible rays of light in a space free from occluders. A popular light field parameterization is the two-plane parameterization, also called the *light slab*. Two parallel planes separated by a finite distance describe all the rays between them. A ray of light exists between all points on the first plane (u, v) and all points on the second plane (s, t) (Figure 4.2(a)). This 4D parameterization, $L(u, v, s, t)$, is frequently used to describe and understand image formation in cameras and image-based rendering techniques.

As discussed in Chapter 2, a traditional film or digital camera (and even the human eye) captures only a 2D slice of the 4D light field entering through the lens. The rest of the information is lost as the sensor integrates over the lens aperture. The exact 2D slice captured by the sensor depends on the focal length of the lens, the distance between the scene and the lens, and the distance between the lens and the sensor. Changing the plane of focus on a camera (or the eye) results in a different 2D projection, but at any given instant we only get a single 2D projection. As early as 1992 [35] several researchers recognized the value of sensing the direction of incident light at each point on the focal plane behind a lens. Adelson's camera system combined a large front lens and a field of microlenses behind the lens. This design gathered what is now known as a 4D light field estimate, and Adelson used it for single-lens stereo reconstruction. More recently Ng et al. [303] refined the idea further with an elegant hand-held digital camera (currently marketed as the Lytro camera) for light-field capture that permits digital refocusing and slight changes of viewpoint computationally.

This chapter describes light fields or the 4D ray space analysis of optical setups. We discuss manipulation and capture of light fields and their use for applications such as refocusing, 3D shape recovery, glare removal, and multispectral imaging. While discussing ray space analysis, it sometimes helps to simplify the problem and consider the easier problem in a *flatland* scenario[1]. Flatland is simply a hypothetical 2D plane where the light field reduces to a 2D quantity. Much of the flatland analysis directly extends to the real world or a 4D light field. We can parameterize the rays in a 2D light field using two parallel lines separated by a finite distance, as shown in Figure 4.2(b). A ray of light exists between all points on the first line (u) and all points on the second line (s) in this *light slab* parameterization. Another parameterization for the light field is the *spatio-angular parameterization* introduced by Georgiev et al. [156] (Figure 4.2(c)). A ray of light is described by the point where it intersects a given line (x), and by the angle it makes with that line (θ). The light field representation using either parameterization, $L(u, s)$ or $L(x, \theta)$, describes all the rays in the 2D space. It is fairly straightforward to convert from one parameterization to another using simple trigonometry.

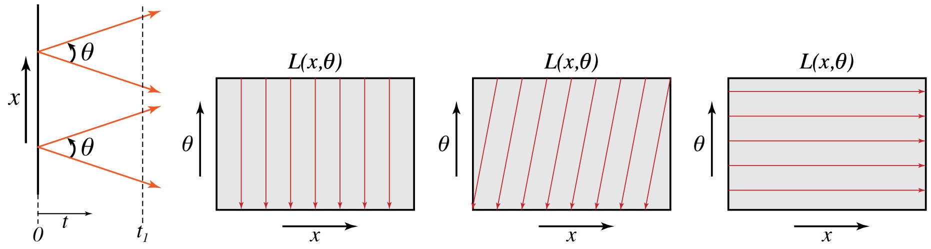
4.2.1 Light Field Projections

Figure 4.3 shows the 1D projection obtained on a screen placed at different points in the 2D light field. As shown in Figure 4.3(b), a screen at $t = 0$ captures the projection of the light field along the θ-axis,

$$I_0(p) = \int L(p, \theta) d\theta.$$

The signal, $I_0(p)$ is essentially the image a camera or the eye focused on the x-plane captures on its sensor or on the retina (shown in Figure 4.4(a)). The cone of rays

[1] The term *flatland* comes from the book entitled *Flatland: A Romance of Many Dimensions*, by Edwin A. Abbott.



(a) $L(x, \theta)$ (b) Screen $t = 0$. (c) Screen at $t = t_1$. (d) Screen at $t = \infty$.

Figure 4.3: Placing a screen at different points perpendicular to the t-axis results in various 1D projections of the 2D light field, $L(x, \theta)$. The projection angle depends on the distance t from the x-axis.

coming out of any scene point (A) combine back to a single, unique point (A') on the sensor plane, and the angular information is completely collapsed by the integral. As we move the screen away from the x-plane to $t = t_1$, the projection direction changes from the vertical to a slant as shown in Figure 4.3(c). Finally, moving the screen to $t = \infty$ (Figure 4.3(d)) results in a projection along the x-axis of the light field,

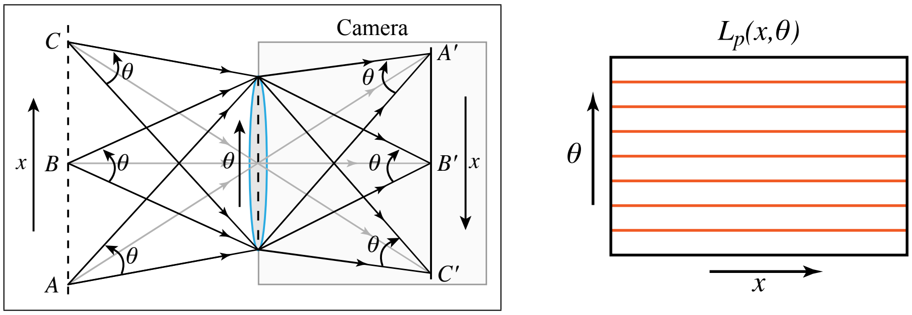
$$I_\infty(p) = \int L(x, p)dx.$$

The signal $I_\infty(p)$ is an integral over all the spatial points for each angle θ.

The *Fourier Slice Theorem* [65] states that the Fourier transform of the projection of a 2D function on a 1D plane is exactly equal to a slice through the origin and perpendicular to the projection direction of the 2D Fourier transform of the original function. Ng applied the Fourier slice theorem to light field projections and demonstrated its use for fast refocusing [301]. So a vertical projection like the one in Figure 4.3(b) is equivalent to the inverse Fourier transform of a horizontal slice through the center of the Fourier transform of the 2D light field. Similarly the horizontal projection as shown in Figure 4.3(d) is equivalent to the inverse Fourier transform of a vertical slice through the center of the Fourier transform of the light field.

Light field of a pinhole camera. An infinitely small pinhole allows a single ray for every scene point to pass through. We assume that the scene is the u-plane, and the pinhole is the s-plane of the light field. The pinhole reduces the light field $L(u, s)$ to $L_{s_0}(u)$, where s_0 corresponds to the pinhole position. The image captured on a film placed behind the pinhole only contains a single degree of freedom, and the remaining information in the light field is lost to the occluder surrounding the pinhole.

Light field of a Lambertian fronto-parallel scene. Consider the special case of a Lambertian fronto-parallel scene (called a *painting*) placed in front of a camera as shown in Figure 4.4(a). Since the scene is diffuse, the outgoing rays in all directions (θ) from each point on the scene are exactly the same. The resulting light field has a single degree of freedom, $L_p(x, \theta) = I_0(x)$. Each row in Figure 4.4(b) is an exact copy of the other. Figure 4.4(a) shows a camera lens (or the lens of a



(a) (b)

Figure 4.4: (a) A fronto-parallel scene, called a *painting*, placed in front of a camera with the lens focusing on the painting, and (b) the corresponding light field when the painting is Lambertian. Each row in $L_p(x, \theta)$ is exactly the same, as indicated by the horizontal orange rows in the light field visualization.

human eye) focusing a scene in the x-plane onto the camera's sensor (or the eye's retina). In the special case when we have a Lambertian painting in the x-plane, all rays coming out of a point A on the painting are identical. The sensor (or the eye) captures a vertical projection of the light field $L_p(x, \theta)$. This special projection is typically referred to as the *image* of the scene, and this is what all film cameras and digital cameras capture. This projection is simply a scaled version of the image captured by a pinhole.

4.3 Transformation of Rays in the Incident Light Field

The goal of optics is to guide rays through the lens assembly so that they form a high quality, invertible image on the sensor. In addition to the traditional optical process, which involves bending rays at the main lens, researches in computational photography attempt to transform a set of rays through additional *bending* as well as *attenuation* and *scattering* [2]. The additional transformation can be classified into three stages, based on where the rays are manipulated.

- **Transform First** optical elements manipulate rays before they are passed through the focusing mechanism.
- **Transform Middle** optical elements manipulate rays while the image is being formed as part of the traditional focusing mechanism.
- **Transform Last** optical elements intercept and manipulate the rays after they have been guided by the traditional focusing mechanism to form an otherwise ordinary image.

[2] Prisms and polarizers cause wavelength-dependent and polarization-dependent bending or attenuation respectively

In this chapter we show how transformations take place at these three stages, with examples in bending (microlens array for light field capture), attenuation (masks for encoding intensities for linear combination), and scattering (color dispersion for multispectral imaging). The concept of stages of first-middle-last transform comes from computer graphics triangle visibility calculations using multiple pipelines, where the triangles are sorted first-middle-last, depending on the assignment of processing units to different depth slabs or display area.

A fourth stage for transforming rays is located between the illumination source and the scene. This stage, usually referred to as *computational illumination*, will be studied in the next chapter.

4.4 Lenses and Focus

Traditional cameras use a complex sequence of lenses to produce a high-quality image on the film or sensor plane. These lenses are designed to minimize image artifacts such as chromatic aberration, spherical aberration, vignetting, and various forms of geometric distortion. The lens is the primary ray bender that transforms the incoming light field so that a sharp, focused image is formed on the sensor. The focal length of the lens determines the effective field of view for a given sensor size. This section describes the application of computational techniques to generalize the concept of lenses and focusing.

4.4.1 How Autofocus Works

As discussed in Chapter 2, a lens focuses a plane in the scene onto the image sensor. Scene points away from this plane are blurred and appear out-of-focus. The photographer or the autofocus camera decides which plane to bring in focus by changing the distance between the sensor and the lens before the photo is captured. This is typically accomplished by manually sliding or rotating the lens in a manual focus camera, or by a motor in an autofocus camera. The first point-and-shoot autofocus camera was the Konica C35 AF, introduced in 1977. It used a phase-based autofocus module designed by Honeywell. The Minolta Maxxum 7000, introduced in 1985, was the first popular autofocus single lens reflex (SLR) camera. The phase-detecting autofocus sensors and drive motor were both embedded in the camera body, and the camera used a mechanical link in the lens mount for focusing and aperture control. Modern autofocus cameras today use clever optics and algorithms to provide almost instantaneous focusing. Most SLR cameras now embed the motor for translating the lens in the lens itself, and use electronic coupling between the lens and camera body.

Phase-based autofocus is most common in digital SLR cameras that use lenses with reasonably large aperture. The camera uses a beamsplitter to direct light to an autofocus sensor under the main reflex mirror on the camera. Two optical prisms capture the light rays coming from the opposite sides of the lens aperture. This creates a rangefinder with a baseline equal to the aperture size of the lens. The two images are usually sensed by 1D sensor arrays. The phase difference between the readings from the two arrays gives the direction and distance of lens motion to

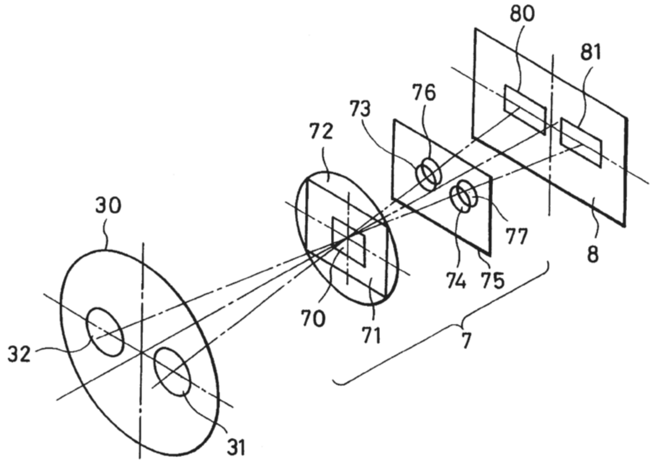


Figure 4.5: Phase-based autofocus, illustrated in this figure from US Patent 5589909 (Nikon), is used in most digital SLR cameras. A beamsplitter directs light to an autofocus sensor. Two optical prisms capture light rays coming from opposite sides of the lens aperture, which produces a rangefinder with a baseline the size of the lens aperture. Phase differences from the two light rays give the direction and distance needed for autofocus. [NOTE: THIS FIGURE NEEDS LABELS FOR THE FOUR PARTS.]

achieve focus. Figure 4.5 illustrates the process of phase-based autofocus.

Contrast-based autofocus is commonly used in smaller digital point-and-shoot cameras, video cameras, some cell-phone cameras, and some digital SLR cameras in live-view mode. Contrast measurement is achieved by measuring the contrast within the image as seen through the lens and captured by the sensor. Measuring the local intensity difference between adjacent pixels gives an estimate of the degree of defocus. Unlike phase-based autofocus, the camera needs to search for best focus by moving the lens in the direction that increases contrast. As a result, focusing is usually slower. Since contrast-based autofocus does not use a separate sensor, it is easily implemented in software on less expensive cameras. Both the phase-based and contrast-based methods require reasonably high contrast in the image at the selected focus point. For example, neither technique works well on a uniformly lit textureless wall.

Active autofocus systems use ultrasonic sound waves (measuring the delay in their reflection), or infrared light (triangulation or amount of reflected light) to measure the distance to the subject independently of the main optical system of the camera. While active methods do not require a level of contrast in the scene, their accuracy is typically lower than passive systems, and they do not work through windows and glass.

A second and equally important aspect of autofocus is choosing where to focus. Most cameras have a mode where the camera automatically determines the best plane of focus based on some heuristics involving contrast, distance, and scene brightness. Newer cameras use face detection to focus on the closest face. Most cameras also allow the photographer to manually select the focus point in the viewfinder or the LCD display. This is usually done by using a touch interface on the display, or a set of arrow buttons, or even by tracking the eye gaze in the viewfinder.

4.4.2 Light Field Capture and Post-Capture Refocusing

As discussed in Section 4.2, a camera based on the traditional *camera obscura* design only captures a 2D projection of the 4D light field incident on the camera lens. Information is lost in the process of obtaining this 2D projection, and as a result we get only one plane in perfect focus in the resulting photograph. The other image planes are blurred by the point spread function (PSF) corresponding to the degree of defocus. Several people have proposed modified camera designs that capture the complete 4D light field on a single 2D image sensor. Most of these techniques trade off spatial resolution for angular resolution, and encode higher dimensional angular information on the same sensor. Consequently the final image resolution is typically an order of magnitude lower than that of the sensor.

In its simplest form, light field capture can be thought as a form of epsilon photography, as discussed in Chapter 3, where the camera position is changed over sensors (using a camera array), over pixels (using a lenslet array), or over time (using a changing aperture mask).

Integral imaging is an auto-stereo imaging technique that displays a 3D image without requiring any special 3D glasses. The technique was first proposed by Gabriel Lippmann in 1908 [248]. An array of small convex microlenses or pinholes is placed over the image plane so that the viewer sees a different view through each eye, thus giving a sense of depth. The image behind each individual microlens element represents the various angular (θ) components for that macro pixel. A similar setup with the microlens array in front of the film plane is used to capture the image of a 3D scene. Ives [196, 197] added a large aperture lens in front of the microlens array in the camera setup. A thorough history of integral imaging and photography is given by Okoshi [307], and Roberts and Smith [342]. The book edited by Javidi and Okano [200] provides a thorough survey of 3D video and 3D TV technologies based on integral imaging.

Light field cameras build on the basic idea of integral imaging to capture a complete 4D light field. Adelson and Wang [35] used a similar optical design to estimate the depth for each pixel in a scene, and called a camera of this design the *plenoptic camera*. Figure 4.6 shows how their camera uses a single main lens along with a pinhole array or a lenticular array placed in front of the sensor plane. The light arriving at each pinhole gets broken up into pixels corresponding to the incoming angle of incidence. Each pinhole can be thought of as a macro-pixel, with corresponding sub-pixels capturing the angular information. Their prototype used a ground-glass diffuser placed behind a lenticular array, which is imaged by a second video camera. The resulting optical setup provides information about how the scene would look when viewed from a continuum of possible angular viewpoints bounded by the main lens aperture.

Ng et al. [303] eliminated the diffuser from the optics and built a handheld

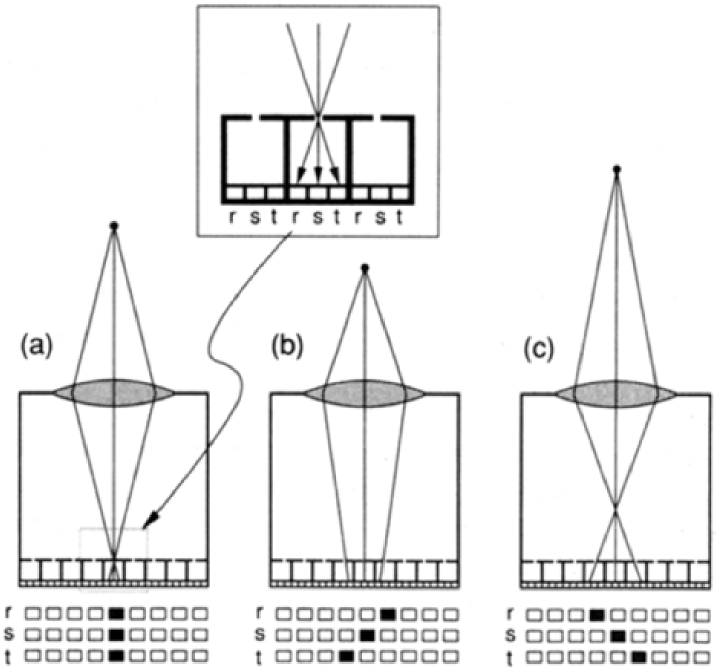


Figure 4.6: Light field capture using a pinhole array placed at the image plane next to the sensor. (The figure is from Adelson [35]) ADD CAPTION TEXT THAT EXPLAINS THE THREE DIFFERENT PARTS OF THIS FIGURE.

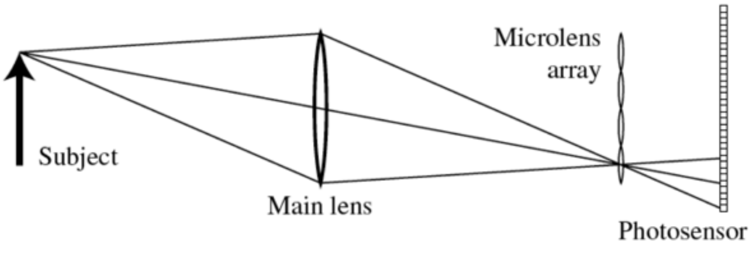


Figure 4.7: Light field capture using a microlens array placed at the image plane next to the sensor. Each macropixel captures the angular image information otherwise lost in a standard camera. (Figure from Ng [303])

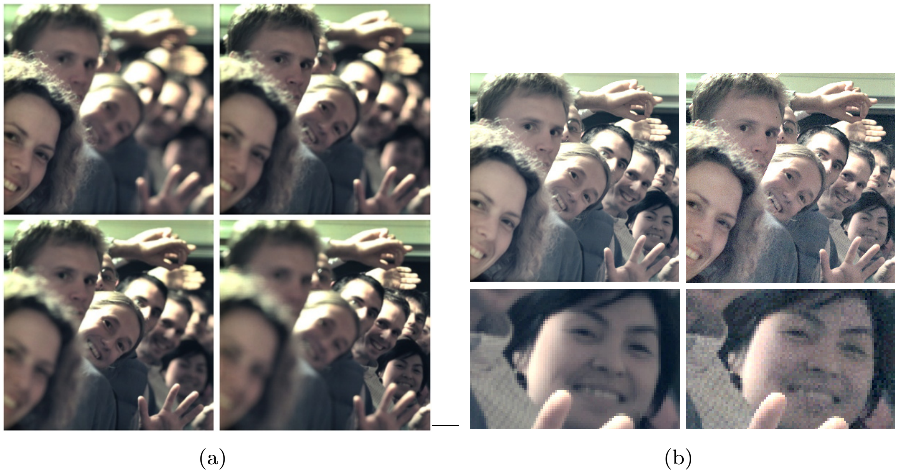
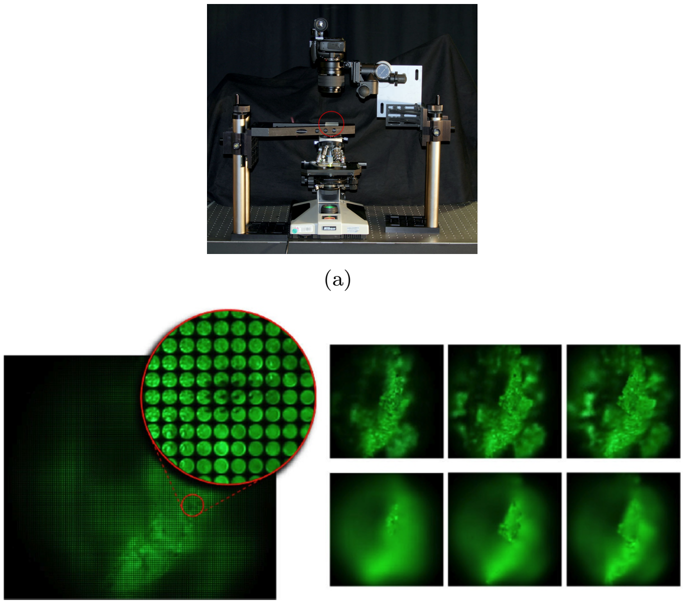


Figure 4.8: (a) Refocusing example from the Stanford plenoptic camera. (b) An all-in-focus image created from the focal stack produced by the light field data has less noise than an image captured by a smaller aperture size to get the same depth of field. (Figure from Ng [303])

plenoptic camera by placing a microlens array immediately above the sensor of a medium format camera (Figure 4.7). They used the captured light field to arbitrarily refocus at different planes in the scene as a post-process (Figure 4.8(a)). They perform real-time refocusing using the Fourier Slice theorem [301] discussed in Section 4.2.1. They also demonstrate changing the camera's center of projection to any point within the aperture of the lens as a post-process. Their prototype gives a spatial resolution of 292 × 292, and an angular resolution of 14 × 14. The ability to change the plane of focus and effective aperture size after the photo capture process is extremely appealing in terms of the flexibility it offers. Like several other light field capture techniques, the resulting images suffer from a significant loss in spatial resolution.

Levoy et al. [241] implemented a light field microscope by inserting a microlens array in the optical path of a conventional microscope (Figure 4.9). While the setup is similar to earlier light field cameras, microscopes are inherently orthographic devices, and the perspective views offered by this setup represent a new way to look at microscopic specimens. Furthermore, the ability to create focal stacks from a single photograph allows moving or light-sensitive specimens to be recorded in a single shot. Unlike traditional photographic applications, however, diffraction is a major limitation on the resolution obtainable by such a setup.

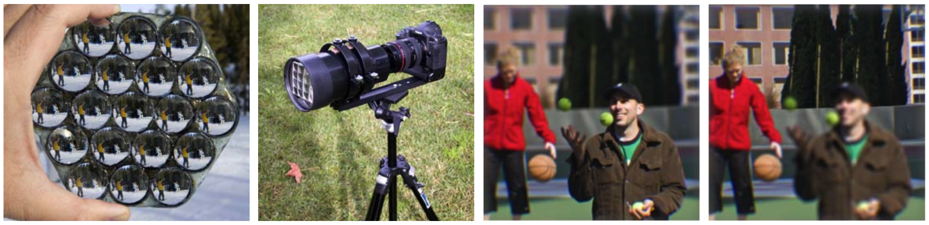
Georgiev et al. [156] used a sort-first approach to capture the 4D light field. They placed an array of lenses outside the camera's main lens to capture the incident light field. Unlike previous approaches, they captured a complete image, each from a slightly different center of projection, behind each lens element. They explore the tradeoff between spatial and angular resolution in light field capture. Their setup sacrifices angular resolution for higher spatial resolution, and relies on computer



(a)

(b)

Figure 4.9: (a) Prototype of a light field microscope (MUCH BIGGER FIGURE (a) NEEDED HERE). (b) A light field captured by the optical setup, and a perspective pan and a focal stack computed from a single shot. (Figure from Levoy [241])



(a) (b) (c) (d)

Figure 4.10: ((a) Optical “lens” consisting of lens-prism pairs. (b) Sparse light field camera with an array of 20 negative lenses. (c) and (d) Refocusing results from a single captured image. (Figure from Todor [156])

vision techniques for interpolating the limited angular resolution. Figure 4.10 shows a prototype system of lenses and prisms that is attached external to a conventional camera. This is very similar to a camera array setup discussed in Section 4.5.6.

All the lens-based techniques discussed thus far spatially re-bin rays of light to capture the 4D light field on a 2D sensor. Veeraraghavan et al. [414] placed a mask close to the image sensor to perform a similar re-binning in the frequency domain. Section 4.5 discusses this and other mask-based techniques for capturing the 4D light field.

Zhang and Chen [440] proposed using a bare sensor to capture multiple pictures for different positions of the sensor to estimate the light field of the scene.

Recently, Levin et al. [232] proposed a Bayesian inference-based approach to compare and contrast the performance of various light field cameras, and analyze the limitations and advantages of the various designs. They also propose a prior that models a real world light field better than the band-limited assumption and significantly reduces sampling requirements.

4.4.3 Extending the Depth of Field

[NOTE: THIS SUBSECTION HEAD IS IDENTICAL TO SECTION HEAD 3.6 IN CHAPTER 3]

Section 3.6 discussed the problem of limited depth of field with a finite-sized aperture, and the use of focal stacks to extend the effective depth of field. Here we discuss some more advanced techniques, most of which either work by capturing the full light field, or use ray-space analysis to integrate the optical setup and computation.

Extending the concept of focal stacks to work with a light field is straightforward. A focal stack is obtained from the light field data by computationally refocusing at different depths. These images are then combined by using a technique similar to the digital photomontage technique of Agarwala et al. [37]. The resulting image is less noisy than that obtained by simply extracting a single sub-aperture image because it integrates more light (Figure 4.8(b)).

As described in Chapter 2, the circle of confusion is the disc-like spot produced on the sensor by an out-of-focus point source. More generally, we call the image produced by a point source on a sensor the *point spread function* (PSF) of the imaging system. The PSF of an in-focus point source of an ideal lens is an infinitely sharp point. The PSF becomes a disc of increasing diameter as the point source moves away from the plane of focus. The PSF is clearly dependent on the depth of the point source, and is often considered a 3D function (two dimensions for the image sensor and a third dimension for the depth of the scene point). For a general out-of-focus scene, the defocus effect is the same as that of a blur filter that is the PSF corresponding to the depth. Using deconvolution to recover a sharp focused image of an out-of-focus plane is non-trivial because

- Inverting the the disc-like blur PSF is ill-posed due to the presence of zeroes in its frequency domain representation. High frequency information is irrecoverably lost in the image formation process.

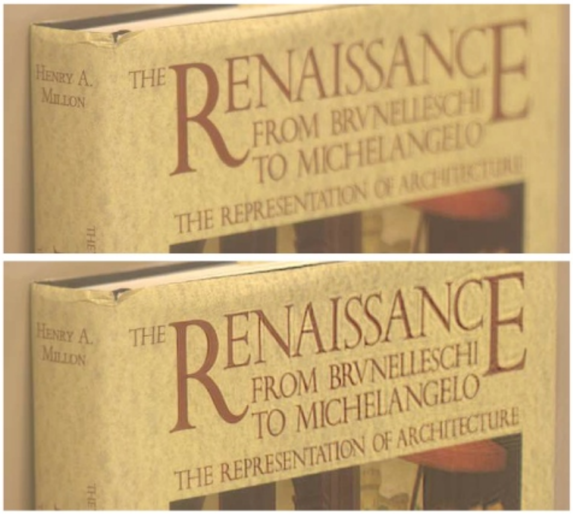
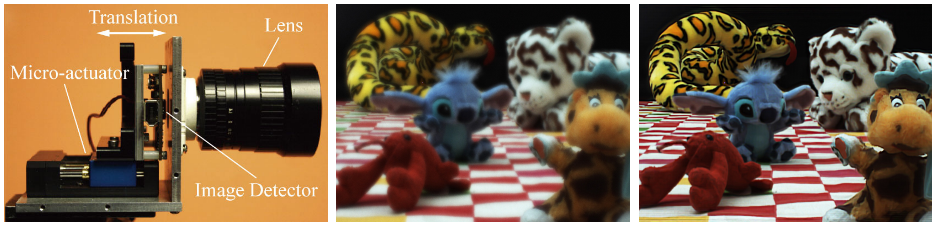


Figure 4.11: Simulated extended depth of field using a broadband coded aperture. (Figure from Veeraraghavan et al. [414])

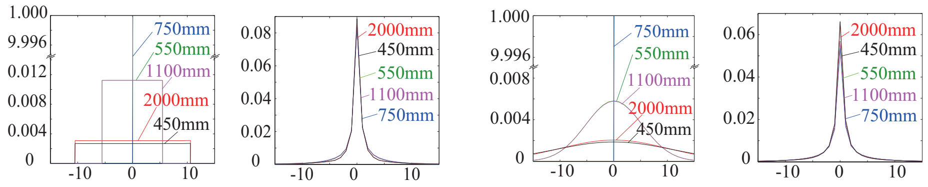


(a) Prototype setup (b) Photo captured at $f/1.4$ (c) Extended depth of field

Figure 4.12: Flexible depth of field setup and extended depth of field results. (Figure from Nagahara et al. [284])

- The scale of the PSF (dependent on the scene depth) is generally not known and hard to estimate.

Section 4.4.5 discusses the use of special aspheric optics in place of a traditional lens to extend the depth of field. The optics are designed to make the point spread function independent of the degree of defocus or scene depth. With a fixed defocus kernel size, they use deconvolution techniques to obtain a sharp image. Similarly, Veeraraghavan et al. [414] and Levin et al. [231] used an amplitude mask placed in the aperture of the lens to engineer the effective defocus point spread function. In the work of Veeraraghavan et al. [414], this amplitude mask makes the blur kernel well conditioned so that its inversion is no longer ill-posed. They demonstrated recovering an extended depth of field by combining images obtained by deconvolving with different scales of the blur kernel (Figure 4.11). These mask-based methods are discussed in more detail in Section 4.5.



(a) Normal Camera PSF (Pillbox) (b) EDOF Camera IPSF (c) Normal Camera PSF (Gaussian) (d) EDOF Camera IPSF

Figure 4.13: (a) and (c) Simulated normal camera point spread functions. (b) and (d) Extended depth of field camera impulse point spread functions (IPSF), obtained by using pillbox and Gaussian lens point spread function models for five scene depths. The IPSFs are almost invariant to scene depth. (Figure from Nagahara et al. [284].) [THIS FIGURE AND CAPTION IS THE ONLY PLACE IN THE BOOK WHERE YOU MENTION 'IMPULSE POINT SPREAD FUNCTIONS.']

Nagahara et al. [284] achieved an approximately depth independent blur, similar to that obtained by wavefront coding, by translating the sensor relative to the lens during the integration time of a single exposure (see Figure 4.12). Figure 4.13 compares the PSF of a traditional camera (pillbox and gaussian) to that produced by one with a translating sensor. Since a typical camera produces a greatly minified image of the scene, very small (order of microns) sensor translation is adequate to cover a large range of scene depths. Since the resulting PSF is nearly constant over the range of depths that the sensor sweeps through during the exposure, the captured photo is deconvolved with a single blur kernel to recover an image with an extended depth of field.

Häusler [177] used a very similar technique to extend the depth of field in a microscope. He moved the specimen under observation along the optical axis as its magnified image was filmed, and the resulting PSF was invariant over the depth range of the specimen. Mohan et al. [**?**] used synchronized sensor and lens translation in planes parallel to one another in order to achieve approximately depth invariant blur size (though their PSF is harder to invert).

4.4.4 Reducing the Depth of Field

The problem of reducing the depth of field in a photograph is perhaps the exact opposite to that of extending the depth of field. Once again, the problem has a parallel in microscopy where confocal imaging is used to illuminate and focus on a very small part of the specimen to minimize scatter from the out-of-focus parts of the specimen. Another approach involves the use of a camera array to simulate a larger synthetic aperture. See Section 4.5.6 for more details on confocal imaging and synthetic aperture.

In photography, a shallow depth of field is often used for artistic and creative effects. The defocus effect produced by a lens is often referred to as the *bokeh* of the lens. A pleasing emphasis is often more important than the sharpness and detail in

Figure 4.14: Defocus magnification by Bae and Durand [51]: (left) input image, (center) estimated blur map, (right) result with magnified defocus. (Figure from Bae and Durand [51])

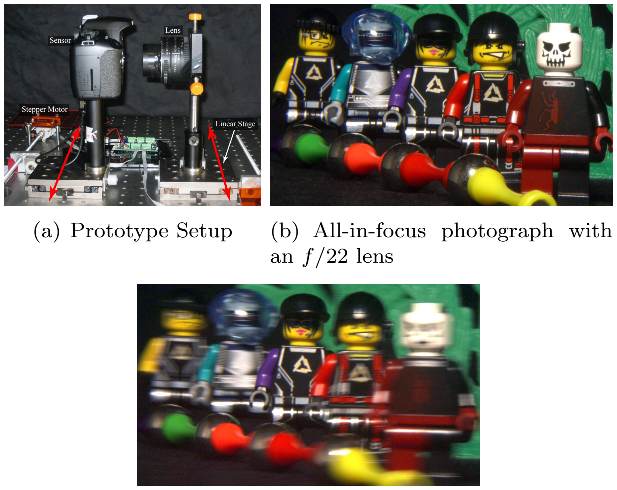
the focused parts of the image, especially for portrait photography. The depth of field and the bokeh are closely related to the shape and size of the aperture. Lenses with large apertures are frequently desired as much for their shallow depth of field as for their light-gathering capability. Unfortunately, such lenses tend to be bulky and expensive. Small point-and-shoot cameras usually have more modest aperture sizes, greatly limiting their use for professional or creative photography.

Bae and Durand [51] proposed a technique to estimate the spatially varying amount of blur over the image, and then magnify the existing blur of the out-of-focus regions, while maintaining the sharpness of the focused regions (see Figure 4.14). Since the blur (and depth) estimation from a single photo is not robust, their technique may suffer from incorrect blur estimation at sharp foreground-background edges. Hasinoff and Kutulakos [175] demonstrated a technique for combining multiple images with varying aperture diameters to simulate a larger aperture.

Mohan et al. [**?**] translated the lens and sensor of a camera parallel to one another in a synchronized fashion during the integration time of an exposure. This *destabilization* of the standard alignment of the sensor and lens allows them to introduce programmable defocus effects, including simulating a lens with a larger effective aperture size (see Figure 4.15). This technique of creating a *lens in time* allows the use of cheap and small lenses to produce defocus effects otherwise possible only with a larger lens on an SLR camera. While the technique works well for 1D translation, it only allows discrete sampling of the virtual aperture plane when used to simulate a 2D lens. This technique is similar to laminography, a technique historically applied in X-ray imaging to focus at distinct layers of a subject without using refractive elements. In contrast to more modern methods of computed tomography, laminography directly forms a sharp cross-sectional image by using synchronized motion rather than post-capture computation.

4.4.5 Wavefront Coding and Phaseplates

Geometric aberrations in lenses cause image distortions, but these distortions can be modeled, computed, and in some cases robustly reversed. In 1995, Dowski and Cathey [100, 76] used special aspheric optics (a cubic phase plate) instead of the lens to extend the depth of field of their imaging system (Figure 4.16). This technique forms images that are intentionally distorted, but makes the PSF independent of the



(a) Prototype Setup

(b) All-in-focus photograph with an $f/22$ lens

(c) Destabilized photograph by shifting an $f/22$ lens and sensor

Figure 4.15: Image Destabilization for reducing the depth of field. (Figure from Mohan [**?**])

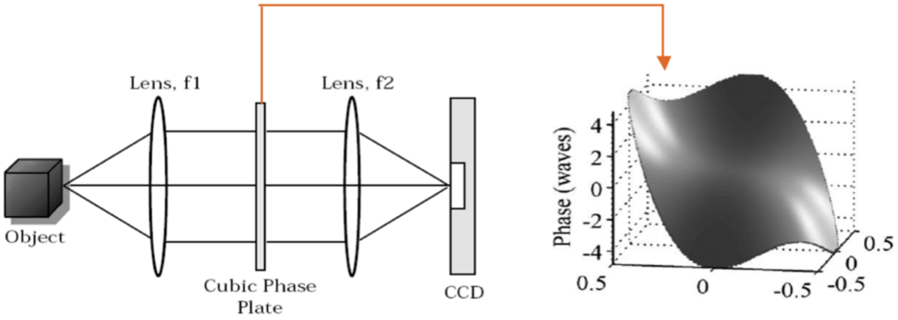


Figure 4.16: Wavefront coding using a cubic phaseplate. (Source probably one of the WFC papers)

degree of defocus or the scene depth. With a fixed PSF size, simple deconvolution techniques can be used to recover an in-focus image over an extended depth of focus.

[THERE IS MORE TO DO IN THIS SECTION ON WAVEFRONT CODING. ORIGINAL MS SAYS "TO DO—MAROUM'S THING]

4.5 Masks and Aperture Manipulation

A *mask* usually refers to a thin planar element that attenuates rays of light in a spatially varying fashion. In some applications masks may change or move during an exposure, and in some cases LCDs and DLPs [325] are used to create time-varying masks. While masks are used extensively in imaging applications such as astronomy, microscopy, and spectroscopes, their use in photography is relatively new. Coupled with appropriate computational methods, masks are becoming an important optical element, just like a lens or a prism is important in traditional optics.

(QUESTION IN MS FOR THE FOLLOWING PARAGRAPH—Do motion detectors really use masks? I think they take the gradient electronically rather than mechanically?) THIS PARAGRAPH ON MOTION DOESN'T BELONG IN THIS SECTION ON MASKS! (RW)]

Motion is detected when an infrared emitting source with one temperature, such as a human body, passes in front of a source with another temperature, such as a wall. You have probably noticed that your light is sensitive to motion, but not to a person who is standing still. That's because the electronics package attached to the sensor is looking for a fairly rapid change in the amount of infrared energy it is seeing. When a person walks by, the amount of infrared energy in the field of view changes rapidly and is easily detected. You do not want the sensor detecting slower changes, like the sidewalk cooling off at night. Your motion sensing light has a wide field of view because of the lens covering the sensor. Infrared energy is a form of light, so you can focus and bend it with plastic lenses. But it's not like there is a 2-D array of sensors in there. There is a single (or sometimes two) sensors inside looking for changes in infrared energy.

For ease of presentation, we categorize masks on the basis of the position of the mask relative to the camera's lens. The three positions are (1) a mask near the scene, (2) a mask at the limiting aperture of the lens, and (3) a mask near or on the image sensor. These three positions are illustrated in Figure 4.17. The boundaries between these categories are vague, however, and Gao et al. [147] proposed a simple optical relay that gives greater freedom in mask placement for most applications.

4.5.1 Lensless Imaging

The idea of putting a mask in the aperture of an optical system has been well explored in the fields of astronomy and scientific imaging [370]. Refined coded aperture methods using Modified Uniformly Redundant Arrays (MURA) [163] enabled lensless gamma-ray imaging systems for distant stars. Zand [**?**] has a nice survey of coded aperture techniques in astronomy. Figure 4.18 shows simulated

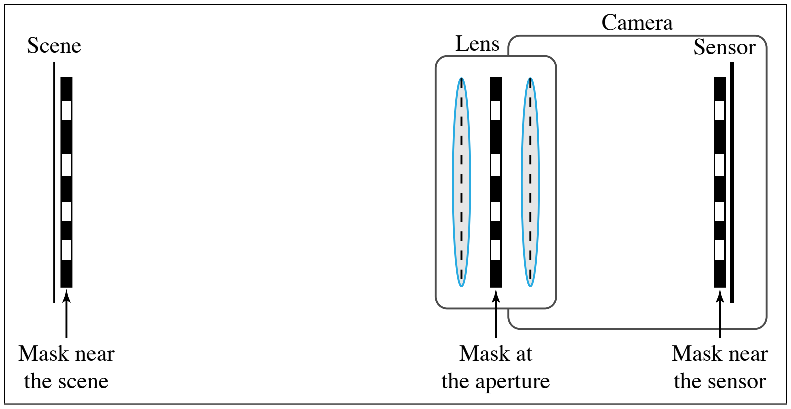
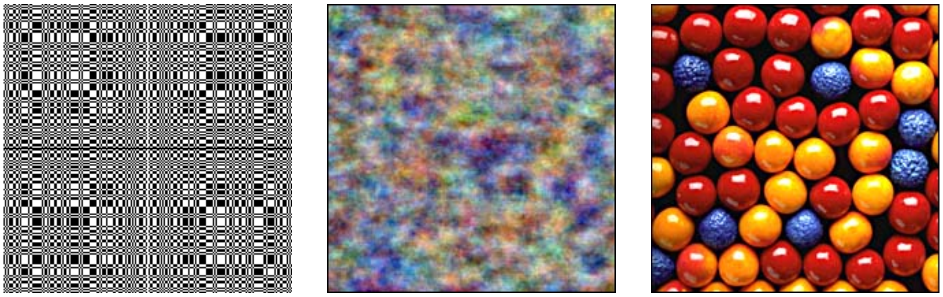


Figure 4.17: Mask positions with respect to the camera: a mask near the scene, a mask at the limiting aperture of the lens, and a mask near or on the image sensor.



(a) MURA pattern used as the aperture of a lensless camera

(b) Image captured by the camera with a MURA mask at the aperture

(c) Reconstructed image by performing correlation of the smeared image with the decoding pattern

Figure 4.18: Lensless imaging using a MURA mask at the aperture. (Figure from www.paulcarlisle.net/old/codedaperture.html.) [SECTION 4.5.1 SAYS THESE IMAGES ARE SIMULATED, BUT THEY LOOK AUTHENTIC. IF THEY ARE SIMULATED, THE CAPTION SHOULD INCLUDE THAT FACT.]

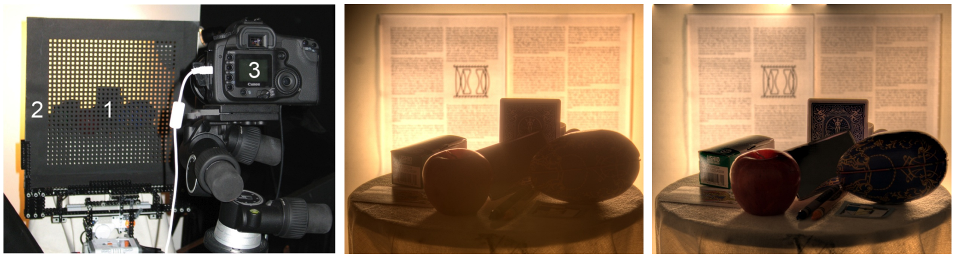


Figure 4.19: (left) Setup used by Talvala et al. [389] for glare reduction. (center) Glare reduces the contrast in the photo of a scene with backlight; (Right) Glare reduced and contrast increased using the technique described in the paper. (Figure from Talvala et al. [389])

images obtained when using a lensless camera with a MURA mask at the aperture. Such techniques are largely limited to scenes with point sources, such as in astronomy. The captured image has a much lower contrast than the scene, and this somewhat limits its applicability.

Lensless imaging is also used in diffraction-free microscopy. The camera consists of an image detector and a special aperture, but no lens. The aperture is a set of parallel light-attenuating layers whose transmittances are controllable in space and time. By applying different transmittance patterns to this aperture, it is possible to modulate the incoming light in useful ways and capture images that are impossible to capture with conventional lens-based cameras. Zomet and Nayar [**?**] proposed a lensless camera made up of a stack of parallel light attenuating masks in front of a bare image sensor. They modulated the incoming light by applying different transmittance patterns in space and time to the various layers. This general design gives the camera several novel capabilities such as capturing disjoint regions in a scene on a single photograph, and panning and tilting the field of view without using any moving parts.

4.5.2 Mask and Gratings outside the Camera

Placing a mask near a scene is closely related to the use of structured illumination as discussed in Chapter 5. Both of these cases attenuate rays from different parts of the scene differently, and this can give more useful information in the captured images. Recently Nayar et al. [292] used photos of a scene illuminated with high frequency striped patterns to separate the direct and global illumination components. Talvala et al. [389] used a similar technique to reduce the effect of veiling glare in a photograph. They captured multiple photos with slightly different positions of a checkerboard mask occluding parts of the scene, and used these to separate the direct and indirect components of the intra-camera light transport. Since glare is a global illumination effect, separating it from the scene data is greatly simplified. Figure 4.40 illustrates these results. Their capture process takes around an hour and is limited to static scenes. Furthermore, since the mask needs to be nearly in focus, they need to place it close to the scene or use a very small aperture, thus

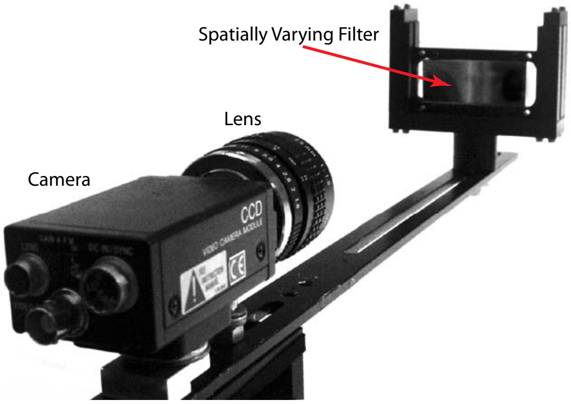


Figure 4.20: Setup used by Schechner and Nayar for generalized mosaicing [356]. (Figure from Schechner and Nayar [356])

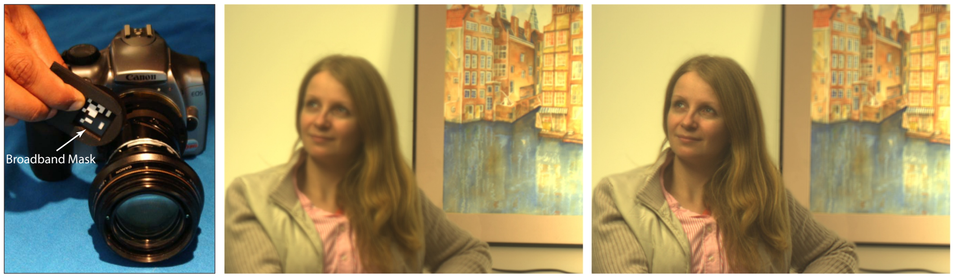
limiting the technique to studio settings.

Schechner and Nayar [356] rigidly attached a spatially varying mask some distance in front of the camera lens (as shown in Figure 4.20). Since the mask is not the limiting aperture of the optical system, different scene points are attenuated differently by the mask. Moving the camera-and-mask setup yields multiple measurements for each scene point under different optical settings, resulting in image mosaics with additional scene information such as extended dynamic range and multispectral data. This technique is called *generalized mosaicing*. The registration algorithm is non-trivial due to spatially varying effects of the filter. They also proposed a vision-based algorithm [358] to synchronize a changing mask in the optical system to the corresponding acquired image, thus allowing for uncontrolled modulation of the imaging system.

4.5.3 Mask at the Aperture

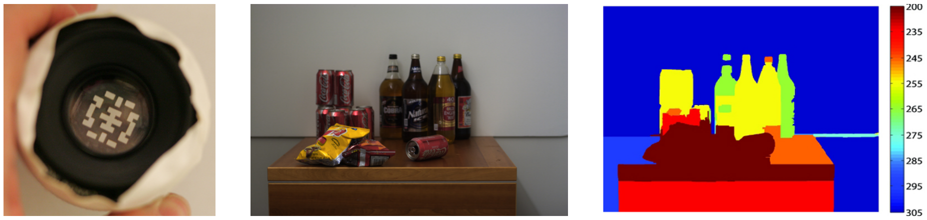
Placing a mask at the limiting aperture of a camera is considered a special case because the point spread function of such a camera is a scaled copy of the aperture mask (ignoring any diffraction effects). The image of an out-of-focus point light source in the scene is a scaled copy of the mask itself, and since the mask is also the limiting aperture, the point spread function is spatially invariant for all points in the field of view. Most techniques using the mask in the aperture assume the scene is planar and Lambertian; extending them to work in the general case is usually non-trivial.

Farid and Simoncelli [127] exploited defocus by using two calibrated masks to optically measure the differential variation in image intensities with respect to a change in camera position. They used this differential change for range estimation. Hiura and Matsuyama [185] introduced a multi-focus camera that captured three images with different focal planes simultaneously. They placed a mask with four pin holes in the aperture plane and estimated depth from defocus from the three captured images.



(a) (b) (c)

Figure 4.21: Broadband mask placed in camera aperture to make the deconvolution process well conditioned. The image obtained after deconvolution is sharp and without any ringing artifacts (right). (Figure from [414])



(a) Mask placed in lens aperture.

(b) Input (single image)

(c) Estimated depth

Figure 4.22: Depth estimates from a conventional camera with a coded aperture. (Figure from Levin [231])

Veeraraghavan et al. [414] put a broadband mask at the lens aperture to make deconvolving the out-of-focus blur a well-posed problem. This allows them to recover refocused images at full resolution for layered Lambertian scenes (see Figure 4.21). Levin et al. [231] placed a similar mask at the lens aperture that helps to discriminate the depth of the out-of-focus parts of the photo. They use this technique to obtain an all-in-focus image, together with a layered depth map of the scene (see Figure 4.22).

Liang et al. [245, 244] proposed a programmable aperture camera that captured the light field by taking multiple photos, each with a different aperture shape. This can be thought of as re-binning the 4D light field onto a 3D sensor, where time is the third dimension. The resulting light field has higher spatial resolution than that obtained with other approaches. This approach may not work for dynamic scenes, however, since it requires multiple photos. Figure 4.23 shows two prototypes that they proposed: the first uses a scrolling pattern in the aperture plane of the lens, and the second uses a custom LCD module placed in the aperture plane. The LCD offers near-real-time control of the aperture shape.

[NOTE: MS INCLUDES THE FOLLOWING TO-DO ITEM HERE]

Superresolution. Mohan et al. [277]. Talk about Roarke's paper. Horstmeyer et al. [189]

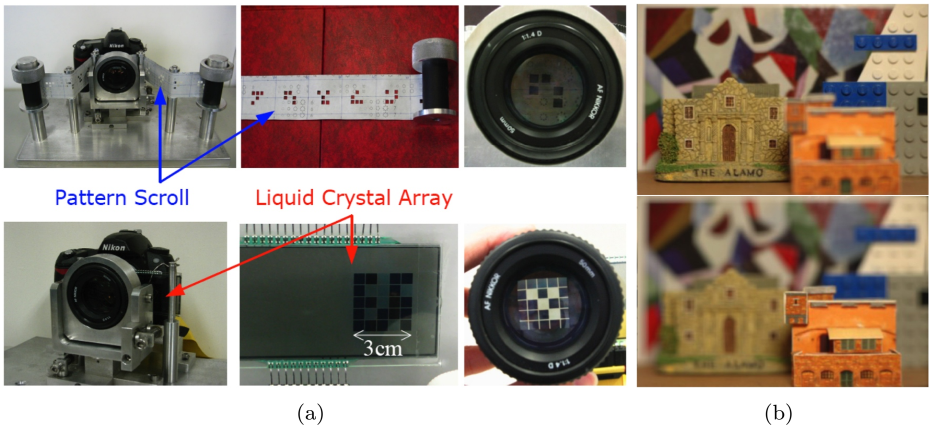


Figure 4.23: (a) Prototypes of the programmable aperture cameras with (first row) aperture patterns on an opaque slip of paper, and (second row) on an electronically controlled liquid crystal array. (b) High resolution refocusing results. (Figure from Liang et al. [244])

4.5.4 Mask near or on the Sensor

Since the sensor plane is conjugate to the plane in focus outside the camera, placing a mask near the sensor is very similar to placing a mask near the photographed scene. However, placing an arbitrary mask next to the sensor can be much easier than modifying the scene, and this makes masks near or on the sensor more interesting.

The most common example of a mask on the sensor of a conventional digital camera is the Bayer filter [55]. The Bayer filter allows a monochrome sensor to capture a full color image in a single shot by what can be thought of as spatial re-binning of rays based on wavelength. Narasimhan and Nayar [**?**] took this basic idea further and proposed assorted pixels, where they use similar structured interpolation between neighboring pixels for applications such high dynamic range color imaging using a mosaic of overlapping color and exposure filters.

Nayar and Mitsunaga [293] placed a spatially varying optical mask on the camera's sensor. This gives adjacent pixels of the detector different exposures of the scene, and they used image reconstruction to obtain a high dynamic range photo. Nayar and Branzoi [290] extended this idea by placing an LCD next to the sensor or at the lens aperture, which adaptively controls the exposure for each pixel based on the radiance value of the corresponding scene point in real time.

Nayar et al. [291] used a digital micro-mirror device (DMD), as illustrated in Figure 4.24, to build a programmable imaging system. The imaging lens focuses the scene onto a programmable array of micro-mirrors, and a re-imaging lens images the DMD onto the final sensor. The DMD is optically in a plane conjugate to the sensor plane. The orientations of the mirrors of the array can be controlled with high precision over space and time, and the DMD allows flexible pixel-wise modulation of the sensor array. This enables the system to select and modulate

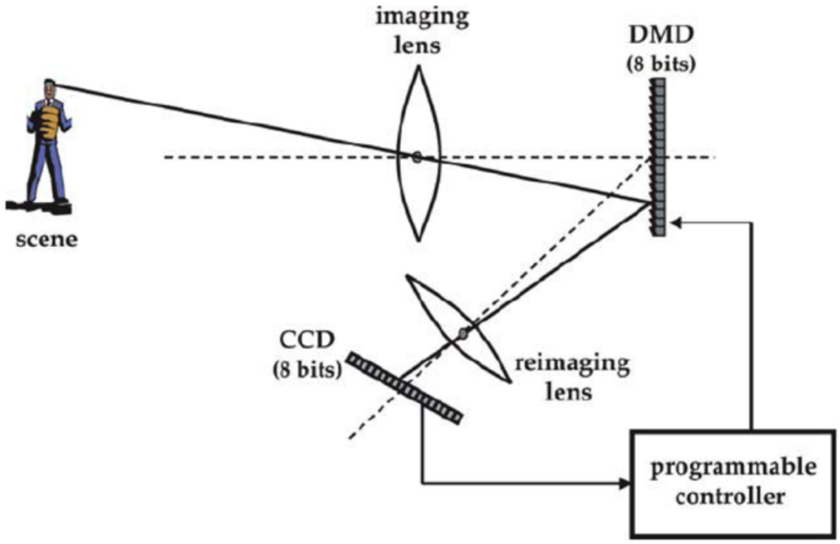


Figure 4.24: Imaging using a digital micro-mirror device (DMD). The scene image is focused onto the DMD plane. The image reflected by the DMD is re-imaged onto a CCD. The programmable controller captures CCD images and outputs DMD (modulation) images. While technically the mask (DMD) is not near the sensor, it is optically in the sensor's conjugate plane. (Figure from Nayar [291])

rays from the scene's light field based on the needs of the application at hand. In addition to applications in high dynamic range imaging, they also demonstrated simple image processing in the optical domain, such as feature detection and object recognition.

Takhar et al. [388] used *compressive sensing* to construct an efficient single pixel camera. They obtained a series of pseudo-random linear projections of all scene intensities by imaging the scene onto a DMD array and integrating the reflected rays onto a single photon detector (see Figure 4.25). Unlike Hadamard coding, the number of samples needed is fewer than the number of pixels. [DEFINITIONS AND MORE DETAIL NEEDED IN THIS PARAGRAPH]

[MS SAYS THE FOLLOWING MATERIAL IS "ASHOK'S STUFF" ON FREQUENCY DOMAIN RE-BINNING]

Designs based on the idea of integral photography [IN SECTION 4.4.2 THIS IS CALLED INTEGRAL IMAGING] spatially re-bin the 4D light field onto a 2D sensor. Veeraraghavan et al. [414] replaced the microlens array with a sinusoidal mask placed close to the image sensor, and achieved similar re-binning, but in the frequency domain. Their design builds on the basic ideas of heterodyning from signal processing and communication theory, and can be thought of as *spatial optical heterodyning.* The 4D light field is recovered from the Fourier transform of the captured 2D image. In microlens-array-based design, each pixel effectively records light along a single ray bundle. With patterned masks, each pixel records a linear combination multiple ray-bundles. Georgiev [155] proposed a generalization

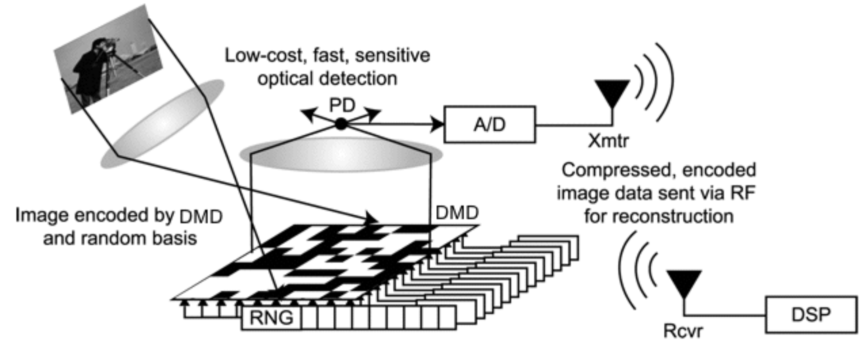


Figure 4.25: Single pixel camera. Multiple pseudo-random linear projections of all the scene intensities are captured on a single pixel, and combined to form a high resolution image. (Figure from Takhar [388])

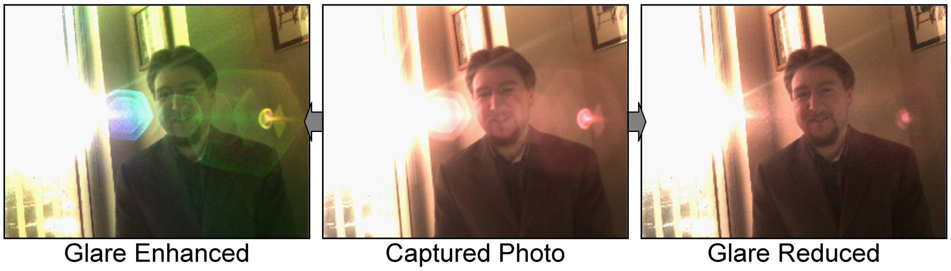


Figure 4.26: Mask-based glare reduction. The glare shows up as a high frequency artifact in the 4D light field analysis, and is easy to remove. (Figure from Raskar [332].)

of frequency domain multiplexing using masks, meshes, and pinholes. and Veeraraghavan et al. [412] analyzed the effect of a non-refractive ray filter on a light field.

Raskar et al. [332] analyze veiling glare in 4D ray space. They use the fact that glare is essentially high frequency noise in the 4D space to reduce its effect without reducing the image resolution significantly (see Figure 4.26). They do not capture the full 4D light field, but their approach is heavily based on the light field analysis.

[TO DO NOTE IN MS: Raskar et al. is discussed in epsilon photography chapter as well; remove details from here.]

4.5.5 Aperture Manipulation

Several researchers have investigated modifying the lens aperture in clever and interesting ways other than simply placing a mask in the aperture or in the optical path.

Pentland [316] estimated range by using the finite depth of field for a given

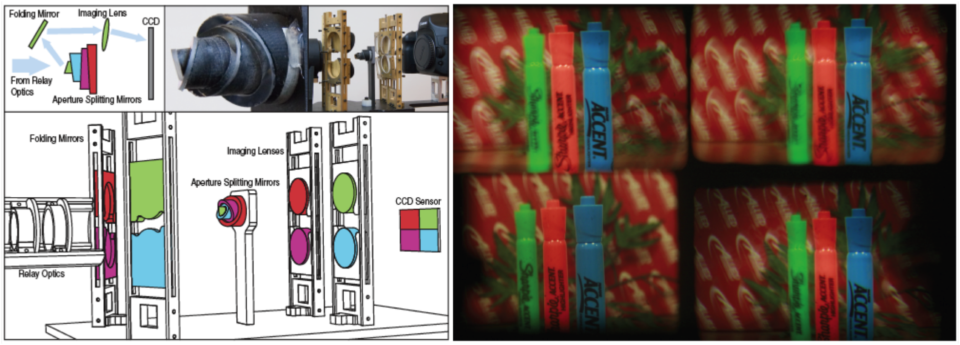


Figure 4.27: (left) Photographs and schematic diagrams of an optical system used by Green et al. [164] to capture multi-aperture images. The system is designed as an extension to a standard digital SLR camera and consists of a main photographic lens imaged through relay optics and split by using a set of tilted mirrors. (right) Sample images captured by the system. (Figure from Green et al. [164])

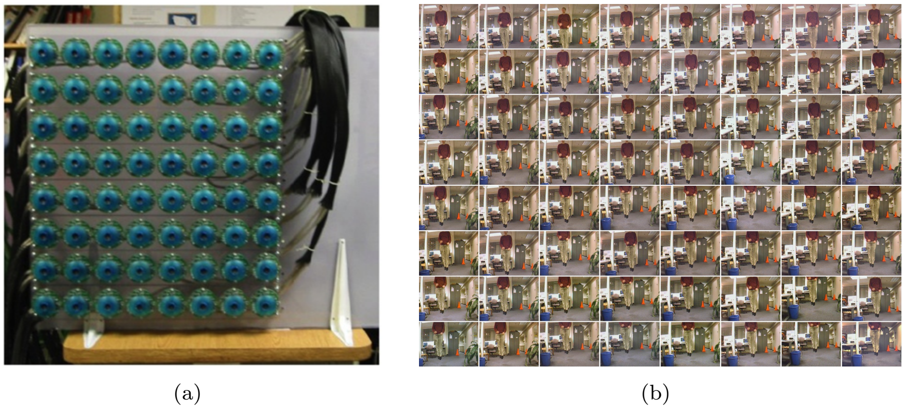
aperture size. Hasinoff and Kutulakos [174] captured several hundred photos of a scene with all possible combinations of aperture size and focus settings, and used this extensive data to compute the 3D shape of the scene. Gao and Ahuja [146] placed a rotating glass plate in front of the camera lens. Each rotation angle provides a stereo pair, and the large number of stereo pairs from various rotation angles gives a robust depth estimate for the scene.

Aggarwal and Ahuja [38] split the aperture into multiple parts, and used an assembly of mirrors to direct the rays arriving at each part onto different sensors. They used a different exposure setting on each sensor, and merged the captured images to construct a high dynamic range image. McGuire et al. [267] performed real-time video matting by using a setup of multiple cameras to capture multiple video streams, each with different aperture size and plane of focus. Later McGuire et al. [266] proposed a general solution, using beam-splitters, to construct efficient cameras that capture an arbitrary number of pixel-aligned images in one shot.

Fergus et al. [133] used an arrangement of random mirrors in place of a lens to acquire several random projections of a light field, and reconstructed the scene by using basis pursuit from compressive sensing and a machine learning model. They explored applications in 3D imaging and superresolution.

Hasinoff and Kutulakos [175] captured multiple photos of a scene with varying aperture size, and used these photos to change the focus setting and depth of field in a post-processing environment. [MS CONTAINS A NOTE HERE— ”WRITE MORE.”]

As shown in Figure 4.27, Green et al. [164] proposed an optical system that captures four images of the scene with different aperture settings in a single shot, and uses these images to modify the depth of field and the focus.



(a)

(b)

Figure 4.28: Light field camera proposed by Yang et al. [434]. (left) A photo of a 64-camera light field camera array. The cameras are arranged in rows of eight. (b) An array of photos captured by the system. Figure from Yang et al. [434]

4.5.6 Camera Arrays and Synthetic Aperture

An approach closely related to aperture manipulation in a single camera is the use of multiple cameras in camera arrays, such as by Wilburn et al. [427], Yang et al. [434], and others. They used a square array of off-the-shelf video cameras to capture the light field with a very large synthetic aperture. The basic idea of synthetic apertures in computational photography is similar to synthetic apertures in radar technology, where a highly directional conventional rotating antenna is replaced with many low-directivity small stationary antennas scattered over some area near or around the area of the object being imaged,

Synthetic aperture focusing consists of warping and adding together the images in a 4D light field so that objects lying on a specified surface are aligned and thus in focus, while objects lying off this surface are misaligned and hence blurred. This provides the ability to see through partial occluders such as foliage and crowds (see Figure 4.29), making it a potentially powerful tool for surveillance [410]. Vaish et al. [409] later proposed a generalized synthetic aperture approach for tilted focal planes and arbitrary camera configurations. However, by their very nature, camera arrays are cumbersome and have somewhat limited applications. Synthetic aperture using camera arrays also suffers from aliasing issues due to inadequate sampling of the virtual aperture space.

Confocal microscopy is a family of imaging techniques that employ focused patterned illumination and synchronized imaging to create cross-sectional views of 3D biological specimens. It was first proposed by Marvin Minsky [272] in a 1957 patent (Figure 4.30). Levoy et al. [239] adapted confocal imaging to large-scale scenes by replacing the optical apertures used in microscopy with arrays of real or virtual video projectors and cameras. A dense array of projectors simulates a wide aperture (synthetic aperture illumination) projector, which can produce a real image with small depth of field. By projecting coded patterns with an array of virtual projectors and combining the resulting views, we can selectively image

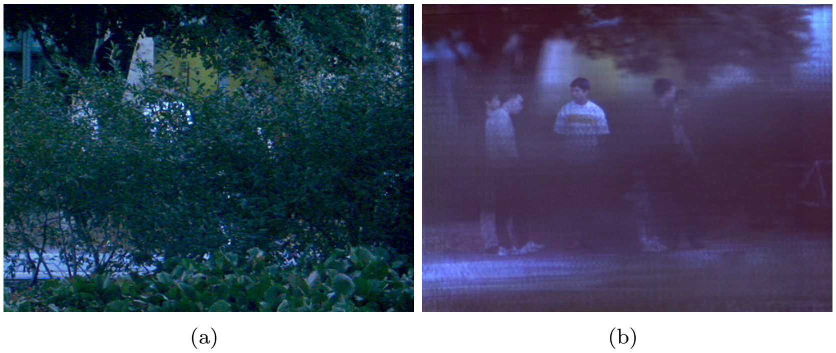

(a) (b)

Figure 4.29: Synthetic aperture using camera arrays. (a) Photo from a single camera. (b) Image obtained by combining photos from each individual camera in the array. (Figure from Vaish [410])

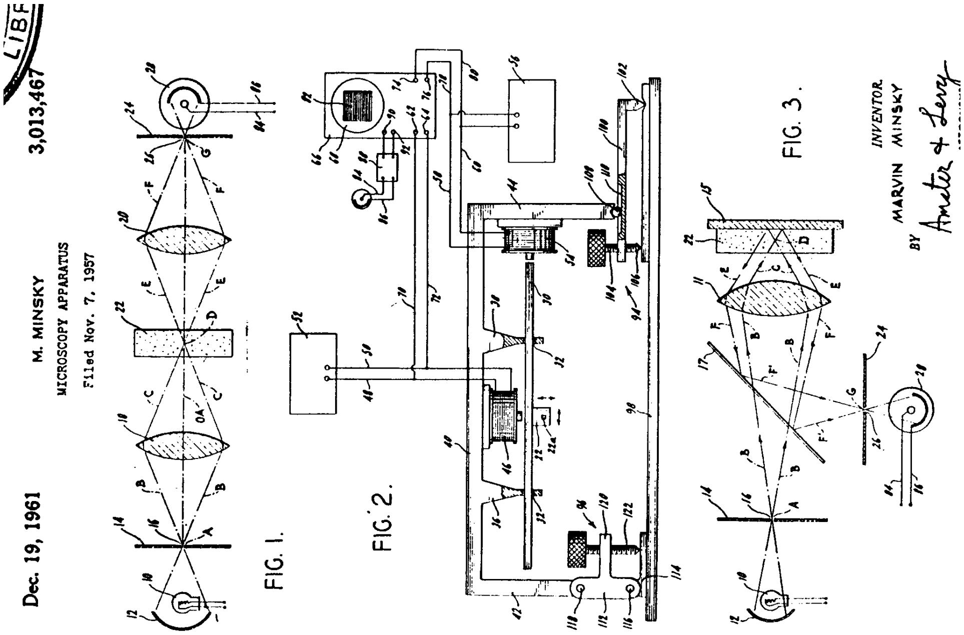


Figure 4.30: Page from Marvin Minsky's patent on confocal imaging, filed in 1957.

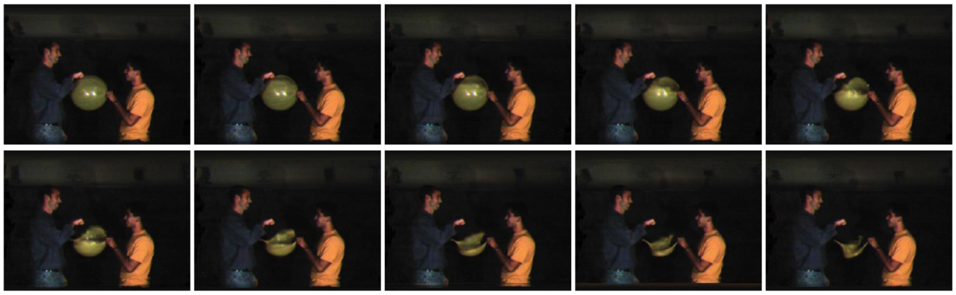


Figure 4.31: High speed photography using camera arrays. (Figure from Wilburn [426].)

any plane in a partially occluded environment. These ideas were demonstrated on enhancing visibility in weakly scattering environments, such as murky water, to compute cross-sectional images and to see through partially occluded environments, such as foliage.

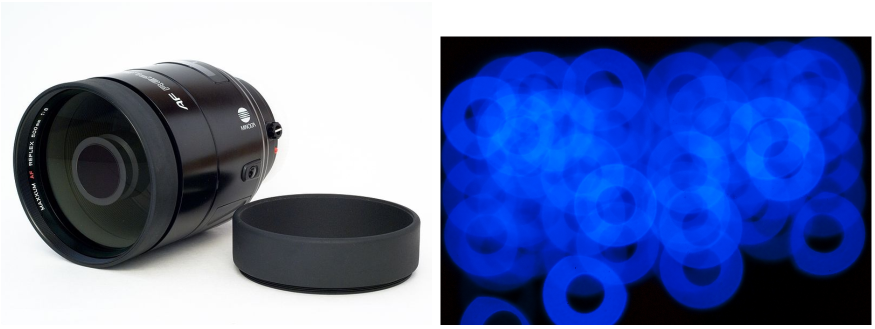
Wilburn et al. [426] proposed a system for capturing high speed videos (multi-thousand frames per second) by using a dense array of relatively cheap off-the-shelf cameras. The camera array allows researchers to capture higher speeds by adding more cameras. The fundamental limit to the scalability of the system is determined by the minimum integration time of the camera. Unlike a single camera, the dense array system allows the possibility of overlapping exposure intervals, which could allow for temporal superresolution. It is assumed that the scene is either relatively planar or reasonably far from the camera. Projective transforms align the pictures captured by the various cameras. The cameras are also color calibrated to get consistent color matching in the resulting video. Joshi et al. [210] used a camera array and synthetic aperture refocusing for robust real time natural video matting.

4.6 Catadioptric Imaging

To bend light and form an image, optical systems often use reflective mirrors in addition to refractive elements such as lenses. An optical system that uses both lenses and curved mirrors is called a *catadioptric* optical system. While bending light by using a mirror results in physically smaller optical systems, aberration is often a problem with catadioptric imaging (though chromatic aberration is minimal).

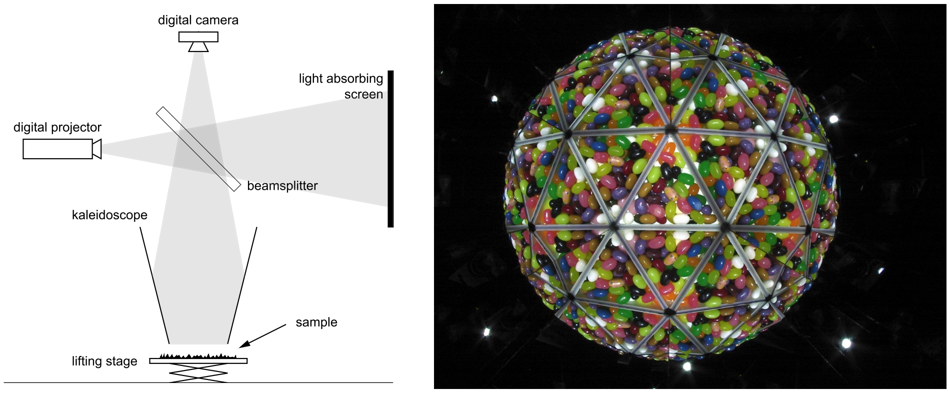
Catadioptric optical systems have traditionally been used in telescopes and long focal length lenses in photography. Figure 4.32 shows the Minolta 500mm Reflex lens, which uses a curved mirror to reduce the physical length of the lens. While the lens is smaller and lighter than a refractive-only lens of a similar focal length, it produces a donut-shaped bokeh (point spread function for an out-of-focus point source) because of the mirror that occludes light in the middle of the front lens element.

While not strictly a catadioptric system (it uses flat mirrors and no lenses), Han and Perlin [169] used a tapered kaleidoscope with a single camera to view the same



(a) (b) Donut shaped bokeh from the lens

Figure 4.32: (a) Minolta AF 500mm Reflex lens. (b) A donut-shaped bokeh image from the Minolta lens, from www.flickr.com/photos/26068133@N07/3122669401. [WHO OWNS THIS IMAGE?] The bokeh is due to the mirror in the reflex lens, which occludes light in the front lens element.



(a) (b)

Figure 4.33: Kaleidoscope optics to measure the bidirectional texture function (BTF) of a sample surface. (Figure from Han [169])

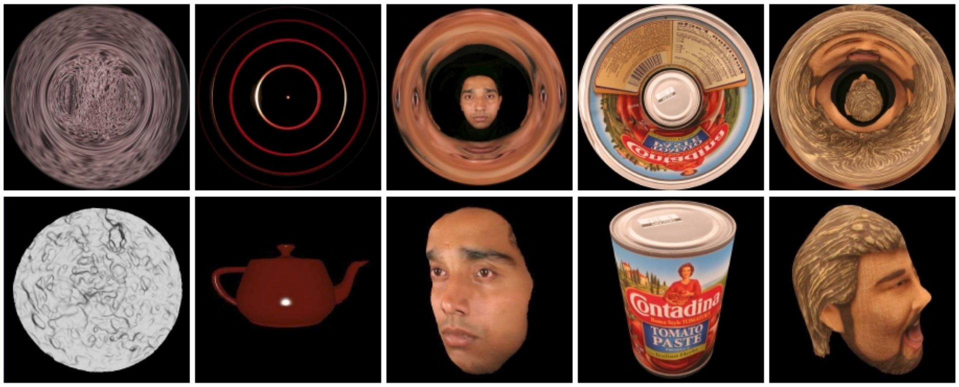
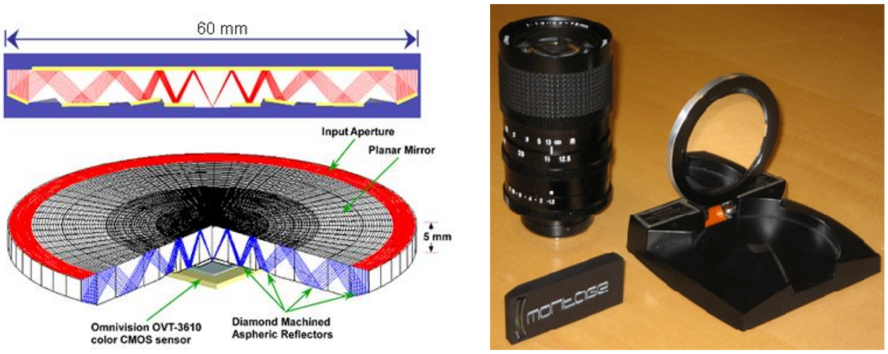


Figure 4.34: Radial catadioptric optics (Figure from Kuthirummal et al. [222])

surface simultaneously from many directions (Figure 4.33). They coupled this with the ability to illuminate the surface simultaneously from many directions by using a single light source to measure the bidirectional texture function (BTF) of the surface *in situ*. The BTF is a 6D function that contains the variation in texture (x, y) with the illumination (θ_i, ϕ_i), and viewing (θ_o, ϕ_o) directions [88]. Capturing the BTF typically requires hundreds of images, and can take several hours. Han and Perlin used a tapered kaleidoscope to simulate the effect of an entire array of cameras pointing toward the sample from different viewing directions. The strength of their method is the single-shot *in situ* capture of the BTF.

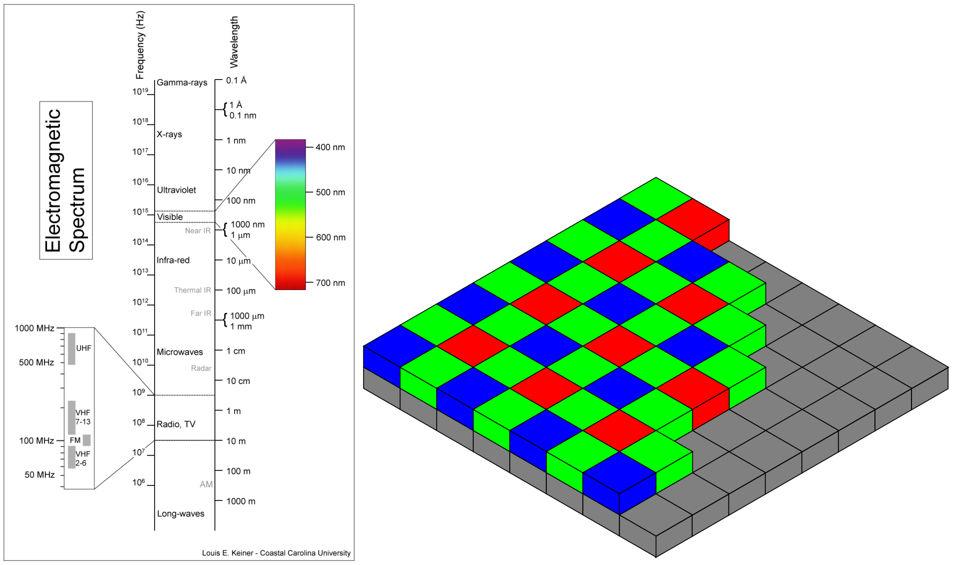
Kuthirummal and Nayar [222] describe a class of imaging systems, called radial imaging systems, that use a camera and a curved mirror to capture a scene from a large number of viewpoints within a single image. These systems can recover scene properties such as geometry, reflectance, and texture, and they can be used to derive analytic expressions that describe the properties of a complete family of radial imaging systems, including their loci of viewpoints, fields of view, and resolution characteristics. Some of these radial imaging systems can, from a single image, recover the frontal 3D structure of an object, generate the complete texture map of a convex object, and estimate the parameters of an analytic BRDF model for an isotropic material. In addition, one of the systems can recover the complete geometry of a convex object by capturing only two images. Radial imaging systems such as these are simple, effective, and convenient devices for a wide range of applications in computer graphics and computer vision.

Tremblay et al. [**?**] proposed an *origami lens*, which is an ultrathin telephoto lens that bends light on multiple surfaces, as shown in Figure 4.35. Light enters through an annular aperture and bounces back and forth between reflecting surfaces to finally produce a focused image on the sensor.



(a) (b)

Figure 4.35: Folded optics origami lens. (Figure from Tremblay et al. [?]



(a) (b)

Figure 4.36: (a) Electromagnetic spectrum including visible light. (b) Bayer filter over a sensor. (Figures from wikipedia.)

4.7 Color and Wavelength

Most color films have three layers of emulsions, each sensitive to a different wavelength range. The *spectral response curve* provides a measure of relative sensitivity of the film emulsion to the different wavelengths of light. Films based on Agfacolor-Neu contain the color couplers along with the emulsion layers, while Kodachrome film required the addition of color dyes during film processing [85]. Film photographers commonly use films with very different spectral response curves for different applications. For example, Fuji Reala is used extensively for portrait photography because of its accurate skin tones, while Fuji Velvia is an obvious choice for landscapes and sunsets because of its saturated reds and greens. Joseph Friedman's classic book on the history of color photography [141] gives an excellent overview of some very clever techniques used for color scene capture over the past century.

In a digital camera, the Bayer filter [55] is a three-color (red, green, blue) filter array placed over the sensor. The purpose of this array is to separate light intensities into three distinct color channels in order to obtain color photographs. Different colored filters are placed over adjacent sensor sites in a checkerboard-like pattern (50% green, 25% red, 25% blue). Figure 4.36(b) illustrates how each sensor site captures light of only one of the three primary colors. A demosaicing algorithm reconstructs a high resolution image from this sparse sampling [33].

The Foveon X3 digital sensor [364] uses a different structure to produce color photographs. Each pixel has three layers, and each layer is sensitive to one of the three primary colors—red, green, and blue. This structure is similar to the multilayer color emulsions used in film, and demosaicing is not required to produce the resulting photograph. DLP projectors [325] use a color wheel with three or more color filters that act as color primaries for the projector. In all these cases, the color filters are fixed at the factory and are not user-replaceable. Moreover, the exact spectral response on the filters is adjusted so that they work reasonably well for most scenes. It is not possible to optically tweak the spectral response, depending on the scene. Cameras do allow limited post-capture processing, but this is limited by the dynamic range and the bit-depth captured by the sensor.

4.7.1 Imaging Spectrometers and Multispectral Cameras

The idea of dispersing light to measure its spectral components is certainly not new. Spectrometry has been an area of research ever since Newton discovered dispersion of light [299]. The book by John James et al. [199] provides an excellent survey of sophisticated spectrograph designs perfected over the years. The basic idea of a spectrograph is simple. Light enters through a slit and a collimating lens and is passed through a prism or a diffraction grating. Rays of different wavelengths are bent by different amounts because of dispersion, and a lens then focuses these rays onto a digital sensor, film, or direct-view optics. Since the power on narrow wavelength bands might be extremely small, most spectrometers typically use more sophisticated basis functions, such as the Hadamard or S-matrices [173], for better efficiency.

Spectrometry traditionally analyzes the spectra of point sources, or spatially uniform diffuse sources. The field of *imaging spectrometry* or *multispectral photog-*

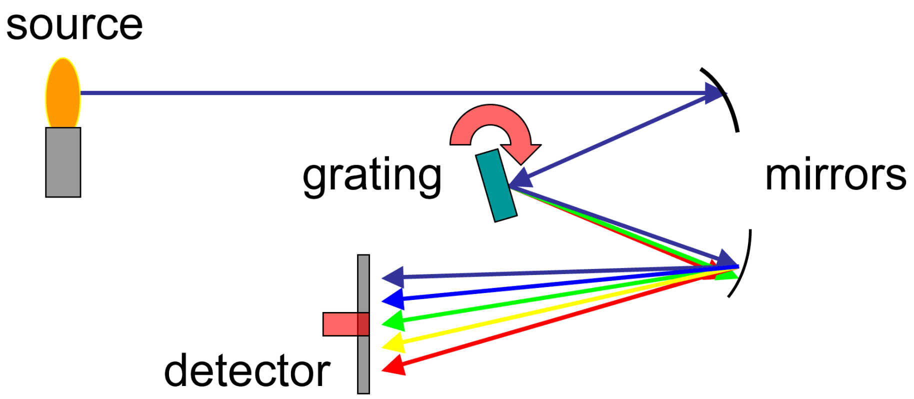


Figure 4.37: Ray diagram showing basic idea behind a diffraction-grating-based spectrometer (Figure from wikipedia).

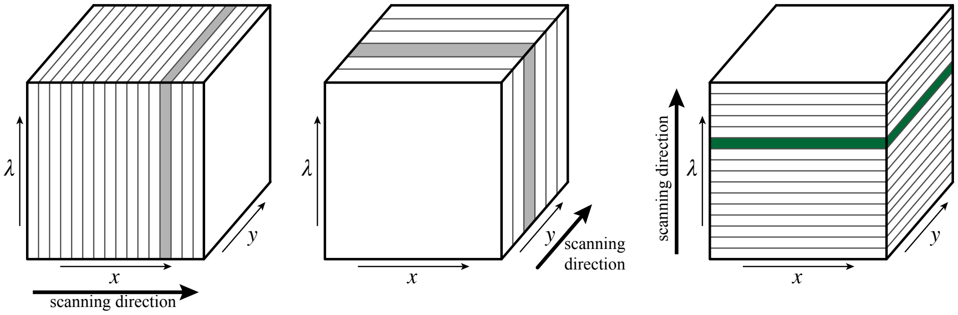
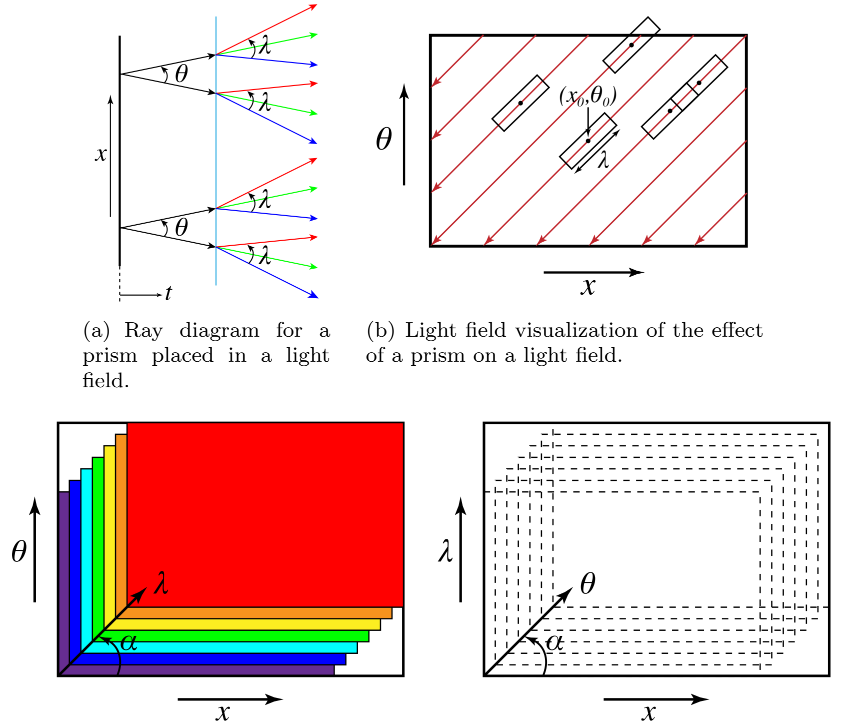


Figure 4.38: Scanning techniques used for acquiring the spectral data cube (x, y, λ). (left) Scan along the x-direction; (center) scan along the y-direction; (right) scan along the wavelength, λ.



(a) Ray diagram for a prism placed in a light field.

(b) Light field visualization of the effect of a prism on a light field.

(c) Overlapping light fields of different wavelengths, $L_o(x, \theta)$.

(d) Blurred spectral light field for a fronto-parallel Lambertian painting.

Figure 4.39: Placing a dispersing prism in a light field gives us many shifted copies of the light field, resulting in a set of *overlapping spectral light fields*. A fronto-parallel Lambertian painting results in a *blurred spectral light field*.

raphy is relatively new. A multispectral camera captures the *data cube*, (x, y, λ), which is equivalent to a spectrometer for each spatial scene point. Since image sensors are only two dimensional, capturing a 3D data cube typically requires scanning over one of the three dimensions. Figure 4.38 illustrates three ways to scan the data cube: along the two spatial dimensions (x, y), and along the spectral dimension (λ). Harvey et al. [172] compared the SNR for various imaging spectrometers. This work is an excellent source for more information about multispectral cameras.

Gat [149] provides a nice review of liquid crystal tunable filters (LCTF), acousto-optical tunable filters (AOTF), and interferometers and their applications in imaging spectroscopy. Placing one of these filters in front of a camera allows a controllable wavelength of light to pass through. A series of images are captured at different wavelengths, and then merged to form one multispectral photograph. Unfortunately, these filters are rather expensive, and usually allow only a single wavelength of light to pass through (these are also called *notch pass* filters).

A classic spatial scanning imaging spectrometer accepts polychromatic input and disperses spectral information across one dimension and spatial information across the orthogonal direction [308]. A diffraction grating or a prism performs the dispersion, with the grating having more throughput. Much of the research in the area of imaging spectrometers was done for airborne remote sensing applications. A spectrometer that scans a scene by using a viewing platform's forward motion, without any active scanning motion, is called a *pushbroom camera*. A *whiskbroom camera* scans a linear array across a scene by using a gimbal mechanism.

A linearly variable interference filter (LVIF) transmits light by using interference films that vary in thickness along one dimension. The filter acts as a spatially varying spectral notch filter, and transmits a narrow spectral notch centered around a wavelength that is a linear function of the spatial position on the filter. A wedge imaging spectrometer [96] consists of a LVIF mated directly to the image sensor. This gives a single, compact, integrated assembly, and avoids the use of gratings and prisms. Detected information varies spatially in one dimension and spectrally in the other. Forward motion of the viewing platform builds a complete multispectral image. Schechner and Nayar [357] rigidly attached a similar LVIF to a camera and moved the camera to capture a mosaic. They then combined the multiple photos to generate a multispectral mosaic.

Li and Ma [243] proposed a novel design for an imaging spectroscope. They used a monochromator to disperse the rays of specific wavelengths from a scene, and then they used a moving slit to select a single wavelength to pass through. Then they used another monochromator to recombine the rays. [EXPLAIN THE ADVANTAGES OF THIS TECHNIQUE]

Recently Gehm and Brady [151] proposed a pushbroom scanning or tomographic (rotational) scanning imaging spectrometer that uses a coded spectrometer proposed earlier by Gehm et al. [152] as the spectral engine. Willett et al. [429] used compressive sensing techniques for multispectral imaging. Gat et al. [150] used reformatted fiber optics to map a 2D image to a linear array, and fed that linear array to the input slit of a 1D imaging spectrometer. This technique allows for very fast multispectral capture, which is ideal for rapidly changing phenomena.

Chromotomography methods reconstruct a data cube (x, y, λ) by capturing multiple 2D projections of a 3D cube. They then use tomography to combine the collection of projections to get an estimate of the unknown data cube. Since all the incoming light is used, the acquisition process is fast and efficient. Levin and Vishnyakov [236] were probably the first to discover this technique. Several other researchers independently came up with similar designs with minor changes. Okamoto and Yamaguchi [306] merged five projections obtained with a 2D transmission grating and a TV camera. Bulygin and Vishnyakov [71] used prisms with different dispersion factors. Betremieux et al. [62] constructed a 2D spectral imager by rotating a 1D spectrometer and by using the same tomograph-based reconstruction technique. Descour and Dereniak [97] proposed the Computed Tomography Imaging System (CTIS) by using crossed gratings, and later a more efficient computer generated hologram [98]. Mooney et al. [280] used a rotating direct vision prism (Amici prism) and captured multiple images. Unfortunately all these methods suffer from the *missing cone problem* because physics restricts the directions in which we can obtain projections. Johnson et al. [203] proposed the use of spatially varying masks to solve this problem. Their setup better recovers energy in the high spectral and low spatial frequencies.

4.7.2 Multispectral Projectors

Controlling the spectrum of a light source is traditionally done by using various types of lamps in combination with diffraction-grating monochromators and color filters. Recently several people have used a digital micromirror device (DMD) along with a spectrometer and a broadband lamp to create a programmable spectrum light source [417, 251, 69, 128]. These methods typically use a diffraction grating to disperse white light, and focus the dispersed wavelengths onto a DMD. The DMD programmatically bends the desired wavelengths toward an exit lens, followed by an integration sphere or another prism to recombine the dispersed rays.

Nelson et al. [298] described a system to project a linear (1D) image with controllable spectrum for each point along it. They positioned a DMD at the sensor of a spectrograph. Light first traverses through the spectrograph in the forward direction, and then again through it in the reverse direction. The 2D DMD controls the wavelength spectrum for each point along the outgoing slit. Rosenthal et al. [346] used a diffraction grating to disperse light, modulate it differently for each pixel in a scanline, and then project one scanline at a time by using a scanning mirror arrangement to form the image.

Among the rapidly expanding choices in illumination sources is the use of narrowband LEDs to illuminate an object and acquire multispectral images [271]. Fryc et al. [142] proposed a spectrally tunable light source using a large number of LEDs and an integrating sphere. Similarly, several new projectors use more than three LEDs to get better color rendition [382]. Park et al. [313] proposed a spectral multiplexing scheme to illuminate a scene by using multiple spectral sources to recover spectral reflectances in the scene.

Rice et al. [341, 339, 340] proposed a hyperspectral image projector design that is perhaps the most similar to the design proposed in Chapter **??**. [LATEX INDEX IS "CHAP:AGILE"—WHERE IS THIS PROPOSAL? WHAT CHAPTER?] Their system has two parts—the spectral engine and the spatial engine. They use a reverse spectrograph [341], a forward spectrograph with a spatial integrator [339], or a double subtractive spectrograph [340] for their spectral engine. All cases essentially obtain a tunable light source by using dispersion, a DMD, and a broadband light source. This tunable source is fed into a spatial engine, which is essentially a DLP projector. By controlling the wavelength emitted by the source *before* the spatial modulation of a DLP projector, they can control the color displayed at each pixel on the screen. This technique requires modifying the light source inside a projector. Unlike their approach, I use light field analysis and modulate the image spectrum *after* the spatial modulation. [WHO IS THE "I" HERE? RASKAR? TUMBLIN? DO WE HAVE A REFERENCE?] This greatly simplifies the projector design since we can easily retrofit a traditional RGB projector by placing lenses and a diffraction grating outside it. In addition, this modification allows us to use the same basic optical setup both as an image projector and as an image capture system.

4.8 Suppression of Glare

A scene with a bright light source in or near the field of view is difficult to photograph because of glare, which occurs because of multiple scattering of light inside

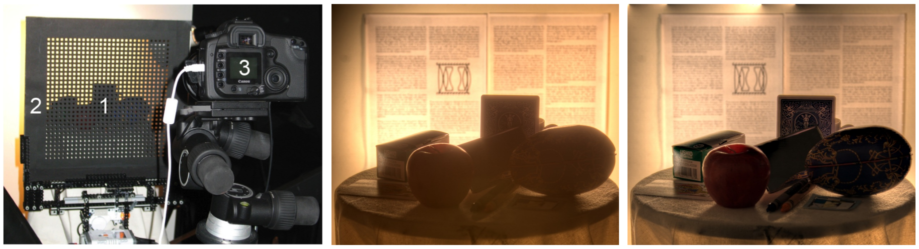


Figure 4.40: (left) Setup used by Talvala et al. [389] for glare reduction. (center) Glare reduces the contrast in the photo of a scene with backlight. (right) Glare reduced and contrast increased by using the technique described in the paper. (Figure from Talvala et al. [389])

the lens optics and the body of the camera. A primary result of glare is reduced image contrast. Glare is unavoidable; it disrupts every optical system, including the human eye. Glare can be broadly classified in two ways: glare due to reflection (Fresnel reflection at lens surfaces) and glare due to scattering (diffusion in lenses). The reflection of the retro-reflective camera sensor on the rear lens element can also result in additional glare.

High-end lenses use special optical design and materials to reduce glare. Lensmakers' strategies include coating and lens shaping. The 4% to 8% transmission loss due to reflection at each glass-air interface means that a 5 to 10 element lens can lose half the incident light and instead create significant reflection glare. Anti-reflective coating films make use of the light-wave interference effect. Vacuum vapor deposition coats the lens with a 1/4 wavelength thin film using a $\sqrt{n}$ refractive index substance, where n is the lens glass index. Multilayered coating can bring down the reflection to as low as 0.1%. But this is not sufficient to deal with light sources, which may be 4+ orders of magnitude brighter than other scene elements. Ancillary optical elements such as filters also increase the possibility of flare effects. Meniscus lenses with a curved profile act as a spherical protective glass in front of the lens assembly and prevent unwanted focused reflections from the sensor. The curved profile defocus creates large area flare rather than ghosts. Lens makers use an electrostatic flocking process to directly apply an extremely fine pile to surfaces requiring an anti-reflection finish. The pile stands perpendicular to the wall surfaces, thus acting as Venetian blinds—an effective technique for lenses with long barrel sections. Structural techniques include light blocking grooves and knife edges in lenses to reduce the reflection surface area of lens ends. Hoods or other shading devices are recommended for blocking undesired light outside the picture area.

Talvala et al. [389] presented a technique that prevents glare-producing light from reaching the sensor pixels. They selectively block glare-producing light by using a structured checkerboard occlusion mask, combined with a new direct-indirect separation of lens light transport to eliminate glare. They captured multiple photos with slightly different positions of the checkerboard mask occluding parts of the scene. Since glare is a global illumination effect, separating it from the scene data is greatly simplified. Their capture process takes around an hour and is limited to

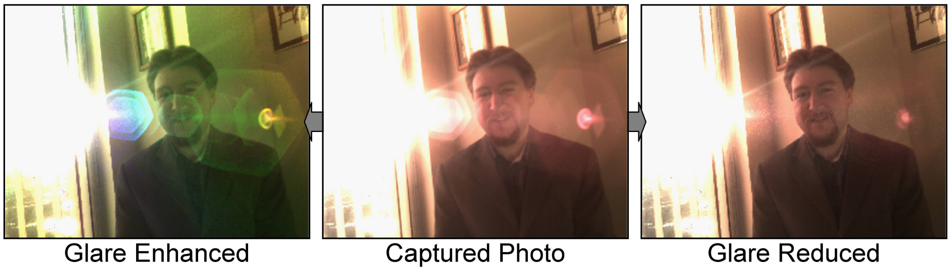


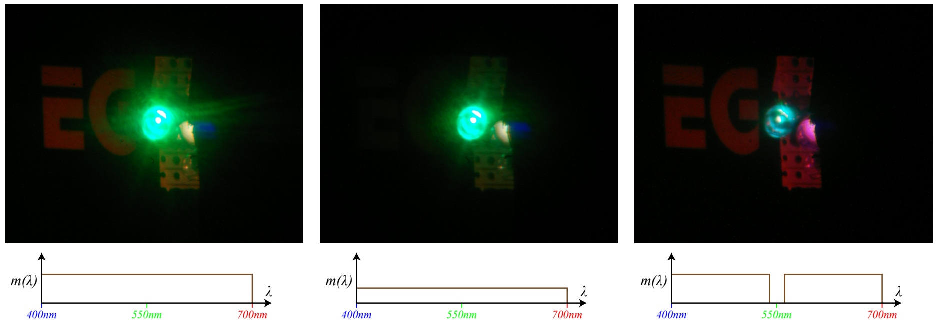
Figure 4.41: Mask based glare reduction. (Figure from Raskaret al. [332]) [THIS FIGURE IS ALREADY INCLUDED EARLIER IN THIS CHAPTER AS FIGURE 4.26

static scenes. Furthermore, since the mask needs to be nearly in focus, they need to place it close to the scene or use a very small aperture, thus limiting the use of this technique to studio settings. The size and focus requirements make it difficult to photograph a scene several meters away from the camera, such as sunlit buildings. Their method is best suited for extended area sources and cannot handle point and small area sources very well. [THIS SAME RESEARCH BY TALVALA IS MENTIONED EARLIER—ALMOST WORD FOR WORD—IN SECTION 4.5.2]

Raskar et al. [332] analyzed veiling glare in 4D ray space. They used the fact that glare is essentially high frequency noise in the 4D space to reduce its effect without reducing the image resolution significantly (see Figure 4.41). The procedure is easier to explain using the terminology of a traditional light field camera. [THIS FIGURE AND THREE ACCOMPANYING SENTENCES WERE INCLUDED EARLIER IN THIS CHAPTER.] A light field camera records the spatial and angular variations of rays incident at each location on the sensor (see Chapter 2). For an unoccluded Lambertian scene patch in sharp focus, the incoming rays have no angular variations. Reflection glare causes a bright light source in the scene to make a stray contribution to the sensor, but only along a specific angular direction (Figure YYY) [SPECIFY FIGURE]. These outliers appear as high frequency noise in 4D although the projection of ray-space onto a 2D sensor creates an apparent low-frequency glare. The high frequency noise is easily removed in higher dimensions using outlier rejection. Unlike light field cameras, they do not reversibly encode the spatial structure of the ray space, leading to a simpler design, and ability to recover a glare-free image without loss of resolution.

The prototype was a modified handheld camera with a mask placed very close to the camera sensor. The method is suited for isolated bright narrow-area light sources (e.g., bright sun or isolated room lights), and is tolerant to pixel saturation due to glare. It works without the need for multi-exposure HDR capture. Furthermore, the technique partitions glare into different types, thus providing easy opportunities for resynthesis. Unlike the method of Talvala et al., however, the technique does not work well for extended-area light sources, or with highly scattering lenses.

Mohan et al. [279] use an agile spectrum setup to acquire high dynamic range images and glare reduction by varying the exposure in the spectral or wavelength



(a) Exposed to get detail in background text

(b) Lower exposure

(c) Green wavelength attenuated

Figure 4.42: Spectral high dynamic range photography and glare removal. (a) Photo with a traditional camera of a green LED next to the red letters “EG.” (b) Photo of the same scene at a lower overall exposure. The wavelength modulation function $m(\lambda)$ is uniformly reduced. The text is too dark to read, and the glare still exists. (c) Agile spectrum camera is used to attenuate the green wavelength throughout the photo. The glare is removed, and more detail is visible.

dimension. Figure 4.42(a) shows a photo of a scene containing the text “EG” and a bright green LED next to it. The LED is too bright and produces a glare that renders part of the text unreadable. Reducing the exposure does not help because it makes the text darker as well (Figure 4.42(b)). Blocking the green wavelength by using an appropriate mask in the so called rainbow plane (see Section AAA) [SPECIFY SECTION—NOTE: RAINBOW PLANE IS NOT MENTIONED ANYWHERE ELSE IN THIS CHAPTER] leaves the red text unaffected, but greatly reduces the intensity of the LED and the glare (Figure 4.42(c)). While this method does not require placing a mask in front of the scene, it does assume that the glare is due to a narrowband light source whose spectral profile is known.

[THE FOLLOWING SECTIONS ARE FRAGMENTARY IN THE MS]

4.9 Polarization

talk about Stokes vectors, Mueller matrices??

[QUOTE] “Natural light has no particular polarization—it is composed of roughly equal amounts of all possible polarizations. When light is reflected from a reflective or specular surface, such as a pane of glass, a mirror or the surface of a body of water, light with one polarization is reflected more than light with the orthogonal polarization. In other words, the light reflected from a specular surface is polarized in one chief direction. A polarizing filter placed in front of the camera lens can remove that polarized light and so ”block out” the reflected light.” www.bobatkins.com/photography/tutorials/polarizers.html

** how to measure polarization (talk about some of the CTIS stuff)

next talk about use of polarizers in photography (wikipedia photos?)

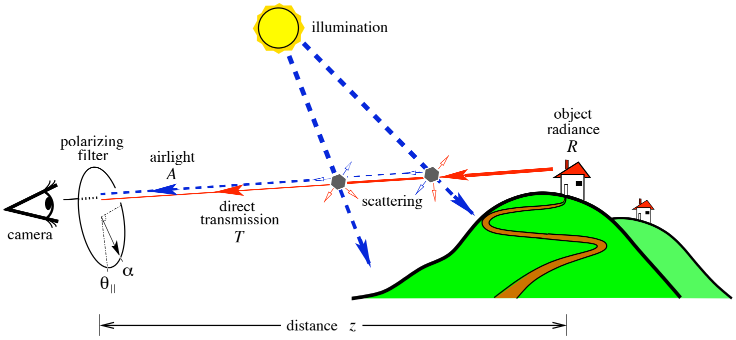


Figure 4.43: Light scattering in air (Figure from Schechner [355].)

TO DO: put some figures of with and without a polarizer Polarizers have long been used in conventional photography to - bluer sky - eliminate the reflection of light on tiny water droplets in the atmosphere, makes the color of the sky darker more saturated. - reduce reflections from surface water bodies, making them more transparent; useful to photograph the river bottom, for example. - more saturated foilage - similar to sky enhancement (reduce reflection) - reduce the effect of haze, fog - reduce the reflection arriving from praticles suspended in atmosphere.

Schechner et al. [355] proposed a technique to remove the effect of haze in images using polarization.

Tali and Schechner [401] - dehazing - underwater imaging

4.10 Non-standard Perspective

Gupta and Hartley [165] describe the pushbroom camera, motivated by the geometry of satellite imagery. Zomet et al. [444] generalize this notion to the cross slit camera, which selects a family of rays passing through two lines in space. Pajdla [310] describes oblique cameras, in which no two rays intersect (linear oblique cameras are also known as bilinear cameras). The work of Yu and McMillan [436] then collected and generalized these cameras, classifying them as two-dimensional slices of the four dimensional space of rays passing between two planes."—from Adams and Levoy [31].

Adams and Levoy [31] reformulated and extended the concept of general linear cameras to include focus.

Yu et al [437] give a thorough overview of the area of multiperspective modeling, rendering and imaging.

Multiperspective Rendering: Yu and McMillan [**?**]

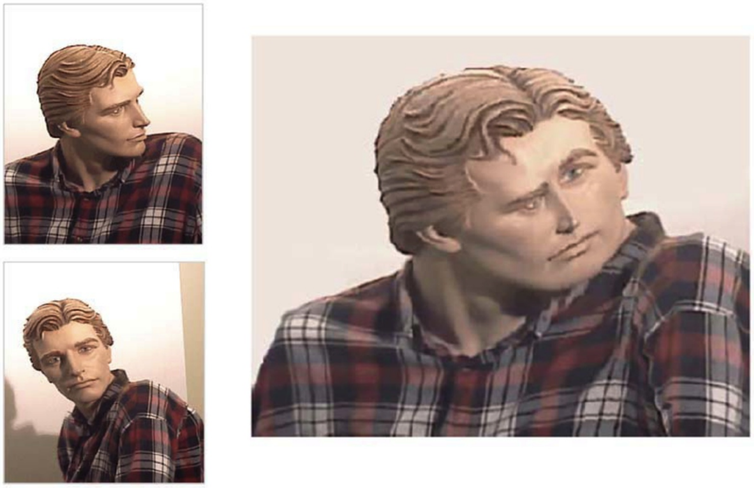


Figure 4.44: Multiperspective rendering (Figure from Yu [**?**].)

4.11 Shadowgrams and Schlieren Photography

4.12 Etcetera—MISC STUFF LOOKING FOR A NICE HOME

Compressive sensing [99]

McMillan and Bishop [268]

Diffraction limit for lensless imaging [270]

Ng aberrations [302]

The Plenoptic function [34] is a 5D function (ignoring wavelength, polarization, and time) that represents the radiance in every direction (θ, ϕ), at every point (x, y, z) in space. This function is redundant in a space free of occluders (outside the convex hull of an object, for example), and reduces to a 4D function called the *light field* [240, 162] as defined in Chapter 2.

Several people have theoretically analyzed the properties of light fields. Chan and Schum [78] analyzed the light field in the frequency space, and used it to estimate the spectral support of a light field. Chai et al. [77] used a frequency space analysis for a scene bounded by a depth range to estimate the minimum light field sampling rate for avoid aliasing. Zhang and Chen [441] provide an excellent summary of the work done in the area of light field sampling. Isaksen et al. [195] proposed dynamic 4D interpolation and filtering of the light field for interactive rendering of photographic effects such as variable focus and depth-of-field. Finally, Ng [301] applied the Fourier slice theorem to light field projections and demonstrated its use for fast refocusing.

Color has been an important part of computer graphics research as for a while. Salesin and others [324, 380] investigated the use of arbitrary ink pigments to reproduce the right color in a printout; Wilkie et al. [428] proposed a BRDF model for diffuse fluorescent surfaces; Hersch et al. [183] presented algorithms for printing images with fluorescent inks that are visible only under ultraviolet illumination; and Gooch et al. [159] proposed an algorithm for perceptual conversion of color images to gray-scale. We propose applications of color modulation in the areas of metamer detection, glare removal, and high dynamic range imaging, which have not been explored previously.

Chapter 5

Illumination

The camera has evolved from the *camera obscura* to a cumbersome view camera on a tripod to an easily portable hand-held imaging device. The technology of lighting the photographic subject, however, remains traditional and problematic, and the lighting equipment is often bulky, awkward, and expensive (see Figure 5.1). Because of the sophistication of modern consumer cameras, we can argue that the only characteristic that distinguishes an amateur photographer from a professional is the use of auxiliary lighting. What can we learn about lighting from the expert photographer? What can we create that goes beyond a professional's traditional lighting techniques? Can we create programmable lighting that minimizes critical human judgment at the time of capture? Can we create computational lighting that manipulates the lighting in images after they are taken?

Every traditionally trained photographer knows how to capture a variety of subjects under different lighting conditions. This is done for any scene by carefully measuring the light intensity with a light meter (hand-held or in-camera) and by manipulating the exposure variables of the camera—ISO sensitivity (or film speed), lens aperture, and shutter speed—to record the light accurately. In modern digital cameras these exposure functions can be automated, or programmed, by camera electronics, so the photographer doesn't have to think about exposure choices. This is good, because it frees the photographer to concentrate on subject and composition, but occasionally it is not so good because for some scenes the camera's programmed choices aren't accurate.

Many situations and scenes, such as traditional portraits or weddings, require a photographer to add auxiliary lighting to create the most pleasing picture. A photographer has many choices for how the auxiliary lighting affects the look of a picture. The following auxiliary photographic lighting choices are considered programmable:

- Duration and intensity;
- Presence or absence of auxiliary lighting;
- Color, wavelength, and polarization;
- Position and orientation;
- Modulation in space and time

Computational photography
Computational illumination

Figure 5.1: The photographic camera has evolved from a cumbersome view camera on a tripod to a portable hand-held imaging device. Computational photography is the on-going next stage in the development of digital imaging. Even though the technology of lighting has not progressed into new forms as quickly as digital photography, can lighting and computational illumination become similarly flexible and programmable?

We will see later in the chapter how a photographer can also exploit the change in natural lighting.

In the early days of electro-chemical flash photography, controlling the duration and intensity of flash was challenging. But today's illumination sources are sophisticated and highly programmable, thanks to advances in solid state lighting, light emitting diodes, sophisticated temporal modulation via strobes, and spatial modulation via spatial light modulators and video projectors. This chapter describes how these choices and advances all contribute to computational illumination. For an ultimate level of control and programmability, researchers have recently developed illumination domes in which hundreds of programmable lights or projectors surround a subject, allowing the researchers to synthesize any desired type or quality of illumination, including from other locations. [REFERENCE TO DEBEVEC ET AL. NEEDED HERE.]

5.1 Modifying Duration and Intensity

A traditional shutter exposure time on a modern digital camera can be as short as 1/8000th of a second. While this shutter speed seems extraordinarily fast, and can freeze most moving objects or athletes in action, it still isn't nearly fast enough to record many physical phenomena in nature. How can we create faster shutter speeds? Similarly, we can carefully control the flash output from small electronic flash units, but can we scale the intensity of the flash illumination to the size of a city or time the release and duration of the flash to the nearly instantaneous burst of a speeding bullet or an explosion? The answer to both of these questions is found in making electronic devices—strobes—that emit bright bursts of light of extremely short duration.

5.1.1 Stroboscopic Freezing of High Speed Motion

In the 1930s, MIT researcher Harold Edgerton and photographer Gjon Mili revolutionized instantaneous photography by employing ultra-short electronic strobe flashes to illuminate transient phenomena. These photos captured the beautiful intricacy and the graceful flow of movement that was too rapid or too complex for the human eye to discern or for traditional cameras to capture. Edgerton and Mili used these techniques to capture dramatic images of bursting balloons and, most famously, of a bullet at the moment of impact with an apple (see Figure 5.2). A key challenge Edgerton faced was how to trigger the flash at the appropriate instant, so he developed a flash release circuit timed to sound. To this day, audio and laser triggers are commonly used to provide a similar synchronization in engineering and scientific applications. We can look at these developments in strobe photography in the 1930s as an early example of computational illumination, even though the term had not yet been invented.

Since the pioneering work of Edgerton and Mili, the technology of high-speed photography has evolved and entered the consumer market. Film-based high-speed photography was facilitated by the transition to stronger film substrates, including mylar and acetate bases introduced by Kodak in the 1960s. The introduction of CCD and CMOS digital sensors in the 1980s allowed for the possibility of ultra-short exposures. Today, modestly priced high-speed digital cameras are available, such as the Casio EX-F1, capable of 300 frames per second (fps) at 512 × 384 pixel resolution and up to 1,200 fps at 336 × 96 pixel resolution. More expensive high-speed digital cameras such as the Vision Research Phantom Flex can shoot 2570 fps at high-definition television resolution (1920 x 1080). With cameras such as these, consumers can capture their own images of balloons bursting and water drops frozen in time. Important challenges remain in high-speed photography, including providing sufficiently bright illumination for ultra-short exposures, as well as providing sufficient storage for the massive data collected during extended high-speed photography sessions.

5.1.2 Sequential Multiflash Stroboscopy

Certain motions are more effectively recorded by using multiple flash illuminations during the shutter exposure. This allows a temporal sequence of motion to be summarized in a single photograph. This technique works well when the subject is photographed against a dark background and when subsequent frames have limited overlap. A good example is a golfer swinging a golf club perpendicular to the camera's optical axis, as shown in Figure 5.2. The narrow golf club appears at distinct non-overlapping positions in successive frames. The results are typically less compelling when the scene is not filmed against a high-contrast background or the motion is toward or away from the camera. Today, high-speed video sequences can be processed to produce similar composite photographs, although the short exposure times of individual frames may lead to images with excessive noise.

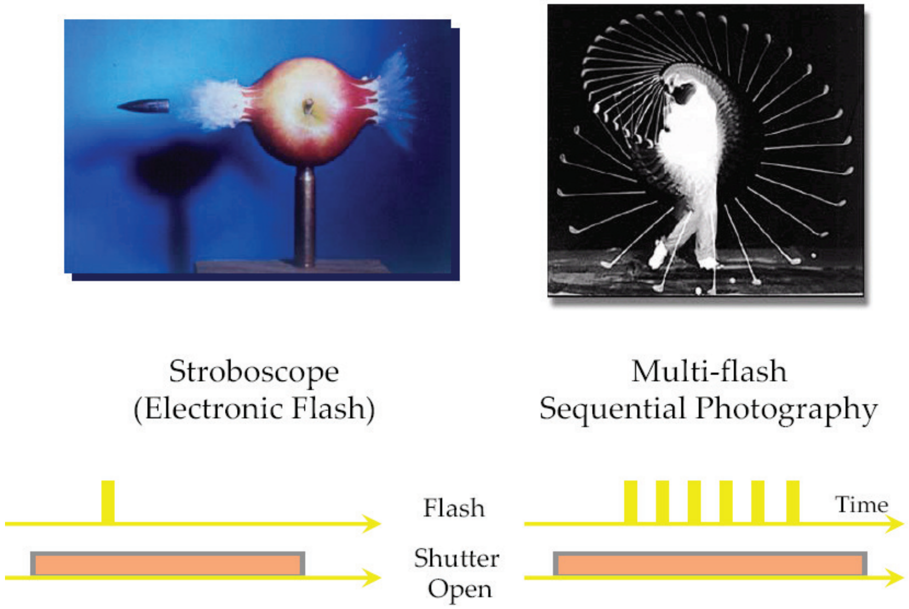


Figure 5.2: Early instances of computational illumination. (left) In the 1930s, MIT researcher Harold Edgerton and photographer Gjon Mili used ultra-short-duration electronic strobe flashes to illuminate transient phenomena that couldn't be photographed with a traditional camera. Their photo of a bullet moving through an apple is world famous. (right) Certain actions, such as a golf swing, are best photographed by a sequence of multiple flashes triggered during the camera exposure. In both of these examples the flash intensity is much greater than the ambient room illumination, and the duration of the camera shutter is significantly longer than the flash duration.

As a result, time-sequential multi-flash stroboscopic flash illumination remains the preferred method to produce such compelling imagery.

5.2 Presence or Absence of Auxiliary Lighting

The simplest form of computational illumination is the ubiquitous and simple camera-based flash unit, whether built into the camera or attached to the camera hot shoe. Electronic circuits in both the camera and the flash unit compute the proper amount of flash intensity to illuminate a dark scene. This simple and direct flash illumination can be used to gather information from a scene. Di Carlo et al. [94] first explored the idea of capturing a pair of images from the same camera position, one image illuminated only with ambient light and the other image using the camera's flash as an auxiliary light source. They used this image pair to estimate object reflectance functions and the spectral distribution of the ambient lighting. Hoppe et al. [187] took multiple photos at different flash intensities, allowing the user to interpolate among them to simulate intermediate flash intensities. In both of these cases, the scene is assumed to be sufficiently close to the camera, so that the flash will produce a detectable change in surface brightness. This close distance requirement is a fundamental limitation for all such methods requiring active illumination.

5.2.1 Flash/No-Flash Pair for Noise Reduction

Petschnigg et al. [112] and Eisemann et al. [108] concurrently proposed similar strategies for combining information contained in the flash/no-flash image pair to generate a single image with enhanced aesthetics. The no-flash photograph captures the large-scale illumination and overall ambiance of the scene (see Figure 5.3). But in low light, the no-flash photo also exhibits excessive noise. In contrast, the flash photograph exhibits lower noise and greater high-frequency detail, yet this image also appears unnatural and fails to convey the mood of the scene (see Figure 5.4). A simple technique combines the photos and decouples the high-frequency and low-frequency components in the photo pair, and then recombines them in a manner that preserves the desired characteristics—high-frequency detail, color, and low noise from the flash photo, and overall ambiance from the no-flash photo. Such decoupling is achieved using a modified bilateral filter called the *joint bilateral filter*.

The flash image is used to perform a guided smoothing and to reduce noise in the no-flash image without excessive blurring of sharp features. In traditional image processing pipelines, smoothing is performed directly on an image by using information available only in that image. Smoothing an image, for example with a Gaussian filter, reduces high-frequency noise, but also blurs sharp edges. By using a bilateral filter [399], the image filtering process can be controlled to preserve sharp edges, while reducing noise by smoothing in regions with slowly varying texture (see Figure 5.5). In this manner, the bilateral filter performs smoothing based on both spatial extent as well as intensity similarity within the kernel filter support. The intensity similarity term "stops" the kernel influence at the intensity edge. By exploiting intensity similarity, the bilateral filter can reduce image noise while preserving sharp details. Nevertheless, bilateral filtering still causes unnecessary suppression of weak details along with noise. Similar methods for general anisotropic diffusion [318] are also subject to these limitations, motivating the development of the joint bilateral filter and the inclusion of auxiliary data from additional images. In addition to these limitations, both bilateral filtering and anisotropic diffusion are non-linear algorithms, significantly increasing the computational cost of evaluating a smoothed image. [AN EXAMPLE IMAGE OF THE FLASH–NO-FLASH TECHNIQUE COULD BE INCLUDED HERE.]

With the joint bilateral filter, smoothing is also influenced by high-frequency detail in a companion image. For example, we can use a high-quality flash image to reduce noise in a no- flash image. The kernel influence in the no-flash image "stops" at locations corresponding to the intensity edge in flash image. This enhances the ability to find and preserve weak details (i.e., low confidence edges) in the presence of noise. The basic idea is to smooth the no-flash image while preserving all edges that are detected in the flash image. The spatial kernel remains the same within

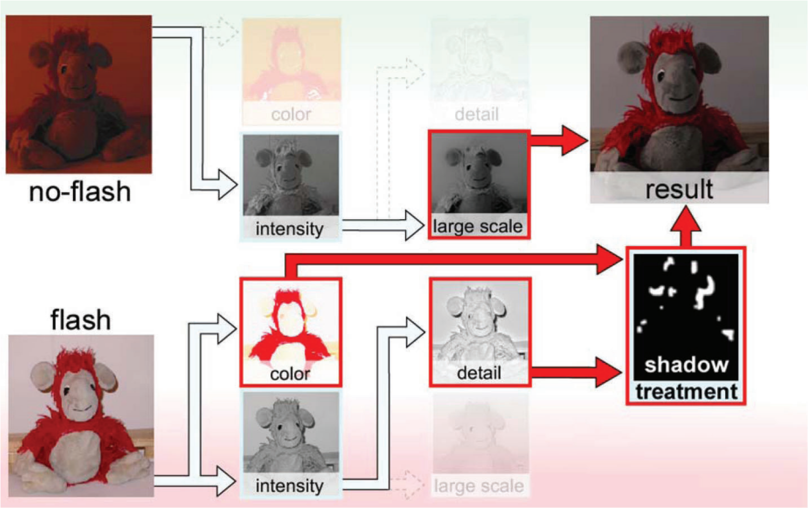


Figure 5.3: The process for reducing noise in a no-flash ambient-light image by using an auxiliary flash image. A flash-illuminated image and a no-flash ambient-light image are recorded. The image is decomposed into independent-color, large-scale, and fine-detail channels using traditional bilateral filtering. Special processing is used to treat cast shadows in the flash image. Finally, the various layers are combined by using a joint bilateral filter. The final merged image preserves the ambiance created in the no-flash image, while retaining the color and detail present in the flash image. (Image courtesy Elmar Eisemann and Fredo Durand.)

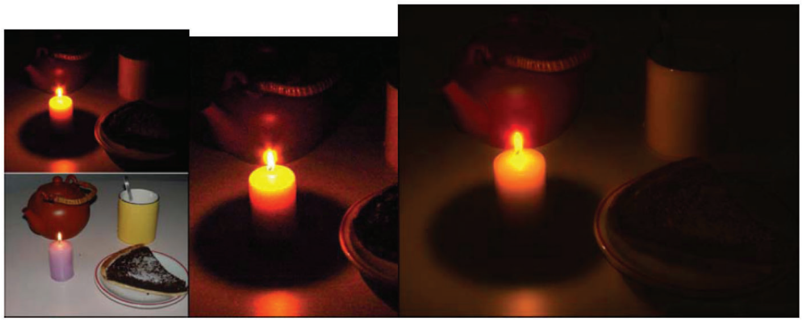


Figure 5.4: Merging a flash/no-flash image pair to enhance image aesthetics. (top left) A photograph taken in a dark environment; is noisy and/or blurry. (bottom left) A flash photograph yields a sharp but flat image with distracting shadows at the edges of objects. (middle) A scaled region shows the noise of the no-flash ambient-light image. (right) The merge technique fuses the two images to transfer the ambiance of the no- flash image. Note the shadow of the candle on the table in the merged result. (Image courtesy Elmar Eisemann and Fredo Durand.)

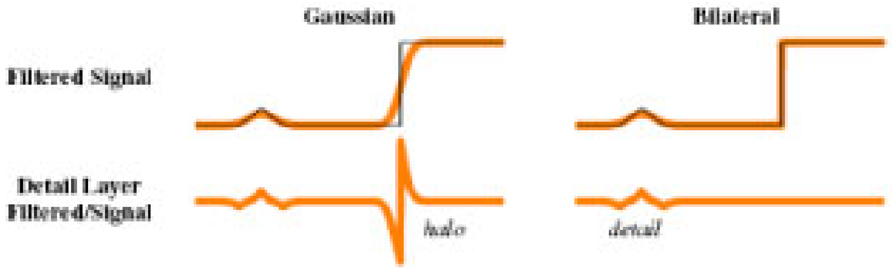


Figure 5.5: Comparison of Gaussian and bilateral low-pass filtering. The intensity profile of a horizontal set of pixels is in black. (left) A Gaussian low-pass filter blurs over sharp image discontinuities near edges, including intensity ridges and valleys, and leads to high-frequency artifacts, such as halos in the“detail” layer. (right) A bilateral filter is locally tuned to prevent smoothing across strong intensity discontinuities, preserving sharp details and minimizing high-frequency artifacts such as halos. [112]

the no-flash image, but the intensity similarity is measured with respect to the corresponding flash-image pixels. Since the flash photo exhibits lower noise, a better result is achieved and over- or under-blurring is avoided. [THIS PARAGRAPH REPHRASES THE PRECEDING PARAGRAPH.]

Finally, to create a noise-free no-flash image, an edge-preserved low-frequency component from the no-flash image, which preserves the overall lighting, is combined with a high-frequency component from the flash image, which preserves sharp details. The problem becomes challenging when dealing with errors due to overexposure or shadows in the flash image. Overexposure leads to a flat detail layer. In this situation, the detail information is neither in the no-flash image (due to noise) nor in the flash image (due to saturation) [105]. Similarly, special efforts must be made to address cast shadows, which introduce artificial high-frequency content into the scene, or light flares, which similarly introduce artificial high-frequency content and brighten pixels in shadowed regions. Taking two photos requires a static scene. However, the flash duration is usually just a few milliseconds, and hence the second photo would add negligible time to the total joint capture time.

The process for reducing noise in a no-flash image using an auxiliary flash image. A flash and no-flash image pair are recorded, using the flash and the available ambient light, respectively. Afterward, the image is decomposed into independent color, large- scale, and fine detail channels using traditional bilateral filtering. Special processing is used to treat cast shadows in the flash image. Finally, the various layers are combined as shown using joint bilateral filter. Note that the final merged image preserves the ambiance created in the no-flash image, while retaining the color and detail present in the flash image. [105] [THE LAST FOUR PARAGRAPHS IN THIS SECTION APPEAR TO BE SEPARATE DESCRIPTIONS OF THE SAME FLASH–NO-FLASH TECHNIQUE.]

5.2.2 Flash, Exposure, and Dynamic Range

Present-day digital cameras use built-in sensors and algorithms to approximate the correct flash intensity and proper exposure setting for flash illumination. But these estimates, based on aggregate measurements, often lead to underexposure or overexposure (saturation) of people or objects in the scene, depending on their distance from the camera. A single setting for the flash intensity cannot simultaneously illuminate distant or dark objects without simultaneously saturating, or "blowing out," nearby or bright objects. Image quality is also affected when the range of light values in a scene exceeds the dynamic range of the camera. The individual images in Figure 5.6 are examples of such an HDR scene. [A SEPARATE IMAGE EXAMPLE MIGHT WORK BETTER HERE.] For such situations, Agrawal et al. [120] suggest merging multiple images captured under varying flash intensities and camera exposures to construct an accurate HDR image. Figure 5.6 shows an example of this strategy in which the flash intensity and exposure parameters are varied.

Given a three-dimensional scene, the requisite flash brightness is a function of the scene depth, the natural or ambient illumination, and the surface reflectance of the various scene elements. For example, a distant object with low reflectance will require a bright flash, whereas a nearby point or an area well lit by the ambient illumination will be overexposed by a flash, even at a low flash intensity. In addition, the scene might extend far into the distance and would not be illuminated even with an intensely bright flash. Only a longer exposure for the ambient light would properly capture such distant regions. To capture such challenging, yet commonly encountered, scenes, we can collect multiple exposures, each at a different setting along the exposure and flash intensity axes. Figure 5.6 tabulates photos taken at six different exposure settings and at four different flash brightness settings—a total of 24 exposures. Many consumer and professional cameras offer manual setting of flash intensity. Though making 24 captures to achieve a single image may be excessive, Agrawal et al. [120] present a greedy approach [DEFINE 'GREEDY' HERE]; pixel values of each capture are analyzed for overexposure or underexposure, which suggests the optimal exposure and flash brightness settings for the subsequent capture. A greedy algorithm then makes the locally optimal choice, before each new image is captured, in order to calculate the global optimum. By using such adaptive sampling of the flash-exposure space, the number of captured images required for any given scene is minimized.

As with any such greedy strategy, the resulting solution is not guaranteed to be the globally optimal result—had all flash and exposure settings been sampled. Furthermore, the proposed method requires multiple exposures to capture a single HDR image, increasing the image capture time and limiting the solution to static scenes. Chapter 6 describes how next-generation sensor technology that allows HDR capture in a single exposure will reduce the number of required flash/exposure sequences.

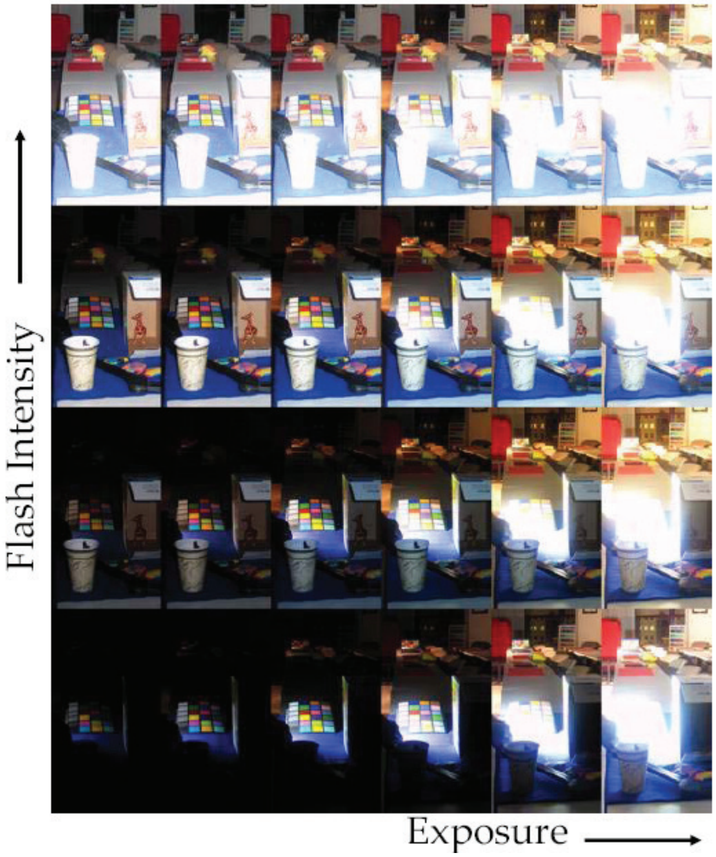


Figure 5.6: Flash-exposure high dynamic range imaging is performed by sampling the two- dimensional space of flash and exposure parameters. In a scene with large variations in depth, illumination, and reflectance, multiple pictures are required to estimate a high- dynamic range image. Instead of directly sampling along the exposure and flash intensity axes, adaptive sampling is used to minimize the number of required samples using a greedy sampling scheme [120].

5.2.3 Removing Flash Artifacts

Flash images notoriously suffer from several problems, including overexposure of nearby objects, poor illumination of distant objects, reflections from objects strongly lit by the flash, and strong highlights produced by reflections of the flash on glossy surfaces. Flash/no-flash image pairs, discussed previously, can also be applied to address the particular artifacts introduced by strong highlights and reflections on glossy surfaces.

Agrawal et al. [120] use a technique based on image intensity gradients. The orientation of the image gradient vector, defined at each pixel in the rasterized image, corresponds to the direction along which the change in image intensity is maximum. The magnitude of this vector is the rate of that change. For example, along an intensity edge the gradient vector orientation is perpendicular to the edge, and the gradient vector magnitude is the strength of the edge. Agrawal et al. observe that the orientation of image gradients due to reflectance or geometry variation are illumination invariant, whereas those image gradients corresponding to artifacts due to changes in lighting are not. Hence, a "gradient coherence" model indicates that, in the absence of artifacts, the gradient vector orientation in the flash and ambient (no-flash) images should be the same. On the other hand, a change in gradient vector orientation between the pair of images indicates the presence of an artifact. By exploiting this gradient coherence, they propose a gradient projection scheme to decompose the illumination artifacts from the rest of the image.

Central to the gradient projection scheme is the ability to reconstruct an image from its gradient vector field. For example, the gradient field $G(x, y)$ for an array of pixels $I(x, y)$ can be discretized and represented simply as the forward difference; i.e.

$$G(x, y) = (I(x + 1, y) - I(x, y), I(x, y + 1) - I(x, y)).$$

Many techniques have emerged in the computer graphics literature since 2002 to compute image I from the given gradient field G by using a 2D integration of the gradient field with certain boundary conditions. This general problem has been solved historically for photometric stereo and shape-from- shading but more recently for mesh smoothing, HDR image compression, image editing, and multi-image fusion. A key challenge in practice is that the provided gradient field is not consistent; specifically, the integral of the gradient field along any closed path in the image should be equal to zero, so that the reconstruction is independent of the choice of integration path. As a result, the measured gradient field must be rendered integrable by further processing. Agrawal et al. [42] present a survey of such methods.

Figure 5.7 shows the flash and ambient light (no-flash) images of a painting. The ambient light image is characterized by distracting reflections of the photographer lit by room lighting. The flash image has a very short exposure duration which overpowers the intensity of anything lit by ambient room lighting, including the photographer's reflection. The flash illumination, however, produces a hot spot in which the reflection of the flash is clearly visible in the image. The gradient projection scheme removes reflections in the ambient image by subtracting the component of the ambient image gradient field perpendicular to the flash image

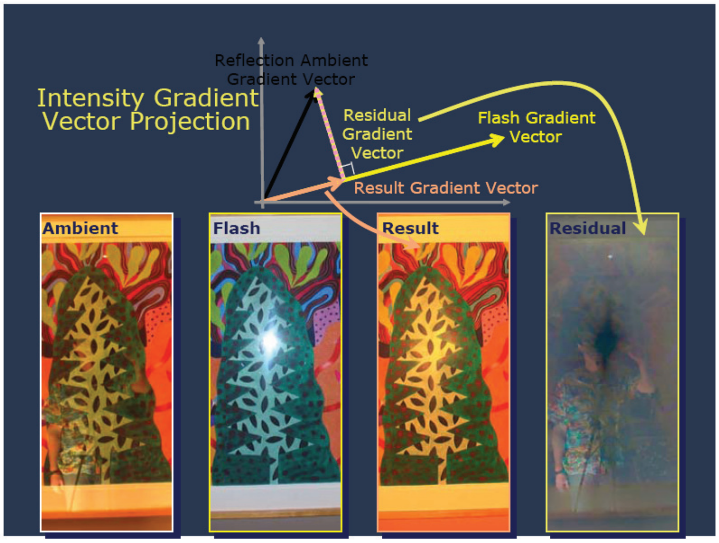


Figure 5.7: Removing flash artifacts with gradient vector projection. (top) Undesirable artifacts in photography can be reduced by comparing image gradients at corresponding locations in a pair of flash and ambient images. (bottom) Ambient and flash photos are captured in a museum setting. Ambient photo shows reflected photographer while the flash photo suffers from highlights. The reflections are removed in the ambient image by subtracting the component of the ambient-image gradient field vector that is perpendicular to the flash-image gradient's vector. Alternatively, the reflection layer is recovered by integrating the subtracted (residual) component. [28] [ARE THE 'TOP' AND 'BOTTOM' DESIGNATIONS IN THE CAPTION NECESSARY, SINCE THIS IS A SINGLE FIGURE?]

gradient field. Reconstruction from the projected gradients produces a reflection-free image.

Interestingly, reconstruction from residual gradients recovers the reflection layer. The gradient orientation is not available, however, when both images have co-located artifacts (for example when the photographer's reflection as well as the hot spot from the flash are visible in the same part of the photo). In addition, gradient orientation is unstable in homogeneous flat regions, so the photographer's reflection in such parts will be difficult to recover. In later research work, Agrawal et al. [42] introduced a gradation projection tensor that is more robust compared to the simple gradient projection procedure. This work also shows how to compensate for the extinction of flash intensity along the optical axis by exploiting the ratio of the flash and ambient light images.

5.2.4 Flash-Based Matting

The problem of extracting a foreground subject from its background, known as *matting*, can be made more precise by combining flash/no-flash imaging with Bayesian matting [123]. Sun et al. [384] make the simple observation that the greatest difference between a flash image and an ambient (no-flash) image is the change in brightness of the foreground subject, provided the background is sufficiently distant. Their approach is readily applicable to images produced by off-the-shelf flash-equipped cameras. For example, a pixel-wise ratio of the flash and no-flash images will be close to unity for distant points (background) but significantly higher for near-field points (foreground). Using joint Bayesian matting, even foreground subjects with complex shape boundaries, such as those made by fur or hair, can be precisely extracted with an alpha matte and placed into a new image context (see Figure 5.8).

Unlike traditional Bayesian matting, which works on a single input image, this technique works even when the foreground and background have similar colors. The technique fails, however, when the flash image does not encode the intensity fall-off with distance in the expected manner. For example, when the background is not sufficiently far away, or when the object is rounded. Furthermore, since matting is typically applied to dynamic scenes, further efforts are required to allow single-shot capture to eliminate distracting strobing sequences (for example, by using IR flashes).

5.3 Modifying Color, Wavelength, and Polarization

The scene radiance is a product of incident illumination and reflectance. By changing the wavelength profile of the incoming light (often simplified as a color profile) or by capturing specific wavelength channels, we can perform programmable color manipulations of images.

By changing the spectrum of incident illumination, it is possible to probe a scene and create multispectral photos or overcome confusion due to metamers (colors that

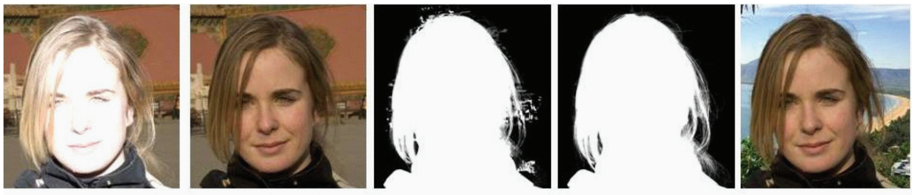


Figure 5.8: Flash matting allows the extraction of foreground from background in an ambient light image by using an additional flash image. (left to right) The (1) flash and (2) no-flash images, (3) Bayesian matting results performed independently on the no-flash image and (4) jointly on the flash/no-flash image pair, and (5) compositing result using the joint Bayesian matte. (Jian Yin Sing and Shum "Flash Matting" 2006)

have the same visual appearance for a given illuminant color). Fluorescence photography, commonly used in medical and scientific imaging, exploits the color shift between incident illumination color and the resultant radiance. Many naturally occurring substances fluoresce, including rocks and minerals, fungi, bacteria and most body tissues. When the scene is illuminated with high-frequency (short-wavelength) illumination, the resulting emission is in lower frequencies (longer wavelengths).

Thus, for example, subjects irradiated with ultraviolet light may release, green, yellow or pink light, and subjects irradiated with visible light may emit infrared fluorescence. Household fabrics are routinely treated with fluorescent dyes to make them look whiter. When illuminated with ultraviolet light (in dimly lit discos, say), clothes treated with fluorescent dyes emit lower frequencies and appear bright. In most fluorescence photography, an ultraviolet-selective filter is placed at the light source. Another filter of a different (visible) wavelength is placed over the camera lens to absorb the reflected ultraviolet rays, permitting only the visible light (fluorescence) from the object itself to be sensed.

Fluorescent marker dyes are used to image objects inside scattering media, such as biological samples in microsocopy. By using a wavelength-rejecting optical filter in front of a camera, all scattered light with the same wavelength can be rejected. The induced fluorescence, however, has a different wavelength and can be imaged by the camera. In these examples, fluorescence is exploited to capture otherwise unseen or difficult-to-image phenomena, at the cost of additional photographic components, including various wavelength-selective filters.

This wavelength manipulation can be done in the post-capture stage by using conventional photographic equipment. Paul Haeberli [167] showed that using multiple exposures of the same subject with different lighting schemes allows the lighting of the scene to be modified after it has been photographed. Figure 5.9 shows the technique with a scene lit with two lamps, one to the left of the subject and one to the right of the subject, in addition to ambient lighting. Three photographs are taken, the first with only ambient light, the second with only the lamp on the left plus ambient light, and the third with only the lamp on the right plus ambient light. The ambient light image is subtracted from each of the images lit by the lamps.

This technique creates an image that shows exactly what light is contributed

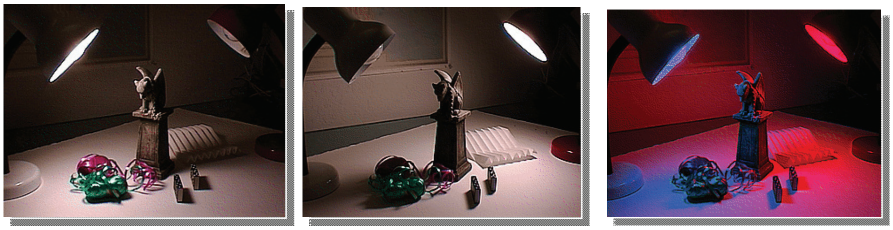


Figure 5.9: Programmable combination of colors. (left) Scene illuminated by the lamp on the left. (middle) Scene illuminated by the lamp on the right. (right) Synthetic lighting achieved by linearly combining the image pair so that the left-lamp image is tinted blue and the right-lamp image is tinted red. [167]

by which light source, allowing wavelength manipulation of each light. Figure 5.9 illustrates what a scene would look like if the lamp on the left was blue (for example) instead of white. Applying a similar process to the lamp on the right in the figure synthetically illuminates the scene with multicolored lamps. The technique can be extended to any number of light sources to control brightness and color. This strategy also allows for negative lighting by subtracting light coming from a particular lamp.

As we describe in Section 5.4.2, this method can be extended to sets of photographs collected by using hundreds of individual light sources, allowing complex environmental lighting conditions to be synthesized in post-processing. Such techniques have found wide-application in the film and interactive entertainment industries, where environmental lighting can be applied in post-production so that blue-screened performances can appear realistically lit in artificial environments. As with many methods considered in this chapter, multiple exposures necessarily limit the method to either static scenes or to situations in which high-speed cameras are used with synchronized lighting.

5.4 Modifying Position and Orientation

Light sources don’t have to be static or in simple configurations. The placement and orientation of lighting can be altered, thus modifying shadows and shading throughout a scene. Note that changing the orientation of a light source will change its absolute intensity at any given point in the scene, yet the angle of incidence of light rays at each point will remain the same. Such invariant properties are often exploited in computational photography and illumination, to facilitate novel or efficient data capture and post-processing. The following examples show how simple modifications of light source position and orientation lead to new techniques for scene understanding and post-processing control.

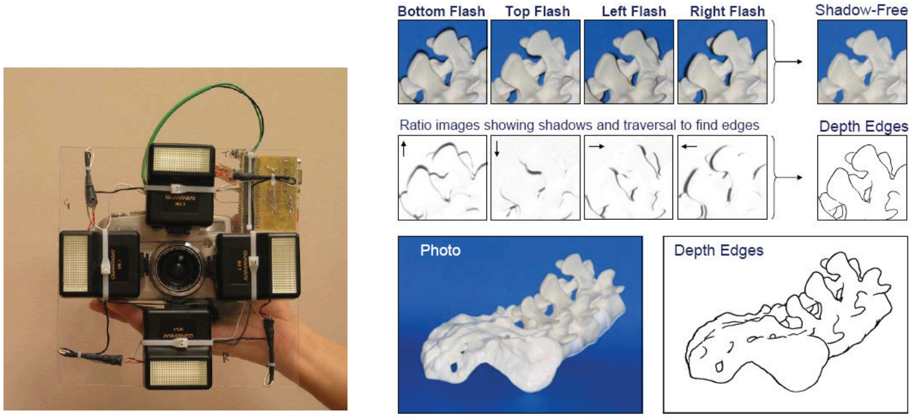


Figure 5.10: Multiflash camera for depth edge detection. (left) A camera with four flashes. (top right) Multiflash image sequence produced by using four individual flash exposures. A shadow-free image is computed by evaluating the per-pixel maximum intensity over the multiflash sequence. (middle right) Each multiflash image is divided by the shadow-free composite to amplify cast shadows. Depth edges are detected by using epipolar traversal; specifically, negative intensity transitions are computed along the direction in each image from the corresponding flash through the center of projection. (bottom right) The shadow-free photo and depth edge image can be used to stylized deption and scene understanding. [329]

5.4.1 Shape and Detail Enhancement Using Multiposition Flashes

A moving light source can be used to inspect and extract subtle surface details and also to distinguish object silhouettes (the boundary curves separating the image of a foreground object from the background). A traditional edge-detection filter can detect reflectance discontinuities, such as those due to texture variation, but it does a poor job in estimating edges due to shape discontinuities. Shape discontinuities occur due to depth differences between a foreground and a background patch or due to sharp changes in surface orientation (e.g., along an intensity ridge or valley). By observing the sequence of images made while moving a light source, and noting the variation in shading and shadows, we can distinguish such shape discontinuities from reflectance discontinuities.

Raskar et al. [329] employed a camera equipped with multiple flashes to find the silhouettes in a scene and create stylized or cartoon-like images. Their multiflash camera employs four strategically placed flashes to cast shadows along the depth discontinuities of a scene. Depth discontinues are edges in the scene formed at shape boundaries or silhouettes, where the depth value of neighboring pixels is different. More precisely, depth discontinuities are “depth edges” due to C0 discontinuities in the depth map with respect to the camera. The flashes individually illuminate the scene during image capture, creating thin slivers of shadow along the depth discontinuities. The position of shadows is determined by the position of the flash on the camera. When the flash is on the right, shadow slivers are created on the left; when the flash is on the left, shadows slivers are created on the right, and so on. In Figure 5.10, we see how the shadows on the subject move in each of the four positions, above, below, to the left, and to the right of the lens. The shadows encode the position of depth edges.

The shadows visible in each image are detected by first computing a shadow-free image, approximated with the max-composite image. The max-composite image is assembled by choosing from each pixel the maximum intensity value from the image set. Then the shadow-free image is compared with the individual shadowed images identifying the shadow regions. The correspondence between the position of light and shadow region boundaries produces the depth edges.

This technique fails to mark a depth edge when it is difficult to detect the shadow slivers attached to the image of a depth edge. For example, the shadow detection fails when the background is too far away, relative to the depth edge. If the foreground object is too narrow (for example, a nail), the shadow observed in the image is detached from the object. Since specularities from shiny surfaces can confuse the max-composite image, a method using an intrinsic image (described below in the subsection on natural illumination variation) can replace the max-image.

The detected silhouettes are then used to stylize the photograph and highlight important features. Raskar et al. [329] demonstrated similar silhouette detection in video sequences by using a high-speed flash sequence. Since its introduction, the multiflash camera has been applied to capture the 3D shape of objects from turntable sequences by Crispell et al. [122]. Feris et al. [118] have also used a video-rate multiflash system for decoding sign language input. Tan et al. [119] have applied a similar multiflash camera to enhance laproscopic image sequences in surgical environments. Alternate methods for active detection of depth discontinuities using structured illumination have been proposed by Kim et al. [124].

By using a larger number of images captured with varying light positions around the photographic subject in a studio or laboratory setting, we can enhance the subtle surface features observed through grazing angle illumination, in shadows due to complex geometry, and in specularities and subsurface scattering. Akers et al. [116] use spatially varying image weights on images acquired with a light stage similar to that in Debevec's group [113]. A painting interface allows the artist to locally modify a relit image as desired. Although the spatially varying mask offers greater flexibility, it can also produce physically unrealizable results that appear unrealistic. Anrys et al. [117] and Mohan et al. [121] used a similar painting interface to help a novice in photographic lighting design. A target image is sketched, and the system is allowed to find optimal weights for each input image in order to achieve a physically realizable result closest to the target.

5.4.2 Relighting Using Domes and Light Waving

The goal of image-based relighting is to create arbitrary novel lighting through post-capture editing. Instead of building an accurate 3D model of the scene, including the material properties of each element, image-based relighting relies on the simple observation that light interacts linearly with material objects [26, 166]. If the scene is lit by one light, then doubling the pixel intensities in a photo will achieve the same effect as doubling the brightness of the light source. This of course assumes that the camera response is linear, without underexposure or saturation.

As we saw in the work of Paul Haeberli earlier, adding two photos, each taken with only one light illuminating the scene, is equivalent to capturing a photo with two lights. More precisely, if a fixed camera records an image I_i from a fixed scene lit only by a light L_i, then the same scene lit by many lights, scaled by weights w_i, will produce an image $I_out = \sum_i (w_i \; I_i)$. Adjusting these weights allows us to synthesize any output image corresponding to a linear combination of lights. However, due to linearity, the effective output image is the same as if the light sources had been modulated (turned brighter or dimmer). By using such a data-driven approach, we can achieve flexible post- capture digital relighting of an image.

In classic computer graphics, material properties and light transport are simulated for a virtual scene (possibly modeled to correspond to a real-world environment). Data-driven approaches, on the other hand, can easily capture details that are difficult to model, including global illumination, caustics, and the appearance of human skin. Of course, such methods also generate enormous quantities of data. Efficient storage and processing mechanisms are required to allow such methods to be used in practice.

For accurate relighting of a scene to synthesize arbitrary virtual lighting conditions, we ideally need to photograph the scene by moving the light through every possible position of the lighting fixture. This is typically a challenging task. For example, consider the case when the light positions are limited to a region within a square flat. The data collection process involves taking successive photos as the light is moved thought a discrete set of positions within the square. From this dataset, we can synthesize photos only from virtual light sources lying within that square. As with any discretely sampled signal, the sampling density of lighting positions will determine the accuracy of synthesized results for arbitrary lighting conditions.

To overcome this limitation and to reduce the number of lighting variations required, we can exploit the fact that all light rays traveling within a bounded volume can be geometrically parameterized by a 5D plenoptic function, defining the irradiance along any given optical ray [35]. Conceptually, to capture the effects of such a 5D function, we would need to record a photograph by illuminating the scene one ray at a time. This is a daunting task, both in terms of the data storage required as well as the prohibitively long time required to sample such a function. If we limit ourselves to synthesizing novel light sources constrained to be outside the convex hull of a given set of objects, the problem is slightly simplified. In this case, we can represent the incident illumination by using a 4D ray-parameterization—the "light field"—rather than the more general 5D plenoptic function. To explain this insight, we need to consider the higher dimensional properties of incident and scattered light and their impact on the recorded intensities in a photograph.

We discussed light fields earlier in Chapter 2. Light fields [240] and the lumigraph [162] reduced general 5D plenoptic functions to four dimensional functions $L(u, v, s, t)$ that describe the presence or absence of light in free space along any given optical ray, ignoring the effects of wavelength and time. Here (u, v) and

(s, t) are the parameters of intersection with two parallel planes, respectively, that describe a given ray of light in free space. To represent an incident light field, a slightly different parameterization can be used.

Imagine an object surrounded by a spherical dome with projectors aimed inward. Parameters (θ_i, ϕ_i) describe the angular position of the projector on the unit sphere, and (u, v) the pixel position in the projected image from a given projector. Thus, the function $L_i(u, v, \theta, \phi)$ gives complete control over the incident light rays on an object in free space. Similarly, an inward-facing array of cameras on the spherical dome would capture the entire radiant light field created by an object, $L_r(u, v, \theta, \phi)$. Such a system would allow us to both capture and display the radiant and incident light fields, respectively. To describe the full light geometric light transport in such systems, Debevec et al. [113] introduced the 8D reflectance field, which describes the irradiance recorded at a given camera pixel, due to any incident light ray displayed by a given projector. An additional dimension of time is sometimes added to describe the changing interaction of light with a moving object. Further dimensions can be introduced by considering multiple color channels.

For image-based relighting, a fixed viewpoint is often used to reduce the dimensionality of the full 8D reflectance field, leading to a 2D radiant field captured in a single photograph. Along with a general 4D incident light field, we are left with the problem of estimating a 6D reflectance field. While the reflectance field gives a complete description of the interaction of light with a scene, its acquisition would require photographing the scene by turning on one ray at a time. This would obviously require inordinate quantities of time and storage. Significant strides have been made toward acquiring lower dimensional subsets of this function and using it for restricted relighting and rendering.

Debevec et al. [113] developed a *light stage*, comprised of a light mounted on a rotating robotic arm, to acquire the non-local reflectance field of a human face. As previously mentioned, a data-driven approach for capturing human skin overcomes the challenges of a model-based approach. In this case, the point-like light source can be thought of as a simplified projector containing a single pixel and translated over the surface of a sphere by the robotic arm. Thus the incident light field is reduced to a 2D function. In total, the 2D projection of radiant field plus the 2D incident light field requires capturing a 4D function. The authors demonstrate the generation of novel images under arbitrary lighting conditions.

More specifically, image-based relighting was accomplished simply by adjusting the weights w_i to match the desired intensity of illumination from various directions, corresponding to positions of a point light on the virtual dome. Going beyond relighting, the authors added a small number of temporally synchronized cameras to capture images of the object from neighboring viewpoints. By using the captured data, and by exploiting the linearity of light, they were able to simulate small alterations of viewpoint with a simple model for skin reflectance. Hawkins et al. [Hawkins et al. 2001] employed a similar configuration to digitize cultural artifacts by capturing reflectance fields. In contrast to previous cultural heritage preservations efforts, such as the Digital Michelangelo Project [237], which used 3D scanners and model-driven approaches for relighting, such methods can capture challenging translucent materials. As a result, they argue for the use of reflectance field capture in digital archiving, rather than geometric models and reflectance textures.

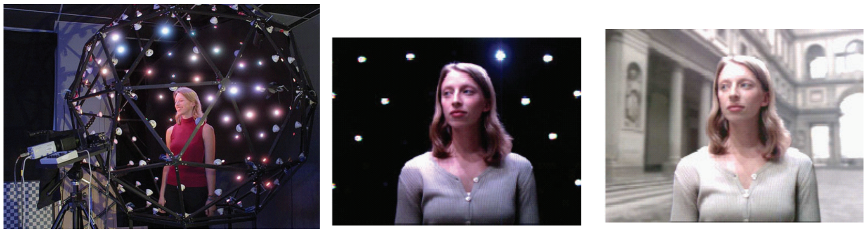


Figure 5.11: The light stage for light-aware matting. Novel lighting conditions can be synthesized by using a linear combination of basis-lighting images collected with a set of translated point sources. (left) A set of high-lumen LEDs are positioned at the vertices of a rigid dome. (center) A performance is captured while the LEDs are continuously strobed to capture the basis-lighting imagery. (right) The actor is inserted into a synthetic scene by using the estimated background matte. Consistent lighting is achieved by weighting the basis lighting by the environment map for the virtual scene. [115] [THE CAPTION TALKS ABOUT BASIS-LIGHTING IMAGES, BUT THE TEXT DOESN'T.]

Koudelka et al. [114] captured a set of images from a single viewpoint as a point light source rotated around the photographic subject, and estimated the surface geometry by using two sets of basis images. From their estimation of the apparent BRDF for each pixel in the images, they could render the subject under arbitrary illumination.

In subsequent years, Debevec et al. [115] proposed enhanced light stages. For example, a later-generation light stage contained a large number (156) of inwardly oriented LEDs distributed over a spherical structure approximately two meters in diameter around the photographic subject—in this case an actor (Figure 5.11, left). Each light was set to an arbitrary color and intensity to simulate the effect of a real-world environment around the actor (Figure 5.11, center). The images gathered from such a light stage, together with a mask of the actor captured under infrared sources and detectors, were used to composite the actor seamlessly into a virtual set, while maintaining consistent illumination (Figure 5.11, right).

Malzbender et al. [254] employed 50 inwardly oriented flashes distributed over a hemispherical dome, together with a novel scheme—called the *polynomial texture map*—for compressing and storing the 4D reflectance field (see Figure 5.12). They assumed that the color of a pixel changed smoothly as the light moved around the object, and they stored only the coefficients of the biquadratic polynomial that best modeled this change for each pixel. Such a highly compact representation allows for real-time rendering of the scene with arbitrary illumination, and works fairly well for diffuse objects. Specular highlights, however, are not modeled well by the polynomial model and result in visual artifacts. While efficient light field compression was considered in the initial publication [239], it remains an open problem in the field. MPEG standardization efforts are currently underway to compress the large multi-view datasets produced by closely spaced camera arrays.

To avoid the mechanical complexity of a light stage, we can employ a more flexible setup and use, say, a handheld light source freely moving around the pho-

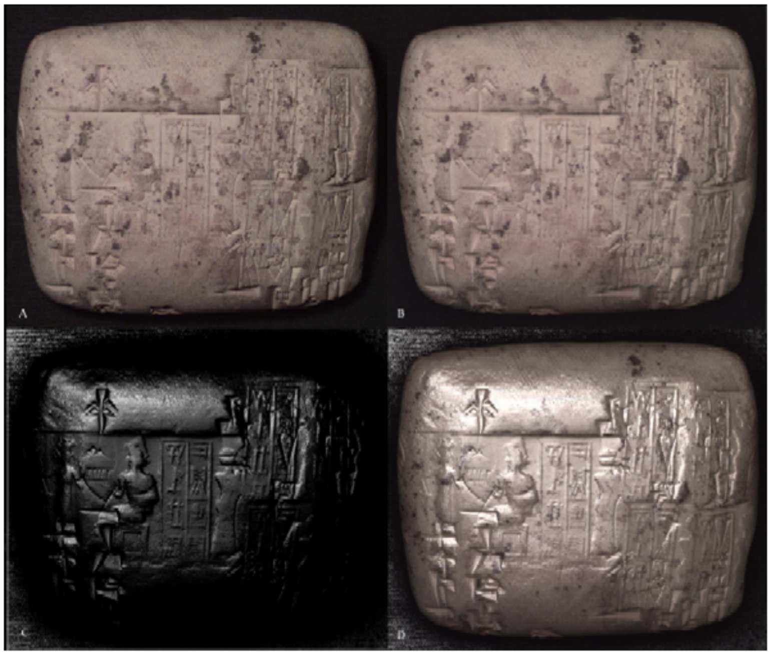


Figure 5.12: Polynomial texture maps store the material appearance under varying illumination by using just six coefficients at each pixel. A user can perform interactive relighting for desired visual effects [254].

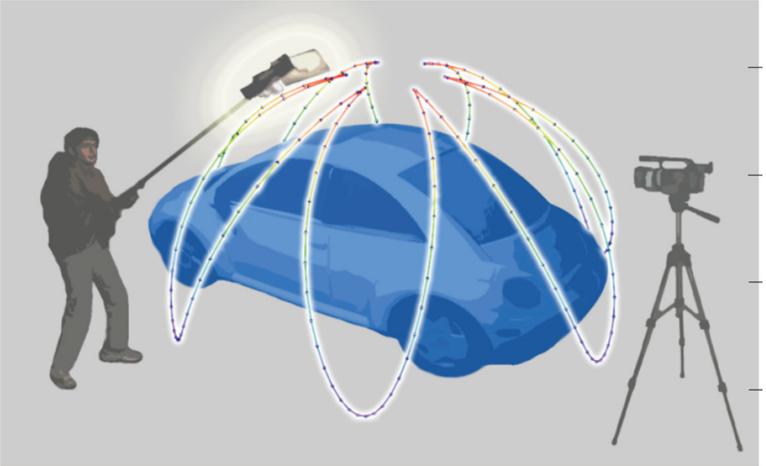


Figure 5.13: The light-waving approach can capture a 4D reflectance field for post-capture relighting without a lighting dome [431].

tographic subject (see Figure 5.13). In this case, the task is to estimate the light positions directly from the recorded imagery. The free-form light stage [27] presented a strategy in which the position of lights was estimated automatically from four diffuse spheres placed near the subject in the field of view of the camera. Data acquisition time was reported as 25 to 30 minutes. Winnemöller et al. [431] used dimensionality reduction and a slightly constrained light scanning pattern to estimate light source position without the need for additional fiducials in the scene.

Winnemöller et al. [431] argued that accurate calibration of light positions is unnecessary for the application of photographic relighting. They proposed a novel reflector-based acquisition system. They inserted a moving-head gimbaled disco light inside a diffuse enclosure, together with the subject to be photographed. The spot from the light on the enclosure acts as an area light source that illuminates the subject. The light source is moved by simply rotating the light and capturing images with various light positions. The concept of area light sources is also used in Bayesian relighting [29]. In both of these cases, an inexpensive gimbaled disco light is used in place of more expensive robotic arms or lighting domes employed in previous systems. The primarily limitations of such designs, excluding long capture times, is the inability to capture sharp shadow details and specularites, since the efficient point light sources created by such designs are larger than the point sources used in the competing systems.

The key disadvantage of many of these reflectance field capture techniques is that they can be used mainly for scenes that are static while multiple photos are captured under varying lighting conditions from a fixed camera viewpoint. Any relative motion among the three elements—the scene, the camera, and the lighting—will introduce artifacts. Some of these limitations can be addressed by using motion compensation via image registration. but often the motion of any one of the elements creates two different relative motions. Thus it is quite challenging to use such methods for traditional photography. Nevertheless, in many controllable settings these methods can be applied successfully.

5.4.3 Toward Reflectance Fields Capture in 4D, 6D, and 8D

The most complete image-based description of a scene for computer graphics applications is its 8D reflectance field [93]. The measurement of reflectance fields is an active area of research. The 8D reflectance field describes the transfer of energy between a light field of incoming rays (the illumination) and a light field of outgoing rays (the view) in a scene, each of which is 4D. As we saw earlier, this representation can be used to synthesize images of the scene from any viewpoint under arbitrary lighting, subject to sampling constraints. The synthesized results accurately capture global illumination effects such as diffuse interreflections, shadows, caustics, and sub-surface scattering, without the need for an explicit physical simulation.

Most of the prior research, however, has focused on capturing meaningful lower dimensional slices of the 8D reflectance field. We saw examples earlier of capturing 4D datasets for relighting from a fixed viewpoint and variable lighting. If the illumination is provided by an array of video projectors and the scene is captured as illuminated by each pixel of each projector, but still as seen from a single viewpoint, then we obtain a 6D slice of the 8D reflectance field. If we use k projectors, each with a million pixels, we need to capture k-million photos for this 6D dataset, since we can measure the impact of only a single projector pixel in each photo. Masselus et al. [260] captured such datasets by using a single moving projector positioned in the k positions.

Sen et al. [362] exploited Helmholtz reciprocity to develop a “dual photography” approach. The Helmholtz reciprocity allows them to interchange the projectors and cameras in a scene. Instead of one camera and k projectors, they used k cameras and one projector. Unlike an array of (lights or) projectors, an array of cameras can operate in parallel without interference. By turning on each projector pixel, one for each photo, but simultaneously capturing k photos, the authors improved on the capture times of these datasets. An earlier method for capturing the full 8D reflectance field [148] exploited the data-sparseness of the 8D transport matrix to represent the transport matrix by local rank-1 approximations. With the sparsity observations, the authors developed a hierarchical parallelized acquisition technique that significantly sped up the process for capturing the reflectance field.

More recently, Peers et al. [315] and Sen and Darabi [363] have used compressive sensing to rapidly acquire reflectance fields by using structured illumination. Sen and Darabi capture a 6D transport matrix by photographing a static scene illuminated by a single projector displaying a sequence of Bernoulli noise patterns. Peers et al. project a similar set of high-frequency noise patterns. In their case, however, the patterns are first projected into a compression basis (e.g., wavelets). In both cases, the compressibility of the reflectance field is exploited to reduce the number of observations required to estimate the elements of the corresponding transport matrix.

At the moment, such efforts to exploit compressibility in light field and reflectance field capture are just beginning to emerge, particularly because of the introduction of compressed sensing itself during the last decade [87]. We expect such efforts will be necessary to reduce the prohibitive capture times currently required for data-driven image-based relighting. Whether light field and reflectance field capture occurs best through clever optical configurations, such as dual photography, or through insightful uses of compressive sensing remains an issue for future research.

5.5 Modulation in Space and Time

The capacity to modulate the flash intensity over both space and time provides additional control of the resulting image. An intelligent camera flash would behave much like a projector. A projector allows modulation of ray intensities in each direction by changing pixel intensities, and is an ideal programmable spatio-temporal light emitter. Hence projectors are commonly used in computational illumination research, although they are inconvenient for incorporation in a practical camera. Using a projector-like light source as a camera flash, which allows for varying not only the overall brightness but also the radiance of every emitted ray, is a powerful alternative to a conventional flash. As a result, an ideal next-generation camera flash would provide similar control over the full 2D set of emitted rays via manipulation of pixel intensity.

Shree Nayar coined the term “CamPro” to designate a projector that supports the operation of a camera [294]. A projector can project arbitrarily complex illumination patterns onto the scene, capture the corresponding images, and compute scene information that is impossible to obtain with traditional flash. Captured images are optically coded by the patterned illumination of the scene. In the future, the unwieldy projector may be replaced by smart lasers or light sources with highly programmable mask patterns.

5.5.1 Structured Light Projection

For scanning the 3D surface of opaque objects, coded structured light is considered one of the most reliable techniques. This technique is based on projecting a set of coded light patterns and imaging the illuminated scene from one or more cameras. Such a scenario can be used to simplify the well-known “correspondence problem” in stereo photography. Given a pair of images of the same 3D scene, captured from two different points of view, the purpose of the correspondence problem is to find a set of points in one image identical to points in another image. For an arbitrary 3D point, its projection in the two images is defined by a pair of corresponding pixels. In turn, given the pair of corresponding pixels in two images, we can compute the 3D location of that point by triangulation (see Figure 5.14). Of course, such methods rely on accurate calibration of the cameras, including the lens distortion, projection matrices, and relative position and pose of each camera with respect to a world coordinate system.

In the case of projected structured light, a single camera view can be used along

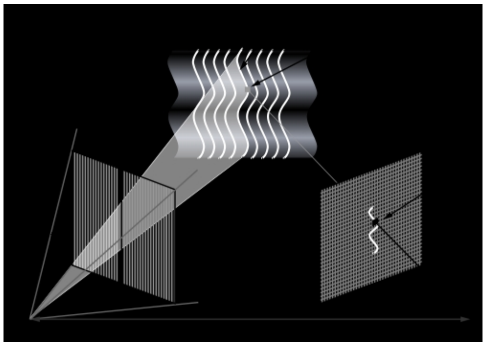


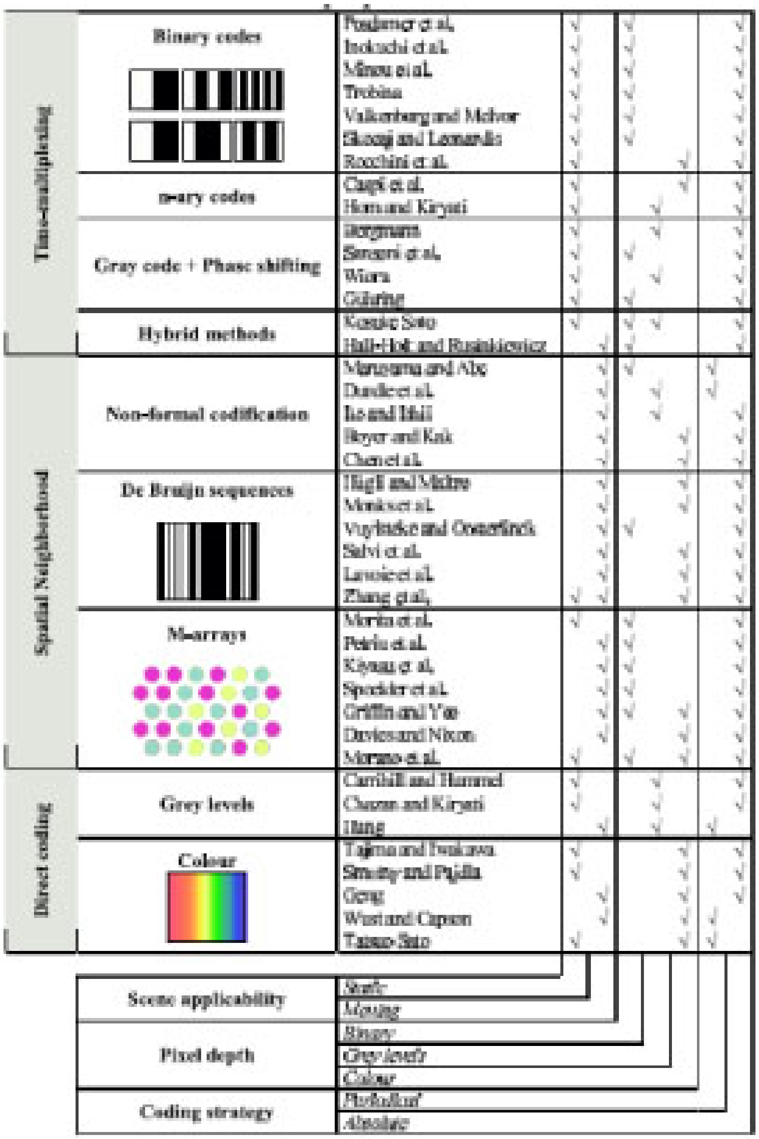
Figure 5.14: Structured light scanning using triangulation between camera and projector.

with a single projector view. By using temporal multiplexing, the projected pattern is coded so that over time each projector pixel is assigned a unique code. Because of this temporal coding, correspondences between camera image points and points of the projector pattern can easily be decoded. In practice, Gray codes are routinely used, yet even simple binary encoding would be sufficient. By triangulating the decoded points, 3D information is recovered in a manner equivalent to stereo triangulation with a pair of cameras. In place of a second passive camera, the projector actively encodes the space via illumination. Hence, this is known as active stereo triangulation.

Coding schemes continue to evolve. The number of projected patterns required to encode the projector pixel can be reduced by exploiting color. Rusinkiewicz et al. exploited modest assumptions about local smoothness of surface and reflectance, and derived a new set of illumination patterns based on coding the boundaries between projected stripes. [349]. In practice, both spatial (within a frame) and temporal (between frame) modulation has been proposed. While the introduction of structured light with binary codes was proposed by Posdamer and Altschuler in 1981 [323], the field of structured lighting remains an active area with new codes emerging every year, tailored for specific applications. The table in Figure 5.15) from Salvi et al. [351] provides a good overview of pattern codification strategies in structured light systems. A good survey is available at www.cs.cmu.edu/~seitz/course/Sigg00/notes.html.

5.5.2 High Spatial Frequency Patterns

Active illumination approaches have been used to analyze multipath light scattering, and to compute the inverse of the light transport. Consider a scene lit by a point light source and viewed by a camera. The brightness of each scene point has two components: direct and global. The direct component results from light received directly from the source. The global component results from light received from all other points in the scene. It turns out that individual materials exhibit unique and fascinating direct and global illumination properties. A traditional camera receives a sum of the two. But a programmable flash can be used to separate a



			Scene applicability		Pixel depth			Coding strategy	
			Static	Moving	Binary	Grey levels	Colour	Periodical	Absolute
Time-multiplexing	Binary codes	Posdamer et al.	√		√				√
		Inokuchi et al.	√		√				√
		Minou et al.	√		√				√
		Trobina	√		√				√
		Valkenburg and McIvor	√		√				√
		Skocaj and Leonardis	√		√				√
		Rocchini et al.	√				√		√
	n-ary codes	Caspi et al.	√				√		√
		Horn and Kiryati	√			√			√
	Gray code + Phase shifting	Bergmann	√			√			√
		Sansoni et al.	√		√				√
		Wiora	√			√			√
		Gühring	√		√				√
	Hybrid methods	Kosuke Sato	√		√	√			√
		Hall-Holt and Rusinkiewicz		√	√				√
Spatial Neighborhood	Non-formal codification	Maruyama and Abe		√	√			√	
		Durdle et al.		√		√		√	
		Ito and Ishii		√		√			√
		Boyer and Kak		√			√		√
		Chen et al.		√			√		√
	De Bruijn sequences	Hügli and Maître		√			√		√
		Monks et al.		√			√		√
		Vuylsteke and Oosterlinck		√	√				√
		Salvi et al.		√			√		√
		Lavoie et al.		√			√		√
		Zhang et al.	√	√			√		√
	M-arrays	Morita et al.	√		√				√
		Petriu et al.		√	√				√
		Kiyasu et al.		√	√				√
		Spoelder et al.		√	√				√
		Griffin and Yee		√	√		√		√
		Davies and Nixon		√			√		√
		Morano et al.	√		√		√		√
Direct coding	Grey levels	Carrihill and Hummel	√			√			√
		Chazan and Kiryati	√			√			√
		Hung		√		√		√	
	Colour	Tajima and Iwakawa	√				√		√
		Smutny and Pajdla	√				√		√
		Geng		√			√		√
		Wust and Capson		√			√	√	
		Tatsuo Sato	√				√	√	

Figure 5.15: Pattern codification strategies in structured light systems, which have evolved over the last twenty years. Simple temporal sequences, such as binary and Gray codes, have been enhanced by using N-ary, phase shifting, and colored sequences. Single- shot methods, including spatial coding with M-arrays, have been used for scanning dynamic scenes. (Table is from Salvi et al. [351].)

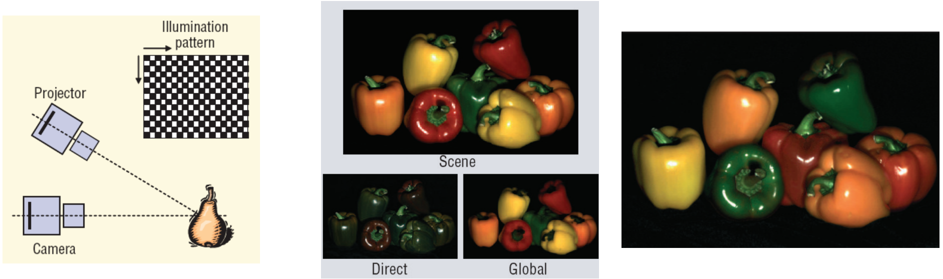


Figure 5.16: Fast separation of direct and global illumination using high-frequency projected patterns. (left) The projector is used as a programmable camera flash. A sequence of shifted high-frequency checkerboard patterns illuminates the scene. (middle) Direct and global illumination components can be easily extracted from the shifted checkerboard sequence. (right) Altering the colors of the peppers in the global component image allows novel image synthesis.

scene into its direct and global components. The two components can then be used to edit the physical properties of objects in the scene to produce novel images.

The image on the left side of Figure 5.16 shows a scene captured by using a checkerboard illumination pattern. If the frequency of the checkerboard pattern is high, then the camera brightness of a point lit by one of the white squares includes the direct component and exactly half of the global component because the checkerboard pattern lights only half of the remaining scene points.

Now consider a second image captured by using the complement of this checkerboard illumination pattern. In this case, the point does not have a direct component but still produces exactly half of the global component. This occurs because the complementary checkerboard pattern lights the remaining half of scene points. Since the above argument applies to all points in the scene, the direct and global components of all the scene points can be measured by projecting just two illumination patterns. The middle image in Figure 5.16 shows separation results for a scene with peppers of different colors. The direct image includes mainly the specular reflections from the surfaces of the peppers. The colors of the peppers come from subsurface scattering effects captured in the global image. Altering the colors of the peppers in the global image and recombining it with the direct image yields a novel image, like that shown on the right in Figure 5.16.

In addition to subsurface scattering, this separation method can be applied to a variety of global illumination effects, including inter-reflections among opaque surfaces and volumetric scattering from participating media. Thus we can distinguish the first bounce direct illumination effect from the multipath scattering caused by global illumination. In practice, additional shifted checkerboard patterns can be used to improve the accuracy of the direct-global separation achieved with this method. As with many examples in computational illumination we've reviewed in this chapter, this leads to a natural trade-off in separation accuracy versus the number of frames. For dynamic scenes, only a few projected patterns could be used in practice, unless special efforts are made to allow high frame rates. Thus the results are best for static scenes.

Direct-global separation is only one such decomposition that can be performed to understand the properties of reflected light fields. Seitz et al. [361] go beyond this partial inversion of light transport. They developed a mechanism to represent the impact of individual bounces of an optical ray. For a purely diffuse scene they also devised a practical method to capture and invert the light transport. They used the same 4D transport matrices we discussed above to model the light transport from a projector to a camera, but their work provides a theory for decomposing the transport matrix into individual bounce light transport matrices.

5.5.3 Modulation in Time

As with the early work in multiflash imaging, the pattern of the flash can also be changed over time. We can synchronize strobes with activity in the scene. For example, high temporal frequency strobes can be used to "freeze" periodic motion. The idea is to create a new low "apparent frequency" for a high-frequency periodic motion. When the periodic scene motion and strobed flash have slightly different frequencies, the perceived rate of periodic motion is the difference between the two frequencies.

For example, vocal folds moving at 1000 Hz can be viewed with a laryngoscope with auxiliary lighting.[1] If the strobe is also at 1000Hz, the vocal folds appear frozen, as long as the person maintains a continuously pitched sound. If the strobed frequency is 999Hz, the strobe creates a 1 Hz apparent frequency so that the vocal folds appear to move only once per second. This makes it easy for the observing physician to see and evaluate the correctness of vocal fold movement. In addition, he can detect any distortions of the vocal fold shape. Sometimes the strobes are colored with different phase delay, or with different frequencies. If anything is static, the two colors just add up. If the object is moving, the moving object appears to have colored trails. [THIS SHORT SECTION COULD USE OTHER EXAMPLES OF MODULATION IN TIME.]

5.6 Exploiting Natural Variations in Illumination

Sometimes we cannot actively change the illumination of a scene for photography, usually because of limited access or proximity to the scene. We can still exploit natural variations such as changes in sunlight throughout the day.

5.6.1 Intrinsic Images

With intrinsic image decomposition, the goal is to decompose the input image I into a reflectance image and an illumination image.

An image is produced because of additive or multiplicative components of a scene. Two of the most important components of the scene are its shading (due to incident illumination) and reflectance (due to material). The shading of a scene is the interaction of the surfaces in the scene and its illumination. The reflectance

[1] Optiview System; www.divop.com/downloads/SS109BOV.pdf

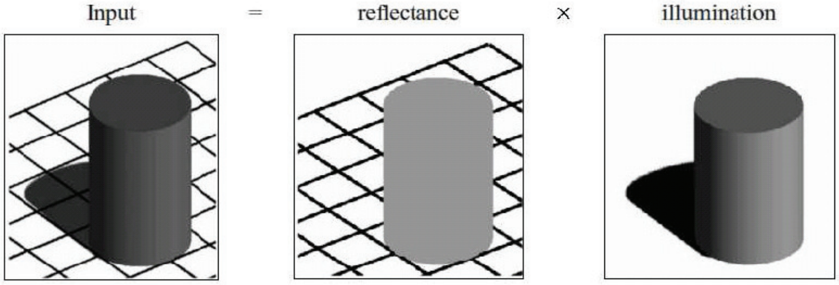


Figure 5.17: An image can be decomposed into a multiplicative combination of a reflectance (intrinsic) layer and an illumination (shading) layer. Such decompositions, like the example of direct-global separation, can be used for applications in computer vision, where illumination variation can confuse recognition algorithms. They can be used as well for computer graphics applications, where the reflectance layer can be edited to insert synthetic objects.

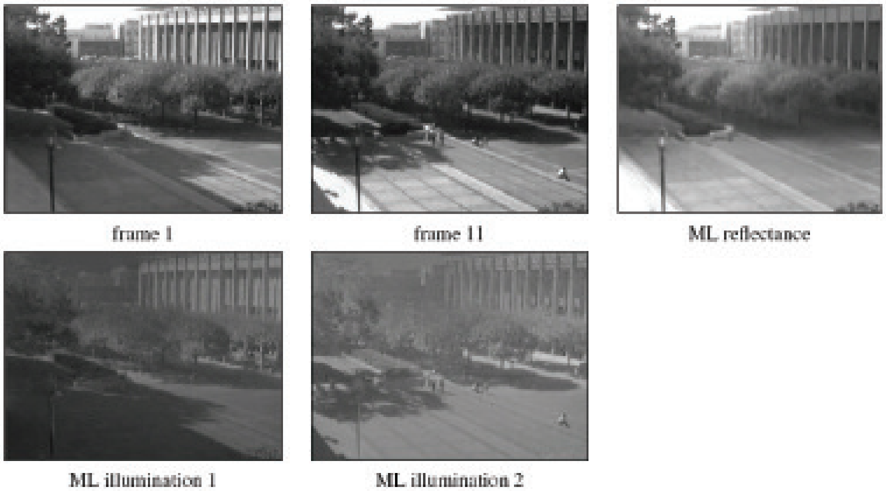


Figure 5.18: Intrinsic images from a webcam sequence. A sequence of 35 webcam images captures the natural variation in illumination through the course of several hours. The maximum-likelihood (ML) reflectance image, free of cast shadows, can be estimated from such sequences [423].

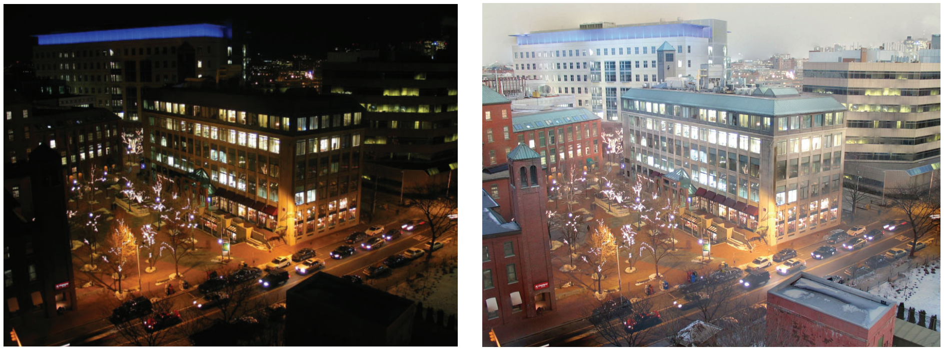


Figure 5.19: A night-time photo is enhanced by using a day-time photo. (left) The input night-time photo. (right) The enhanced photo created by using gradient-domain fusion to enhance the visibility of the night-time photo. [330]

of the scene describes the pattern and material and how each point reflects light (see Figure 5.17). The ability to find the reflectance of each point in the scene and how it is shaded is important because interpreting an image requires the ability to decide how these two factors affect the image. For example, segmentation would be simpler given the reflectance of each point in the scene.

Yair Weiss [423] showed a method to compute the reflectance components, or the *intrinsic image* by using multiple photos where the scene reflectance is constant, but the illumination changes. Even with multiple photo observations, this problem is still ill-posed, and Weiss suggests approaching it as a maximum-likelihood estimation problem. He computes the gradient (forward differences) in each photo. The median of gradient over time at each pixel gives the estimated gradient of the intrinsic reflectance image. The intrinsic image is estimated by performing 2D integration of the 2D gradient field. He also shows that such a reflectance-only layer can be manipulated by inserting new materials. By multiplying by the illumination layer, augmentation of real scene photos can be done. In the example of a webcam sequence shown in Figure 5.18, a novel image decomposition can be achieved, but because natural illumination is exploited, the data collect time can be prohibitively long. In this case, several hours may be required to observe sufficient gradient variation to estimate an intrinsic image.

5.6.2 Context Enhancement of Night-Time Photos

Natural illumination changes are most prevalent over the day-night cycle. Night-time images such as the one shown in Figure 5.19 (left) are difficult to understand because they lack background context due to poor illumination. If this photo is taken from an installed camera, we can exploit the fact that the camera can observe the scene all day long and create a high-quality, well-illuminated background. Then, we can simply enhance the context of the low quality image or video by fusing the appropriate pixels, as shown in Figure 5.19 (right).

Raskar et al. show that an image fusion approach is based on a gradient domain

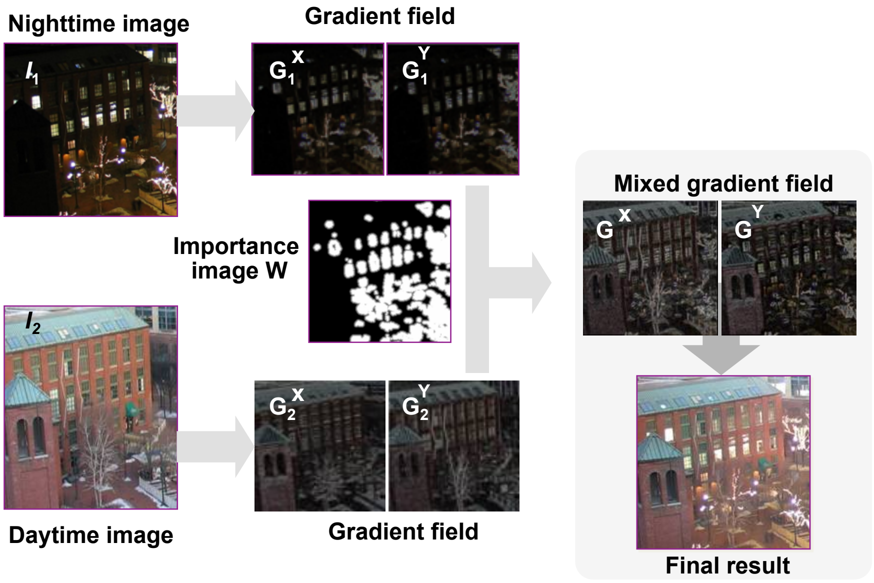


Figure 5.20: Gradient domain fusion for context enhancement. [330]

technique that preserves important local perceptual cues while avoiding traditional problems such as aliasing, ghosting and haloing [330]. They first encode the pixel importance based on local variance in input images or videos. Then, instead of a convex combination of pixel intensities, they use linear combinations of the intensity gradients where the weights are scaled by pixel importance. The image reconstructed from integration of the gradients achieves a smooth blend of the input images, and at the same time preserves their important features. The dark regions of the night image in Figure 5.19 (left) are filled in by day image pixels but with a smooth transition. Figure 5.20 illustrates the steps in the gradient domain fusion process.

Similar to other time-lapse methods, data collection can be prohibitively long for some applications. The results would be difficult to achieve, however, by using active programmable illumination or direct vision-based approaches, given the scale and complexity of outdoor environments.

Chapter 6

Modern Image Sensors

Digital image sensors convert incident photons, which contain a wide spectrum of intensities and colors, into arrays of electric charges. These charges are then processed into image pixels that are perceived by the human eye as photographs. We begin this chapter with a description of the parameters of digital image sensors, followed by an examination of the human retina. We then return to the subject of image sensors for a summary of more sophisticated sensor operations.

The physical design and software processing of modern sensors contribute greatly in solving important challenges in photography. In the following sections we examine a sensors basic ability to record dynamic range, resolution, color, depth estimation, and motion. We then discuss how enhanced sensor design and software can improve image output. These enhancements include varied pixel layouts, additional color filters or transparency filters, structured masks and lenses on the sensor surface, and irregular integration time of the photon intensities. Additional enhancements can include sampling from neighboring pixels and synchronization with active illumination.

6.1 How Image Sensors Work

The surface of a digital sensor contains millions of photosensitive diodes, or photosites, which are designed to capture photons. The accumulated photons at each photosite generate electrons in the semiconductor material of the sensor, and these electrons represent the relative brightness of the incident light at that site. The more light that hits a photosite, the more photons it records. Photosites capturing highlights in a scene will absorb many photons, while photosites capturing shadows will absorb few photons. When the exposure is complete, the accumulated electrons at each photosite are counted, processed, and converted into a digital number that represents the intensity and color of that pixel in the digital image. [Reference is www.shortcourses.com/sensors/)

Sensors that use this process, called photo detection, are classified by the following parameters: pixel size, fill factor, full well depth, dynamic range, spectral quantum efficiency, sensitivity, and noise.

Pixel Size

Larger photosites allow more light to be collected in an exposure interval. A typical photosite size in a commercial digital camera is 3–10 μm. For special applications such as astronomy the photosite size can be as large as 20 μm, while smaller and cheaper sensors with photosites as small as 2 μm are currently in demand for cell phones and compact digital cameras. Unfortunately, sensor miniturization can go only so far. At a certain point, diffraction-limited optics in a given camera and lens system, not the pixel size and overall sensor resolution, becomes the limiting factor in image quality.

Fill Factor

This parameter, also know as the optical fill factor, is the percent of the photosite area that actually captures photons. In current sensors it is well below an ideal value of 100% and typically close to 25% because of the space required for non-light-gathering substrate material and additional circuitry. Microlens arrays (also called microlenticular arrays or lenslet arrays) on the surface of some sensors can increase the fill factor. These tiny lens systems focus photons onto the active photosite area instead of allowing them to fall on non-light-gathering areas where they would not be collected. [from learn.hamamatsu.com/articles/microlensarray.html]

Full Well Depth

Text to be added.

Dynamic Range

This parameter specifies the maximum achievable signal strength divided by the camera noise, where the signal strength is determined by the full-well capacity, and camera noise is the sum of dark and read noises. Dynamic range is further defined as intrascene and interscene. As the dynamic range of a device is increased, the ability to quantitatively measure the dimmest intensities in an image (intrascene performance) is improved. The interscene dynamic range represents the range of intensities that can be accommodated when detector gain, integration time, lens aperture, and other variables are adjusted for differing fields of view. The full-well capacity, which corresponds to the saturation charge and depends on the pixel size, limits dynamic range. In case of CCDs, once the finite charge capacity of the well fills up, accumulation of additional photo-generated charge results in overflow, or blooming, of the excess electrons into adjacent sensor photosites.

Spectral Quantum Efficiency

This parameter is the number of electron-hole pairs created and successfully read out by the sensor for each incoming photon. It is especially important for low-light imaging applications. In addition, conversion gain defines volts per electron. Sensor sensitivity, which is quantum efficiency × conversion gain, is often measured in volts/lux, where lux is measured in watts/m^2 (watts per meter squared).

Sensitivity

Text to be added.

Noise

Text to be added.

6.1.1 CCD and CMOS Image Sensors

[Material below is from www.shortcourses.com/sensors/sensors1-1.html] All image sensors capture light by exposing a grid of small photosites on the sensor surface. The two primary types of sensors—CCD and CMOS—differ from each other in how they are manufactured and how they process the image.

CCD Image Sensors

A charge-coupled device (CCD) is characterized by how the charges on its photosites are read after an exposure. Charges in the photosites are transferred, one row at a time, to a sensor storage unit called the read-out register. The charges are then amplified and sent to an analog-to-digital converter. After conversion, the charges in the read-out register are deleted and the next row is processed. In this manner all the rows in the sensor grid are processed in sequence. Each row of charges is thus "coupled" to the row above it, and each row of image pixels is determined one row at a time.

CMOS Image Sensors

Image sensors are manufactured in wafer foundries, or fabs, where tiny circuits are etched onto silicon chips. The biggest problem with CCDs is that they use specialized manufacturing processes that are costly and dont allow economies of scale. Larger foundries use a more efficient process called complementary metal oxide semiconductor (CMOS), which is the most widespread and highest yielding chip-making process in the world. Millions of computer processors and memory chips are made yearly as CMOS devices. By adapting the CMOS process and equipment to make CMOS image sensors, manufacturing costs are dramatically reduced because the fixed costs of fabrication are spread over a larger number of devices. As a result of these favorable economies of scale, the cost of fabricating a CMOS wafer is significantly less than the cost of fabricating a similar CCD wafer. Overall costs are further lowered because CMOS image sensors can have processing circuits on the same chip, while CCD sensors require processing circuits on separate chips.

Regardless of their manufacturing differences, CCD and CMOS sensors can produce excellent images, and both types of sensors are found in successful cameras from major camera companies.

6.1.2 Noise and Resolution Limits

astro.union.rpi.edu/documents/CCD%20Image%20Sensor%20Noise%20Sources.pdf
theory.uchicago.edu/~ejm/pix/20d/tests/noise/

(From www.dpreview.com/learn/?/key=noise) Each sensor photosite contains light-sensitive photodiodes that convert incoming photons into an electrical signal that is subsequently processed into the intensity and color value of the pixel in the final image. If the same pixel is exposed several times by the same amount of light, the resulting color values would have small statistical variations, called noise, in the final intensity and color values. Even when no light is striking the sensor, the inherent electrical activity of the sensor generates some background signal, equivalent to the faint hiss of audio equipment that is on but not playing music. This additional noise signal, called the noise floor, varies over time from pixel to pixel, increases with temperature, and adds to the overall noise of an image. Clearly, the output of a photosite has to be larger than the noise floor in order to distinguish the signal from noise.

[Resolution Limits. Add a paragraph here on this topic, since it is in the section subheader.]

6.2 The Human Retina

(Most text below is created by taking individual sentences spread throughout the article at faculty-web.at.northwestern.edu/med/fukui/Human%20eye.pdf)

The interscene dynamic range of photoreceptor cells in the human eye is more than ten decades, ranging between 10^{-6} and 10^5 lumens/m^2 (lux) of light intensity. Cones in the retina are responsible for photopic vision (color vision in the upper seven decades of photoreception), while rods are responsible for scotopic vision (grayscale vision in the lower three decades). Photopic color vision occurs because of a photochemical reaction of red, green, and blue pigments, while scotopic vision is caused by the presence of rhodopsine. In general, the human eye requires about thirty minutes of adaptation to adjust from photopic vision to scotopic vision. During the interval of time before full adaptation is acquired, the intrascene dynamic range is only about four decades.

Spatial Resolution

In the center of the retina (the central fovea), the distance between adjacent cones is 1.5 to 2 mm. By taking the Nyquist factor of two, we would need a separation of 3–4 mm for stimuli to be accurately resolved. In contrast, if we assume the diameter of the pupil is 2 mm, then the estimated radius of the first-order diffraction pattern for a 555-nm point light source formed at the retina would be 4.6 mm. The visual accuracy is closely related to the ability of involuntary rapid movement of eyes. Once scanned, the signal is processed by a complex neuron network. The threshold for photo sensitivity is 100–150 photons (measured at 507 nm light entering a pupil with a diameter of 2–8 mm), which is equivalent to a luminance of smaller than 10^{-6} cd/m^2. The wavelength of light we are sensitive to changes with the eyes adaptation to light levels in bright or dark scenes. Light-adapted eyes are sensitive to 400–700 nm (peak at 555 nm, or yellow light), while dark-adapted eyes are sensitive to 380–650 nm (peak at 507 nm, or green light). In terms of contrast, for reasonably high luminance and for a test object that is fairly large, human eyes can detect about a 2% difference in gray levels. However, at very low luminance, a contrast of 100% (double the luminance) or more is needed to be distinguished.

According to the Rose model [345], the human eye as a sensing device has the following properties: storage time = 0.2 seconds; signal-to-noise (S/N) ratio = 5; quantum efficiency at low luminance = 5%; and quantum efficiency at high luminance = 0.5%. A quantum efficiency of 5% indicates that only five out of a hundred photons that enter the eye are detected by the retina. Modern digital sensors have a much higher quantum efficiency (around 30%) and a higher signal-to-noise ratio. Despite these two parameters, however, digital image sensors still dont come close to the capability of the human eye in dynamic range, spectral wavelength range, noise performance, and other parameters. Perhaps this is true because the evolution of the human eye has tuned these parameters more effectively, so that the eye is much more compatible (compared to digital sensors) with the natural limits of optics, sensing and processing.

We now look at how modern sensors deal with issues such as dynamic range, resolution, color, and motion (e.g., camera shake by the photographer and object motion in the scene). We also examine how modern sensors enable a new range of capabilities, such as 3D range sensing.

6.3 Extended Dynamic Range

Creating accurate high dynamic range (HDR) images from natural scenes remains a major challenge for solid-state image sensors. The human eye can perceive a much greater range of scene intensities than any current sensor can record, which means digital images can only approximate human vision. Scenes observed by the human eye span over eight decades of illuminance, ranging from $10^{-3)}$ lux in starlight to 10^3 lux for indoor lighting, 10^5 lux for bright sunlight, and even higher illuminance levels for specularities or direct viewing of bright sources (such as oncoming headlights or the sun) [20]. At any one moment in time, within this range, the human eye can perceive illuminances spanning five decades. Typical APS-sized linear CCD and CMOS sensors in most consumer cameras can capture three decades of dynamic range, while logarithmic CMOS sensors can capture over five decades.

Human perception roughly approximates Weber's law, which states that the threshold to sense a difference between the illuminance of a fixation point and its surroundings is a fraction, about 1–10%, of the surrounding illuminance. It would take 23 bits to quantize illuminance on a linear scale with 1% accuracy throughout a five-decade range. This is currently beyond the capability of most solid state sensors, which typically record only 8 to 12 bits. At the sensor level, various approaches have been proposed for high dynamic range imaging. These will be summarized in the subsections that follow. In general, high dynamic range imaging sensors must typically solve two main challenges in creating additional digital levels: (i) adding quantization bits to highlights and (ii) adding quantization bits to shadows.

6.3.1 Space Domain: Multisensor Pixels

One approach to high dynamic range imaging uses multiple sensing elements with different sensitivities within each photosite cell [381, 171, 424, 168]. Multiple scene measurements of different light intensities are made by the sensing elements, and these measurements are combined on-chip to produce a high dynamic range image. Unfortunately, because of the presence of these multiple elements, the spatial sampling rate is lowered in these devices, and spatial resolution is sacrificed in the final image. Research in improved high dynamic range sensor design is currently in progress, but the implementation is usually costly.

A novel approach called *assorted pixels* has been proposed by Nayar and Narasimhan [293, **?**]. In this design the effective exposure varies across the spatial dimension of the imaging sensor. A pattern with varying sensitivities is applied to the photosite array. This pattern resembles the Bayer color mosaic pattern, but rather than using color filters, the pattern changes exposure sensitivity instead. The particular form of the sensitivity pattern, and the methods of implementing it, are both quite flexible. One implementation places a mask with cells of varying optical transparencies in front of the sensing array. Here, just as with a Bayer mosaic, spatial resolution is somewhat sacrificed and aliasing can occur. Measurements under different exposures (sensitivities) are accumulated and spatially interpolated, and then combined into a high dynamic range image.

Unfortunately, a major limitation of creating spatially varying sensitivity by manipulating pattern transparency is that the pattern reduces the amount of light striking the photosites, and we lose photons. A current series of implementations in a commercial camera, the Fuji Super CCD (HR, SR, and SR II) [144], uses two photosites per image pixel, where one photosite is larger than the other. The pixels are octagonal rather than rectangular with the larger photosite at the center of the pixel and a set of smaller photosites between the other pixels. [****NOTE: insert a figure here to illustrate this configuration.] The larger photosite accumulates photons more quickly than the smaller sites, and hence it saturates more quickly and provides greater quantization in the highlight areas of the scene.

6.3.2 Time Domain

Another approach to high dynamic range imaging adjusts the well capacity of the sensing elements during photocurrent integration [217, 353, 95]. Unfortunately, this approach produces higher noise [****NOTE: Why?]. In a different approach [67], the time needed for each photosite to reach saturation is measured by a computation element attached to each sensing element. This time encodes high dynamic range information because it is inversely proportional to the brightness at each pixel [****NOTE: explain further?].

Brightside Technologies [18] exploits the interline transfer of a CCD sensor to capture two exposures during a single mechanical shutter timing. [****NOTE: more description of this sensor is needed here. Plus, according to Wikipedia, Brightside was acquired by Dolby Systems in 2007, and is no longer operating under the name Brightside. Their website no longer exists.]

Pixim Digital Pixel System (DPS) cameras use a unique chipset consisting of a CMOS image sensor and a companion image processor at each photosite. The key innovation in the sensor is the digital pixel read-out technology. Analog-to-digital conversion is achieved directly at the photosite by using a scheme that processes the conversion immediately at the moment of light capture [Pixim 1999]. Because the readout is all digital, and it is performed in parallel for the entire array, very high frame rates are possible. The Pixim's first 0.18 μm implementation of this DPS design included 4 MB of embedded memory, allowing multiple reads to be buffered on chip (up tp 370k bitplanes/sec). For video generation, high-speed LVDS (low-voltage differential signaling) processing moved the data off-chip during the vertical blanking interval. The DPS sensor captures high dynamic range video by sampling each frame multiple times for each global reset, and saving pixel data before saturation at the highest SNR. The Pixim D2500 sensor achieves high sensitivity partially because of the high fill factor (50% in a 7 × 7 μm pixel) and the decorrelated nature of the noise. Fixed pattern noise is the dominant noise, and the companion chip performs a two-point noise correction for each pixel, plus dark signal non-uniformity (DSNU) scaling and subtraction.

The companion image processor generates standard definition video (NTSC or PAL) by decoding the HDR sensor data, converting it to a linear base, and processing it through the HDR color image pipeline. Automatic control of white balance and exposure is achieved by an algorithm running on an embedded advanced RISC machine (ARM) processor. Unlike most conventional capture systems, exposure control is performed by changing the rendering of the HDR data, and not by changing the capture. The DPS chipset is capable of up to 17 bits (102 dB) of dynamic range at video rates, which is the practical limit of most 1/3 inch optics.

Pixim Eclipse cameras achieve complete ambient light rejection by subtracting successive frames taken with and without synchronized illumination (provided typically by LEDs). Because the frames are computed directly on the Pixim DPS sensor, the two exposures can be made short (global shutter) and can be read out with a very short interval because of the in-pixel analog-to-digital converter and on-chip memory. The typical exposure time is 500 μsec, and the exposure interval is 50 μsec.

6.3.3 Logarithmic Sensing

Sensors with a logarithmic response [359], which roughly approximates human perception, have also been proposed to increase dynamic range. Linear-response CCD and CMOS sensors integrate the charge produced by photon absorption over an interval of time, which results in images characterized by about three decades of illumination. In contrast, logarithmic sensors continuously convert incident photons into a voltage that is proportional to the logarithm of the light intensity, which results in images characterized by over five decades of illumination. The expanded dynamic range is achieved by exploiting the logarithmic currentvoltage relationship of a MOS transistor operating in the weak inversion region. This design creates signal compression at the pixel level, but it loses the benefit of integrating the small photocurrent over time. Unfavorable consequences include low voltage swing (0.2–0.3 V), poor resolution, low signal-to-noise ratio, and a difficult back-end design for the analog-to-digital converter. In effect, logarithmic sensors increase dynamic range as they sacrifice resolution, because a large input range is encoded into a limited voltage swing.

Using fewer bits than required by Webers law could result in a failure to capture perceptible detail, especially at dim illuminances. A good solution is to encode illuminances on a logarithmic scale so that a fractional threshold becomes a constant threshold, suitable for uniform quantization over a high dynamic range. On a logarithmic scale, capturing five decades of illuminance with 1% accuracy requires only ten bits of quantization. (Note that, normally, the dynamic range of a linear pixel is taken to be the ratio of the maximum voltage signal to the noise voltage level, whereas the dynamic range of a logarithmic pixel is taken to be the ratio of the maximum to minimum current signal in between which the pixel circuit maintains a logarithmic current-to-voltage relationship.) The disadvantage of logarithmic sensors is that this preserves details in low intensity regions, but washes out details in high intensity regions because of quantization.

[NOTE IN MS: This paragraph originally came from www.ee.ust.hk/~eebermak/papers/logPixel.pdf.]

6.3.4 Gradient Camera

Tumblin, Agrawal and Raskar proposed a gradient camera that measures the difference in intensities between neighboring pixels rather than absolute intensities at each pixel [**?**]. The concept is similar to differential encoding used in signal compression, where fewer bits are required to represent difference values than are required to represent absolute values. For the gradient camera, the difference is measured even before it is converted into quantized values. By quantizing the sensed intensity differences between adjacent pixel values, an ordinary analog-to-digital converter can be used to measure detailed high dynamic range scenes.

Two solutions have been proposed; one is based on gradients (i.e., forward differences) of intensities and the other is based on gradients of intensities measured by a logarithmic sensor. In images of natural scenes, the distribution of the gradients has a strong peak at zero. Thus most of the gradients in natural images are low valued. By measuring these differences and summing them up (in 1D) we can measure the entire dynamic range. In 2D, we perform a 2D integration of the estimated gradients by solving a Poisson equation.

By taking differences, the quantization levels are optimally used. A disadvantage of logarithmic sensors (or linear sensors with few quantization bits) is that fine details (corresponding to small intensity variations) in high intensity regions are washed out by large quantization steps. In a gradient camera the zero of the analog-to-digital converter is tied only to the intensity of the neighboring pixel. Thus we don't have to increase the quantization step to measure a high dynamic range signal. By mimicking the differential nature of the human retina, a gradient camera can accurately capture fine details in both high and low intensity regions.

To compute difference values and produce the best exposure the gradient camera must take two photos. One photo is for differences in the vertical direction and the other photo is for differences in the horizontal direction. This exposure dilemma is solved by creating alternating mphcliques, or small groups, of sensors that locally determine their own best exposure, and then reconstructing the image by using a

Poisson solver.

6.4 Resolution

Engineers continually strive to increase the resolution of digital sensors. For example, the field of superresolution is a collection of processing techniques that improve image resolution by combining multiple low-resolution images to recover higher spatial frequency components that would otherwise be lost to undersampling. Some of these techniques estimate the relative motion between the camera and the scene, and then register all images to a reference frame, after which images are fused by interleaving filtered pixels to obtain a high-resolution image. Keren et al. [212] and Vandewalle et al. [411] used randomized or jittered sensor positions that were estimated by using sub-pixel image registration (one commercial example is Sinar). Komatsu et al. [219] integrated images taken by multiple cameras with different pixel apertures to get a high-resolution image. Joshi et al. [209] merge images taken at different zoom levels, and Rajan et al. [327] investigate the use of blur, shading, and defocus for achieving superresolution of an image and its depth map.

Most authors also applied modest forms of deconvolution to boost the images high spatial frequency components, which were reduced by the box filter convolution induced by square pixels. Park et al. [314] and the book by Chaudhuri [80] provide a unified survey and explanation of current superresolution methods.

Recent superresolution research has raised significant doubts about the usability of reconstruction-base superresolution algorithms (RBA) [335] in the real world. Baker and Kanade [52] showed that the condition number of the linear system and the volume of solutions both grow quickly as the magnification factor increases incrementally. Lin and Shum [246] provided a comprehensive analysis of RBA and showed that the effective magnification factor can be at most 5.7.

An approach that overcomes some of these limitations is to design a sensor that uses non-square photosites rather than square photosites. Some implementations of this technique include Fuji Finepix Super CCD cameras [144] and Penrose tiles by Ben-Ezra. Penrose tiling is an aperiodic tiling of the plane, first presented by Roger Penrose in 1973. Rhombus Penrose tiling consists of two rhombuses, each placed at five different orientations by specific rules. The ratio of the number of thick-to-thin rhombuses is the Golden Number $(1 + \sqrt{5})/2$, which is also the ratio of their area. Unlike regular tiling, Penrose tiling has no translational symmetry; it never repeats itself exactly. For superresolution it is theoretically possible to integrate and sample the infinite Penrose-tiled plane indefinitely without repeating the same pixel structure. In practice, this design allows the capture of a significantly larger number of different images than is possible with a regular grid. Moreover, all images can be optimally displaced approximately half a pixel apart and still be different. In contrast, a regular square tiling forces the maximal delta between different displacements in x and y to be at most M, where M is the linear magnification factor. The rhombus Penrose tiling is a good candidate for hardware color sensor realization because it is 3-colorable (for an RGB color space) and has simple tiles. This is the primary reason we selected this particular aperiodic tiling. [****NOTE: Who is “we” here? And what were the rhombus Penrose tiles selected

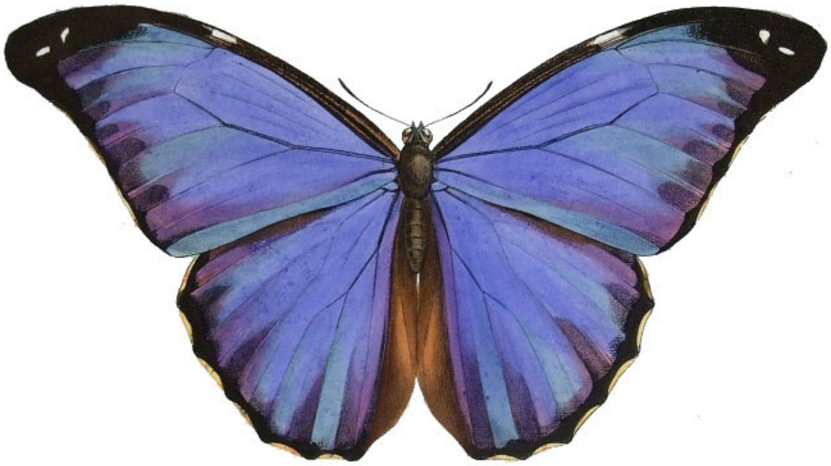


Figure 6.1: The deep iridescent cyan colors of the morpho butterfly are visible to the human eye but difficult to capture with a digital sensor.

for? This last sentence needs clarification.]

6.5 Color Sensing

We start by summarizing how color is determined in the human visual system. Then well discuss three techniques of digital sensing: Bayer mosaic, Foveon, and 3ccd, along with an earlier scheme of sequential color. Well look at some of the limitations of these digital sensing techniques, and then explore some computational approaches that are seen mainly in satellite line scans and tomography-based systems.

Rods in the retina are responsible for human vision; they respond to color by using three pigmentsat red, green, and blue wavelengthsas the basis wavelengths for photopic (color) vision. How do modern digital sensors compare with the human retina? Digital sensors mimic the 3-rod design by creating red, green, and blue sensor photosites rather than sensors that measure a full spectrum of colors. This approach is sufficient to capture a majority of typical images in photography, but it cannot reproduce all the colors perceived by human vision. Because of a mismatch between wavelength profiles of rods in the retina and corresponding red, green, and blue sensor photosites, many vibrant colors are difficult to capture accurately. A good example is a morpho butterfly with deep iridescent cyans, as shown in Figure 6.1.

The photosites on an image sensor capture brightness, not color. Digital camera sensors typically use a color mosaic pattern, or Bayer pattern, of red, green, and blue filters to sense three separate spectral bands, which then form the basis for color reproduction. So-called *demosaicing* methods, which are widely varied and often proprietary, convert raw interleaved color sensor values from the Bayer mosaic grid into red, green, and blue estimates for each image pixel, while including as many luminance details and as few chrominance artifacts as possible. Unfortunately, the demosaicing process is not perfect; its limitations force designers to make tradeoffs that affect image quality. The search for improvements in the demosaicing process is an ongoing area of design and innovation.

Sony's four-color CCD adds emerald pixels that correct defects in the rendition of red tones at certain frequencies. The novel Foveon sensor, used in some Sigma digital cameras, avoids the Bayer mosaic pattern entirely. It stacks three layers of photo detectors in a figurative sandwich, with each layer above, below, or in between the others. This design, which detects wavelength bands for color according to photon penetration depths, eliminates the potential errors and artifacts of the demosaicing process, and substantially reduces post-processing requirements.

Current color sensors are not only limited in how they represent visible wavelength profiles, such as the cyan in the morpho butterfly or certain ocean colors. They also cannot capture light in ultraviolet and infrared frequencies. This underrepresentation of color values causes metamers, color bleeding, and color quantization contouring artifacts.

6.6 Three-Dimensional Range Measurements

The goal of range-sensing cameras is to estimate the depth, or distance from the camera, of a scene point at each pixel in an image. Several companies offer 3D cameras that determine these values. All range-sensing methods used in 3D cameras can be grouped into either of two fundamental techniques: time of flight and triangulation.

6.6.1 Time-of-Flight Techniques

Systems by Canesta and Zcam precisely measure the time of flight required for modulated infrared illumination to leave the camera, reflect from the scene, and return to fast camera sensors. Several earlier laser-based time-of-flight systems (e.g., Cyberware) used flying spot scanning to estimate depth sequentially. Without scanning, newer systems apply incoherent light (e.g., from infrared LEDs) and electronic gating to build whole-frame depth estimates at video rates. Systems from Canesta include the emitters in the same chip substrate as the detector, enabling a compact single-chip sensor unit. The Zcam device is an augmented professional television camera unit that provides real-time depth keying and 3D reprojection. This approach is similar to how radar works in RF and ultrasound works in the audio domain. Unfortunately, the active illumination used in time-of-flight techniques is significantly limited. It will not work outdoors, it could impact the scene [****NOTE: you could explain how it impacts the scene], and it depends on scene reflectance. Specular or mirror-like surfaces may not reflect any illumination at all (a principle used in designing stealth aircraft). In addition, because of distance-albedo ambiguity [****NOTE: explain this further], these illumination techniques require a normalizing second image to nullify the effect of varying albedo. [****NOTE from Dan Raviv for this paragraph: "add resolution reframe rate."

6.6.2 Triangulation Techniques

When our eyes observe a scene point in 3D space, our human vision system uses triangulation to process the parallax and estimate the distance to the point. This distance is defined as the difference between the image coordinate of the scene point in the left eye and the image coordinate of the scene point in the right eye. Triangulation-based range sensors use the same technique. If we photograph a scene with two cameras and back-project the light rays from the two pixels that correspond to the same 3D scene point, the distance to the point can be estimated by triangulation (i.e., by creating a triangle formed by the centers of the two cameras and the two light rays to the scene point).

Unfortunately, triangulation is not an easy problem to solve. The challenge is finding the corresponding scene points in a pair of images captured from two different points of view. Matching a pixel in one image to a pixel in another image is quite challenging. This requirement leads to the need to solve the well-known correspondence problem. Given a pair of images of the same scene, captured from two different points of view, the correspondence problem requires we find a set of points in one image that are identical to points in the second image. The projection of an arbitrary 3D scene point in the two images is defined by the pair of corresponding pixels. In turn, given the pair of corresponding pixels in the two images, we can compute the 3D depth of that point by triangulation.

Depending on the content of a scene, the correspondence problem can be difficult, time consuming, or impossible. For example, matches are missed when a scene point in one of the images is partially occluded, or when regions with no texture or repeated texture cannot be matched easily. Some of these problems can be overcome by using active illumination, or by projecting a random texture to improve the likelihood of finding a unique match. Structured-light techniques get around the correspondence problem by projecting a set of coded light patterns and imaging the illuminated scene from one or more cameras. Other depth estimation procedures involve image phase estimation (which is also used in autofocus), depth from defocus or focus, and light-fieldbased depth recovery.

6.6.3 The Importance of Accurate Range Measurements

How will depth recovery impact computational photography? If ordinary cameras could use triangulation or time-of-flight techniques, the shape of the scene would be easy to determine. Most current depth-estimation schemes are cumbersome, however, which is why these techniques havent been implemented in consumer photography products. Even a 3D camera that only approximates depth at each pixel would revolutionize computational photography because it would allow the insertion of objects with virtual geometry. Such a 3D camera would be popular in video gaming and interaction, and would lead to better estimates for regions of focus. With a measure of depth we could also compensate for the fall-off in flash intensity as the square of the distance by simply amplifying pixel intensities proportionally to the square of the distance. Finally, with more precise depth estimation, we could take multiple 3D photographs and merge multiple estimates to create a 3D model.

6.7 Encoding Identifiable Information

Determining the identity of objects in an image is useful in creating labeled photos. The Sony ID CAM camera system [****NOTE: I can find no information online about 'Sony ID' cameras, as originally written, but I did find a link to the ID CAM] consists of beacons, which are high-speed blinking light sources, and an ID camera that decodes the blink pattern of the beacons by analyzing all the pixels [261]. An ID camera is called a smart camera because it decodes beacon IDs and also captures a scene image like an ordinary camera. Beacons blink at high speed and transmit IDs as a packet via a blink pattern (Figure XXX) [****NOTE: the MS states this is "Figure 2," but no such figure is available]. The ID camera captures the image with an array-like image sensor operating at a higher speed than the blink frequency of the beacons for every frame, and then outputs the beacon IDs in each frame. Because all the photosites of an image sensor can independently decode the blink pattern, even if two or more beacons are blinking simultaneously, the ID camera can recognize every blink pattern. Accordingly, any beacon can blink asynchronously with an ID camera.

The ID CAM contains an EVIS chip, developed by Sony and the Sony Kihara Research Center, which functions as a fast CMOS image sensor [379]. The chip can quickly detect a change in brightness in both visible and infrared light. With a resolution of 192 by 124, it can detect a change in the brightness of individual pixels as well as capture scenes like an ordinary camera. Each pixel of the EVIS chip contains a photo diode, four analog memories, and a simple comparator. A pixel detects a change in brightness by comparing analog memories that save the output of the photo diode without an analog-to-digital converter. The pixel then outputs the binary result of the comparison as HI/LO data. This function makes it possible to detect weak differences in light faster than with ordinary image sensors that use an analog-to-digital converter.

6.8 Handling Camera and Object Motion

An ongoing problem in photography is how a camera handles the relative motion between the camera and objects in a scene. The most traditional solution is to use a very short exposure interval, which freezes the motion, but which in many circumstances results in an underexposed and noisy image. Other solutions include reducing the relative motion between an object and the camera, either by image stabilization in camera hardware or by decoding the blurring image in software.

6.8.1 Line-Scan Cameras

High-speed narrow-view or line-scan cameras, designed exclusively for critically timed sporting events such as track meets and horse races, offer more opportunities for capturing accurate visual appearance. The FinishLynx camera from Lynx System Developers, Inc., is an example. It views the finish line of a race through a narrow vertical slit, and assembles an image whose horizontal axis measures time instead of position (see Figure 6.2). Despite occasionally strange distortions of the

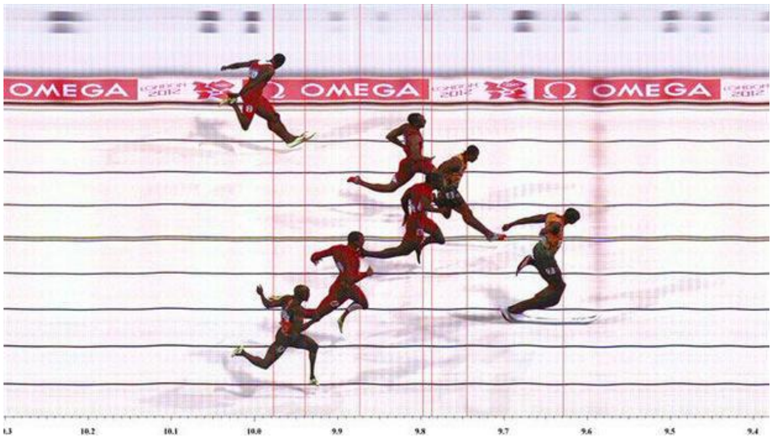


Figure 6.2: The line scan camera takes an image whose horizontal axis measures time instead of position.

racers, the camera reliably depicts the first body part to cross the finish line as the right-most feature in the time-space image.

6.8.2 Image Stabilization

A hand-held camera in unsteady hands is subject to camera shake, which leads to blurry photographs, especially at long focal lengths. Modern cameras can compensate for camera shake by including optical-mechanical systems that correct for camera motion and reduce image blur. Note that these mechanisms are not the same as steady-cam video systems, for example, which stabilize the camera body so that successive frames are taken on a smooth trajectory. These steady-cam systems do not prevent motion within an individual frame. For any camera that is subject to shake during an exposure interval, optical-mechanical stabilization can be achieved by two methods: displacement of the optical elements in the lens or displacement of the sensor in the camera.

Optical image stabilization in the lens works by using electromagnetic sensors to move a floating lens element orthogonally to the optical axis of the lens. Two piezoelectric angular velocity sensors, or gyroscopic sensors, detect horizontal camera movement and vertical camera movement, respectively. Note that this kind of image stabilization, implemented primarily in Nikon and Canon lenses, corrects only for pitch and yaw axis rotations, and not for rotation around the optical axis. Some lenses have a secondary mode that counteracts only vertical camera movement, which is particularly useful when a photographer pans the camera. How this secondary mode is activated depends on the lens design; sometimes it is done automatically and sometimes it is done manually by a switch on the lens. Many of Nikon's recent vibration reduction (VR) lenses have an active mode that is designed for shooting from moving vehicles. This mode is designed to correct for larger degrees of camera movement than the normal mode of VR operation. Unfortunately, using the active mode in normal shooting conditions often results in poorer image quality than using the normal mode of VR operation.

Most camera manufacturers recommend that photographers should turn off image stabilization when a lens is mounted on a tripod, primarily because stabilization is erratic and unnecessary when a camera platform is fixed. Some image stabilization lenses (such as Canon's IS lenses) can measure extremely low vibration readings and automatically detect when the camera is mounted on a tripod. The lens then disables image stabilization and prevents a reduction in image quality. In general, image stabilization is an important feature, but it isn't perfect. The addition of two or more gyroscopic sensors to a lens obviously increases the cost of the lens. Also, light passing through a floating lens element shifts from its true optical path when it projects onto the sensor, resulting in poor Bokeh. This visual look of out-of-focus areas in an image is a subjective and subtle experience that is strongly sought by professional photographers, and more difficult to achieve with image-stabilized lenses.

Optical image stabilization in the image sensor, rather than the lens, avoids some of the problems of lens-based stabilization. In this technique, the sensor's physical position in the camera, not the path of the light, is moved in order to counteract camera motion. This technology is called mechanical image stabilization. As a camera moves, gyroscopes encode movement information to an actuator that moves the sensor and maintains the proper projection of the image onto the image plane. This stabilization method is implemented differently by different manufacturers. Konica Minolta uses a technique called "anti-shake," which is now marketed as SteadyShot in some Sony cameras and *shake reduction* (SR) in several Pentax cameras. This technique uses a precise angular-rate sensor to detect camera motion. Olympus introduced mechanical image stabilization with a system called Supersonic Wave Drive in their E-510 digital SLR body. Other manufacturers use digital signal processors to analyze the image on the fly, during camera motion, and then move the sensor appropriately.

The primary advantage of sensor stabilization is that the image is always stabilized, regardless of the choice of lens. This allows stabilization to work with any lens a photographer chooses, including older lenses, manual lenses and lighter lower-cost lenses. One disadvantage of sensor stabilization is that the image projected to the viewfinder is not stabilized. (Cameras with electronic viewfinders do not have this problem because the image projected on the viewfinder is taken directly from the image sensor.) A second disadvantage of sensor stabilization is that the imaging sensor is moved but the autofocus sensor is not moved. Camera shake can therefore lead to lower performance of the autofocus system in low light. Note that this problem occurs only with digital SLR cameras that have a dedicated phase-detection autofocus sensor. It is not a problem with smaller cameras that use the main sensor itself for contrast-detection autofocus. Note also that sensor stabilization does not function in digital SLR cameras that can record video because the sensor must lock in place during video recording. Lens-based stabilization systems don't have this limitation, and they can function in all imaging modes.

(The Wikipedia site en.wikipedia.org/wiki/Image_stabilization is the original source text for the four highly modified paragraphs above.)

6.8.3 Hybrid Imaging

Motion blur due to camera movement can significantly degrade the quality of an image. Since camera movements are typically arbitrary, and follow a random path, computational attempts to remove motion blur can be difficult and problematic. Previous methods to correct motion blur have included blind restoration of motion blurred images, optical correction with stabilized lenses, and use of special CMOS sensors that limit the exposure time in the presence of motion. Ben-Ezra et al. exploited the fundamental trade-off between spatial resolution and temporal resolution; they constructed a hybrid camera that can measure its own motion during image integration [57]. The acquired motion information was used to compute a point-spread function (PSF) that represents the path of the camera during integration. This PSF was then used to deblur the images made by using long exposure and complex camera motion paths in several indoor and outdoor scenes.

A hybrid imaging system proposed by the author [****NOTE: which author? Ben-Ezra? Nayar? Raskar? Tumblin?] consists of a high-resolution primary detector and a low-resolution secondary detector. The high-resolution detector records the image information while the secondary detector computes the motion information and the PSF. The motion between successive frames is limited to a global rigid transformation model, which is computed by using a multi-resolution iterative algorithm that minimizes the optical-flow-based error function. The Richardson-Lucy algorithm is then used to process the resulting continuous PSF to remove motion blur. The authors used a 3-megapixel Nikon digital camera as the primary detector and a Sony digital video camcoder as the secondary detector. The two detectors were calibrated offline. Deblurred results were demonstrated on several real sequences with exposure times ranging from 0.5 seconds to 4 seconds, and with blur ranging up to 130 pixels.

Recently, Fergus et al. have shown that, in case of camera shake, the point spread function can be estimated from a single image [132]. They exploit the natural image statistics on image gradients and then use the probability blur function [****NOTE: do you mean probability density function?] to deblur the image.

Blur due to camera shake is different from blur due to object motion. While camera shake can be estimated by on-board accelerometers, it is difficult to estimate object motion from a single frame. Let us study some recent approaches for linear object motion.

6.8.4 Coded Exposure via Fluttered Shutter

In a conventional single-exposure photograph, the movement of objects in the scene or the movement of the camera by the photographer can both contribute to motion blur in an image. The exposure time interval defines a temporal box filter that smears the moving object across the image by convolution. This box filter destroys important high-frequency spatial details so that deblurring via deconvolution becomes an ill-posed problem. Raskar et al. have proposed to flutter the cameras shutter open and closed with a binary pseudo-random sequence during the chosen exposure time interval, instead of leaving the shutter open for the entire interval, as in a traditional camera [331]. The flutter changes the box filter to a broadband filter that preserves high-frequency spatial details in the blurred image, and the corresponding deconvolution becomes a well-posed problem.

Results were presented for several challenging cases of motion-blur removal, including extremely large object motions in outdoor scenes, with textured backgrounds and with partial occluders. The authors assume that the PSF is given or can be obtained by simple user interaction. Since changing the integration time of conventional CCD cameras is not feasible, an external ferro-electric shutter is placed in front of the lens to code the exposure. The shutter is driven alternately opaque and transparent according to the binary signals generated from PIC [define this acronym] using the pseudo-random binary sequence. The code is chosen to minimize the deconvolution noise, assuming a specific amount of motion blur in the image. Unfortunately, a coded exposure also reduces the light entering the camera. The chosen code was 50% on/off, so half the light was lost, compared to a traditional camera with the same exposure time.

6.8.5 Motion Invariant Photography via Sensor Motion

In motion invariant photography (MIP) as presented by Levin et al., the motion blur (or motion PSF) is invariant to object speed within a certain range [235]. Thus objects moving with different speeds within that range would result in the same motion PSF. To deal with motion in a known direction, you must move the camera with a constant acceleration during a single image exposure. Knowing the direction of the object motion is important, since the camera should be moved accordingly, but knowing the magnitude of the object motion is not required. An obvious disadvantage of MIP is that the static parts of the scene are also blurred during capture, which leads to deconvolution noise on those scene parts.

If the direction of object motion is exactly known, and the motion magnitude is unknown within a range, MIP should be the solution used for capture. However, if the direction of object motion is unknown, then coded exposure is the optimal choice. In addition, as the object speed and direction differ from assumed values, performance degrades slowly for coded exposure but sharply for MIP. For MIP, estimation of the PSF is not required for static scene parts, but they are also blurred due to camera motion, leading to degradation in SNR [41].

Sensor motion is useful for some other applications. Instead of x-y motion in the sensor plane, motion in the direction of the optical axis during a single exposure produces a rubber-focus [****NOTE: describe this effect more clearly, and link the concept of optical axis motion to your next sentence]. Wavefront coding [75] modifies the defocus blur to become depth-independent by using a cubic phase plate with lens, while Nagahara et al. [284] move the sensor in the lateral direction during image capture to achieve the same [modification?]. [****NOTE: slim paragraph could use some expansion]

6.8.6 Performance Capture via Markers

Motion capture is an increasingly important component in film and television special effects, as well as in the development of accurate motion-based user interfaces and the analysis of body movement for injury rehabilitation. Optical sys-

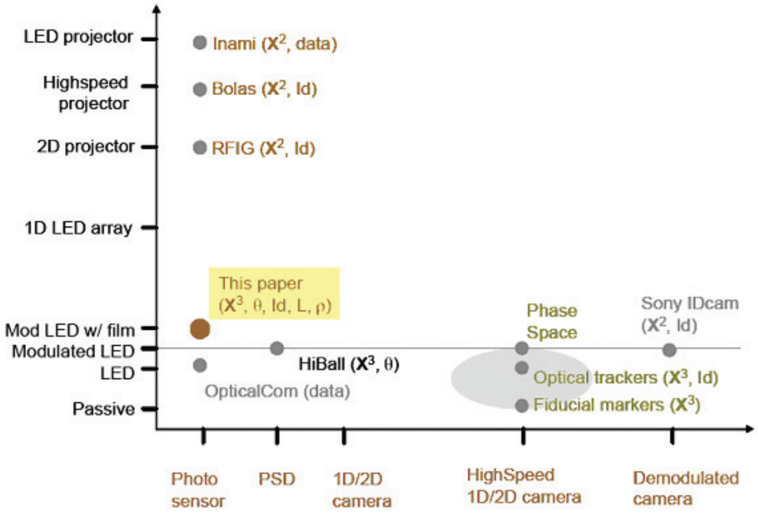


Figure 6.3: Communication choices between light sources and receivers, plotted in terms of complexity of receivers (x-axis) versus complexity of transmitters (y-axis).

tems are typically more effective at motion capture than magnetic, acoustic or mechanical systems because optical systems experience lower latencies and provide greater accuracy and precision in recording motion. Motion capture systems used in movie studios commonly employ high-speed cameras to observe passive visible markers or active light emitting diode (LED) markers. These expensive camera-based systems have special high-bandwidth sensors, and they require a controlled environment to maintain high contrast between the marker and its background. [343, 415, 24, 30, 25, 309, 79]

For example, the Vicon MX13 camera can record 1280 × 1024 full-frame grayscale pixels at frame-rate speeds up to 484 frames per second, with onboard processing to detect the marker positions. These cameras, which have been developed over the last three decades, provide highly reliable output data for special effects. Expensive high-speed motion-capture cameras, however, dont always scale easily as demand for more precise data increases. Bandwidth restrictions limit image resolution as well as frame rate. Higher frame rates (i.e., shorter exposure times) require either brighter controlled scene lighting for passive markers or the use of power-hungry active LED markers. To segment the markers from the background as robustly as possible, these systems also use methods for increasing marker contrast, which usually requires an actor to wear dark clothing and perform under carefully controlled lighting.

Update rate is affected by camera frame rate. Active beacons must use time division multiplexing [25] so that only one LED can be turned on at a time. Each additional tag requires a new time slot. [****NOTE: Update rate should be clarified here. Also, are beacons and tags identical to active markers? If so, the terminology should be consistent. If not, the new terms should be defined.] Hence, the total update rate is inversely proportional to number of tags. Such systems are ideal for applications requiring high-speed tracking of a small number of tags, e.g. head or hand tracking in virtual reality systems. Passive markers need to resolve correspondence to avoid the marker swapping problem. [****NOTE: slim paragraph could use some expansion. Where does callout for Figure fit in?]

Chapter 7

Processing and Reconstruction

The previous chapters focused primarily on system-level solutions, combinations of new optical hardware and novel algorithms that together help us overcome some of the limitations of traditional photography. As our goals are more ambitious than simply taking better pictures, we can broaden the processing steps to incorporate more standard techniques and recent advances in image processing and computer vision. For example, as "cheaper, faster, and better" digital cameras and storage let even the most casual photographer gather abundant collections quickly, novice and professional users alike must confront the daunting task of organizing and browsing massive personal and on-line photo collections. Current methods are quite varied and disjoint; most people patch together their own haphazard mix of methods and competing software packages, and resort to some combination of building hierarchies of directories, manually renaming each photo, tagging individual photos or each group of photos with thoughtful metadata, uploading, downloading to online repositories, and using images as raw materials for website-building tools. Can't we find a better, more forward-looking approach? How might we adapt existing and new image-processing methods to computational photography tasks?

Several image processing and computer vision techniques already provide new ways to interact with existing photos, such as automatically generated suggestions for metadata. Some recent 'smarter' camera designs adopt fast image filtering, recognition and metering methods to adjust their settings automatically to match the scene content, detect smiles, and make sensible low-level decisions to let photographers devote more of their attention to a photo taken in less time. Advances in geometric operations on large sets of photos now allow even novices to explore their image content in 3D. In such a data-rich environment, in which millions of photos of any object can be archived and retrieved at little cost, automatic processing must become a central feature of computational photography.

7.1 Filtering and Detection

The traditional field of image processing is primarily concerned with image-based object detection and recognition. In part because classic film-like cameras duplicate the lens-and-sheet-like sensor arrangement of the human eye, researchers have been inspired to further replicate the performance of the human visual system, starting from simple low-level image acquisition all the way through high-level understanding of the scene. For example, human's quick and robust text-reading abilities have long inspired optical character recognition (OCR) systems in image processing. These systems optically scan or photograph printed characters to capture their image, and then convert the printed characters from noisy bitmaps to a machine-readable format, typically in the form of an ASCII character sequence.

Despite decades of active study, image-based object recognition also remains a challenging area of computer vision research, due in part to the visual complexity of the physical world. In practice, changes in illumination, shading, occlusion, and inter-reflections as well as the position and pose of objects in a world are difficult to detect and assess reliably. As a result, systems that perform even the most basic visually guided tasks by using a camera to control a robotic arm, remain tailored to niche applications, where the visual complexity of the environment is kept small and manageable (e.g., under controlled lighting and known viewing conditions).

All of these new and previous software-based methods can be seen as maximizing the information content that can be achieved with a given imaging system. With additional information supplied from novel devices for optics, illumination, sensors, and with active communication between these devices, what additional capabilities are possible? Our exploration is organized into (1) filtering to reduce the impact of noise, (2) detection to localize desirable or important image features, and (3) recognition of these features to categorize images with discrete sets of higher-level labels.

7.1.1 Detection and Recognition

Image-based methods for object detection and recognition are too numerous and varied for a comprehensive survey here, but a few promising methods recently appeared in innovative digital camera designs, including face detection system in both the Sony and Fujifilm point-and-shoot camera lines (see Figure 7.1). As amateur photographers sometimes have difficulty choosing the best aperture, exposure, and focus settings, many cameras now include pre-set modes for portraits, night-time, sports, back-lit or other scenes where well-chosen settings make good pictures more likely. Detecting human faces in the scene allows such cameras to make better choices for settings to ensure everyone's face is properly focused, color-balanced, and well exposed in the final photograph. Going even further, Sony recently introduced their "smile shutter" mode that defers triggering the shutter until the system determines that everyone is smiling. As the quality of results depends on the quality of the underlying computer vision algorithms, we must note that face detection is a notoriously difficult problem, in which even lighting changes, such as illumination from multiple light sources, can confuse many existing methods. While the current generation of camera-based face detection has difficulty detecting faces

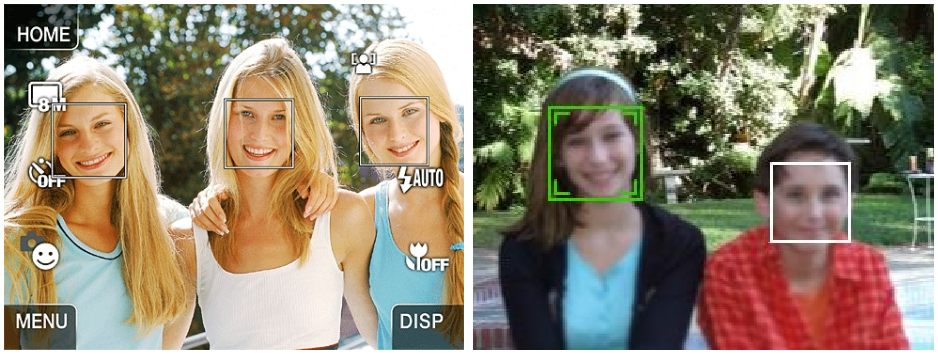


Figure 7.1: Automatic face detection in consumer digital cameras. (Left): Sony's smile shutter feature (from their Cyber-shot DSC-T200) delays the shutter release until the main subject smiles. (Right): FujiFilm FinePix S600d camera's Face Detection Technology, identifies up to 10 faces to prioritize its focus and exposure settings. (sources: tinyurl.com/4sl8xc and facedetection.fujifilmusa.com/)

in profile or when partially occluded, and cannot currently distinguish a printed 2D face from a live human subject, the results are still quite suitable and helpful for casual photography of friends and family. Like any 'automatic assist' function, smile detection might not always be appropriate, as sometimes an enigmatic expression is more desirable than a smile; a broad smile would not improve Leonardo da Vinci's work on the Mona Lisa.

As nearly every mobile phone now includes both a small built-in digital camera and growing computation and storage, both users and manufacturers are adapting them to new everyday uses. Several vendors have begun offering automatic mobile translation software; for example, Linguatec's "Shoot & Translate" software combines four key technologies: optical character recognition, automatic translation, voice output, and a dictionary [247]. When confronted with a sign in an unfamiliar language, users just take a picture of it: the phone then extracts the text via OCR, identifies the language, performs simple machine-translation to the user's language, displays the translated text, and even reads aloud the translation using text-to-speech conversion. Already included in several Nokia phones and slated for release for others, this and related translation features from Moka LLC, Ectaco Inc., and Transclick Inc. (see en.wikipedia.org/wiki/Mobile_translation) may aid greatly anyone visiting an unfamiliar city or cultural region. As with face detection software, these systems apply decades-old computer vision research, but adapted to new uses and very lightweight platforms. These photo-assisted mobile translation systems can trace their origins to similar text-to-speech systems designed for assisting the visually-impaired, such as the K-NFB device invented by Ray Kurzweil (see news.bbc.co.uk/1/hi/technology/5088464.stm). In these systems, simple compositional transformations are fairly straightforward, but translations that accurately interpret idioms and that provide suitable responses to slang and culturally-dependent idioms remain difficult. While at first these image processing applications to consumer electronics may appear as gimmicks, each address a new problem in a new way and suggests entirely new directions and everyday applications for computing and photography.

7.1.2 Noise Reduction

Digital image sensors in their current form may never entirely escape from noise or the need for noise removal: as Chapter 6 explains, even a perfect noise-free electronic system will still suffer from photon-arrival noise in low-light conditions (from limited light in the scene, limited exposure time, and limited apertures and lens' ability to gather enough light for each pixel). However, if we can assume that the underlying noise processes are spatially-uncorrelated (or at least high in frequency) and the underlying image is slowly-varying, then filtering methods can help us remove or hide that noise without unduly damaging the image content. For example, we might apply a low-pass filter to pixel values in a small neighborhood in an attempt to preserve the original features while 'smoothing away' most of the noise.

However, this method can fail badly at sharply detailed texture edges, and has inspired many ingenious approaches to developing visually suitable edge-preserving filters, including multi-scale, PDE-based, histogram-based, iterative and non-iterative approaches. Inspired by early scale-space work, Perona and Malik's work in anisotropic diffusion [319] found wide application and development of many variants. This method iteratively reduces the differences between each pixel and its four neighbors, but weights those reductions according to a local 'edginess' measure, which they chose as a function of gradient magnitude. High gradients earned weights at or near zero, 'freezing' these edge-like features in place on every iteration, while low gradients gained high weights, smoothing away these small differences. In addition, anisotropic diffusion exhibits 'shock-forming' behaviors that, with enough iterations, smooth images towards piecewise-constant results. Shocks are the self-reinforcing formation of discontinuities, where smoothing on either side of a preserved edge sharpens a previously smooth step-like feature into a discontinuity. Many follow-on papers present improvements or refinements, examine shock-forming and its consequences, accelerate convergence, and identify mathematical links with other PDEs and the bilateral filter. While well understood, this and many other iterative PDE-based methods have limited practical utility for real-time noise reduction.

As presented in Chapter 4, the bilateral filter has emerged as one of the most popular methods for non-iterative edge-preserving image smoothing. This rather simple but very effective idea was invented independently several times, perhaps first by Aurich and Weule in 1995 [48], then again by Smith and Brady [374] as part of their SUSAN framework, then again by Tomasi and Manduchi [22] who gave it its current name.

The bilateral filter can be viewed as an adaptive low-pass filter: it replaces each pixel with a weighted sum of its neighbors, but those weights depend on both the intensity and position of that neighbor. The key idea is to take an ordinary Gaussian filter that assigns weights to neighbors according to their spatial *domain* filtering, and multiply it by a second Gaussian filter, one that filters intensity differences with neighbors. This second, *range* filtering assigns large weights to neighbors with intensities similar to the central pixel, but low weights to neighbors with very different, 'outlier' intensity values.

As defined by Tomasi and Manduchi, a bilateral filter can be applied to an image $\mathbf{f}(\mathbf{x})$ as

$$\mathbf{h}(\mathbf{x}) = k^{-1}(\mathbf{x}) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbf{f}(\xi) c(\xi, \mathbf{x}) s(\mathbf{f}(\xi), \mathbf{f}(\mathbf{x})) d\xi$$

with this normalization to ensure all weights for a pixel sum to 1.0:

$$k(\mathbf{x}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} c(\xi, \mathbf{x}) s(\mathbf{f}(\xi), \mathbf{f}(\mathbf{x})) d\xi,$$

where $c(\xi, \mathbf{x})$ measures the geometric distance between the neighborhood center $\mathbf{x}$ and a nearby point ξ and $s(\mathbf{f}(\xi), \mathbf{f}(\mathbf{x}))$ measures the photometric similarity between points $\mathbf{x}$ and ξ in the image. As shown in Figure 7.2, a typical choice for the similarity functions corresponds to shift-invariant Gaussian filtering, in which

$$c(\xi, \mathbf{x}) = e^{-\frac{1}{2}\left(\frac{|\xi - \mathbf{x}\|}{\sigma_d}\right)^2} \quad \text{and} \quad s(\mathbf{f}(\xi), \mathbf{f}(\mathbf{x})) = e^{-\frac{1}{2}\left(\frac{|\mathbf{f}(\xi) - \mathbf{f}(\mathbf{x})\|}{\sigma_r}\right)^2}.$$

Note that the geometric and photometric standard deviations, given by σ_d and σ_r, control the amount of averaging in the spatial and range dimensions, respectively.

The bilateral filter has proven useful for a wide variety of computational photography applications beyond image filtering, and recent acceleration methods make it suitable for interactive image editing on multi-megapixel images (for more information see the SIGGRAPH course notes on the subject [312]). As noted by Paris et al., applying the bilateral filter independently to individual R, G, B channels can cause unusual visual anomalies; strong but nearly isoluminant edges preserved in one color channel might be smoothed away in another. Bilateral filtering in CIE-Lab color space, or applying cross-bilateral filtering from luminance to each chrominance channel, will keep the same edges sharp in all color channels. While non-iterative and visually appealing for many applications such as tone mapping and multi-scale image-detail editing, the bilateral filter still has notable limitations for noise reduction, as outlined by Buades et al. [70].

Video noise reduction presents a substantially different problem. Similarities between adjacent frames provide more opportunities to distinguish between the photographed scene and the noise that obscures it, but camera motion or movement of objects in the scene makes identification of these frame-to-frame similarities far more difficult. As described in Chapter 3, most video cameras currently capture a low dynamic range (LDR) sequence, consisting of a set of uniformly-exposed digital images. In practice, a high dynamic range (HDR) sequence is required to capture natural scenes. Bennett and McMillan [59] proposed "virtual exposures" to enhance underexposed, low dynamic range videos using adaptive spatial and temporal post-processing. As shown in Figure 7.3, the algorithm begins by estimating the exposure setting using a spatially-uniform tone mapping for each pixel. They then synthesize a corresponding gain ratio by combining uniformly-exposed frames for each pixel via averaging temporal samples of static scene elements and spatial samples of dynamic elements. Such virtual exposures reduce sensor noise and enhance

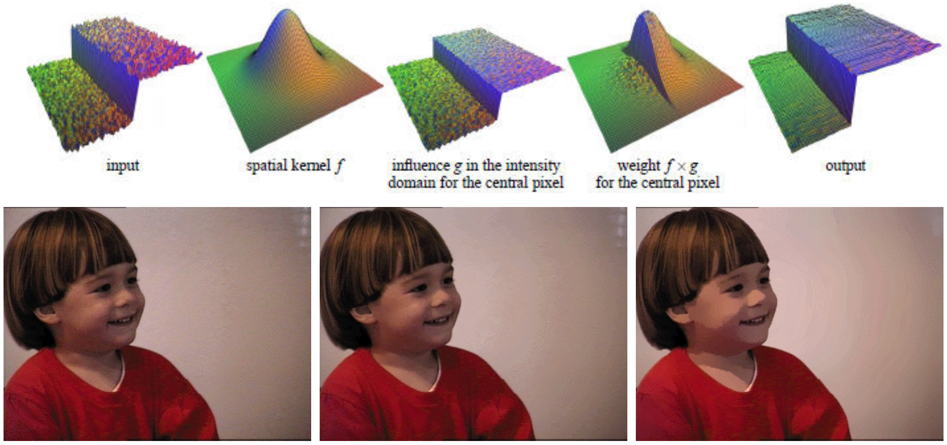

input
spatial kernel f
influence g in the intensity domain for the central pixel
weight $f \times g$ for the central pixel
output

Figure 7.2: Bilateral Filtering [22]. (Top) The bilateral filter is centered on a pixel on the "bright" side of a step edge. The geometric distance function exponentially decreases from the center pixel, whereas the photometric similarity function only includes points on one side of the discontinuity. As a result, the bilateral filter kernel only averages values on the bright side of the edge. (Bottom) From left to right, an input image, the output after one iteration of bilateral filtering, and the output after five iterations. Note that a cartoon-like appearance results when multiple iterations are applied, corresponding to the elimination of shading variation between sharp intensity discontinuities. (sources: tinyurl.com/c8l92a and tinyurl.com/cnxsuy)

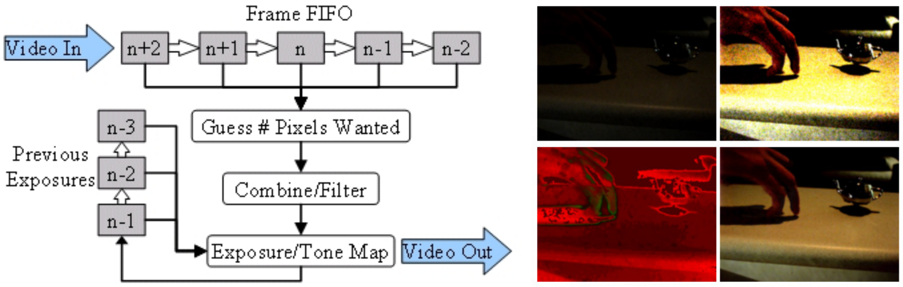


Figure 7.3: Video Enhancement using Per-Pixel Virtual Exposures [59]. (source: ericpbennett.com/VideoEnhancement/index.htm)

previously unseen details in shadowed regions, under the assumption of zero-mean underlying noise processes. The proposed system enhances raw uncompressed video streams off-line and cannot currently operate in real-time, as it requires about one minute of processing time per 640×480 frame. Furthermore, the performance depends on the systems' ability to track objects in the video, and artifacts may result from independently-moving regions that are too complex, occluded or numerous for accurate tracking. This paper's promising results suggest that further work on computational methods for denoising videos could prove fruitful.

7.1.3 Colorization and Color-to-Gray Conversion

"Colorization" or "colorizing" describes any process that adds color to an existing monochromatic image or video, usually by adding hue and saturation without modifying luminance. Adding color has deep historical roots: hand-tinting methods perfected for postcards, advertising, figures and prints in books hundreds of years ago, and included everything from crude color washes to meticulous water-painting after printing. For example, John Jay Audubon's landmark "Birds of America" books, first printed in 1827, include 435 of the huge 'double-elephant folio' (26 x 39 inches) versions. Each page was a printed engraving then water-colored by hand by teams of artists supervised by Audubon himself.

Few manual colorization methods were as meticulous as Audubon's, but many were adapted to even the earliest of photos, including Dageurreotypes, 'tin-types'. Usually applied as a soft 'wash' of color to make cheeks look bright and rosy and skies look blue, these methods showed that even very faint amounts of color greatly enhanced the appeal of the photos. Very broad, gentle blobs of faint color with very blurry transitions of hue and saturation gave pleasant results, as the luminance edges in the photo would suffice to indicate visually important scene boundaries. With the advent of offset printing capable of half-toning and precisely aligned multiple inks, hand-tinting disappeared rapidly in the early 1900s.

Widespread colorization for films began in the 1980s as computer-assisted color video processing, editing, digital film scanning, and digital film-printing became practical. Printing became practical. In the mid-1980s companies such as American Film Technologies in San Diego, CA began computer-assisted colorization for black-

and-white films, including parts of the MGM back-catalogue of films purchased at that time by Ted Turner, who became a strong advocate for film colorizing (though many film critics such as Roger Ebert and many directors such as John Huston protested these modifications: Huston specifically prohibited colorization of his own works).

Methods for computer-assisted colorization were at first very closely-held trade secrets and were quite labor-intensive, requiring manual selections from hue/saturation color palettes, automatic image segmentation with motion tracking whose results were manually verified and corrected frame-by-frame by artistically skilled operators using pen-like user interfaces. Most, but not all early colorization methods used intentionally under-saturated images to help minimize visual distractions, including occasionally jarring color choices, ambiguous segmentation at complex or transparent boundaries (fog, feathers, fur) inconsistencies between scenes, or tracking mistakes where moving colors did not match moving objects exactly. Results improved greatly in the 1990s with the advent of 3D texture-tracking methods and better tool-automation and user-interfaces for finding and correcting mistakes. Some movie colorization projects such as Frank Capra's "It's a Wonderful Life" were quite meticulous, using historically researched color palettes, and incorporated a broad range of film-restoration techniques that accounted for the film's sensitometric response curves, suppressed film noise, removed film scratches, corrected frame-to-frame mis-registrations, and restored contrast lost in aging film stocks.

Computer graphics researchers have been investigating stroke-based, manually-guided and semi-automatic colorization methods since the late 1990s. At present, state-of-the-art colorization uses computer vision algorithms to segment and track individual scene components. Users typically provide coarse color strokes in their target regions to initialize an iterative color-filling process (e.g., see Figure 7.4). Levin et al. [233] use a quadratic cost function to identify neighboring pixels in space-time (i.e., throughout a given video sequence) and automatically propagates users' initial color strokes throughout the space-time volume by solving a well-formulated optimization problem, with object boundaries automatically detected and tracked over the image sequence. In their work, as in other state-of-the-art colorization algorithms, the ability to preserve color boundaries depends on the quality of the space-time segmentation, and the color-strokes act as a form of manual assistance to the method's semi-automatic image segmentation. By casting colorization as an optimization problem, Levin et al. enables this sub-field to benefit from advances in other computer vision problems, including more sophisticated affinity functions for segmentation and faster optimization techniques.

Black-and-white or grayscale printing is still much cheaper than color and usually offers both higher resolution and better shading results. However, many black-and-white renditions of color images do not adequately convey all the visual information of the original, because the color-to-grayscale conversion process still imitates black-and-white photographic film. Like modern 'panchromatic' film, these well-established methods compute an approximation to luminance (spectral power of light weighted by the luminosity curve of Figure 2.13 in Chapter 2), but are not required to do so—they do not share the restrictions of film. Complex chemistry determines the spectral response curves of photographic film during manufacturing, but digital color-to-gray conversions are free to use any method we can devise that

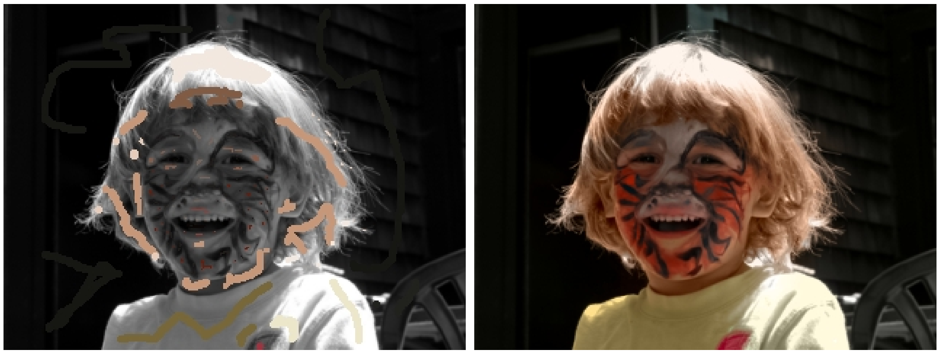


Figure 7.4: Colorization by Optimization [233]. (Left): Users annotate a grayscale with a sparse set of target-color scribbles. (Right): Propagating those user strokes to fill the space-time volume creates a fully colorized image. (source: www.cs.huji.ac.il/~yweiss/Colorization/)

might produce better results.

By far the most common grayscale renditions for printers are made by simple luminance conversions meant to approximate the 'luminosity' spectral response curve of Section 2.2.4). Conversions range from averaging ($L = 1/3(R + G + B)$) to misuse of the now-obsolete NTSC analog television standard: $L = 0.299 * R + 0.587 * G + 0.114B$, or the more suitable gamma-corrected version: $L = 0.3086 * R + 0.6094 * G + 0.0820 * B$ (from www.graficaobscura.com/matrix) or by using only the luminance (L) channel of the CIE La*b* color space. Each of these methods are only projections; they discard two of the three dimensions we have available to describe human-perceivable colors. Not surprisingly, this discarded information often held important visual features and distinctions that were plainly visible in the original color image. For example, in Figure 7.5, the background sky and fading dusk-orange sun are isoluminant; they differ only in hue and saturation, and if we display only luminance information, the sun vanishes.

Pioneering work by Gooch et al. [160] introduced the "Color2Gray" algorithm to reduce such losses by attempting to preserve the salient features of a color image. They present a 3-step process: (1) conversion of RGB inputs to a perceptually uniform CIE-Lab color space, (2) use of chrominance and luminance differences to assign scalar distance values to all their neighboring pixels, using metrics that combine spatial and color-space differences, (3) solving an optimization problem to determine the grayscale values for each pixel whose differences from their neighbors best match the originals in the least-squares sense. (see Figure 7.5).

Their approach works well to preserve visible details on a wide range of challenging images, but can be quite slow to compute due to the large and slow $\mathcal{O}(N^4)$ optimization required, and the method needs a small amount of user-intervention to decide which hues should split the hue-circle into two portions, and which one should map to darker and which to lighter tones. As this and several follow-on methods attempt to map three dimensions of color variations to just one dimension of luminance, it cannot guarantee that all marginally-visible color variations in an

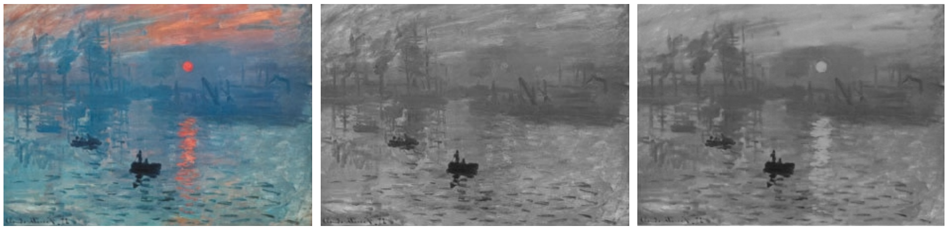


Figure 7.5: Color2Gray [160]. A color image (Left) often reveals important visual details missing from a luminance-only image (Middle). The Color2Gray algorithm (Right) maps visible color changes to grayscale changes. (Image: Impressionist Sunrise by Claude Monet, courtesy of Artcyclopedia.com). (source: tinyurl.com/d4f2df)

image will still be visible in the grayscale rendition (especially for images with a very broad distribution of colors and color differences), but does ensure a reasonably proportional mapping, where the size of the color differences between pixels in any local neighborhood correspond to the size of the luminance pixels in the same neighborhoods in the result. It also guarantees that the result fits the available contrast abilities of the display without over- or under-exposed regions. Contemporaneous work by Rasche et al. [328] addressed the same problem using multi-dimensional scaling (MDS) on a set of 256 colors found by quantizing the source image, mapping these to display luminance, and then mapping scene colors onto them. In addition, they presented a very worthwhile expansion of the problem into visual prosthetics; how can we re-map all the color differences of an ordinary color image to differences that people with color-perception deficiencies (e.g., deuteranopes) could discern reliably? For both tasks their method provided substantial improvements, and also showed promise as an aid for viewers with color-deficiencies. However, the method's performance degrades for images with very broad, uniform color distributions and is insensitive to spatial variations included in the Gooch approach. Both papers inspired a steadily improving sequence of methods, both faster and more perceptually valid; for a good overview see [373].

7.1.4 Motion and Defocus Deblurring

As described in Chapter 2, digital imagery is often blurred due to two factors: (1) motion of the camera or scene objects during the exposure, and (2) defocus of objects located outside the camera's depth of field. Image processing methods have been proposed to sharpen such imagery by compensating for these sources of blur in post-processing. As previously described, the imaging process can be modeled as a linear shift-invariant (LSI) system, in which a lens creates a magnified copy of distant scene planes. Each scene plane is subject to defocus blur which can be modeled by convolution by a spatially-invariant kernel whose width is proportional to the distance from the plane of focus. Similarly, camera and object motion can be modeled as a linear superposition of images, corresponding to the instantaneous frames captured over the exposure time. The cumulative result of motion and de-

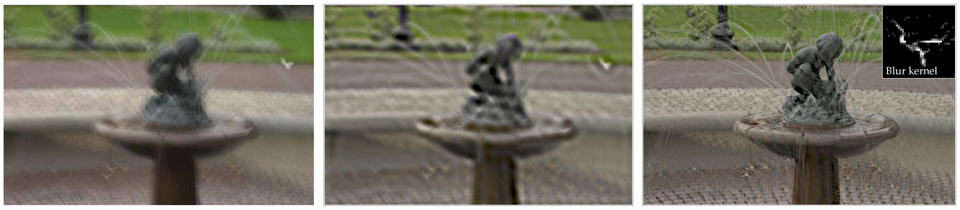


Figure 7.6: Removing camera shake from a single image [132]. (Left) An image degraded by camera shake. (Middle) Result after applying Photoshop's unsharp mask. (Right) Result from the blind deconvolution method of Fergus et al. [132], with the inset showing the recovered PSF. (source: people.csail.mit.edu/fergus/research/deblur.html)

focus blur is to produce a single point spread function (PSF). This PSF defines a depth-dependent convolution kernel that approximates the image blur under the LSI model. As a result, image deblurring involves inverting that process by *deconvolving* the point spread function. Motion and defocus deblurring algorithms can be classified as either blind or non-blind; the 'blind' methods begin with little or no knowledge of the point-spread function, and the non-blind methods begin with accurate *a priori* estimates. The Richardson-Lucy algorithm, perhaps the best-known non-blind deconvolution method, calculates the most likely image given the observed pixel values, the known point spread function and the 'non-negative constraint' that all pixel values are greater than zero. In practice, iterative methods are often used to find a maximum-likelihood deconvolution result. The more challenging blind image deconvolution methods must simultaneously estimate a deblurred image and the point spread function.

Image deconvolution is a well-studied topic in image processing and computer vision, but most approaches assumed that both the PSF and the camera's aperture are unavailable for manipulation to help in the deblurring process. As we saw in Chapter 4, recent coded aperture methods broke these assumptions, and show that simple masks and shutter modulations ensure that the point spread function is invertible—eliminating strong ringing artifacts typically produced from deconvolving PSFs from an unmodified circular aperture. In recent years, several new approaches emerged for deconvolution without masks, methods that exploit image pairs and the statistics of natural imagery.

Fergus et al. [132] addressed the problem of removing motion blur due to camera shake in casual photographs, where they know neither the camera's lens PSF nor its trajectory. They assume the camera shake is pure translation in the plane of the camera's sensor, and describe its effect on the image as convolving the desired unblurred image with a single shift-invariant blur kernel that resembles a scribble—a thin, curved, possibly tangled tracing of the camera's displacement measured against the sensor's image plane. The goal of their blind-deconvolution method is to simultaneously determine the motion-blur kernel and find an estimate of the blur-free image that, when convolved with the blur-kernel estimate, will faithfully recreate the blurred source image.

As they note, this problem is severely under-constrained, with far more un-

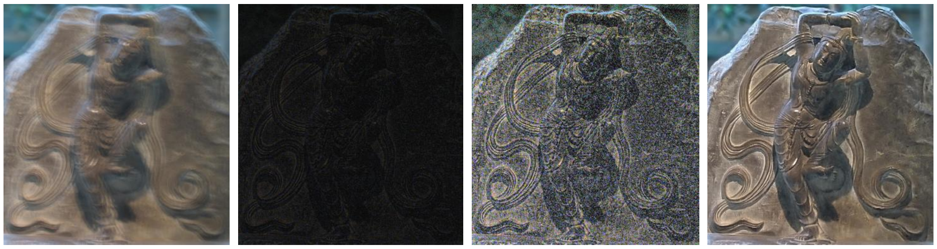


Figure 7.7: Image Deblurring with Blurred/Noisy Image Pairs [438]. (Far left) An image degraded by camera shake (with a 1 second exposure and ISO 100), (Center left) Noisy image (with 1/100 second exposure and ISO 1600). (Center right) Noisy image after levels adjustment and gamma correction. (Far right) Deblurred result from Yuan et al. [438]. (source: research.microsoft.com/en-us/um/people/jiansun/)

knowns (pixels in the blur kernel and pixels in the de-blurred result) than known values (pixels in the given blurred image). To make the problem tractable, they impose new sets of 'image priors'; while previous methods usually imposed constraints in the frequency domain that sometimes admitted strong ringing artifacts in the solutions, they imposed two new priors. First they noted that a histogram of gradients for photographs of natural outdoor scenes as well as many others is sharply peaked at zero and has distinctive statistics; they require that their resulting 'deblurred' images match those statistics. Second, they build on work by Miskin and MacKay [273], applying a Bayesian approach that takes into account uncertainties in the unknowns. (see: people.csail.mit.edu/fergus/papers/deblur_fergus.pdf). While quite elaborate to implement, the results showed substantial improvement over long-standing previous methods, and have inspired follow-on works that further improve and simplify the approach; for a current survey, see the recent work by Levin et al. [230]

Figure 7.6 shows typical deblurring results from Fergus' pioneering efforts. While effective for a broad class of photos, the method does not address camera rotations or out-of-plane motions, but these are small and mostly-uncommon errors for hand-held cameras. In addition, some artifacts of camera shake remain too difficult to reverse by this method, including blur from bright glints and light sources that caused saturated pixels along their trajectories, additional blur from object motion within the photographed scene, and distinguishing between artifacts from camera shake and unrelated errors from aggressive image compression (e.g., JPEG ringing). Furthermore, they assume a linear sensor response after inverse gamma-correction, which is not entirely correct in practice (see Chapter 6). Finally, the method relies on a simple additive Gaussian noise model with zero mean and statistically independent from the image. As we have seen before, accurately modeling the sensor noise process can lead to new insights on the benefits and limitations of a computational photography technique (e.g., see multiplexed illumination in Chapter 5).

Motion blur from camera shake is often the result of dim lighting conditions and/or long focal lengths (telephoto) that force an auto-exposure camera to choose a long exposure time. During this time, even the steadiest hand may let the camera shake enough to create large and complex point spread function shapes. Reducing the exposure time severely can capture an image without significant shake, but would yield a dark and noisy image; boosting gain to brighten the image would further increase noise in the captured image.

As shown in Figure 7.7, Yuan et al. [438] proposed using the best of both, fusing a blurred and noisy image pair to synthesize one single high-quality image. Each captured image complements the other: the blurred image has low noise, but lacks high-frequency content, edges, or a known PSF. The noisy image has a simple, sharp PSF and thus contains the scene's high-frequency content, but corrupted by noise. First, the method uses both images to estimate an accurate blur kernel. Next, they devised a residual convolution method that uses components of both image pairs to suppress most the ringing artifacts that usually result from deconvolution, and finally apply a gain-controlled deconvolution in smooth image regions to further suppress remaining ringing.

Capturing a pair of images is a great help to reducing the under-constrained nature of the deblurring problem, but it imposes rather severe practical limitations. Many compelling hand-held photos capture transient moments that pass too quickly for a second photo, such as a child's reaction to surprise, or a bird pausing in a birdbath. However, it may hold promise for new, more suitable hardware that might capture the photo pair simultaneously, and the method captures more exploitable information: the blurred/noisy pair might also be sufficient to estimate space-variant blur from camera rotations or objects with simple movements within the scene. As we will see in Section 7.6, such image fusion methods have become a reoccurring theme is computational photography research.

7.2 Geometric Operations

In this section we review methods to manipulate the geometric composition of image elements, rather than their photometric appearance. Such methods include image warping, recent advances in context-aware image resizing, and passive scene analysis. As in the previous section, none of these methods require customized image-capture hardware, but instead rely only on post-processing of conventional images.

While a single, flat 2D image contains no explicit depth information, humans easily interpret its contents as a complete 3D world. Even without recovering depth explicitly, this section will show that clever algorithms and careful priors on the scene structure permit us to perform a wide range of visually plausible deformations, 3D modifications, and transformations from 2D image manipulations.

7.2.1 Image Warping

Image warping algorithms can deform individual image regions into novel user-defined shapes. These geometric image manipulations can correct errors from optical aberrations in cameras, such as severe radial lens distortion from 'fish-eye' lenses. Warping can straighten curved lines in a photograph to match the known-parallel lines in the 3D scene. Similarly, warping lets us simulate the appearance of fisheye lenses from ordinary lenses or from panoramas constructed from multiple images. We can apply simple warping methods to recover synthetic imagery, corresponding to that produced under an ideal pinhole projection model, even with highly-distorted fisheye lenses. Implementing stand-alone image warping that includes high-quality image filtering can be dauntingly tedious; we recommend a simpler approach that exploits 3D graphics rendering hardware available on almost any computing device. Through either the OpenGL or DirectX API, first create a uniform mesh of quadrilaterals or triangles, attach the source image to that mesh as a simple texture map without lighting or shading, enable the hardware's best built-in texture-filtering capabilities (e.g., MIP-maps, bilinear or trilinear) and then render the textured mesh on-screen. To warp the image, simply compute new positions for each vertex in the mesh, and permit the 3D rendering hardware to create the new warped image on-screen. These warping results are not only quick to render (usually available at interactive rates), but are also visually-accurate if the barycentric coordinate are invariant for the underlying image deformation process. To improve fidelity, simply increase the density of the underlying mesh (smaller quadrilaterals or triangles, and more of them) until the residual error between vertices is small. Most desktop computers now provide sufficiently powerful graphics hardware to render million-vertex regular meshes at interactive rates, enabling full-screen renderings that map no more than one or two pixels to each mesh element. [****NOTE: comment from Dan Raviv, regarding this paragraph: ”why not discuss the methods”?]

As described in Chapter 3, panoramic imagery is commonly synthesized by stitching together many individual photos with moderately overlapped fields of view. Creating panoramic mosaics requires accurate registration and blending of the individual images. Calibration involves correcting for variations in perspective, radial distortion, vignetting, and other image aberrations due to the variation of camera parameters. Image registration requires estimating the aiming direction, orientation and focal length for each photograph. Finally, blending registered photos involves color correction and exposure compensation to ensure photometric consistency between overlapping images and throughout the entire panoramic assemblage. In recent years commercially-produced motorized pan-tilt systems have automated and simplified gigapixel panorama capture. One such system was recently described by Kopf et al. [220] (see Figure 7.8). As they show, image warping must be performed to not only composite images, but also to view the final mosaic. Unlike QuicktimeVR [81], they show that very large panoramas are best viewed by seamlessly transitioning from a perspective projection for narrow fields of view to a cylindrical or spherical projection as the field of view widens past nominal values. While their system explores the benefits of interactive gigapixel mosaics, it requires long capture times (many hours). Illumination changes as the sun moves and large moving objects such as ships and trucks can cause inconsistencies between overlapping photos that complicate the panorama assembly process. In contrast to such scanning sensors, single high-resolution gigapixel images may one day become possible, though a few pioneering systems such as the GigaPxl Project (see www.gigapxl.org/) found adequate lens designs were expensive and very difficult to implement, and required well-lit scenes for best results.

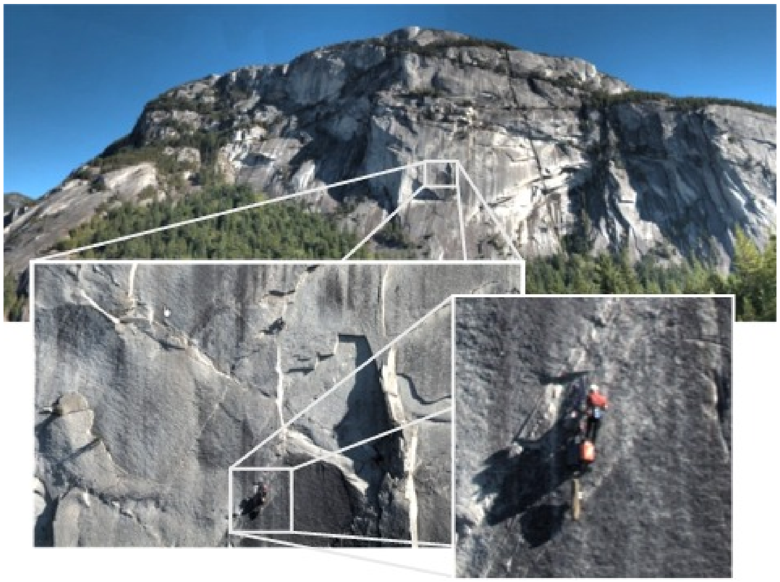


Figure 7.8: Gigapixel image capture and viewing [220]. Gigapixel imagery can be interactively viewed at multiple scales. Wide-angle views, when the scene is zoomed out, are viewed under cylindrical or spherical projection, whereas close-up views are rendered under a local perspective projection. (source: research.microsoft.com/en-us/um/redmond/groups/ivm/HDView/)

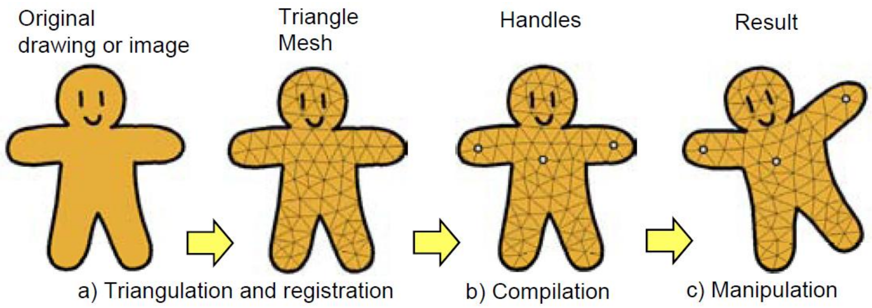


Figure 7.9: As-Rigid-As-Possible Shape Manipulation [191]. The system first creates a triangular mesh to match an input drawing, then the user defines a set of control handles and moves them to any desired new position. The system applies local"as rigid as possible" deformations to each triangle to satisfy the new positions of the control handles. (source: tinyurl.com/cmhqwh)

Nomura et al. [304] use flexible camera arrays to recorder dynamic scene collages. Similarly, catadioptric imaging systems (en.wikipedia.org/wiki/Catadioptric) can be used to achieve the necessary wide field of view for recording large panoramas; for a thorough survey, see [23].

Image morphing is a special effect that creates an animated transition between two images. In early motion pictures such transitions were usually cross-fades implemented by re-photographing each frame on an optical printer, to combine two shots matched as carefully as possible in-camera. Careful registration was necessary because optical printers offered no image warping capabilities, and any mismatches would appear as mis-registration between the "before" and "after" image of actors. With the advent of digital compositing hardware in the 1980s, such as the Pixar Image Computer, simple cross-fades were improved to include matched geometrical warping to minimize mismatches between blended frames, guided by feature points matched for both images (e.g., the position of the eyes, nose, and mouth of two actors). Recently, similar morphing techniques have been applied to allow general user-defined animations. Igarashi et al. [191] present a freeform deformation tool, in which a given shape is manipulated by defining the translation of a sparse set of control points (see Figure 7.9). The system finds the position of the remaining vertices by minimizing the distortion of each triangle using a two-step algorithm, which first refines the rotation and then the scale of each triangle to satisfy the transformed control points. Their algorithm achieves real-time deformations, but yields physically-implausible results for some control point transformations. A similar method for shape interpolation proposed by Alexa et al. [43] offered aesthetically different results. As they observed, no one best solution may exist for such deformations, making a completely automated system neither desirable nor practical. Instead, such systems can find their greatest usefulness by offering intuitive, easy-to-use controls that permit users great flexibility when needed, and a wide variety of easily accessible styles that permit both prompt results and quick explorations to find it.

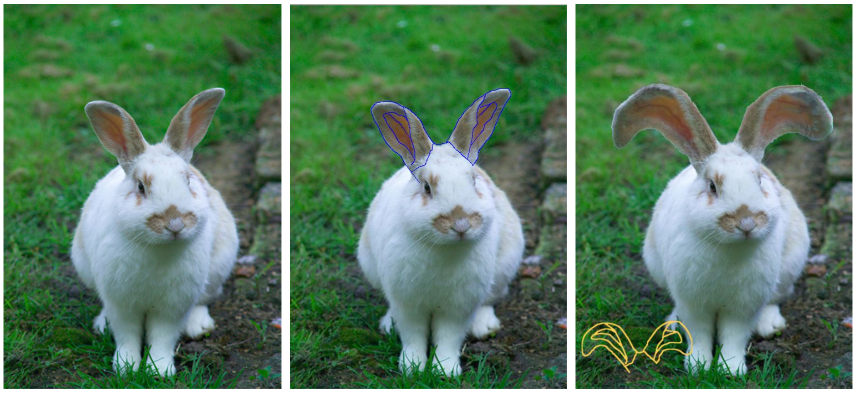


Figure 7.10: Detail Preserving Shape Deformation in Image Editing [126]. (Left) The input image. (Middle) User-drawn feature curves shown in blue. (Right) Final deformed image with user-specified feature curves shown in yellow. (source: graphics.cs.uiuc.edu/~huifang/deformation.htm)

Image warping remains an active area of research. While hardware-assisted textured-mesh rendering schemes are sufficient for some warping and morphing algorithms, such schemes tend to unrealistically stretch or compress texture detail. Fang et al. [126] propose an image editing system that preserves such details while allowing extensive deformations. As shown in Figure 7.10, their method preserves details in deformed regions by synthesizing textures that maintain texture frequency content as well as texture orientation. While the state-of-the-art of image morphing has moved well beyond cross-fading, image morphing has begun to emerge as an enabling technology for both interactive animation design and image editing.

7.2.2 Smart Image Resizing

Important small features in large images intended for large, high resolution displays (e.g., motion-picture screens) may vanish when shown on very small ones, such as a mobile phone or hand-held display device. Cropping these images can focus the viewer's attention on a desired subject, and cropping might be required to display that subject with sufficient resolution to discern its smaller features, or to display an image with a wildly different aspect ratio. Cropping is one of only a select few image manipulations permissible in modern photojournalism, together with content-preserving operations such as color correction and sharpening. Cropping can also be applied to video sequences, producing the "pan and scan" conversions typically used to display widescreen films (typically 2:1 or 16:9 aspect ratios) on standard aspect television screens (4:3). Recently Golub [158] introduced PhotoCropr to assist novices to select image crops using various heuristics, including the "Rule of Thirds" and the "Golden Mean". Such expert systems, however, have difficulty handling general scenes and are inconvenient to apply to video sequences.

Cropping is most appropriate for scenes dominated by a single element or area of interest. With widely-spaced multiple elements, we may need to remove image content from the *m*iddle of the image instead of the edges, 'retargeting' the scene content for a different display. Successful retargeting may require multiple operations, including scaling, shifting, and re-sizing individual scene elements to fit within a target aspect ratio (e.g., to display a family portrait on a mobile phone screen). As shown in Figure 7.11, several methods have recently emerged in the computer graphics and vision communities for automatic image retargeting. Setlur et al. [365] follows a multistep remove-and-replace process: first segment an image into regions, identify the most-important regions and remove them temporarily. Next, fill any resulting gaps via texture synthesis, resize the remaining image, and re-insert the important regions atop the result. Their solution resizes regions independently based on their estimated saliency, but sometimes may convey inconsistent relative proportions between these regions. Furthermore, the run-time of their implementation may make it impractical to complete on a mobile device, as desktop processing times ranged from 5 to 40 minutes for test images.

Seam carving, a popular and conceptually simple image retargeting algorithm recently proposed by Avidan and Shamir [49] may help address such run-time problems. They observe that the aspect ratio of an image can be adjusted by removing or appending minimum energy *seams*, defined by an optimal 8-connected path of pixels on a single image from top to bottom, or left to right, where optimality is defined by an image energy function (e.g., saliency can be approximately by the gradient magnitude). The selection and order of seam removal/insertion is designed to protect the content of the image, as defined by the energy function. Furthermore, object removal can be achieved by modifying the energy function to penalize seams that pass through a user-selected region. A multi-size image can be resized in real-time, by storing a pre-computed ordering of seam deletions. Recent enhancements to seam carving extended the method to video retargeting [348], as well as mesh retargeting [221], and general extensions for any spatial media [366].

7.2.3 3D Analysis

Traditional photography destroys information because it collapses a 3D world onto a 2D image plane. Recent methods in computational photography have explored the ways to capture or preserve more 3D information to permit re-composing the image of the scene, or to capture usable 3D geometry from the scene to enable new methods to improve its depiction. Freeman and Zhang [140] introduced shape-time photography to synthesize novel images that summarize the spatial and temporal characteristics of an interesting motion. For example, the odd-looking (and strange-sounding) acceleration of a spun coin rolling on its serrated edge as it falls flat Figure 7.12 is difficult to convey with just one or even a series of conventional photographs. While a video with sound might depict it well, we can't put the video in a book or on a billboard: how could we adequately convey this motion in a single image? Earlier photographic innovators such as Harold Edgerton and Eadweard Muybridge made multiple-exposure images with strobes or time-release shutters, but these summaries don't work well for in-place rotations. Overlapped exposures that are too numerous or too complex form a jumble that doesn't reveal depth or temporal relationships between the overlapped photos. Instead, Freeman and Zhang captured their photo series with a stereo camera and estimated a depth

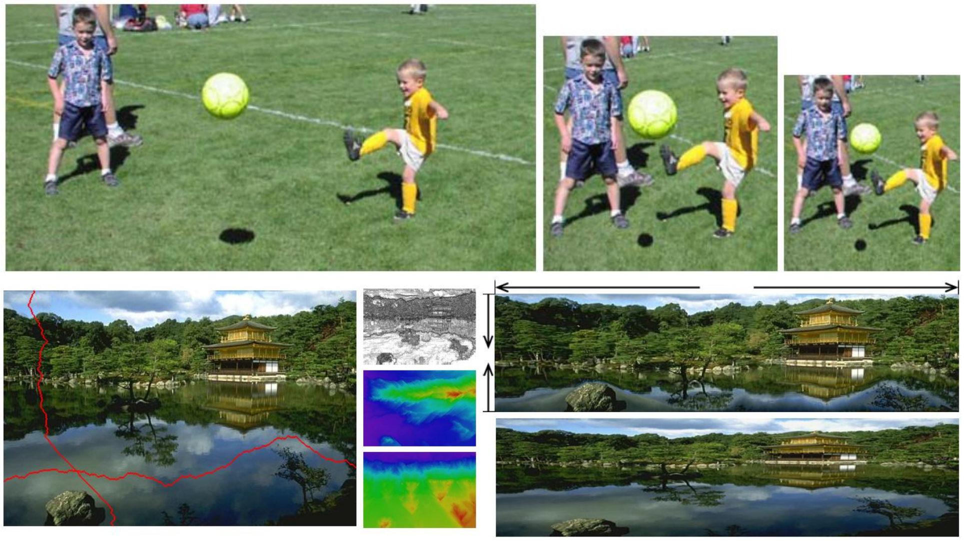


Figure 7.11: Image retargeting [365] and seam carving [49]. (Top) Image retargeting methods segment, transform and re-assemble salient scene features to fit smaller displays or adapt to different display aspect ratios while preserving visually salient features. (Bottom) Instead of re-sampling, “Seam carving” changes image size or aspect ratio by deleting/inserting low-energy 8-connected pixel curves connecting opposite image edges. The energy function assigns high (or low) values to user-selected salient image features to preserve (or remove) them. (source: www.faculty.idc.ac.il/arik/SCWeb/imret)

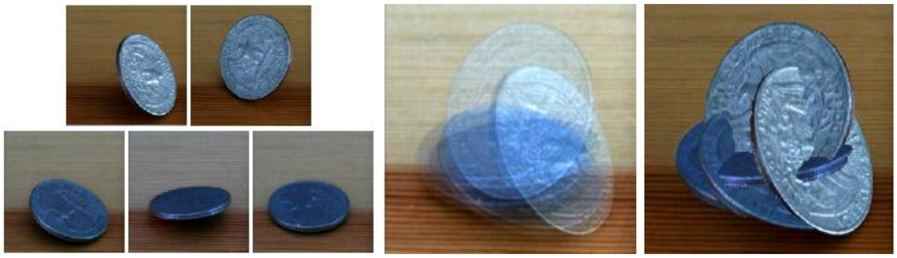


Figure 7.12: Shape-time Photography [140]. (Left) A sequence of images of a coin rolling on its edge as it falls on its side. (Middle) A multiple exposure summary made by averaging the frame sequence. (Right) Shape-time photo is a composite of images layered by depth and time to convey both shape and motion in a single image. (source: people.csail.mit.edu/billf/freemanw_shapetimeRef.pdf)

map for each stereo pair. From these, they created novel "shape-time photos" by compositing the photos in both temporal and depth order, creating a simpler, multi-instant photo that conveys both shape and movement in a more comprehensible form.

No truly robust, reliable, economical and accessible method yet exists for entirely passive 3D photography, despite decades of effort in computer vision and photogrammetry. While humans make qualitative shape assessments quickly and easily, we seem to rely on massive amounts of a-priori knowledge, and learn these 3D assessment skills as we reach out and touch the world around us, and remember surprising results and exceptions. While many researchers have pursued fully-independent systems, others have shown that just a little guidance from interested users can guide mostly-automatic 3D shape, illumination, and reflectance estimators to reliable solutions, even from a single input photograph. As Figure 7.13 shows, the Façade system devised by Debevec et al. [92] creates 3D architectural models from a small set of input photographs. The method is straightforward; first, users electronically mark a few corresponding corner points and dominant lines shared among photos taken from multiple viewpoints, gathered casually by walking around the building and taking sets of overlapping photos. The system uses the corresponding 3D feature marks to estimate the photographer's original 3D positions and construct a coarse 3D geometric model from those marks, using projective texture-mapping to assign the photograph's colors and textures to the model. The system then refines the coarse model, adding finer depth details via conventional stereo techniques at any locations where overlapped projected textures show mismatches or inconsistencies that depth modifications can resolve. Inspired in part by earlier image-based modeling projects such as the "Tour Into the Picture" project by Horry et al. [188] that formed simple but explorable texture-mapped 3D models from uncalibrated photos or paintings that exhibit perspective. To build these models, users marked easy-to-find features such as wall corners, parallel lines or vanishing points, and the system constructed a 3D model as an explorable "spidery mesh". Later systems, including "Automatic Photo Pop-up" by Hoiem et al. [186], further automated the geometry extraction process and are reviewed in

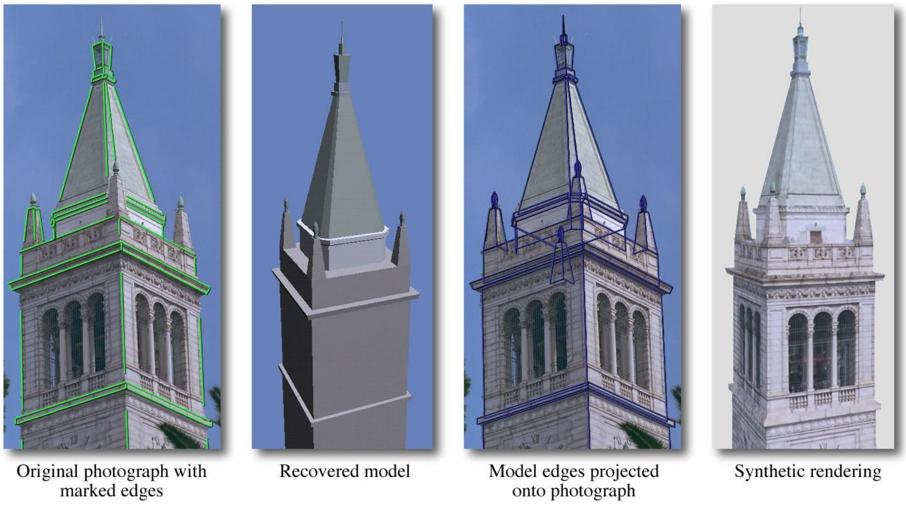

Original photograph with marked edges
Recovered model
Model edges projected onto photograph
Synthetic rendering

Figure 7.13: Modeling and Rendering Architecture from Photographs [92]. A synthetic rendering constructed from just a few overlapped images taken from viewpoints that surround the building. With modest user assistance, Debevec's system matches corresponding points and lines in the photos to construct a coarse 3D model, applies the photos as texture-maps to that model, then further refines the model by stereo reconstruction for regions where overlapped textures show consistent misalignment. View-dependent textures applied to the refined 3D model ensure users see surface colors from the nearest-available source photo. (source: www.debevec.org/Research/)

more detail in Section 7.4.

7.3 Segmentation and Tracking

Digital images stored as nothing more than pixels badly complicate the fundamental task of robust segmentation. Human beings assess the world around them as an assembly of distinct objects, moving, separate or in contact, on, in, around or disjoint from each other, but segmenting a grid of pixels in these same ways remains an unreasonably challenging task, even with substantial user assistance. In new forms of computational photography, we would like to see lights, optics, sensors and processing methods that collaborate to supply us with more information in a picture, information that would dramatically simplify and enhance our ability to segment photographs and other visual records of a scene. What might assist us in identifying connected components and correspondences in images or sets of images? How might we identify individuals, specific physical objects, the material types, their illumination, and the distances to objects or between objects?

In this section we review recent and historic works that address the central problems of foreground/background segmentation which decomposes a scene into at least two distinct depth layers. Essential for automatic scene understanding, this segmentation extracts moving objects close to the camera for separate processing. As demonstrated by many of the works we cover in this section, robust segmentation allows rapid, intuitive and interactive image editing, as well as the synthesis of novel imaging results.

7.3.1 Matching

Chapter 3 describes how to construct a panoramic mosaic by spatial alignment of many photos taken from a fixed viewpoint in a scene. With good geometric registration and a sequence of progressively longer exposure times, we can also construct HDR composites or capture and combine multiple lighting conditions.

Sand and Teller [352] introduced “Video Matching” as a technique for spatio-temporal alignment of multiple video sequences. Their approach matches and merges two different video sequences from spatially-similar camera trajectories (i.e., trace a similar set of spatial viewpoints, although at differing temporal rates) as shown in Figure 7.14. The output consists of a warped version of the second video that is both temporally and spatially aligned with the first. They begin by searching the second video to find a candidate frame that matches the first frame in the primary video. They accelerate this search by testing only a subset of frames (e.g., one frame, two frames, or five frames forward/backward in time at each step). Next, they apply robust image matching and warping to spatially align the image content of the chosen pair of the two video frames, and measure their image similarity as well. Their matching criterion assesses large-scale image differences (e.g., addition or removal of scene components); unlike traditional video registration systems using optical flow [63], their approach handles large scale scene changes well.

Such video matching algorithms have great potential for assisting users in complex image- and video-editing tasks, including removal of support wires (e.g., in special effects sequences where actors are suspended from wires), composition of multiple exposure or lighting conditions, and replacement of stand-ins or other undesired scene elements.

7.3.2 Matting from Colored Backgrounds

Used for everything from inserting a weather map behind a TV announcer to flying a super-hero through the skyscraper canyons of New York City, foreground/background segmentation and compositing for still pictures and video sequences permit us to ‘cut out’ an actor or an object from one photographed setting and composite it into another, to achieve the appearance of editing locations rather than just video sequences. The ability to robustly segment a scene into one or more depth layers allows us to then modify each photographed object separately, and merge or ‘composite’ it with other, independently captured objects or background video sequences.

Such segmentation and compositing has been achieved by many different processes by many different names. In film, they began as ‘matte paintings’ on glass sheets placed between the camera and the scene or movie set to replace static backgrounds. Foreground extraction and compositing, known as a ‘travelling matte,’ was first developed at RKO Radio Studios in the 1930s. A film-based blue-screen

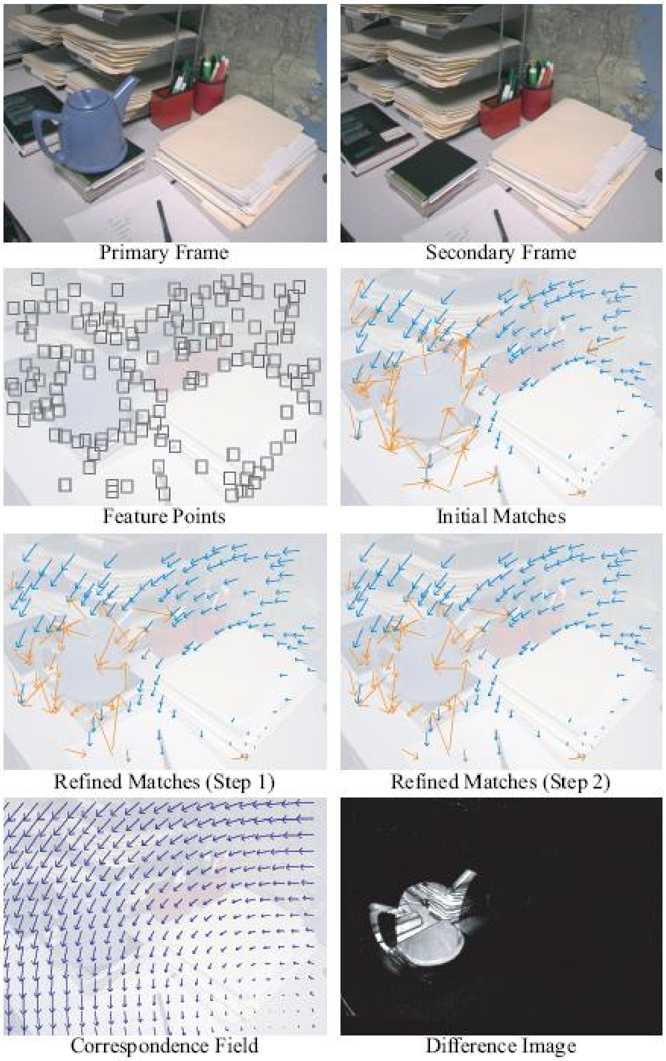


Figure 7.14: Video Matching [352]. (source: people.csail.mit.edu/sand/vid-match/index.html)

process developed by Larry Butler for *The Thief of Bagdad* helped the film win an Academy Award for Best Visual Effects in 1941. Further developments included blue-screen and UV advances in 1950 by Larry Widmer at Warner Brothers, refined optical printers by Richard Edlund in the 1970s (synchronized film camera and projector aimed at each other, mounted on a milling-machine bed and aimed at each other to permit re-photographing film and compositing by multiple exposures), digital camera-motion control in the 1980s by Petros Vlahos, and many others. More information on this topic can be found at the Wikipedia article on Chroma key.

In television, 'keying' and 'chroma-keying' apparently originated as a minor feature first included in a commercial video master-control switcher product for early color-television production. The system switched between two live-video streams depending on a thresholded chrominance value of one of them, but was notoriously difficult to adjust for flawless results, leading to many jokes about TV weathermen with holes in their heads from bluish specular reflections from hair, foreheads, wet lips, pocket-pens, or glasses. In addition to 'chroma-keying', and 'keying', the term 'matting' is also fairly widespread.

In video, segmentation and compositing techniques began as analog *chroma keying* methods. Video cameras photographed actors performing in front of a fixed, uniformly-lit background with strong, easily identified chrominance (e.g., a blue or green scene) values detected independent of luminance values to avoid influence from shadows or lighting changes. Background colors were chosen for high signal-to-noise ratio, high resolution (fine film grain) and large, easy to detect differences from all foreground colors. For NTSC television, strongly saturated blue ensured greatest difference from human skin hues, and in motion pictures, blue or UV films' smaller film grain also aided good matting results. As high-resolution digital video cameras augmented film for special effects, green-screens provided better signal-to-noise ratios than blue. Simple thresholding of hue and chrominance can often separate foreground and background, but without great care the results are often contain a heavy burden of artifacts. Specular reflections from the foreground (a shiny watch, a glinting forehead forms a 'hole' in the foreground), glass objects, transparency, translucency or fuzzy boundaries can make high-quality blue-screen results very challenging to achieve. Skilled, experienced technicians, a growing folklore of tricks and tips, and specialized studios, equipment and software from companies such as Ultimatte Corporation, Primatte, and others now provide the best-quality segmentation results from single-color backgrounds for TV and motion-picture production.

In 1996, James Blinn and Alvy Ray Smith presented a principled analysis of colored-background matting, and showed that foreground-background segmentation against a single-color background is under-constrained, and also explained why additional ad-hoc techniques are both necessary and successful. Further, they showed that capturing an object twice, with two different-colored background enables a simple, well-posed foreground background separation that can capture opaque, translucent, and transparent objects, as well as the shadows they may cast on those backgrounds. [371]. While colored-background matting remains a popular and low-cost segmentation solution, more recent methods permit matting of foreground objects against unstructured backgrounds that may include moving objects. More recent work by Levin et al. [234] extends the hard-edge assumptions

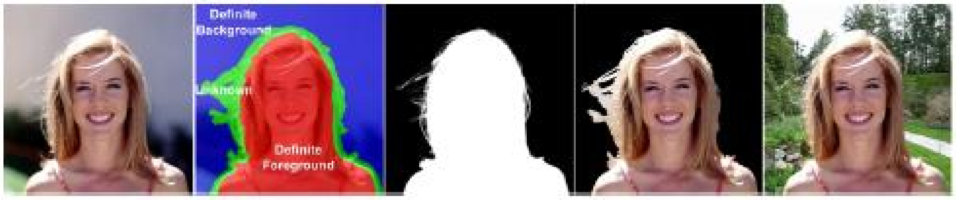


Figure 7.15: Trimap-based matting [418]. From left to right: input image; user-defined trimap; estimated foreground/background matte; extracted foreground colors; background replacement. (source: www.juew.org/publication/mattingSurvey.pdf)

of conventional matting to include soft-edged objects, and may prove suitable for fully-automatic spectrally-guided foreground/background separation.

The matting problem was formally defined for digital computing in 1984 by Porter and Duff [322]. Under their definition, an image I_z, for $z = (x, y)$, is modeled by the convex combination

$$I_z = \alpha_z F_z + (1 - \alpha_z) B_z,$$

where F_z and B_z are the foreground and background images, respectively. In their formalism, $\alpha_z \in [0, 1]$ is the *alpha matte* which defines the local mixing between the foreground and background images. In practice, many modern systems use a coarse user-defined *trimap* to define an initial solution for the alpha matte (see Figure 7.15). The trimap is an auxiliary image, provided by the user, which labels each pixel as either definitely foreground, definitely background, or uncertain. Provided with a trimap, many algorithms can be used to obtain an accurate alpha matte, several of which we briefly review in the following paragraphs.

Chuang et al. [83] proposed Bayesian Matting in 2001. Given a trimap that identifies definite foreground and definite background, and uncertain pixels, they showed that Bayesian methods can be used to model the statistical distribution of colors in these regions. Furthermore, in parametric methods, a low-order model can be fit to these color distributions in each region. Bayesian matting can be classified as a parametric method, which fits a Gaussian distribution mixture to the color histograms of the foreground and background regions bordering an unknown region. Next, they find a maximum *a posteriori* (MAP) solution for the alpha matte by comparing the color histogram of local patches within the unknown region to the estimated foreground and background models. In practice, Bayesian matting produces accurate segmentations when the trimap is conservative and well-defined (e.g., the unknown regions are thin and accurately labeled). But as a parametric method, Bayesian matting often does not capture high-frequency details well if the foreground and background regions are highly textured, due to the use of low-order Gaussian models.

In contrast to parametric color-sampling methods that directly model the alpha matte, such as Bayesian Matting, affinity-based algorithms attempt to recover the gradient of the alpha matte. Such affinity-based methods include Poisson Matting, introduced by Sun et al. [385] in 2004. By taking the gradient of the matting equation, we obtain

$$\nabla I_z = (F_z - B_z)\nabla\alpha_z + \alpha_z \nabla F_z + (1 - \alpha_z)\nabla B_z.$$

If we assume the underlying foreground/background images are smooth, then we can reasonably approximate the gradient matte as:

$$\nabla\alpha_z = \frac{1}{F_z - B_z}\nabla I_z.$$

And thus recognize the matte gradient as linearly proportional to the image gradient. In unknown regions of the user-provided trimap, Poisson Matting assigns the gradient magnitude $(F_z - B_z)$ using the closest available foreground/background pixels. To find the complete matte from its gradients, the authors use a conventional Poisson solver. While quite successful on ambiguous transparencies with simple shapes such as cigarette smoke, Poisson Matting tends to have difficulties with complex shapes; accordingly the method includes tools for user-assistance to improve results in troublesome regions, such as long fur and sparse, feathery hair. Figure 7.15). In addition, Poisson matting tends to be computationally-intensive and may impede practical interactive uses, especially on these troublesome cases.

7.3.3 Smart Region Selection

Image editing that applies interactive segmentation is already widely available in digital image editing software such as Adobe Photoshop and the GIMP (GNU Image Manipulation Program). As Figure 7.16 shows, several available interactive segmentation algorithms can extract foreground, background, and alpha mattes. In Photoshop(c) from Adobe Systems, Inc. the “Magic Wand” tool starts at a user-selected point and finds a closed boundary around a region of connected pixels whose boundary shares similar color statistics. Intelligent Scissors, introduced by Mortensen and Barrett [283] and implemented as the “Magnetic Lasso” in Photoshop, allows the user to select an object boundary by tracing with the mouse. As the user moves the mouse cursor, the system iteratively computes the minimum cost path from the curve’s initial point to the current cursor position. Users can also specify additional seed points in the image if the path makes unwanted deviations.

Recent advances in interactive image segmentation such as ‘GrabCuts’ allow incomplete specification of trimaps, as well as greatly simplified user interaction. As proposed by Rother et al. [347] in 2004 and shown in Figure 7.16, this iterated-graph-cut method often needs little more than a rectangular bounding box around the foreground object to initialize the segmentation and matte construction.

An alternative segmentation method related to both Support Vector Machines and Bayesian Matting introduced Gaussian Mixture Models (GMMs) to find approximations to the underlying statistical distinctions between the foreground and the background colors. The method corrects most segmentation errors by a second phase of user interaction to provide additional background, foreground, or uncertain strokes.

Given the importance of segmentation and matting methods to image editing and the generality needed for their solutions, we expect to see further simplifications

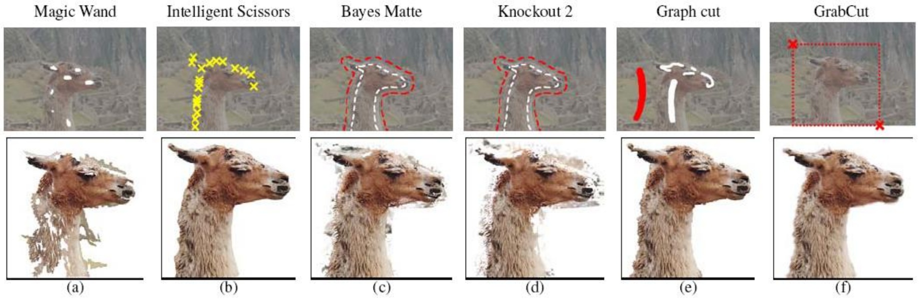


Figure 7.16: GrabCut: Interactive Foreground Extraction using Iterated Graph Cuts [347]. Unlike other recently proposed methods (from left to right: 'Magic Wand', 'Intelligent Scissors','Bayesian Matting', 'Knockout2', 'Graph Cut'), the iterated graph cut or 'GrabCut' method can construct a suitable matting result with substantially less user interaction–just specifying an enclosing rectangle in this case–compared to multiple selected points, tri-map specification, and labeled strokes. In addition, the quality of the resulting foreground separation is in many cases comparable or better on particularly difficult source images such as this one. (source: research.microsoft.com/en-us/um/cambridge/projects/visionimagevideoediting/segmentation/grabcut.htm)

to the user-interactions required, and may eventually see stroke-based and trimap interactions replaced by eye-tracking systems that provide seamless user assistance to segment image features that attract the user's closest attention in displayed images and video.

7.3.4 Tooning

Over the past 100 years, the styles of print, film, and video 'cartoons' and 'graphic novels' have evolved into a clean, streamlined, simple style. Its large, washed regions of constant color chosen from a very limited color palette, separated by thick but deftly-shaped black lines that suggest much more scene content than the artist depicts directly. Cartooning styles that emerged in the early 1900s were at first motivated by necessity: motion picture cartoons pushed artists toward styles they could codify for large teams to draw and paint thousands of animation cels quickly. Cartoons printed on cheap paper pushed artists' towards styles that would ensure their work would remain legible when printed as small panels on coarse, absorbent newsprint in black, or with just a few poorly-aligned color inks and coarse half-toning. Even as low-cost printing quality improved greatly from the 1960s onwards, the styles remained popular and flexible enough to include individual expressiveness. With such objective rules and simple renderings, computer-assisted cartooning seemed a natural fit with computer graphics methods, and Marc Levoy pioneered such a system used for the first computer-assisted cartooning at Hanna-Barbera Inc. in the late 1970s, automating previously hand-drawn pen-and-ink processes. The Levoy-developed system remained in everyday use for producing cartoon shows at Hanna-Barbera until its retirement in 1996, well after better-known computer-animation studios such as Pixar, Rhythm-and-Hues, and PDI-Dreamworks devised new 3D cartooning styles for motion pictures.

As an offshoot of research on non-photorealistic rendering publications that has proliferated since the late 1990s, some researchers have explored both the automatic- or semi-automatic restoration of classic pen-and-ink film cartoons (see Sykora et al. [387]), and others developed systems to construct cartoon-style renderings from video.

After successful experiments in a few commercials and short special-effects sequences in earlier movies, several moviemakers have adapted ideas from earlier non-photorealistic rendering methods to new digital forms of traditional *rotoscoping* (in which artists painted or inked directly on film frames to modify its content) as shown in Figure 7.17. Digital rotoscoping permits particularly accurate abstraction and easy creation of cartoon-like drawings from live-action footage. Done skillfully, it permits artists to capture the complexities of human and animal motion from actual film footage of performances, and to focus more directly on the spatial aspects of shading and lighting, and frees them from the serious challenges that conventional pen-and-ink animators face ensuring natural appearance and temporal consistency for a given motion.

Rotoscoping was used to great effect in *Snow White and the Seven Dwarfs* in 1937, using live-action footage to capture the complex and varied performances of suitably-sized actors and actresses, as well as the mannerisms of animals adapted to animated characters. Historically, rotoscoping requires the intensive efforts of large animation teams, because artists must paint every frame. More recently, computed-assisted digital rotoscoping methods allow animators to specify keyframes, and spline-based interpolation and tracking of painted regions automatically completes the intermediate frames (see Figure 7.17).

More recently, Winnemöller et al. [432] presented a method for fully-automatic video abstraction, including video summarization into strips of cartoons complete with speech bubbles showing text extracted from closed-caption data streams, and GPU-assisted real-time conversion of video streams to cartoon-like renderings. The video abstraction method performs a sequence of fast image operators on a sequence of frames to simplify low-contrast regions while enhancing high contrast regions, apply thresholding, apply color quantization, apply quantization smoothing to avoid posterization-like effects, and smooth black-inked edge features in a manner consistent with image content—operations that mirror the production steps of comic illustrators. As Figure 7.18 shows, they first compute an abstracted color image using a bilateral filter. They then optionally quantize the luminance into a small number of levels, and blur their free-space boundaries to remove the appearance of color contours. Finally, they compute edges by thresholded Difference-of-Gaussian filtering and a subsequent smoothing step, and apply them to the abstracted images to emulate a cartoonist's India ink strokes. While fully-automatic, we observe that their approach does not significantly alter the composition of an image or video sequence, as it lacks the understanding and interpretive abilities that an artist might apply in manual cartooning.

Cartoon motion has evolved its own idioms that help emphasize the story and engage the audience even as they only approximate, exaggerate, or completely ignore laws of physics. Expert animators use motion to guide viewers' perceptions

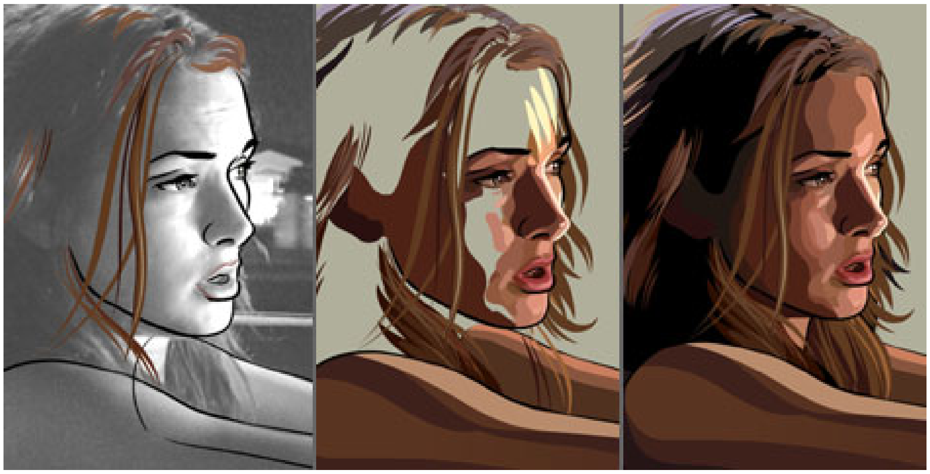


Figure 7.17: Manual rotoscoping in *A Scanner Darkly.* From left to right: an input live-action frame with initial animator strokes; after the addition of additional strokes; final rotoscoped image. (source: Warner Independent Pictures www.usatoday.com/life/movies/news/2006-08-01-rotoscoping_x.htm)

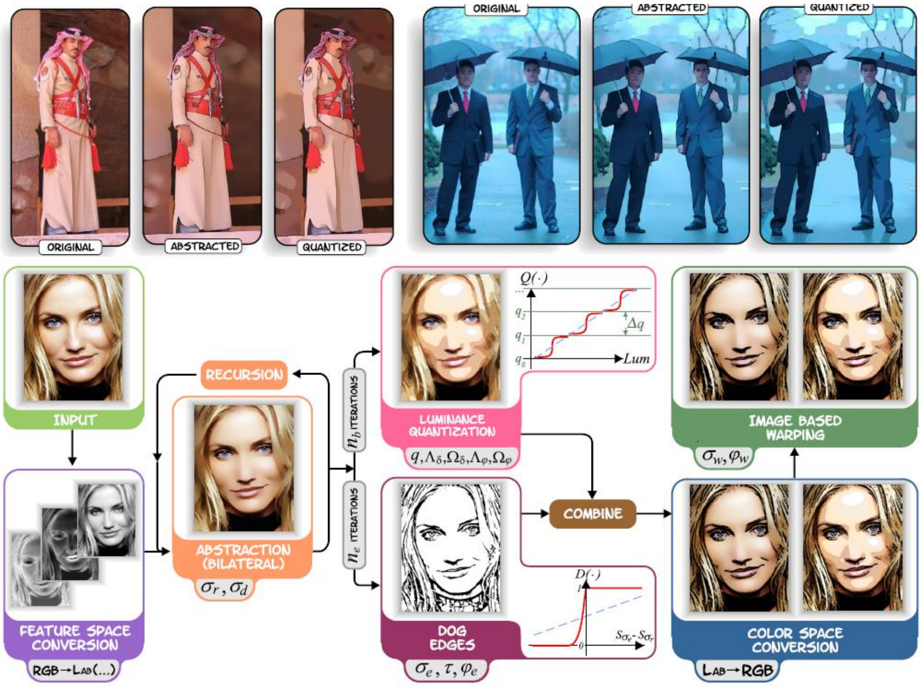


Figure 7.18: Real-Time Video Abstraction [432]. (Top row) Abstraction examples. (Bottom row) Abstraction framework. (source: videoabstraction.net/)

and make the entire scene depicted seem lively and responsive. As outlined in Thomas and Johnston's "The Illusion of Life" [206] and repeated by Lasseter [227], skilled animators use techniques that help direct the viewers eye and attention, and add emphasis through techniques such as "squash and stretch" distortions to shapes that move or are about to move. These skills are not innate: cartoonists and 3D animators learn these grammars and conventions of cartoon movement through study, example, and hard-won experience. While anyone can learn to use keyframing tools for computer animation now widely available, few of us have gained the animators skills needed to produce high-quality cartoon-style animation.

To make such idioms more accessible to a broader set of users, Wang et al. introduced the "Cartoon Animation Filter" to convert physically accurate motion capture data to cartoon-like movements. Mathematically, their filter output $x^*(t)$ is a second-derivative filter:

$$x^*(t) = x(t) - \overline{x''}(t),$$

where $x(t)$ is the input signal (e.g., the position of motion capture vertices over time or the coordinate of an image element) and $\overline{x''}(t)$ is a smoothed and possibly time-shifted second derivative of x(t) with respect to time. As shown in Figure 7.19, this simple filter automatically synthesizes many traditional animation techniques, including anticipation/follow-through and squash and stretch. Though the results lack the quality and refinement of work by a professional animator, the method does synthesize a plausible result consistent with basic animation rules. In the future we expect to see similar "expert filters" for a wider variety of tasks. In the spirit of image analogies described in Section 7.4.3, automatic learning and synthesis of artistic techniques may extend computer-aided animation just as it has extended image editing and manipulation.

7.4 Data-driven Techniques

Consumer data storage and transmission systems have grown so steadily and so quickly over the past decade that terabyte hard drives and multi-gigabyte USB flash memory-drives are commonplace and easily affordable. With cheap, abundant storage and computing, computer graphics and vision researchers can now explore 'brute-force' and data-driven techniques previously considered too inefficient or too extravagant. For example, early light field capture and rendering experiments in the mid 1990s struggled to manipulate hundreds of megabytes of image data, but demonstrated the utility of data-driven rendering. Simply re-binning these images allowed complex optical phenomena to be accurately depicted, including variation of surface materials, scattering, translucency, lighting and inter-reflections. Such light field imagery is unmatched by any real-time results of classic model-based rendering that uses storage in far more 'sensible and efficient' ways; however, with the advent of cheap storage and computation, such savings offer little added utility. Researchers are now exploring other data-driven methods that may provide compelling new answers to long-standing, challenging graphics problems using massive on-line image collections, image metadata, and probabilistic inference algorithms.

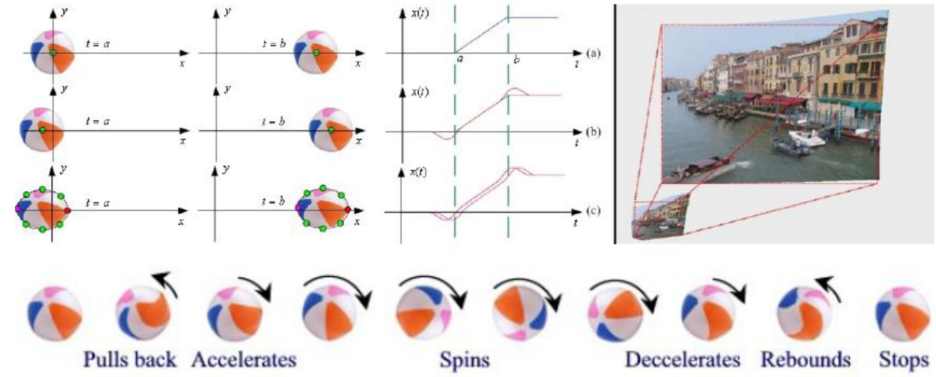


Figure 7.19: The cartoon animation filter [419]. Left: (a): Abrupt start to constant-velocity ball translation and rotation looks unnatural, because it begins with infinite acceleration. (b): Wang et al. filter applied to the ball's centroid emulates cartoon animators' 'anticipation' and 'follow-through' conventions due to filter under-shoot and overshot at start and end of ball motions. (c): Applying the same filter to the positions of the ball's outermost vertices induces 'squash' and 'stretch' effects. Right: animator-like interpolation between offset small and large image files curves transitional images at start and end of the motion. Bottom: filtering a ball that translates and spins at a constant rate creates a smooth movement that begins with an anticipatory pull-back, and ends with a smooth overshoot and correction. (source: www.cs.washington.edu/homes/juewang/juew/JueAniFilter.pdf)

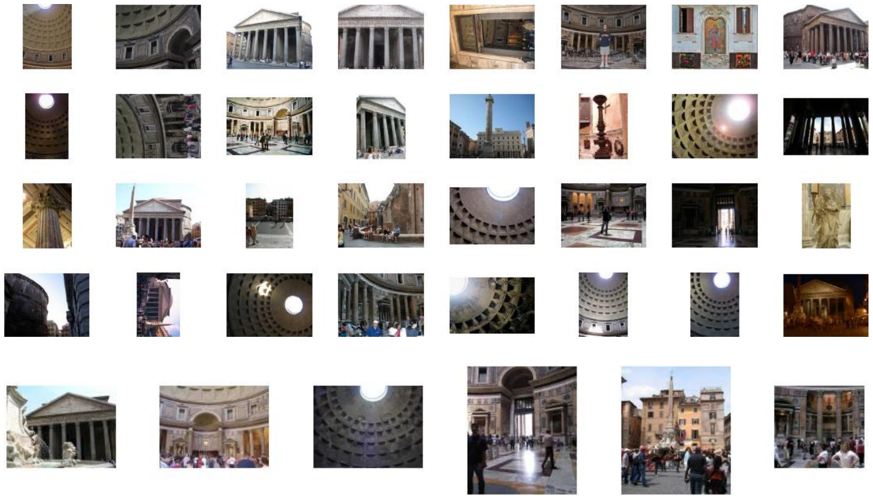


Figure 7.20: Scene summarization for online image collections [369]. (Top rows) A random subset of 32 images of the Pantheon. (Bottom row) A set of output "canonical views" computed using the proposed clustering algorithm on the total set of thousands of similar images. (source: grail.cs.washington.edu/projects/canonview/)

7.4.1 Image Collections

Inexpensive digital photography, while convenient, has given rise to a new problem: how can we organize and view our large personal photo collections? The higher cost of film-based photography encouraged caution and care, and most people's photo collections were precious, accumulated slowly, and could fit neatly into bound, hand-annotated photo albums or boxes of slides. Digital photography imposes no discipline—photos cost us almost nothing but the time required to transfer them from our camera's memory. Instead of a few rolls of film, many amateur photographers may return from vacation with many hundreds of photos to select, label, and organize, a task than can take hours or days.

How to select the most important photos to summarize a vacation, a day, or tell a story? Such editorial decisions require a higher-level understanding of the emotional context of a photo. However, recent advances have been made in (semi-)automatic image collection analysis by applying spatial and temporal constraints derived from computer vision algorithms.

Simon et al. [369] address the problem of scene summarization; as shown in Figure 7.20, the system must represent adequately, but concisely, the important visual contents of a frequently photographed scene (e.g., the Pantheon in Rome) by selecting a small set of exemplar photos, or "canonical views", that capture the key sites of interest (e.g., the Oculus, the entrance of the Pantheon, and various views of its interior). The system divides this task into three sub-problems; first, partition the image set into groups of images that depict a similar scene feature, by finding sets of shared image features using SIFT feature co-occurences and applying a fast clustering algorithm. Second, identify an exemplar member from each group by applying a likelihood measure to co-occurrences within the group. Third, extract a metatag that summarizes the content of each canonical image to assist in browsing and image search. As we will review in Section 7.4.4, such metadata is critical for presenting readily machine-readable content (e.g., ASCII text strings) for organizing and browsing massive image collections.

A wide variety of competing photo management applications have emerged in recent years, including iPhoto (Apple), Picasa (Google), and Photoshop Album (Adobe). These applications often handle direct I/O from a camera or storage media, allow basic resizing, cropping, and retouching operations, and offer organization and navigation services. Recently, some of these applications added automatic face detection and metadata tagging prompting publications that propose new paradigms for photo navigation such as Bederson et al. [56].

Bederson's 'TreeMaps' are an interactive visualization method that presents large amounts of hierarchical data of any kind very efficiently on-screen. It's layout algorithms encourage smooth, continuously variable on-screen display, and are suitable for dynamically-changing hierarchies that gain or lose nodes over time, including very large, dense trees and also highly uneven or unbalanced trees. Instead of conventional plots of on-screen nodes connected by edges, which often leave large portions of a rectangular screen empty and unused for large trees, the TreeMaps algorithm subdivides the screen efficiently into rectangular regions, nesting child-node subregions inside their parent regions, keeping node size large and empty space low. It also controls the aspect ratios (height/width) of these regions to avoid any sub-regions that are too narrow, thus ensuring easy visibility and effective use of that space for text, images, or other annotations. TreeMaps are particularly well-suited for large collections of photographs held in nested directories, and is available as both stand-alone software and applied in the "PhotoMesa" browser (see: www.photomesa.com/).

Treemaps supply users with interactive, continuously variable 'zoom' and 'move' controls. Zoom permits them to see more or fewer sibling nodes, and/or more or fewer levels of the tree, and 'move' lets users scroll across the tree to explore tree portions too large to fit on-screen at the current zoom setting, and can permit users to annotate and rearrange nodes and their positions within the tree. Treemaps can provide an easy, intuitive browser for endlessly vast image collections; it does not require any metadata tagging for source photos, but can respond to them if desired, and can permit easy tagging of individual photos, groups of photos, or nested groups of photos. When coupled with face detection and recognition, it forms a good framework for automatically grouping photos according to the people within them. Treemaps limitations seem to arise mainly from describing photos in a single hierarchy; users may wish to form multiple hierarchies or organize photos in a database-like arrangement that can respond to queries such as 'find all Florida Keys vacation photos that show Aunt Martha near the water but not studying for her real-estate exams'. While TreeMaps might maintain multiple trees with the same nodes, currently it does not organize nodes into a database. (from: www.cs.umd.edu/hcil/photomesa/). We expect that the development of such interfaces will continue to evolve, especially as automatic metatagging, such as face and place detection, becomes commonplace both in-camera and in browsing applications.

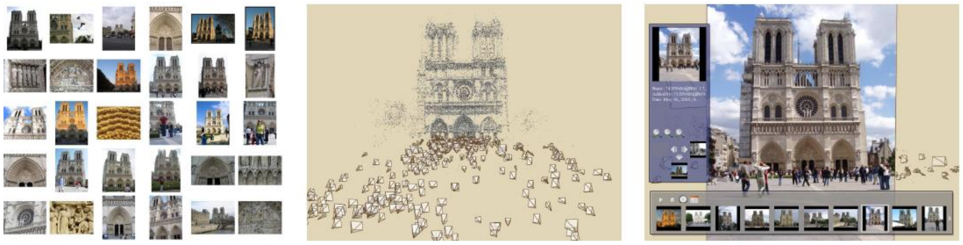


Figure 7.21: Photo Tourism: Exploring Photo Collections in 3D [376]. (Left): An input collection of unstructured photographs of a single scene found by on-line image search. (Center): Structure-from-motion on corresponding points from all pairs of photos in the collection creates a sparse 3D point-cloud of the scene, along with automatic camera viewpoint recovery for each photo. (Right): Interactive 3D image browsing permits users to navigate smoothly among photos by indicating a desired viewpoint or selecting a portion of the 3D point cloud, providing the impression of freedom of movement to tour and explore the scene in 3D. (source: phototour.cs.washington.edu/Photo_Tourism.pdf)

7.4.2 On-line Photo Collections

In recent years, massive on-line photo and video collections such as Flickr, Google Image Search, and YouTube began offering free access and storage for anyone's photos and videos in exchange for viewing advertisements while browsing these collections. While welcome and simple for anyone willing to share their works freely with everyone, these sites greatly complicate legal questions of ownership of images, fair-use versus copyright violations, and protection of valuable works sold for profit, such as clips of television shows, photographs by Ansel Adams, and motion pictures. Several new licensing methods such as the Creative Commons licenses allow novices to contribution to these collections in a manner consistent with their wishes (i.e., to either allow, restrict or prohibit derivative works or noncommercial applications). The advent of such large, legally-accessible image repositories has facilitated data-driven approaches in computer vision and graphics. In this section we examine the more immediate task of organizing and navigating such collections, and in the next we examine how image collections can be used to resolve long-standing challenges in the field.

As most people encounter scenes they photograph by walking around and examining it as a 3D environment, Snavely et al. [376] realized that interactive browsing that recreates the impression of exploring or 'touring' the 3D scene might offer a particularly attractive and intuitive way to explore well-photographed scenes. While no one person might take enough photos of an interesting location to permit 3D reconstructions, many thousands of people visit many popular historic sites and tourist attractions and each take their own set of photos. If enough of these are available on-line and the photos capture the scene from a sufficiently varied set of viewpoints, Snavely et al. reasoned that 3D reconstruction might be still possible even without any other camera calibration, pose, or a-priori knowledge of the scene itself, and developed a novel system for browsing large unstructured image collections in 3D. Now known as Photosynth, available as a free download from Microsoft Live Labs, their system is the first to apply structure-from-motion methods to large community photo collections. While traditional image browsing software uses directory-based hierarchies and slideshows, their viewer presents users with a 3D scene and an interface that lets viewers explore by clicking on anything they'd like to see. As shown in Figure 7.21, their ambitious multi-step optimization process iteratively recovers sparse scene geometry and camera viewpoints simultaneously, and encourages navigation and 3D exploration using an interface similar to current-generation 3D games.

The authors also describe extensions that automatically cluster and group photos, that extract canonical views [369] and that recommend popular walking paths through the scene [375]. Their system relies on SIFT for automatic feature detection, sparse bundle adjustment to recover the 3D scene points, and an aggressive RANSAC strategy to encourage convergence with suitably low error for navigation. Currently, their system's best results require very dense and numerous photo sets with substantial overlap between photos, but on-line photo collections for attractive single-site monuments such as the Statue of Liberty, Eiffel Tower, or Notre Dame Cathedral routinely provide thousands of suitable photographs. They apply camera settings from EXIF data embedded in photos when available, but even then the system sometimes must discard as many as 30 percent of the input photos to find a suitably converged solution, and early versions of their software reportedly took up to a week to assemble the 3D models for photo sets presented in the original paper. Authors noted that corresponding point pairs were deemed too sparse to permit lens distortion corrections, and that the proposed method only constructs maps of sites whose photos' shared features form a single connected component, rather than multiple disjoint groups of photos. However, these limitations may fade quickly as GPS-enabled cameras supply absolute 3D position in their EXIF data, reducing the optimizers' reconstruction errors.

7.4.3 Probabilistic and Inferential Methods

In this section we review data-driven probabilistic and inferential methods that gain leverage from massive on-line photo collections. Such databases let us solve problems by simply searching for suitable answers or examples, copying them instead of computing them. The genuinely massive example-photo databases now available online (and growing quickly) can help us solve classically intractable problems such as 'hole-filling' or scene completion, 3D geometry estimation from a single photo from an unknown camera, and physically plausible texture synthesis and interpolation.

Scene Completion

The novel user-guided image completion method by Hays and Efros [178] shown in Figure 7.22 demonstrates several strong advantages to data-driven methods. Unlike previous image-repair or completion methods this method can suitably 'fill in' or repair tremendously large holes cut out of an image. It does not rely on example-guided texture-synthesis, image quilting, or image inpainting using Beltrami-flow

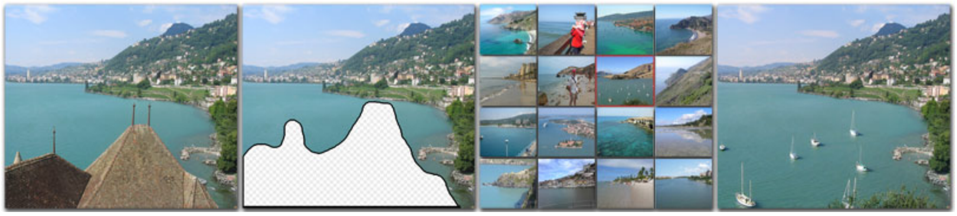


Figure 7.22: Examples from "Scene completion using millions of photographs" [178]. (Far left): The input image and (Middle left): marked region we wish to replace. (Middle right): Searching a large on-line image collection finds several closely-related scenes, (Far right): Use graph-cuts to find suitable boundaries for replacement regions in the source and destination photos, and then combine the pieces using Poisson editing methods to blend them without visible seams. (source: graphics.cs.cmu.edu/projects/scene-completion/)

PDEs, and does not cover those holes and gaps by improvising or copying textures from small nearby texture patches in the same image. Instead, the method searches vast numbers of other images to find a large, directly-substitutable region (see Section 7.4.3), uses graph-cuts to snip out a properly-oriented, well-matched piece of that region big enough to cover the hole, and then fuses it into the original image using Poisson editing methods (gradient-domain image fusion [317]) to hide the seams. As Figure 7.22 demonstrates, any search that finds regions that match enough of the basic image features (e.g., colors, gradients, edges, textures, surface normals, lighting, etc.) is likely to find regions that match more difficult, higher-level semantics as well (e.g., ocean, not lake; calm sea, not stormy) and form a visually plausible substitution.

The paper presents a two-stage approach. First, the system searches the image database using low-level descriptors to find approximately 200 matching scenes, then asks the user to choose semantically appropriate substitutes to fill the image hole. Then they compute a SIFT-like 'gist descriptor' (introduced by Torralba et al. [400]) for the selected patches that describes its aggregate oriented edge responses at multiple scales, sorted into very coarse spatial bins. Second, they create a seamless composite by graph-cut and Poisson editing.

Database size and processing time both limit the systems' current practical abilities: the database size may never grow enough to hold a high quality solution for all possible image completion tasks, yet their preliminary system still requires about one hour on one CPU to complete one image: typically 50 minutes for matching, 20 minutes for local context matching, and 4 minutes for compositing. Continually growing on-line photo databases, multi-core computers and distributed search methods may help make the system more practical soon.

In a similar vein, Lalonde et al. [224] devised a "Photo Clip Art" system to insert new objects into existing images a simple query-based interface. Users choose 3D location for the new object in the image, select an object to insert from an hierarchical menu (e.g., with entries for people, places, household objects, etc.), and then uses gradient domain fusion methods (Poisson editing) to merge a pre-

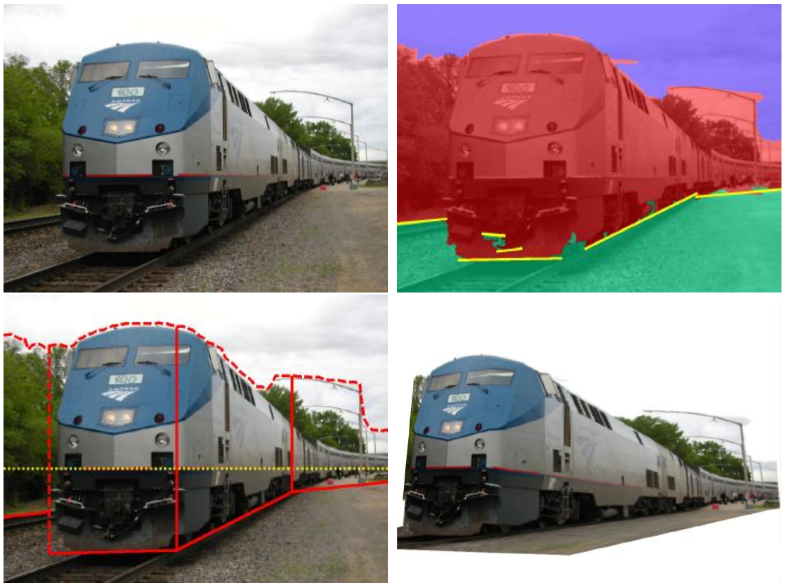


Figure 7.23: Automatic Photo Pop-ups [186] construct a rough 3D model from a single input image (far left). The system automatically detects and labels individual scene planes, such as the foreground, ground, and sky "folds" them to create a 3D pop-up shape, and then applies image textures to the 3D shapes. The system then synthesizes novel views interactively by texture-mapped rendering of the 3D pop-up shapes. (source: www.cs.uiuc.edu/homes/dhoiem/publications/popup.pdf)

stored picture of that object with the target image. The system does not render an explicit 3D model for each object, but instead maintains a large database of object photographs. As it also computes and matches automatic scene lighting conditions, and refines the object segmentation to match any scene-specific occlusions, the final compositing results look quite convincing. As with scene completion, the results replace elaborate synthetic objects whose subtle flaws may lack 'realism' with actual photographs instead; if we can gather enough photographs and access them promptly, we can simply bypass difficult synthesis problems. While results are remarkably good, they cannot entirely replace synthesis because we may never have 'enough photographs' online to represent all objects under all possible viewing conditions.

Coarse Geometry Estimation

While the recovery of dense 3D models from image sequences remains difficult for arbitrary scenes, sometimes coarse geometry is sufficient for simple scenes of man-made objects. As we saw with "Photosynth" and "Tour Into the Picture", just a few well-chosen points are planes can supply enough visual information for convincing 3D navigation, and we do not need dense geometric or photometric models except to render much more geometrically complex scenes. Similarly, Hoiem et al. [186] introduced "Automatic Photo Pop-up" that enables users to move freely in a 3D environment constructed from a single photo. Though limited to scenes with vertical walls, right-angle corners between walls, and flat or geometrically simple floors, their system is suitable for building interiors and exteriors, cityscapes, oceans or lakes and docks, and a very limited subset of natural scenes.

Building on the pioneering work 'Tour Into the Picture" by Horry et al. [188], their system identifies commonly-encountered planes within a scene (e.g., ground, vertical, or sky regions) by integrating multiple geometric and photometric cues. They create a "pop-up", similar to those in children's books, by partitioning the image into a set of texture-mapped billboards. They identify each billboard region by finding line-segments in the image that trace along the boundaries between vertical walls and the local ground plane, construct a vertical wall there, and then 'fold' along the floor seam to form a level ground plane, or a floor consistent with other floor seams in the image, as shown in Figure 7.23. Unlike the manual marking and segmentation required by "Tour Into the Picture", edge-finding in the image is enough to construct a coarse 3D model; users simply load the image, and after the system automatically constructs its system of 3D "pop-up" billboards, users can explore the scene in 3D immediately. While reliable for photos with clear and simple geometric boundaries, the system relies on automatic image segmentation that may fail for some images due to (1) edge labeling errors, (2) polyline fitting errors, (3) model assumption errors, (4) occlusion within the image, and (5) erroneous horizon estimation. However, users can easily identify most of these problems and correct them with manual editing of the segmentation results.

As with many examples seen in this section, computer vision methods are sometimes fragile, and may require manual intervention to correct errors and ambiguities. With suitable user interfaces, these corrections are easy and quick, but might be eliminated altogether as our cameras advance to capture more information about a scene than just a pixel map of the image formed behind the lens.

Texture Synthesis

Inpainting restores missing or damaged portions of images or videos, and borrows its name from centuries-old techniques developed by painting conservation experts (and abandoned in the late 19th century in favor of preserving the artists' original work by preventing new deterioration). When cleaning was not enough, when bugs ate holes in the canvas or the wooden backing, or when old paint curled, blistered, or fell away, early conservators would often maintain works of art by applying color-matched paints to "in-paint" the missing or damaged areas, restoring the painting's original appearance as accurately as their skills permitted, and inferring the missing brushstrokes from their surroundings and from a high-level knowledge of the original artist's method and style. Since the 1980s, researchers have developed semi-automatic digital methods to remove film scratches, repair damaged photographic prints, and remove text, grime or graffitti from digitized images. As these algorithms lack higher-level knowledge of the scene content they can only supplement the work of experts, but these semi-automatic methods may prove suitable for casual users to clean up old photos, typically plagued by scratches, water-spots, tears, cracks, creases and other artifacts. This section briefly reviews some of the

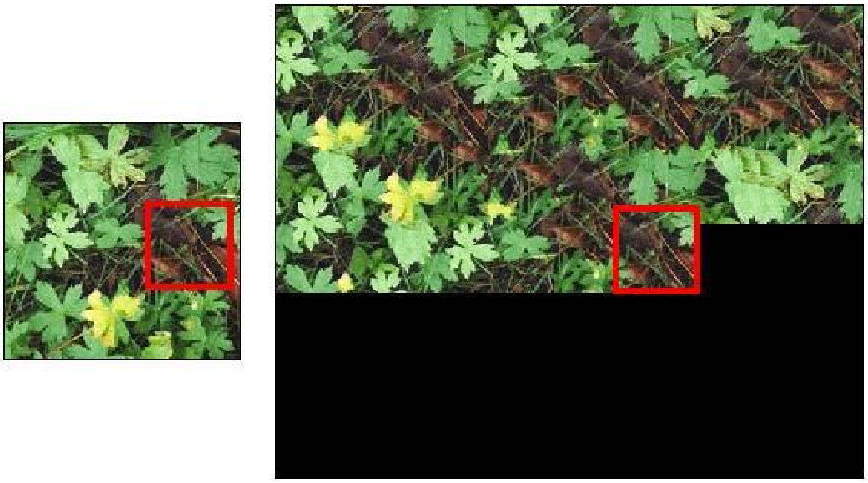


Figure 7.24: Image quilting for texture synthesis and transfer [103]. Image quilting synthesizes large textured regions from a small textured patch by merging multiple patch copies that were offset, overlapped, aligned, and trimmed to fit the region's boundaries and geometric constraints. (left) The input texture, with its small source patch outlined in red. (right) Synthesized result region, made from randomly chosen input blocks. Each block is added in raster order (i.e., from left-to-right and top-to-bottom). Overlapping regions between blocks are blended by finding a minimum cost path through the overlapping region (e.g., minimizing gradient magnitude similarity). (source: graphics.cs.cmu.edu/people/efros/research/quilting.html)

key recent works in inpainting and data-driven texture synthesis.

As defined by Efros and Leung,

> The problem of texture synthesis can be formulated as follows: let us define texture as some visual pattern on an infinite 2D plane which, at some scale, has a stationary distribution. Given a finite sample from some texture (an image), the goal is to synthesize other samples from the same texture. Without additional assumptions this problem is clearly ill-posed since a given texture sample could have been drawn from an infinite number of different textures. The usual assumption is that the sample is large enough that it somehow captures the stationarity of the texture and that the (approximate) scale of the texture elements (texels) is known [104].

A wide variety of texture synthesis algorithms have been proposed over the last decade. Early on, parametric model-driven approaches were used. For example, the inpainting approach by Bertalmio et al. [61] numerically solves a PDE in order to extend image curves that arrive at a boundary into the empty region while preserving their angle of arrival. In contrast, Efros and Leung [103] use a non- parametric, data-driven approach. Their algorithm functions by "growing" a texture, pixel by pixel, outwards from an initial seed by using the seed's surrounding colors

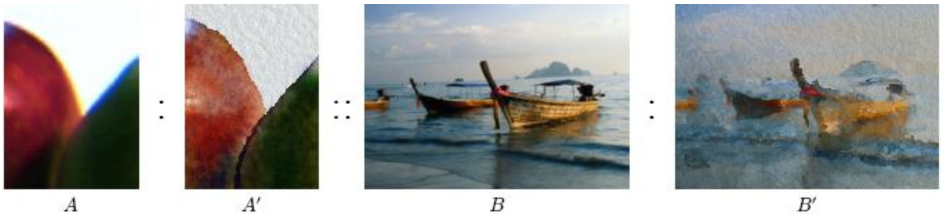


Figure 7.25: Image analogies [184]. An “analogous” image synthesizer makes B' similar to B in the same way that A' is similar to A. (source: www.mrl.nyu.edu/publications/image-analogies/)

as an index into a large, pre-constructed dictionary. For each set of surrounding colors found in the texture sample, the dictionary holds a histogram of the pixels that actually have those surrounding colors. We will choose a color from that histogram at random, but not all colors are equally probable; we set the likelihood of choosing each one to match the histogram, so that as we synthesize new pixels we will not change the histograms in the dictionary.

While effective for textures with small unstructured features, this pioneering texture synthesis approach is quite slow compared to follow-on efforts: within a few years, Ashikhmin [46] devised a simple fast method that could synthesize texture at interactive rates and permitted users to ‘paint’ textures on-screen with mouse-strokes. In addition, the size of the dictionary grows very rapidly with the size of the neighborhoods it catalogues, and the the time required to create these dictionaries can easily stretch to hours as the neighborhood size exceeds a few tens of pixels. Unsurprisingly, the synthesis method does not work well for textures with features substantially larger than the neighborhood used for that dictionary. In addition, their algorithm occasionally “slips” into the wrong part of the search space and starts growing garbage or repetitive structures. Extensive later works addressed most of these limitations, such as the early follow-on work using a multi-scale synthesis by Wei and Levoy [422].

While texture synthesis methods generate more texture area from one small example of the texture we want, Hertzmann et al. [184] proposed “Image Analogies” to generate more stylized images from a single example of a style. As shown in Figure 7.25, image analogies emulate linguistic analogies: given an image A and a stylized version of that image A', find the similarly stylized image B' given an image B. Successful image analogies permit us to emulate what appear to be ‘high level’ abstractions with statistical processes alone, and let us experiment with transferring styles from famous artists to photographs, or even ‘reverse’ stylistic changes by swapping A and A'.

While texture synthesis ‘learns’ the likelihood of various pixel colors given its neighbors, texture analogies learn the likelihood of changes to a neighborhood, given its neighbors. Hertzman et al. measures similarity between input image pair A and A' by approximating a Markov random field model of variations in pixel values and the responses of a multi-scale hierarchy of direction-sensing (steerable) filters. They estimate and store the joint statistics of these values over small neighborhoods around each pixel in the input image pair. To construct the stylized image B' they begin with a copy of B and then modify each pixel with a random perturbation that enables B and B' to satisfy the joint statistics recorded for the A and A' image pair.

7.4.4 Indexing and Search

Metadata tags are added bits of machine readable information that describe digital image features that we can't extract directly from its pixel grid. Just like the digital image formats themselves, metadata tags lack any standards or conventions for formats, tag names and values, or even where to store the metadata tags. Some systems store metadata as part of the image file itself (e.g., EXIF, TIFF), others create a companion file with the same name as the image file but with a different extension, others create a 'thumbnail' file to store tags for all images in a single directory (e.g., Canon cameras), others create a single database of tags that describe sets of images in selected directories, drives, computers, or networks (e.g., Google's Picasa system).

However, nearly all modern digital cameras manufacturers support metadata tagging within the EXIF standard, and their cameras create image files that hold machine-readable detailed information on the camera's settings. In the late 1990s the Japan Electronic Industries Development Association (JEIDA), a forward-looking consortium of digital camera manufacturers, established committees to devise a single industry-wide file format that would ensure compatibility and interoperability for digital photos. Together they devised the 'Exchangeable Image File Format' (EXIF) specification that augments the existing JPEG, TIFF (revision 6.0), and RIFF WAV image file formats with a set of carefully devised metadata tags. The original 1998 specification was updated to version 2.2 in April, 2002; while not maintained by any standards-setting organization, its widespread ongoing use by camera and both commercial and open-source software ensures its stability.

In each file, EXIF tags may describe camera manufacturer, model, lens descriptor, date, time, aperture, exposure, focal length, metering mode, and ISO speed, thumbnail previews, and copyright information for a given image. More recently, cameras with built-in GPS receivers, tilt sensors and compasses record geographical coordinates and camera aiming directions as well. Aggregating tags from large sets of photos can help form initial estimates of scene contents and viewpoints, even without GPS metadata as demonstrated by Snavely et al. [375]. As they discovered, this data is not always reliable, because some photographers may edit their photos by cropping or resizing with software that does not update the EXIF tags set by the camera. We believe metadata tags from EXIF are an excellent start, but could be much richer and more useful for indexing and browsing large photo collections, particularly as cameras gain better abilities to estimate camera pose, lighting, and image content.

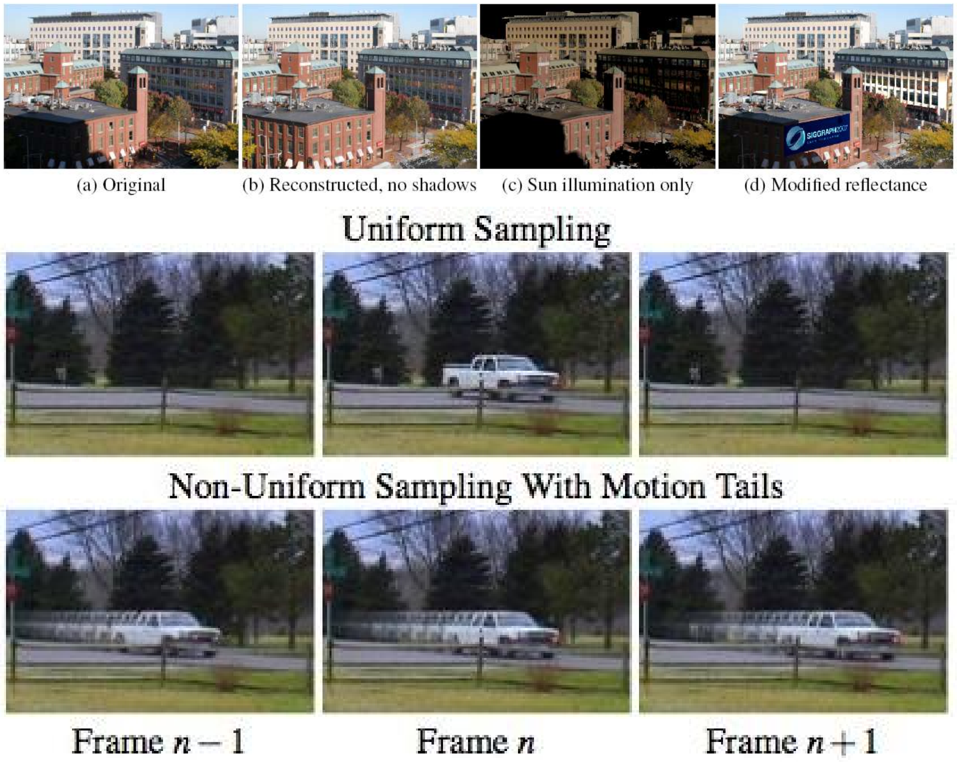
For example, the growing popularity of automatic face-detection (Section 7.1) in current generation digital cameras may eventually permit your camera to recognized frequently-photographed people and tag them by name automatically, as some existing photo-browsing software can do now (e.g., Google's Picasa and Apple's iPhoto). Similarly, cellular towers can triangulate and determine the location of existing camera-equipped ‘Edge’ and ‘G3’ phones even without GPS; why not relay that location information to those phones for use in tagging photos? New biometrics made possible by cameras, such as fingerprints and iris scans could also augment EXIF data with tags to identify the photographer. As both Hays and Efros’ img2gps [179] and Jacobs et al. [198] showed, adding accurate timestamps can also assist in determining geolocation, and the directions of shadows from sunlight can assist in determining camera aiming direction. In the future we expect auxiliary tilt sensors to become more common, facilitating not just user interface improvements, but also post-capture scene recovery in the style of Photosynth.

7.5 Image Sequences

Aggregated on-line photo collections present new sources of weakly-structured image data; if organized according to 3D location (e.g., as with PhotoSynth) we can choose our own paths to explore the scene, selecting photos in any order we wish to learn more about scene features we find particularly interesting. Currently, we lose most of this freedom with video sequences. Video browsing only permits us to select individual video clips, and playback shows us only one fixed temporal sequence; we have no choice of camera locations, movements or viewpoints. At best we can only lengthen an interesting sequence using pause and slow-speed playback, or shorten a boring one using fast-forward or skip-ahead in playback. How might we gain more choices in creating and exploring the contents of one video? How might we better aggregate the visually important contents of multiple video sequences? In this section we review some of the historic and recent work on video processing, particularly for time-lapse imaging, motion depiction, and video summarization.

7.5.1 Time-lapse Imaging

Time-lapse photography summarizes long-term visual changes in much shorter video sequences, enabling us to stretch our useful attention span from tens of seconds to hours, days, months, or years. As video and motion pictures consist of individually photographed frames recorded and played back at a constant update rate, time-lapse sequences are simply those recorded at a slower frame rates than their playback. Setting the ratio between recording and playback rate determines the apparent temporal rate; 30Hz playback of a video recorded at one frame per minute shows changes 1800 times faster than normal. Time-lapse experiments began even before motion pictures with flip-book animations, and appeared at the dawn of commercial motion pictures in Georges Méliès’ motion picture Carrefour De L’Opera (1897). Popularity for educational use began in part due to F. Percy Smith’s stop-motion flower-blooming films and studies of the eerie growth of slime molds (see: www.screenonline.org.uk/people/id/594315/), but many others adopted the method for effective illustration of growth, celestial movements, weather, and more. Although he was a stop-action pioneer (1915-1918), Roman Vishniac’s later contributions to scientific photography advanced stop-action, light-interruption, and polarization methods for living organisms of all kinds (see: en.wikipedia.org/wiki/Roman_Vishniac).



(a) Original (b) Reconstructed, no shadows (c) Sun illumination only (d) Modified reflectance

Figure 7.26: Time-lapse video processing. (Top): Factored Time-Lapse Video [386]. (a) A single frame from an outdoor time-lapse sequence factored into its sunlight(directional), skylight(non-directional), and reflectance components. (b) Shadow removal reconstructed from sky light and reflectance estimates (c) Sunlight-only component. (d) Synthetic markings added by modifying the reflectance estimate, then multiplying by the sunlight and skylight estimates. (source: people.csail.mit.edu/wojciech/FTLV/index.html) (Bottom):from Computational Time-Lapse Video [60]. The first row shows three consecutive frames in a time-lapse sequence with uniform temporal sampling. The bottom row shows three consecutive frames in a time-lapse sequence with non-uniform temporal sampling, used to preserve the key event of a truck passing, with synthetic motion trails added to indicate high speed motion. (source: ericpbennett.com/TimeLapse/BennettMcMillan07.pdf)

During time-lapse photography, exposure time choices determine the amount of motion blur in each frame; more blur helps reveal the velocity of moving objects as they pass through the scene. Long frame times and very short exposure times undersample the scene, and may cause temporal aliasing (e.g., airplane propellers, wheels and tires that stop and spin backwards). Long time-gaps between exposures also enable actors and directors to make seemingly-impossible 'stop-motion' animation films by hiding intervening actions between each frame, such as re-molding plasticine characters (e.g., Aardmann Animation's "Wallace and Gromit" series), bending armatures and replacing molded, pre-painted heads (e.g., Disney's "The Nightmare Before Christmas") and countless entertaining student videos with actors that move slightly for each frame or leap into the air just before each shutter-release (e.g., 'Human Skateboard' on YouTube www.youtube.com/watch?v=MtbQ4J3RfQ8&feature=related).

Time-lapse or frame-by-frame photography for film or video usually relies on either a manual shutter or an 'intervalometer' to take photographs at uniform time intervals, and may miss important events. Digital video provides us with several new kinds of flexibility; as each frame costs almost nothing to capture, we can record full-frame rate (e.g., 30Hz) digital video, then create 'time-lapse' sequences computationally, choosing the best frames to keep, or combining those we captured to construct fewer frames from them. Recently Bennett and McMillan [60] explored some of these methods, and described several ways to construct more useful and aesthetically pleasing time-lapse sequences by adaptive sampling and multi-frame video processing.

First, they address temporal aliasing, (the information loss caused by skipping frames) in time-compression of surveillance videos. At low frame rates, these cameras may capture only a single frame of an important event such as an individual entering or leaving an area. Instead, they propose recording frames at ordinary video frame rates, but then discarding most of the frames that contain no visually-significant changes, forming an event-driven time-lapse sequence that moves rapidly through long stretches of uneventful time. They used a dynamic programming method to optimize video temporal sampling for best match to user's desired frame rates and specifications of visual events. By making weighted sums of uneventful video, they created a virtual shutter effect that adds synthetic motion blur to a time-lapse sequence (see Figure 7.26). Their implementation presumes the surveillance camera is fixed and observes a mostly-static scene, as they use frame-differencing metrics to detect significant scene changes.

Over long time periods, fixed video cameras viewing fixed outdoor scenes gather a surprisingly information-rich collection of images. Natural illumination varies dramatically but predictably over time, due to changes in illumination direction from sun movements (both dawn to dusk and season to season) and also from weather and artificial illumination from street lights, interior lights, and moving lights from cars and other vehicles. Sunkavalli et al. [386] examine visually meaningful compact *representations* for these time-lapse sequences.

As they observe, a camera that takes a picture every 5 seconds yields 17,280 images per day and nearly 1 million per year. Their studies showed that while naïve PCA methods can summarize these images using a few low-dimensional basis images, they can construct a much more meaningful and useful decomposition by

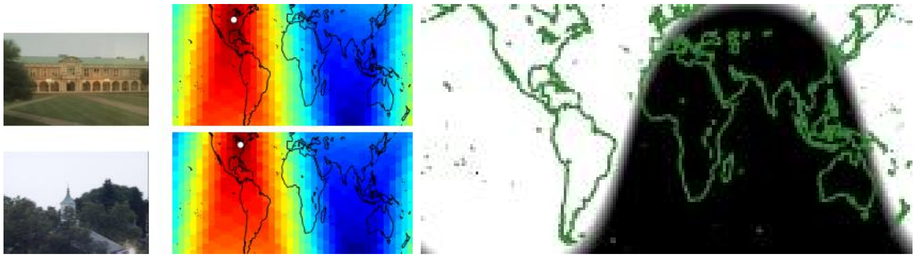


Figure 7.27: Toward Fully Automatic Geo-Location and Geo-Orientation of Static Outdoor Cameras [198]. (source: ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4544040)

factoring the images into estimates of reflectance, shadow, directional (sunlight) and non-directional (sky-light) illumination, whose weighted products formed accurate approximations of the scene under any desired time-of-day or degree of overcast. By using the time-varying intensities of each pixel independently they first identify the onset (the edge) of moving shadow boundaries in the scene to label pixels as in shadow or direct sunlight. They then apply matrix factorization to the volume formed by their time-sequence of labeled images to find a set of basis curves that, together with per-pixel offsets and scales, describe the image set. From these, they can estimate surface normals and surface reflectances with sufficient accuracy to permit some 3D image editing in addition to recreating any desired relighting. Such representations are useful for computer vision tasks such as background modeling, image segmentation, and scene reconstruction, as shown in Figure 7.26.

Time-lapse imagery of outdoor scenes has utility for geo-localization as well. Similar to the img2gps method of Hays and Efros [179], one would expect that the natural changes in illumination captured by a simple webcam would be sufficient to localize the camera with some degree of certainty. Jacobs et al. [198] describe just such an algorithm, with the additional constraints that the images are accurately time-stamped and the camera remains static (see Figure 7.27).

First, they find a sequence of images where the camera has not moved. Second, they compute a "canonical day" decomposition (similar to a PCA decomposition, but consistent across multiple cameras). Third, they create a full hemispherical sky intensity map for a given geo-location and time of day. Finally, an optimization process searches to find the orientation of the camera relative to this hemisphere which maximizes the correlation between the sky pixels in the image and the predicted sky intensity.

While image-based geo-location may not provide the accuracy of true GPS, we believe it is remarkable how accurate state-of-the-art system are—for instance, the recent method of Jacobs et al. already reports a localization accuracy within 50 miles for static cameras.

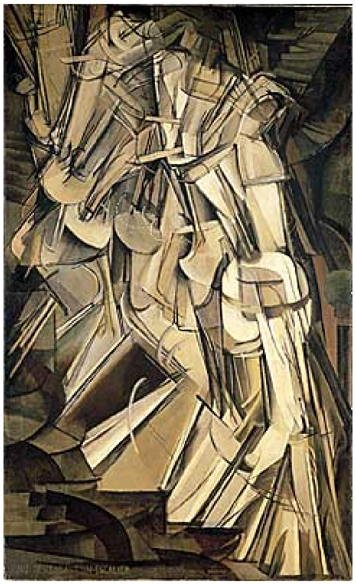


Figure 7.28: Marcel Duchamp. Nude Descending a Staircase, No. 2 (1912). (source: en.wikipedia.org/wiki/Nude_Descending_a_Staircase,_No._2)

7.5.2 Motion Depiction

How can a digital image best depict motion? Video sequences try to record motion objectively with an optical copy, but artists show us many much more inventive methods to depict motion in stationary, moving, or responsive displays from oil paintings to Alexander Calder's mobiles and wire drawings in 3D. Even subtle changes in composition can create or destroy our impression of movement in a single image, yet the sheer diversity we find in the best motion depictions suggest that computational approaches may uncover new ones. The radical yet playful works by Duchamp such as "Nude Descending a Staircase, No. 2" shown in Figure 7.28 show abstractions and multiple copies of moving forms; image sequences by Edgerton and Mili (see Chapter 5) remove abstractions and show us the details of forms that change too rapidly to see with our own eyes. Some recent computational photography-related papers seem to emulate these works, such as the shape-time images by Freeman et al. [140], and others give us new tools for abstraction and selective detail. Can we compute a continuum of motion depictions that let us choose or vary them between detailed and abstract? What varieties of abstraction are comprehensible but still computable?

Our motion-depiction choices are much broader than just variations on motion-blur such as strobed multi-image photographs or trajectory lines. Liu et al. [250] recognized that while 'time-lapse' photography can amplify our ability to comprehend changes over long time-scales, film-like photography offers no ability to amplify motion itself, no sensible way to exaggerate very small-amplitude movements we might otherwise miss. As shown in Figure 7.29, they first group pixels

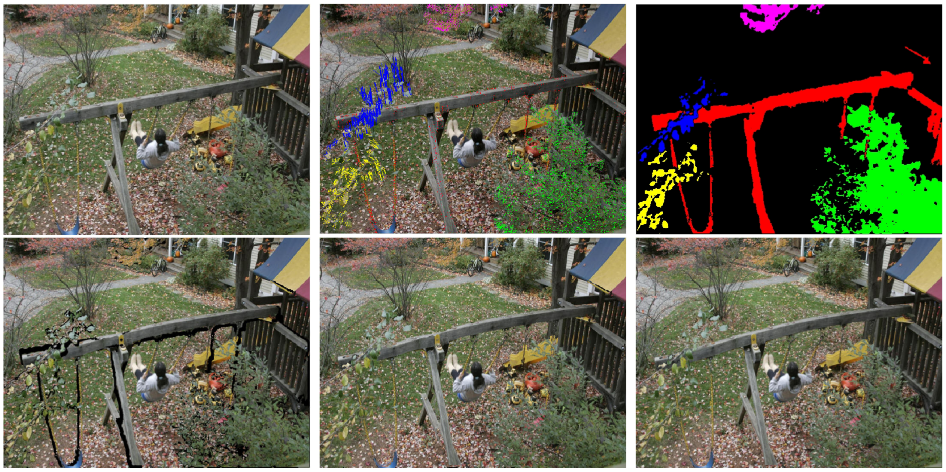


Figure 7.29: Motion Magnification [250]. (top left) One registered input frame. (top middle) Clustered trajectories of tracked features. (top right) Related motion layers (bottom left) Exaggerated motion causes holes (in black) (bottom middle) Texture synthesis fills these holes. (bottom right): User-modified segmentation map corrects errors. (source: people.csail.mit.edu/celiu/motionmag/motionmag.html)

into layers that each contain moving parts, and estimate motion vectors for each pixel in each layer. Users then explore these layers and selectively amplify motion vectors for particularly interesting layers (e.g., boost the bending movements for the swing set's support beam). As amplified movements can displace some pixels substantially and leave behind empty areas in some of the video frames, they fill these holes either by interpolating static features at this location from earlier or later frames, or they apply texture synthesis to create plausible replacements for occluded scene content. The results let us see very subtle movements that are both visually interesting and suggest a structures' tolerance for its current loads; for example, a motion-amplifying surveillance cameras for large highway bridges might help warn engineers of unseen structural weaknesses or resonant movements. While quite effective, camera resolution and noise currently limits the method's practicality for low-light or high-contrast settings, and extensive data fitting may limit its temporal sensitivity as well.

7.5.3 Video Summarization

Just as conventional video provides just one rather limited depiction of motion, we have no one 'best' method for summarizing video's visual contents. While modern graphical operating systems present thumbnail views of images, documents, and other resources, the creation of such preview images for videos remains an open topic. How can we compress a short animation? how can we best represent the movement of the subjects and the camera throughout? Goldman et al. [157] observe that video summarization is closely related to the *storyboarding* process using in film production, and emulate it for summarization. As shown in Figure 7.30,

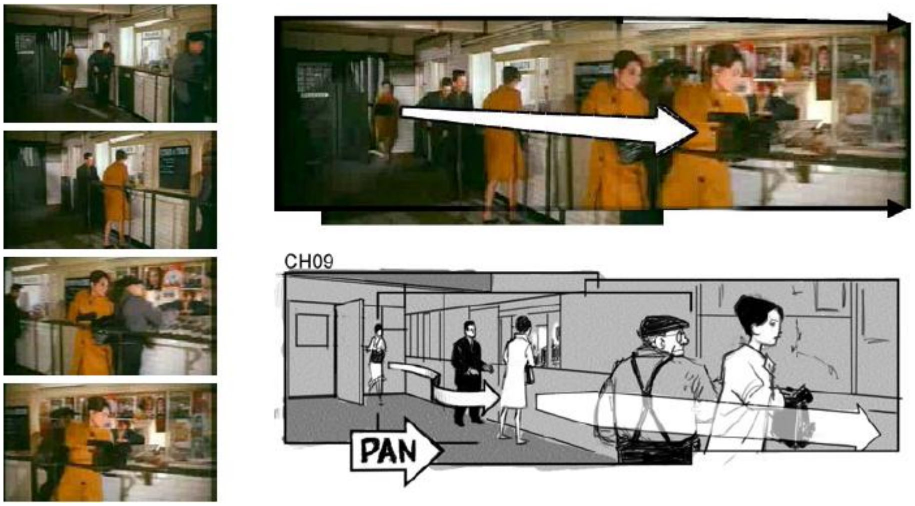


Figure 7.30: Schematic storyboarding for video visualization and editing [157]. “(Left): Four still frames from one shot of the film “Charade”(1963). (Top): Schematic storyboard composed from the frames at left; the large 3D arrow, indicating motion toward the camera, was identified and rendered without determining the 3D location of subject or camera. (Bottom): Professional storyboard artists’ rendering of the shot, composed with Adobe Photoshop and Corel Painter (Image credit: Peter Rubin). (source: grail.cs.washington.edu/projects/storyboards/)

professionally-trained graphic artists create storyboards to help plan the most effective shots to convey a story or a sequence of ideas. The storyboard artist uses a succinct graphic language in a sequence of drawings that describe the composition of each shot in the video. Sparse, deft line drawings depict subjects, sets and backgrounds, and bold, annotated arrows indicate actor and camera movements. Unlike the entirely-artist-driven storyboard method, Goldman et al. render schematic storyboard layouts from input video clips using a minimum of user interaction. They draw upon a number of core computer vision technologies, including tracking, segmentation, and keyframe selection to make similarly sparse, simplified backgrounds and to make action-indicating arrows from photographed motions. As observed by the authors, the creation of such storyboards has wider applications to “...video editing, surveillance summarization, assembly instructions, composition of graphic novels, and illustration of camera technique for film studies.” Others have addressed some of the topics, such as “Salient Stills” [393], one of the earliest computational attempts to depict video content in a single image. Assa et al. [47] also applied keyframe select for video summarization. As these projects demonstrate, content-driven image-processing techniques can already automate some of the tedium of simplified sketching and routine drawings and transcriptions, freeing artists for far more cognitively valuable depiction tasks.

7.6 Multi-image Fusion

Any computational technique that requires more than one photograph employs some form of multi-image fusion to merge their visually significant content, but 'fusion' methods assemble an image from pieces of many others. For example, Chapter 3 describes fusion by photometric blending of overlapping images to stitch together seamless panoramic images of static scenes. Chapter 3 shows how to extend depth-of-field by combining images with different planes of focus, but at the cost of multiple exposures. Chapter 5 explained how to fuse day/night and flash/no-flash image pairs to enhance the overall aesthetics, reduce noise, and increase legibility of the photographed scene. This section briefly reviews other multiple-image fusion methods, including multi-resolution image pyramids, wavelet-domain blending, and gradient-domain blending.

Blending overlapping image regions requires synthesizing a coherent transition from one image to another. Despite potential mismatches between these images, blended results should depict all key local textures and shape cues just once, and maintain consistent color, noise, and texture across the entire transition region from one photograph to another. This easy-to-state problem has proven surprisingly difficult to solve robustly, and routinely requires substantial user interactions for best-quality results. A wide variety of methods rely on multi-resolution image decompositions described well in early papers by Burt and Adelson (e.g. [73]).

First, users supply a pair of aligned images and a boundary that specifies where to make a 'seamless' transition from one image to the other. Before blending, Burt and Adelson first construct a Gaussian pyramid from each image. A Gaussian pyramid is just a source image (level zero of the pyramid) followed by sequence of progressively smaller images or 'levels' stacked on top of it that we can imagine as a pyramid of pixels. Each step up the pyramid shrinks the image size to half its width and height, culminating in a one pixel image at the top. This single apex pixel (or level or image) summarizes the entire image, and with each step down the pyramid we find an image with more details than the last. From this, they construct a Laplacian pyramid: they double the width and height of each Gaussian pyramid level by interpolation and then subtract it from the next-lower level image to remove the shared information and leave behind only the finest details held in that level. The resulting band-pass pyramid holds an orderly sequence of image details from finest (at the pyramid base) to coarsest (at the one-pixel apex). We can then collapse the pyramid to reconstruct the original image perfectly: start at the apex, double the layer's width and height, add it to the next-lower level, and continue towards the base.

Instead of cutting source images along user-supplied boundaries to merge them, Bert and Adelson's multiresolution method cuts the Laplacian pyramids, pieces together each level into a new pyramid, then collapses it to form a merged image. The two images' coarse, low-spatial-frequency components blend together smoothly over a wide transition region around the boundary, while progressively finer details blend over progressively narrower transition regions. Laplacian pyramid blending creates visually pleasing transitions particularly well-suited for dissimilar surfaces on geometrically smooth, well-aligned shapes, such as their celebrated result that blends an apple and an orange.

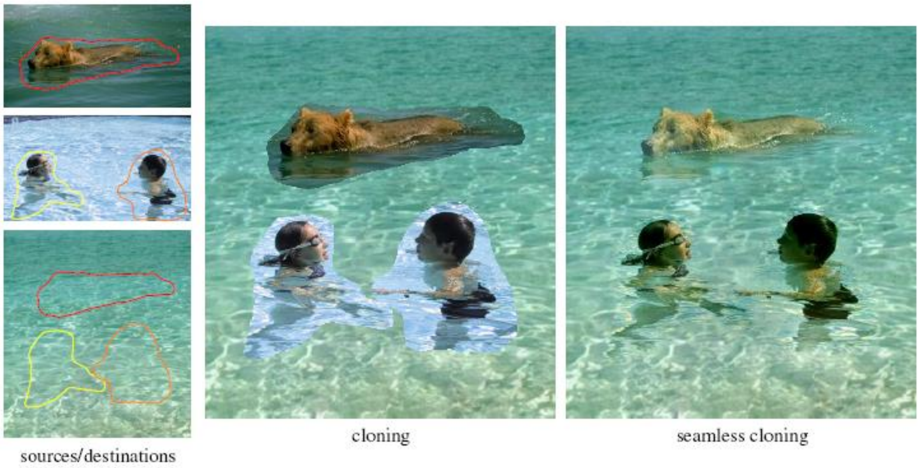


Figure 7.31: Poisson Image Editing [317]. Instead of editing pixel values, Perez et al. showed that editing image gradients offers greater flexibility, with results that are often far more visually plausible. Gradients record only local changes, rather than absolute pixel values or colors, and thus permit users to transfer textures and visually significant variations without the color mismatches caused by copying their pixel values. (source: www.irisa.fr/vista/Papers/2003_siggraph_perez.pdf)

In 2003, Pérez et al. [317] presented Poisson image editing, a less elaborate alternative to multiresolution methods that retains most of its advantages, and also performs well for geometrically complex boundaries. Poisson image editing manipulates only the difference between adjacent pixels instead of an entire pyramid of pixel differences, and then applies a Poisson solver to these modified differences to reconstruct the result image. Instead of pyramids, the method converts each source and destination image into a vector-valued 'gradient' image that holds the forward differences between each pixel and its eastward and northward neighbors (e.g., replace each scalar pixel value $I(x, y)$ with the 2D vector value $[I(x+1, y) - I(x, y), I(x, y+1) - I(x, y))$. As shown in Figure 7.31, users draw a closed boundaries to specify source-image regions we wish to transfer to a destination image (e.g., the red-outlined bear, the yellow- and orange-outlined swimmers). Over-writing destination gradients with source image gradients transfers only their local changes rather their pixel values, and often permit seamless matching of source and destination regions because it ignores globally-defined discrepancies such as lighting or color differences.

As gradients describe only the changes in color and intensity, colors at the boundaries of the pasted region always match the destination image, but achieve this match by re-coloring the regions' interior. The method performs very well for cut-and-paste operations for objects on very similar backgrounds, but dissimilar color or textures can induce objectionable artifacts. For example, a brown bear surrounded dark blue-green river water looks a bit sun-bleached but still retains its bear-like colors if we paste its gradients into light blue-green swimming-pool water as shown in Figure 7.31. However, that same brown bear may turn a sickly gray-green and will swim in a patch of magenta water if we paste its gradients onto a strong red background, because the Poisson solver will apply our pasted water-to-bear gradients to red colors instead of water-like colors as it constructs our output image. Mismatched background textures also induce halo-like artifacts: if we paste our water-and-bear gradients onto the gradients of colored random noise, the Poisson solver may construct rainbow-like fringes around the pasted region's boundaries due to nonzero curl and Dirichlet boundary conditions. [1]

Algorithms for multiple-image fusion continue to evolve with each publication, in part due to the close tie of hardware and software in computational photography. Regardless, certain families of solutions such as Poisson image editing and variants (e.g., Photoshop's 'healing brush' tool, introduced in 2001) have become widespread. We further discuss some of the practical details of such techniques in the Appendices, including Graph Cuts and further gradient-domain image manipulation techniques.

7.7 Conclusion

As we've learned in the previous chapters, conversion of raw sensor outputs into picture values involves sophisticated processing. While existing digital cameras perform 'de-mosaicking', (interpolating the Bayer grid), remove fixed-pattern noise, and hide 'dead' pixel sensors, recent work in computational photography leads further. The main idea is a "co-design" of optics and processing for optimal capture and post-capture resynthesis. Such co-design has emerged as a common theme in coded photography.

In some cases, outputs of ordinary cameras can be manipulated by incorporating recent advances in image processing and computer vision. Advances in geometric operations on large sets of photos now allow even novices to explore their image content in 3D [376]. In such a data-rich environment, in which millions of photos on any object can be archived and retrieved at little cost, automatic processing must become a central feature of computational photography. Modern processing methods use filtering to reduce the impact of noise, detect and recognize important image features, such as faces, as well as categorize and automatically assign higher-level labels. To remove blur due to defocus or motion, recent algorithms solve the ill-posed blind deconvolution problem by enforcing certain natural image statistics on the solution.

The statistical 'priors' exploit the common observation that there are large gradients at sparse image locations or that the histogram of gradients of natural scenes is sharply peaked at zero. Ever-increasing online photo collections are allowing rapid progress in data driven, probabilistic and inferential methods. Cartoon

[1] An intensity image with N pixels always has one unique set of $2N$ forward differences. However, not all sets of $2N$ gradients describe an image, because we can specify gradient field with non-zero curl. In these images, the pixel differences along at least one closed path (e.g., around the edge of the image, or around any loop connecting together a sequence of adjacent pixels) don't add up to zero. For these 'impossible' gradients, a Poisson solver will find an image that matches them best in the least-squares sense, causing smooth, broad, halo-like errors that spread outwards from the curl-inducing regions.

rendering from photos is redefining what it means to be ‘photorealistic’.

Chapter 8

Future Directions

What will a camera look like in 20 years? What will Photoshop look like in 20 years? Will we use gigapixel cameras? How will movie making and news reporting change? Will we be able to monitor our own health with nano-scopes built into our mobile devices? Can cameras protect our elbows and shins from sharp-edged furniture or other sources of painful bumps and bruises? How will we stay in touch (or maintain telepresence) with our loved ones? It is difficult to say how the field will evolve. To help understand the future of Computational Photography, four broad trends deserve your attention: (i) Modern developments in fields other than photography introduce novel improvements to cameras, imaging, and non-imaging sensors. (ii) Future photography will be decided not just by capture-side innovations but also by new methods to share, display and interact with visual results. (iii) Current scientific, industrial and medical imaging that were previously considered exotic and unwieldy procedures such as tomography, confocal and coded-sensing methods will blend rapidly with casual camera use and photography. (iv) Finally, unusual configurations of multiple cooperating devices (cameras, lights, scenes, wireless services, online databases) may dominate or replace the current photographic conventions and how everyone captures a visual experience.

8.1 Modern Imaging

Over the last decade, we have seen a rapid growth in professional and consumer photography. Digital cameras, introduced in early 1990s have already reached over 100 million in sales per year. Thanks to such large numbers, photography research and techniques were mainly driven by consumer demand for smarter cameras and superior imaging. However, camera technology is rapidly developing beyond photography for other applications. On the one hand, applications such as direct remote surveillance, teleconferencing, online live video and sharing pre-recorded video, etc. require little or no additional computational abilities beyond what most of us already have in a desktop computer. On the other hand, industrial vision, computer vision, robotics and mining of online collections involve sophisticated computations that still demand far greater resources, and may remain out of reach without further microprocessor advances, GPU co-processing or distributed cloud computing enabled by widespread availability of high-speed networks. The hardware (cameras, sensors, optics, processing, lighting and displays) and software developed for these non-photographic applications all impose strong influences on computational photography.

In addition, emerging billions of mobile camera phones have spawned a new camera culture that changes the rules of visual communication. Beginning scarcely 10 years ago [137], the number of digital cameras in mobile phones has skyrocketed from zero to over one billion. This is a fascinating time for camera and imaging research. What happens when a billion people worldwide become empowered with tools of visual communication? The goal of Computational Photography research is not simply to use these cameras, but to amplify their shared abilities and develop the next generation of imaging devices, algorithms and software.

8.1.1 Wafer-level Cameras

Unsurprisingly, a main area of innovation is cost reduction. In 2008, about 80% of the 1.2 billion mobile phones sold had a camera inside. A camera has become an indispensable feature of several devices including mobile phones, game consoles such as Xbox, Playstation or Nintendo Wii and notebook computers. Wafer-level optics is a novel technology that is designed to meet the demand for smaller form factors, higher resolution and cost-effective pricing in the next generation of camera phones. The optical components are fabricated on glass wafers in a manner similar to that of fabricating integrated-circuit chips on silicon wafers. The entire camera is aligned and assembled at the wafer level and subsequently diced to form individual camera modules.

In electronics, smaller is better. WLC creates smaller image pixels i.e. higher-resolution and smaller devices. Unfortunately smaller is not better in optics because the wavelength cannot be scaled down leading to diffraction and photon noise. Scaling down the optics translates to a smaller lens diameter (D) and shorter focal length (f). For the same light collection ability (F), the size of the diffraction-limited spots is independent of the lens diameter. However, the number of resolvable image spots—the space bandwidth product—is drastically reduced for very small lens diameters. Nevertheless, WLC approach allows a new opportunity to produce a unified optics-sensor architecture and the co-design can overcome traditional problems in optics (diffraction and aberration) or sensor (noise, well-depth, etc.).

Multi-aperture and Multi-scale Optics

The co-design of optics and sensors has shows ways to unconventional fusion of light bending and sensing architectures. As resolution of pixels approaches below a micron, there is limited gain due to the limits of conventional optics. There are two approaches being explored: multi-aperture and multi-scale optics. A multi-aperture (MA) architecture consists of an array of small submicron pixel imagers (apertures), each with its own integrated optics. By focusing the integrated optics onto an image plane formed by an objective lens in a region above the MA imager, the apertures capture overlapping views of the scene. The correlation and redundancy between

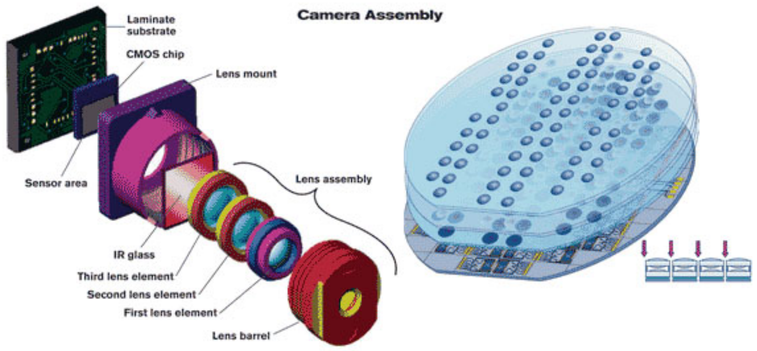


Figure 8.1: Comparison of conventional mobile phone camera fabrication (left) with wafer-level camera fabrication (right). Opto-wafers and CMOS wafers are mounted together and diced into individual camera modules. [Reinhard Voelkel, SUSS MicroOptics SA]

apertures, along with computation, provide several new capabilities, including: (i) simultaneous capture of a 2D image at higher resolution than the aperture count and a 3D depth map without the need for active illumination or calibration; (ii) simplification of the objective lens design; (iii) reduction of color crosstalk via per-aperture color filters; and (iv) increased tolerance to pixel defects.

The multi-scale design discusses strategies for increasing the information capacity of geometric-aberration limited lens systems by adding reprocessing optics near the focal plane. Defense Advanced Research Projects Agency (DARPA) is already planning 1.6 Gigapixel sensors as part of the ARGUS program to support unmanned air vehicle for aerial surveillance/ [1] DARPA also has a call for a 50 Gigapixel project under the MOSAIC initiative [396]. The theoretical diffraction limit for space-bandwidth product (i.e. FOV/angular resolution) is rarely achieved by a practical system due to lens aberrations. For example, a 1 cm aperture with 1 cm focal length (F/1 system) is diffraction limited to 310 Megapixels. But, real systems underperform by an order of magnitude (well below 30 Megapixel). So there is a new need for computational optics as well as defocus deblurring algorithms.

The hope is that with multi-scale design, the optical re-processing strategy will enable multi-gigapixel or even terapixel imaging through a single aperture [66].

8.1.2 Modern Optics

Wavefront coding is an imaging technique introduced by Dowski and Cathey utilizing joint optimization of a coded phase plate and digital postdetection processing. Phase plates originally introduced for extended depth of field using a cubic profile can be used for a variety of purposes. Companies are using it for correcting for chromatic aberration and wide field of view imaging without significant lens aberration.

[1] www.darpa.mil/IPTO/programs/argus/argus_approach.asp

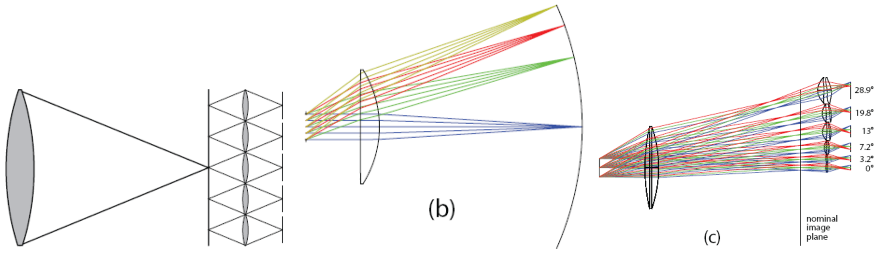


Figure 8.2: Optical pre-processing near the sensor plane (a) Multi aperture design adds new lenslet near the sensor for redundant imaging. (b) Traditional lenses create focused spherical focused image plane. (c) Multi-scale optics uses a location-dependent choice of microlenses to correct for the spherical aberrations. [David Brady Duke U.]

Electroactive Optics

Future lenses may eliminate the need for moving lenses back and forth for focus or zoom. Newer lenses can change shape based on applied voltage. The liquid lenses by Varioptic are based on the electrowetting phenomenon [398]. A water drop is deposited on a substrate made of metal, covered by a thin insulating layer. The voltage applied to the substrate modifies the contact angle of the liquid drop. The liquid lens uses two isodensity liquids, one is an insulator while the other is a conductor. The variation of voltage leads to a change of curvature of the liquid-liquid interface, which in turn leads to a change of the focal length of the lens.

A very similar electromechanical design involves Electroactive Polymers used by Optotune [397]. Several field actuated materials that change shape exist such as piezoelectrics and magnetostrictive materials. Electroactive Polymers, however, are superior in terms of produced strain, actuation pressure and specific energy densities. They use compliant electrodes that enable polymer films to expand or contract in the in-plane directions in response to applied electric fields (or mechanical stresses). The electrostriction of elastomeric polymer dielectrics with compliant electrodes is potentially useful as a small-scale, solid-state actuator technology. Optotune has shown that such polymers are well suited for tunable lenses. Both electroactive lenses described above currently suffer from limited lifetime and show hysteresis. Nevertheless, for Computational Photography these newer electroactive adaptive lenses allows fast change in focal length between successive images or within a single image, foveated imaging (spatially varying resolution) and phase-retardation for controllable phase plates.

Photonic Crystals

Current explosion in information technology has been derived from our ability to control the flow of electrons in a semiconductor in the most intricate ways. Photonic crystals promise to give us similar control over photons - with even greater flexibility because we have far more control over the properties of photonic crystals than we do over the electronic properties of semiconductors. Photonic crystals

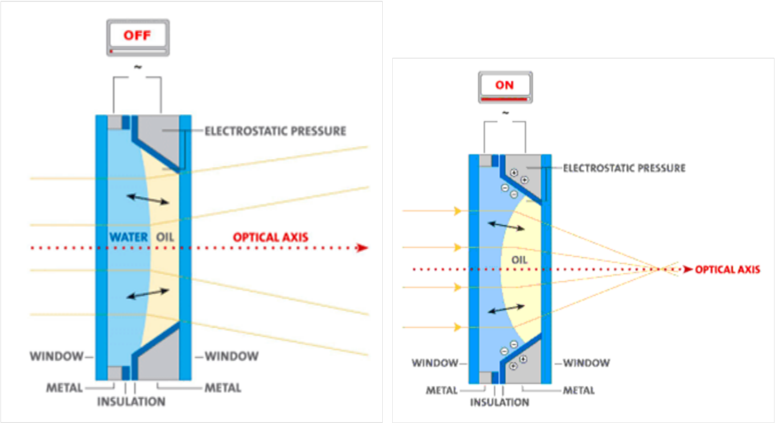


Figure 8.3: Liquid lenses from Varioptic use electrowetting principle to change the focal length.

(PC) are typically nanostructure material with ordered array of holes. A common configuration is a lattice of high refractive index (RI) material embedded within a lower RI material. The behavior is dependent on creating a high index contrast. PC are produced in 2D or 3D periodic structure. The highly periodic structures that blocks certain wavelengths. The gap or notch in wavelength is called Photonic band gap. Hence PC are also called Semiconductors for light and mimics silicon band gap for electrons. With photonic jets, it is possible to focus light to spot sizes below possible with conventional optics and below diffraction limits. Hooman Mohseni and his group at Northwestern University have shown a focusing spot one third the wavelength of light. This has been shown only at 3 micron diameters lenses but the hope is that it can scale to several millimeters.

Computational photography can exploit the programmable and highly selective/rejecting narrow wavelength filters that go beyond the traditional Bayer color grids on pixels. PC are also exploited for light efficient LEDs that could be part of narrow-wavelength flashlights. PC are used in optical communication to support extreme bandwidth via wavelength multiplexing. There is also hype about future terahertz CPUs via optical communication on chip. Optical computation, processing, communication and storage then can be married with optical sensing to create an all-optical compact devices.

Nonlinear Optics

Nonlinear optics has been a rapidly growing field and involves interaction of intense coherent light radiation with matter. Currently, it is mainly observed in intense, pulsed lasers and special crystals. In traditional linear optics, light is deflected or delayed without wavelength change (a linear system). Nonlinear optics (NLO) deals with behavior of light in nonlinear media. In such media the dielectric polarization responds nonlinearly to the electric field of the light. A software analogy would be self-modifying programs. Why do nonlinear effects occur, in general? Imagine

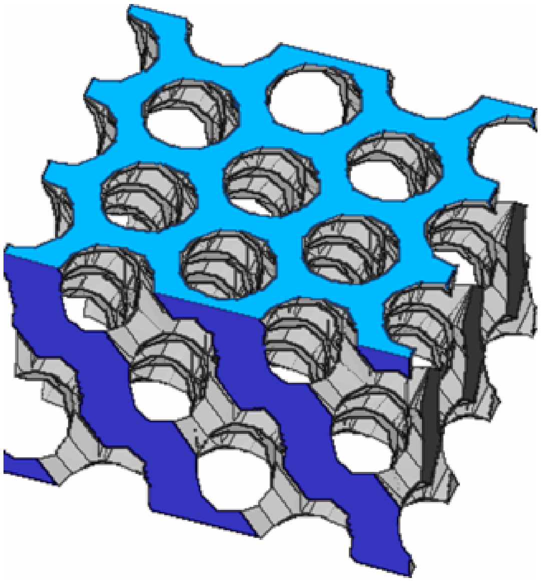


Figure 8.4: Photonic crystals are considered semiconductors of light. A common structure is lattice of high refractive index (RI) material embedded within a lower RI material.

playing music through a cheap amplifier and speaker that fails to reproduce loud notes. The amplifier maps input to output in a non-linear fashion, possibly leading to clamping at high intensities. A mono-frequency note with a single sinusoidal pattern after clamping or any nonlinear scaling produces higher frequencies or harmonics. In optics, those frequencies correspond to modifications of wavelength (or the color of light). [2]

Nonlinear effects are responsible for several non-intuitive effects. In addition to color change, we can change light beam shape in space and time, use light intensity of one beam as a programmable switch to gate propagation of a second light beam, and devise imaging for shortest events—in femtoseconds, 10^{-15} seconds. Although, most effects are for high intensity lasers, scientists are inventing new crystals and materials. For computational photography, an exciting area is programmable refractive index based on the optical Kerr effect discovered in 1960. If the refractive index (RI) of a material is n0, and the nonlinear component of the RI is n_2 , then a beam of intensity I can modify the RI to $n = n_0 + n_2 * I$. Another intensiting effect is self-focussing of light beams. Since RI is higher for higher intensity, a beam with Gaussian cross-section intensity profile creates a convex lens inside a homogeneous material. Unfortunately, in most cases beam collapses on itself converging till the material is damaged!

[2] An excellent overview is available at www.physics.gatech.edu/gcuo/UltrafastOptics/

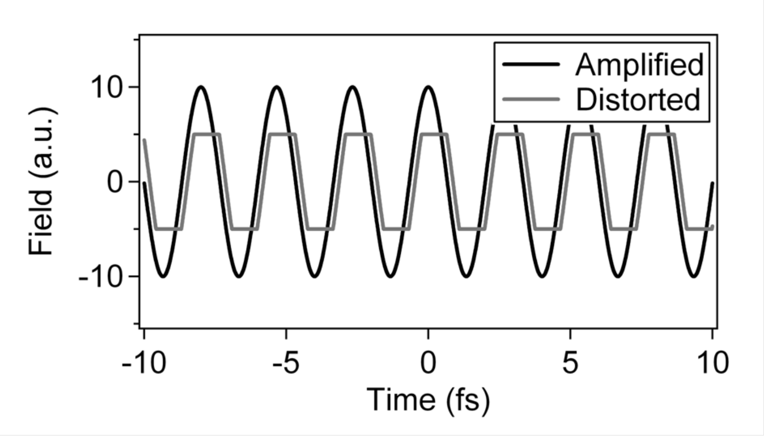


Figure 8.5: Demonstration of the impact of nonlinear response on frequency spectrum. When the input signal intensity is high, the system does not behave as a linear system and the clamping produces higher harmonics.

8.1.3 Modern Sensors

A trend that takes advantage of scaling is increasing pixel-level processing by adding more transistors to each pixel. Additional logic allows image sensor architectures with per-pixel analog to digital converters (ADCs). This in turn means reduced demands on ADC performance, elimination of fixed pattern noise (FPN), and high speed readout.

One of the challenges for photography is low-light sensitivity. However, in scientific and industrial applications often deal with applications that measure and analyze light emitted at levels so low that detection of single photons is required. For decades researchers have relied heavily upon glass photomultiplier tubes and newer channel multipliers. For dealing with motion blur, Time Delay and Integration (TDI) charge-coupled devices (CCD) are widely used. However linear speed and direction are known *a priori*. The strategy is equivalent to image motion compensation by panning, but performed on sensor. Sensed charge patterns and shifted them across the CCD array in sync with the movement of the image, to integrate more light from the scene. TDI provides the technology for capture of moving objects where high output level flash tubes are not available, are too distant, or are insufficient. Examples of such applications are satellite or aerospace imaging, machine vision, or industrial inspection. A TDI clock is used to synchronize the movement of charged packets in the CCD. Unlike a traditional sensor that provides a frame of output, during exposure, the CCD detector reads out the information in the opposite direction at the same speed. Thus, the frame rate is proportional to the speed of the moving object. Computational photography techniques will evolve rapidly as these sensors with additional bandwidth or logic per pixel become available.

Back-side Illumination

Back-side illumination (BSI) in CMOS sensors is a relatively new trend. In traditional front-side illumination (FSI) sensors, the light passes through optics (a microlens and color filter) and pixel well electronics (various metal layers for circuits before reaching the sensing silicon). The decrease in pixel size does not match decrease in the depth of the pixel well. Small pixels create very narrow wells reducing angle of acceptance and reducing light sensitivity. With BSI, the well electronics is upside down. The sensing silicon absorbs light immediately after optics. BSI provides the most direct path for light to travel into the pixel, avoiding light blockage by the metal interconnect. Companies have announced BSI sensor plans for 0.9 micron pixels. With pixel size approaching visible light wavelengths, a completely new imaging architecture is being planned for light field or wavefront detection [136, 392].

3D VLSI Sensors

Three dimensional very large scale integration technology (3D VLSI) technology in the field of optoelectronics provides vertical as well as horizontal interconnects between multiple substrate layers. This is especially beneficial for large image sensor arrays with stacked and interconnected sensing, processing and communication layers. Such stacked processing elements could mimic human retina which is a layered structure with several layers of neurons interconnected by synapses. Traditional 2D VLSI puts constraints on pixel fill factor and readout rates due to space required and distance covered by non-sensing electronics. The tighter integration in 3D VLSI allow significant reduction in size, weight, power consumption and delay. On-board per pixel circuits achieve analog local pre-processing for spatial and temporal filtering and gain control. So it is easy to support high dynamic range imaging, gradient sensing and sophisticated real (positive as well as negative) valued linear filter operations in analog domain. Asynchronous pulse time modulation and address event representation encoding and processing of data in distributed architectures, is an attractive alternative to traditional synchronous digital signal processing in 2D arrays [258, 395].

8.1.4 Metadata and Non-Visual Data

Our own visual sensing and our memory of viewed scenes is deeply impacted by non-visual data. Meta-data captured with additional, possibly non-visual sensors may greatly aid photography. Augmentations may include currently available sources such as automatically-sensed identity of people or objects (from RF, optical, or electronic tags and badges, or even a fingerprint reader on a cameras shutter release button), sound (audio recording) and location and pose sensing (GPS, compass, tilt-sensors, and indoor location tracking). But information about other conditions may also play a role. Sensing temperature, smell and wind will allow us to synthesize and manipulate visual recordings to recreate savory experiences in a restaurant, the breeze on a beach or an exhilarating ride on a rickety roller coaster.

8.1.5 Computation and Optimization

[SUGGESTED ADDITIONAL SUBSECTION BY DI WU—NEEDS TO BE WRITTEN] Can we add one subsection here to talk about the important role of computation/ optimization in imaging, like the nature photonics paper: Looking around corners and through thin turbid layers in real-time with scattered incoherent light. And also perhaps compressive sensing light field capture?

8.2 Displays

Computational photography may create new mechanisms for hyperrealistic synthesis, but ultimately the display medium is critical for realizing the visual experience. Unfortunately, the situation for displays is similar to the situation with modern cameras. Just as digital cameras mimic film-like photography, nearly all current displays mimic back-lit film or film projectors, and offer little more visual information and interactivity than high-quality CRT (cathode ray tube) monitors. However, several pioneering efforts demonstrate very promising new directions, including “Lighting Sensitive Displays by Nayar et al. [289], the extended auto-multiscopic method of Matusik et al. [262], the first true 360-degree 4D light-field display from Jones et al. [208] and dramatically improved dynamic range from Dolby HDR displays [394]. As most displays are tailored to available content rather than new capabilities, change may come slowly to digital displays designed for 2D photos, videos and animations. With more capable modern displays and new ways to share visual information, the corresponding capture and manipulation techniques will change, but new cameras may wait on new displays, and new displays may wait on new cameras! Let us review some of the recent progress and some future directions.

8.2.1 High Dynamic Range Displays

BrightSide technology [394] uses LCD technology. Usually LCD displays have a backlight provided by CCFL (Cold Cathode Fluorescent Light) tubes. Thats why even when the LCD screen is black it is not actually black, it still has some residual light: the tube is still turned on. The main idea of BrightSide technology is to remove this CCFL backlight and use several LEDs instead, a technology they call IMLED (Individually Modulated Array of LED backlights).

The brightness of each LED is controlled by an 8-bit signal, so each LED has 256 brightness steps (zero would mean tuned off while 255 would mean totally turned on; a 128 value would turn on the LED with 50% of its luminance, a 64 would turn on the LED with 25% of its luminance and so on). The first display launched with this technology—called DR37-P (which is a 37 panel)—has 1,380 white LEDs behind the LCD screen.

So, the idea is quite simple. Instead of having just one light source behind the LCD screen that is turned on all the time with the same brightness, BrightSide technology displays use several white LEDs where each one can have its brightness controlled (256 different brightness steps).

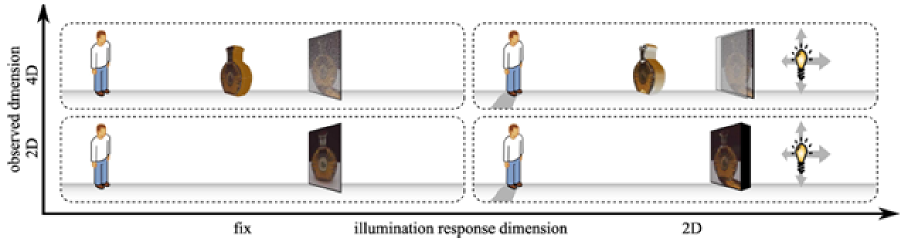


Figure 8.6: Towards a 6D photo-frame: a multi-dimensional display which passively react to the light of the environment. (bottom left) Traditional photo-frames present 2D images. (top left) Displays with horizontal or vertical parallax (3D) or both parallax (4D) use lenslet arrays or holograms to create a fixed outgoing light field. This creates a floating virtual object but it does not respond to ambient light. (bottom right) Fuchs et al. [2008] create a lighting aware 2D photo-frame, i.e., a 4D display that responds to 2D position of light. (top right) Ultimately, the photo-frame should encode a 6D reflectance field and reveal a different 4D light field depending on the 2D environmental illumination.

8.2.2 3D, Volumetric and View Dependent Displays

How can we create photo-frames and displays that display higher dimensional reflectance fields? Such visualizations play an important role in our everyday life: in images, volumes, light fields or reflectance field data sets. However, many optical visualizations and recording techniques are limited to a 2D structure. Therefore, methods have been presented in the past which address this problem by flattening the high dimensional data, embedding it in a planar, 2D representation. Integral photography [248] is an early approach which records a 4D light field on a photographic plate. The main concept is adding an array of lenses to the plate, we can discretize the spatial coordinates into spatial as well as angular dimensions.

Planar encodings of light fields are since the days of integral photography closely coupled to the development of displays which create a 3D impression by projecting a light field into space. Nakajima et al. [285] described a lens array on top of a computer display for a 3D viewing experience. In 2004, Matusik and Pfister [262] presented an end-to-end system which records a 3D light field, streams it over the network and then displays it on a lenticular array screen. Their article also gives a good overview of current multi-dimensional display techniques. Javidi and Okano [201] discuss a range of related techniques.

Compressive Displays

SUGGESTED ADDITION BY DI WU—NEEDS TO BE WRITTEN add compressive display work from our group, from HR3D–¿ layered 3D ? polarization field ? Tensor display

While light fields capture the appearance of a static object, reflectance fields further encode the optical response of an object to illumination. Nayar et al. [289] presented a display which measures the distant room illumination, approximated as environment map, and interactively renders an image in this illumination. Koike and Naemura [218] propose an extension towards emitting a light field in a similar fashion. Both displays are electronic and rely on software and hardware evaluating the illumination and rendering the reflectance field. Scharstein et al. [354] obtained a patent on a device which is passive: it employs optics in order to create a numeral display of the current time. This is achieved by encoding a pattern in a slit mask so that natural sunlight direction produces different symbols. However, this construction inherently blocks the majority of incident light rays. Fuchs et al. [143] follow a passive approach to illumination variant displays using lenses and colored patterns, thus using a larger portion of the available light for a higher contrast display of more expressive patterns. For distant light and a fixed observer position, they demonstrate a passive optical configuration which directly renders a 4D reflectance field in the real-world illumination behind it. Combining multiple of these devices in a 2D array they build a display that renders a 6D experience. The incident 2D illumination influences the outgoing light field, both in the spatial and in the angular domain. Since it is free from any electronic parts, the 6D photo-frame can be potentially printed as a 2D pattern and displayed behind the frames optics.

Some futuristic displays include manufacturing techniques for bionic eyes at microscopic scales to combine a flexible, biologically safe contact lens with an imprinted electronic circuit and lights [214]. Looking through a completed lens, you would see what the display is generating superimposed on the world outside for see-through augmented reality. The work is lead by Babak Parviz at University of Washington. However, the challenge is that the resolution of an image emitted from the contact lens will be low (due to blur). Saccades, which stabilize an image outside but not on retina, may blur everything.

8.3 Great Ideas from Scientific Imaging

Light transport and manipulation is very much part of several non-imaging operations such as optical communication, lithography and medical procedures. Rapid progress in theoretical understanding, hardware and methodologies in those fields will in turn impact cameras and photography. In terms of the responsibility of the task, there is also a need to think about balancing between physical and digital layers in terms of size, cost and power.

Some computation imaging which required supercomputers for scene reconstruction will be now possible via computing power on consumer devices. This removal of computing barrier means these complex information processing techniques will be used for casual problems. Emitters and sensors used in this exotic imaging, e.g. high speed, infrared, ultraviolet, is becoming cheaper and accessible. In a sense, the physical components of a camera, including high quality lenses, may not get cheaper, but silicon sensing is becoming cheaper.

In several scientific imaging scenarios, the sensed image quality is quite poor. So in this fields, there is a lot of emphasis on overcoming the limitations of the physical medium via clever physical architecture or highly sophisticated software. The low quality maybe due to difficulties in dealing with a wide spectrum, backscatter or secondary scatter or low SNR of sensor. For Computational Photography, it may be possible to use low-quality lenses and overcome the physical limitations of lenses with similar scientific computation. On the other hand, in several area of scientific imaging such as astronomy and microscopy, they have already explored (studied and characterized) optical lenses that are well beyond any consumer photography. This includes super-resolution, careful manipulation of spectrum, e.g. in fluorescence microscopy, wavefront phase manipulation and programmable illumination for feature revealing imaging. Scene measurement and representation in 4D and beyond encompasses previously isolated ”Islands” of Ingenious Scientific Imaging Measuring. What can we learn from them? Can we extend their methods? Particularly promising fields include the following.

(i) Tomography: Tomography is a powerful technique that allows us to see inside a volumetric object via cross-sectional imaging. For any penetrating measurements, attenuation along straight-line paths can be used to construct 3D images of internal structures This is currently used measuring sound transmission to electrical capacitance, from seismographic disturbances to ultrasonics to X-rays. We saw earlier that tomographic techniques are used in building multi-spectral imagers.

(ii) Spectrographic methods: complex interdependencies between wavelengths, reflectance, and transmissions are used for image forming, and broad classes of statistical measurements help decipher or identify useful features for land management, pollution studies, atmospheric patterns, wildlife migration, and geological and mineral features..

(iii) Confocal Methods and Synthetic Aperture methods: As described above, one can achieve very narrow depth-of-field image by collecting a widely divergent rays from each imaged point and these methods can extend to macroscopic scales via multiple cameras and multiple video projectors.

(iv) Fluorescence Methods: Some materials respond to absorbed photons by re-emitting other photons at different wavelengths, a phenomena known as fluorescence While very few materials fluoresce in the narrow range (¡ 1 octave!) of visible wavelengths, hyperspectral imaging reveals instructive fluorescence phenomena occur over much wider bands of wavelengths. Many organic chemicals have strongly varied fluorescent responses to ultraviolet light, and some living tissues can be chemically or genetically tagged with fluorescent markers that reveal important biological processes. Accordingly, hyperspectral imaging and illuminants can directly reveal chemical or biological features that may be further improved by 4D methods.

(v) Compressive Measurements: There is a strong interest in Analog to Information (A2I) sensors. Sparse in transformed domain. The idea behind Task-specific Imaging (TSI) is to achieve information optimal projection. For example, the location of a moving vehicle in a surveillance video is only a few bits of information. Can we build a camera that records only the tracking information instead of a high resolution video? Using careful, possibly adaptive, projection of the data, Mark Neifield and his group at University of Arizona have achieved temporal compressive sensing. David Brady and his group at Duke University have built an adaptive multispectral imager.

8.3.1 Other Dimensions

As noted in the Assorted pixels paper [Nayar2003], photographic capture gathers optical data along many dimensions, and few are fully exploited. In 4-dimensional ray space we sense and measure more than simple intensity (or more formally, radiance), but also visually assess wavelength, time, materials, illumination direction and more. Polarization is also sometimes revealing, and the mapping from polarization direction of the illuminant to the polarization of reflected light is not a simple one: for some biological materials, the mappings are nonlinear and unexplored [Wu et al. 2003]. Extended exploration of wavelength dependence is already well advanced. Hyperspectral imaging has already gathered a rich and growing literature for a broad range of applications from astronomy to archival imaging of museum treasures.

8.3.2 Transient Imaging and Ultra-fast Imaging

How can we exploit the finite speed of light to improve image capture and scene understanding? New theoretical analysis coupled with emerging ultra-high-speed imaging techniques can lead to a new source of computational visual perception. We need a new theoretical foundation for sensing and reasoning using transient light transport, and experimentation with scenarios in which transient reasoning exposes scene properties that are beyond the reach of traditional machine vision. In a traditional camera, the light incident at a pixel is integrated along angular, temporal and wavelength dimensions during the exposure time to record a single intensity value. Distinct scenes may result in identical projections (images) and, hence, identical pixel values. Thus, it is challenging to estimate scene properties which are not directly observable. Steady-state light transport assumes equilibrium in global illumination. In transient light transport framework, light takes a finite amount of time to travel from one scene point to the other. Recent advances in ultra-high speed imaging have made it possible to sample light as it travels 0.3 millimeter in 1 picosecond. The dynamics of transient light transport in response to a single ray impulse illumination can be extremely complex, even for a simple scene. Unlike a traditional 2D pixel, which measures the total number of photons. Transient light transport measures photon arrival rate as a function of time. Kirmani et al. [216] show that multi-path analysis using images from a time-of-flight (ToF) camera provides a tantalizing opportunity to infer 3D geometry of both visible and hidden parts of a scene.

Looking Around Corners

[SUGGESTED ADDITIONAL SUBSECTION BY DI WU—NEEDS TO BE WRITTEN] We could add more details from the looking around corner (nature comm. paper), and visualize light in motion project.

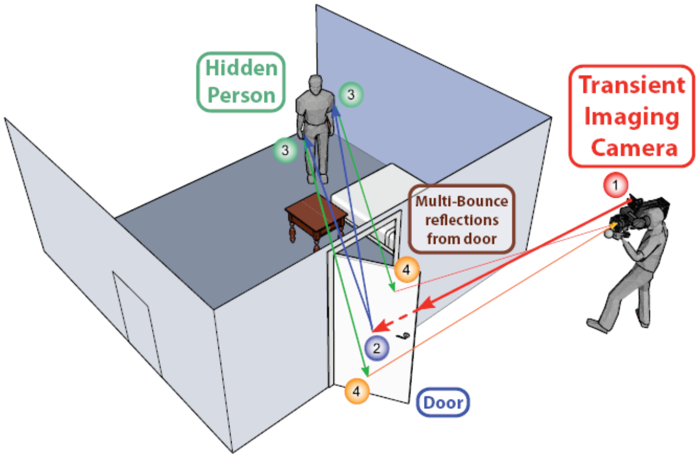


Figure 8.7: Can you look around the corner into a room with no imaging device in the line of sight? Kirmani et al. show that by emitting short pulses (1 —> 2), and analyzing multi-bounce reflection from the door (4 —> 1), they can infer hidden geometry even if the intermediate bounces (3) are not visible. The transient imaging camera prototype consists of a femtosecond laser illumination and picosecond-accurate detectors.

8.4 Fantasy Imaging Configurations

Beyond what we can do now, what would we like to achieve in computational photography? What can we imagine if we set aside current practical limits? By widening the goals of photography to the capture of the visual essence of an object, scene, or event, we can escape the current notion of an ideal photography studio as a room full of lights and box-like cameras, and expand it in wholly different directions.

8.4.1 The Moment Camera

Film-style photography relies on an instantaneous ideal: we attempt “stop time by capturing any photographed scene quickly enough to ignore any movement that happens during the measurement process. Even so-called motion pictures consist of a sequence of individual frames, and current video systems also make these same film-like serial attempts at instantaneous capture, rather than direct sensing of the motions themselves. Harold Edgerton pushed the instantaneous ideal to extremes by using ultra-short strobes to illuminate transient phenomena, and ultra-short shutters to measure ultra-bright phenomena quickly, such as his famous high-speed movies of atomic bomb explosions. Can we work in the opposite direction with new kinds of digital sensors, ones that summarize movement and change rather than incident light averaged over time?

Digital sensors offer new opportunities for more direct sensing, and digital displays permit interactive display of the movements we capture. Accordingly, Michael

Cohen has proposed that the film-rooted distinction between still cameras and video cameras should gradually disappear. He proposed that we need an intermediate digital entity he calls a moment; one visually meaningful action we wish to remembera childs fleeting expression of delighted surprise, a whisper of wind that sways the trees, etc., and it might fit in short video clips [Cohen 2005]. Motion sensing and deblurring itself can improve in the future [BenEzra 2004, Raskar 2006]. Movement also causes difficulties for constructing panoramas. However, if the movement is statistically consistent, it is possible to combine conventional image stitching operations with so-called video texturing methods [Schdl 2000] to create consistent, seamless movement that captures the moment of the panorama quite well. It can be further extended to capture video texture panoramas [Agarwala 2005].

8.4.2 Slow Glass

As pointed out by Rick Szeliski in recent talks, “slow glass is a fictional material in Bob Shaw’s short story ”Light of other days” (Analog, 1966), and several subsequent stories. The glass, which delays the passage of light by years or decades, is used to construct windows, called “scenedows, that enable city dwellers, submariners and prisoners to watch ”live” countryside scenes. In the original story, Shaw implied that slow glass was just a material with an enormously high index of refraction (somewhere in the quadrillions). In a later story, he supplied a more convoluted explanation attributing the delay to photons passing “.through a spiral tunnel coiled outside the radius of capture of each atom in the glass. The high refractive index explanation fails because Fresnel reflection (proportional to the square of the difference in refractive index) from the surface would form a near-perfect mirror; no light would enter the glass!

8.4.3 Blind Camera

With ubiquitous wireless communications, any camera could become an always-connected networked object. How might we exploit its access to vast repositories of online information? In his insightful 2006 art projects, [3] Sascha Pohflepp built a prototype “Blind Camera” that has a viewfinder and responds to its shutter release by taking a picture, but has no optics, sensor or illumination. The connectivity supports another channel for visual information. If the goal of computational photography is hyperrealistic synthesis, the part about machine-readable coded capture could be completely bypassed.

Imagine taking a photo in Times Square in New York city as a tourist. It is debatable whether you should take that picture at all. There are thousands of photos online, probably from the same place from which you plan to snap the photo. Sascha Pohflepp built his “Blind Camera by mounting a mobile phone in a black case to let you snap other people’s pictures. When you click the button to take a picture, the mobile phone connects to the internet, searches for photos from the same location with a similar time stamp and returns a picture to the viewfinder display. In todays world, this brilliant idea can be expanded to grab a photo by trawling a large photo sharing site, such as Flickr, to get a photo

[3] www.blinksandbuttons.net/index.html

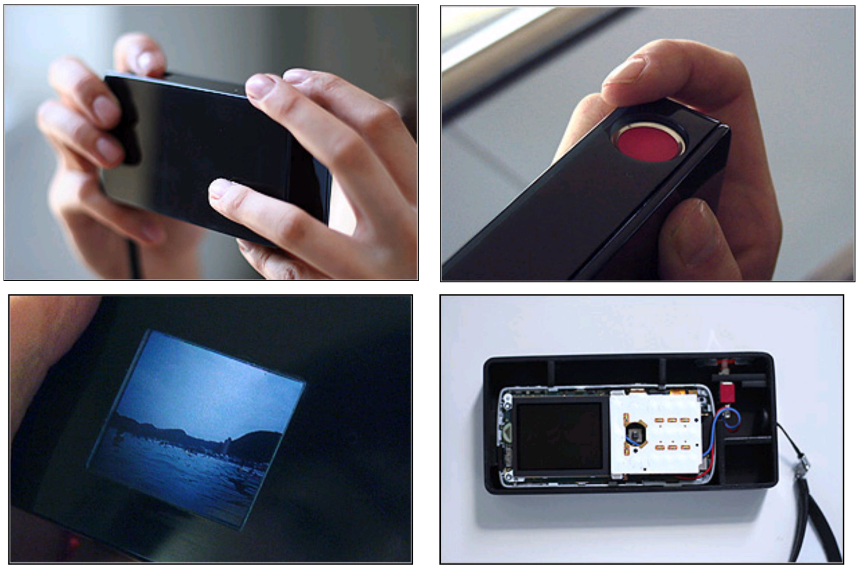


Figure 8.8: The Blind Camera has no traditional camera optics. When you take a picture, the camera connects to the internet, searches for photos from the same location with a similar time stamp, and returns a picture to the viewfinder display.

from roughly the same position (via GPS), same viewpoint (via compass), same time of day and weather conditions (weather records). Photos shared in space-time dimensions are certainly easy for architectural or natural landscapes. The challenge is to insert scene elements seen only at-the-moment including your own friends and family members. Recently Neel Joshi [297] used photos from family album as image priors to enhance low-quality photos of family members. As long as the album has high-quality images, one might get away with low-quality cameras for casual photography. With a sufficiently large database of the preferred appearance of family members and complete visual records of popular vacation sites, then perhaps a consumer on a trip will need only a cute little black box with a big red button!

8.4.4 Sheet-like Cameras

The dream of a flat thin (rigid) camera has many groups chasing an array of techniques. Although miniature cameras such as those found in cell phones are now commonplace, their resolution and light collection abilities compare poorly with their full-sized counterparts due to diffraction-limited optics and small aperture size. Several researchers have developed methods to aggregate many small cameras into more powerful combinations:

The MONTAGE (Multiple Optical Non-Redundant Aperture Generalized Sensors) program was sponsored by the Defense Advanced Research Projects Agency (DARPA) with program managers Ravindra Athale and Dennis M. Healy. The origami lens exploits a unique reflective multiple-fold approach [402]. Light enters the element through an outer annular aperture and is focused by a series of concentric zone reflectors to the image plane in the central area of the optic. The 40mm focal length annular aperture lens is compacted to fit within the 5mm thick volume with a conventional three-megapixel CMOS sensor and the whole assembly is integrated as a flat disk. This is similar to a Cassegrain telescope which exploits additional folding of optical paths.

Another thin camera approach is based on compound lens design that involves a planar array of miniature cameras. In theory, it is similar to a camera array. A compact image-capturing system called TOMBO (an acronym for thin observation module by bound optics) has shown several interesting applications of the thin optical configuration [Tanida 2001]. This form of array imaging employs specially designed array of lenslets to capture an ensemble of images of a subject, enabling the collection of significantly more information than possible with conventional single-lens imaging systems. An integrated array imaging system, dubbed PERIODIC, was recently developed to exploit different dimensions including sub-pixel displacement, phase, polarization, neutral density, and wavelength [Plemmons 2007]. Each camera has its own modified optical filter. Fisher information dictates theoretical upper bounds on the fidelity of reconstruction of high-resolution images from low-resolution image sequences. In general, the recovery involves solving complex ill-posed image registration and reconstruction problems.

8.4.5 Camera Ubiquity and Life Logs

Thad Starner and other cybernauts who began personally instrumenting themselves in the 1990s have experimented with always-on video cameras, and projects at Microsoft Research and the MIT Media Lab have both explored gathering video memories of every waking moment. So called smart dust sensors (extensively explored by DARPA projects) and other unstructured ubiquitous sensors might gather views, sounds, and appearance from anywhere in a large city.

Two entirely fanciful designs (as proposed by Tumblin [333]) suggest alternative approaches to appearance capture and set a a potentially useful distant goal. Suppose we could construct a flexible cloth-like material that holds microscopic, interleaved video projectors and video cameras. As the micro-projectors emit hyperspectrally colorful patterns of light in all possible directions from all possible points on the cloth (a flexible 4D light source), the interleaved micro-cameras would make geometrically calibrated coordinated hyperspectral measurements in all possible directions from all possible points on the cloth (a flexible 4D camera). Wiping the cloth over a surface would illuminate and photograph inside even the tiniest crack or vent hole of the object, banishing occlusion from the data set; a quick wipe would characterize any rigid object thoroughly.

Suppose we wish to capture the appearance of a soft object, without touching it? Then perhaps a notebook-like device made of two plates hinged together would help. Each panel would consist of interleaved cameras and projectors in a sheet-like arrangement; simply placing it around the object would provide sufficient optical coupling between the embedded 4D illuminators and 4D cameras to assess the object thoroughly. What other configurations of sensors, lights, optics, processing

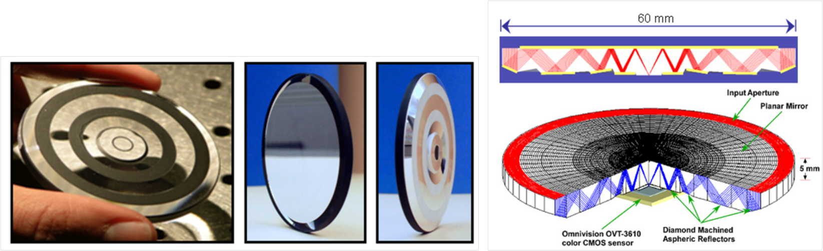


Figure 8.9: The origami lens exploits a reflective multiple-fold approach. Light enters the element through an outer annular aperture and is focused by a series of concentric zone reflectors to the central area of the image plane. The 40mm focal length annular aperture lens is compacted to fit within the 5mm thick volume with a conventional three-megapixel CMOS sensor. The whole assembly is integrated as a flat disk [402].

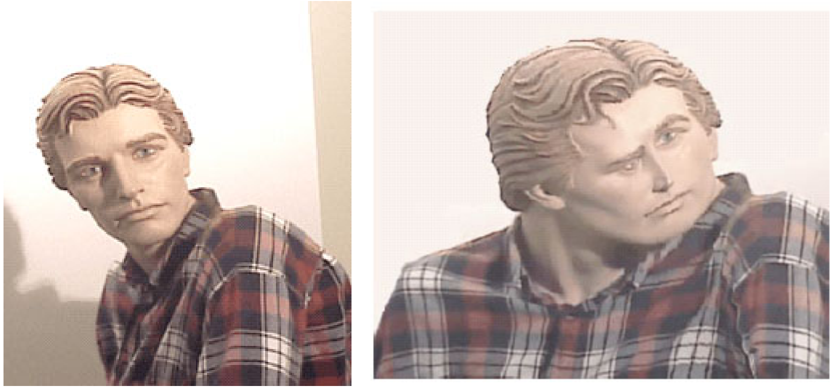


Figure 8.10: Multipespective Views. How can we build a camera to generalize the perspective projection? (Courtesy: Jingyi Yu and Leonard McMillan)

and displays would enable new forms of computational photography?

8.5 Conclusion

The film-like cage around us is gone. Released from 150 years of habits and conventions from film-like photography, we cant help but be a little bewildered right now, like a zoo-raised animal released into the wild. We dont yet see all of the possibilities, and even our fantasies of the ideal forms of photographic equipment cannot yet give us the complete and definitive answer for the best possible forms of photography. Even the close-range cloth camera lacks the long-range abilities of a top-quality telephoto lens, and diffraction limits preclude capturing the subtleties of a majestic mountain-range scene at dusk, or the details in the pale moon near the horizon. If the goal of photography is to capture, reproduce, and manipulate a meaningful visual experience, then the camera cloth is not sufficient to capture even the most rudimentary birthday party. The human experience and our personal viewpoint is missing. Ted Adelson suggested camera wallpaper or the balloon camera, ubiquitous sensors that would enable us to compute arbitrary viewpoints at arbitrary times.

While arbitrary viewpoints at arbitrary times might aid us in capturing that party, they simply expand our choices, without offering any help on making those choices. What makes these moments special? What parts of this video will become keepsakes or evidence? How do we find what we care about in this flood of video? Advanced cameras can supply us with visual experiences, but we need artificial-intelligence-like software or human intervention to decide what to keep, to find what matters most to humans.

This is an exciting time for exploration; every new direction may advance our ability to capture visual experiences, to construct photo-based visual prosthetics, to devise new and widely-available forms of capture, manipulation, and interaction with what we see around us, or would like to see. We hope the new choices and toolsets presented here will stoke your imagination as well, and entice you to seek out wholly new paths to the as-yet unimagined photographic wonders still ahead of us all.

Chapter 9

Appendix

introductory material to the Appendix goes here

9.1 Image Gradients

Consider the image intensities as a two dimensional function $I(x, y)$. The gradient of image at each point is then defined as a 2D vector (I_x, I_y) whose components are given by the derivatives in the horizontal and vertical directions respectively. At each point (pixel), the length (magnitude) of the gradient vector corresponds to the rate of change of intensities in that direction and the direction of the gradient vector denotes the direction of the maximum intensity change.

Figure 9.1 shows a simple example of a bright circle on a dark background. At each point of the circle boundary, the gradient points towards the normal at that point, since that direction corresponds to maximum intensity change. The magnitude of horizontal and vertical gradient components vary continuously across the circle.

9.1.1 Gradient Domain Algorithms

Figure 9.2 depicts the typical flow of gradient domain algorithms. Image gradients are obtained from a single image or set of images. These gradients are then ma-

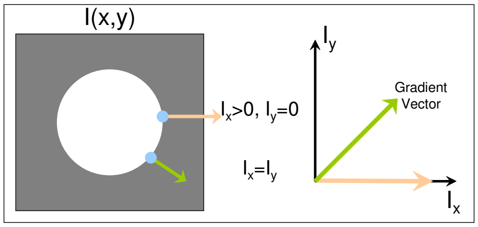


Figure 9.1: (Left) Intensity image. (Right) Gradient vectors at two points on the circle boundary.

nipulated in linear or non-linear fashion to result in an output gradient field. The resultant gradient field is then integrated to obtain the final image.

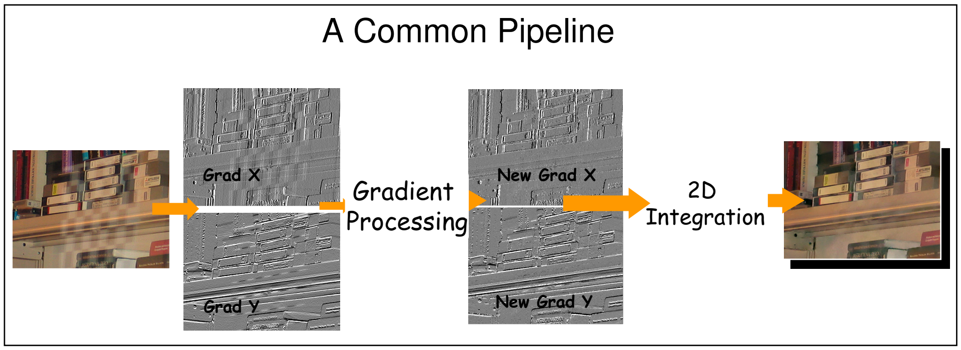


Figure 9.2: A common pipeline for gradient domain algorithms.

9.1.2 Pseudo Code for Removing Reflections

Inputs: Flash/No-flash image pair Output: Reflection removed image and reflection layer

- Compute gradients from the given images
- For each pixel in no-flash image, project the gradient on to the image gradient of flash image. Compute the difference of given gradient in no-flash image and projected gradient.
- Integrate the two resulting gradient fields to obtain reflection removed image and reflection layer.

9.1.3 Matlab Code for Image Reconstruction from Given Gradient Field

Inputs: given gradient field P (horizontal gradients) and Q (vertical gradients)

Size: H by W

Output: Reconstructed image Z (H by W)

- Obtain Differentiation Matrices and Vectors

```
idx = [1:W-1]'; idy = [1:H-1]';
Dx = sparse(idx,idx,-1,W,W) + sparse(idx+1,idx,1,W,W);
Dy = sparse(idy,idy,-1,H,H) + sparse(idy,idy+1,1,H,H);
D = [kron(Dx',eye(H)) ; kron(eye(W),Dy)];
```

- Vectorize gradients
  g = [P(:);Q(:)];
- Poisson Solver
  Z = reshape([0;D(:,2:end)\g],H,W);
  Z = Z − min(Z(:));

9.1.4 Web Links

Software (Matlab codes): www.umiacs.umd.edu/∼aagrawal/software.html, www.umiacs.umd.edu/∼aagrawal/iccv05/Agrawal ICCV05MatlabCode.zip
Gradient Domain Course: www.umiacs.umd.edu/∼aagrawal/ICCV2007Course/index.html

9.2 Bilateral Filter

In several image/graphics applications, edges preserving smoothing or robust filters are required. Bilateral filter is an edge preserving filter whose weights depend on spatial filter and a range filter. In contrast, Gaussian filtering of images solely depends on spatial filter and thus smoothes out sharp edges/boundaries

The equation for Gaussian filtering is given by

$$GF[p] = \frac{1}{W_p} \sum_{q \in S} G_{\sigma_s} \left(\| \, p - q \, \|\right) I[q], \tag{9.1}$$

where GF is the filtered image, I is the input image, S denotes a neighborhood around pixel p and G_{σ_s} denotes a spatial Gaussian filter. W_p is a normalization constant given by $W_p = \sum_{q \in S} G_{\sigma_s} (\| \, p - q \, \|)$.

The equation for bilateral filtering is given by

$$BF[p] = \frac{1}{W_p} \sum_{q \in S} G_{\sigma_s} \left(\| \, p - q \, \|\right) G_{\sigma_r} \left(\| \, I[p] - I[q] \, \|\right) I[q], \tag{9.2}$$

where G_{σ_r} denotes a range filter and

$$W_p = \sum_{q \in S} G_{\sigma_s} \left(\| \, p - q \, \|\right) G_{\sigma_r} \left(\| \, I[p] - I[q] \, \|\right).$$

9.2.1 Web links

people/csail.mit.edu/sparis/bf_course

9.3 Graph Cuts

Graph Cut is a discrete optimization technique for efficiently minimizing energy functions. The technique is widely used in image segmentation, multiview reconstruction, stereo algorithms and for solving discrete labeling problems.

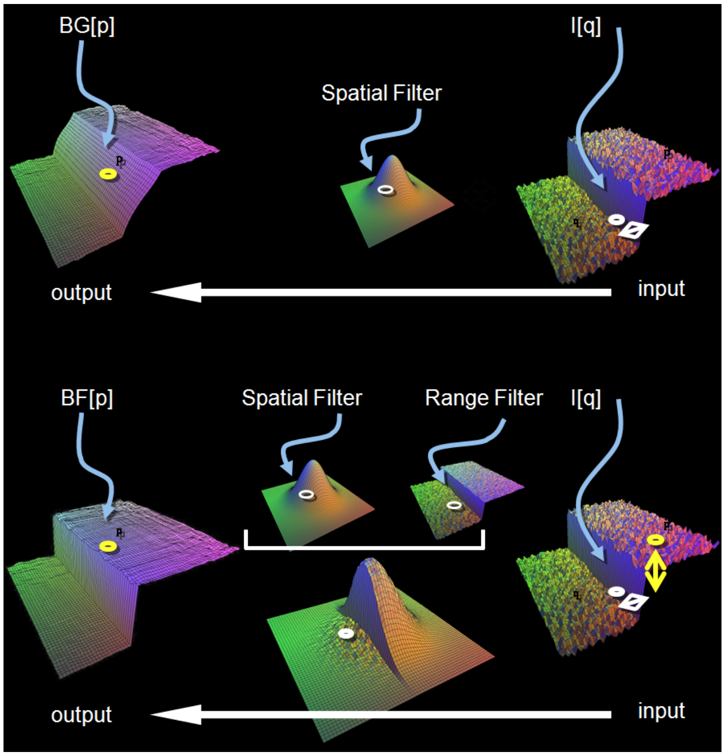


Figure 9.3: Demonstrating the basic concept of Gaussian filter (top) versus bilateral filter (bottom). The bilateral filter at every pixel depends on both the spatial filter and the range filter. At image discontinuities, the filtered output at pixel p is obtained from the same side of discontinuity.

An image segmentation problem is converted into a graph partition problem as follows. A pixel is a graph node and a link between neighboring pixels is an edge connecting two nodes. The pixel similarity is the edge weight. In the simplest case, pixel similarity is difference between the two pixel intensities. In graph partition, our goal is to find the minimum cost cut where the cost is the sum of weights of removed edges. A cut is represented by a set of edges whose removal makes a graph disconnected.

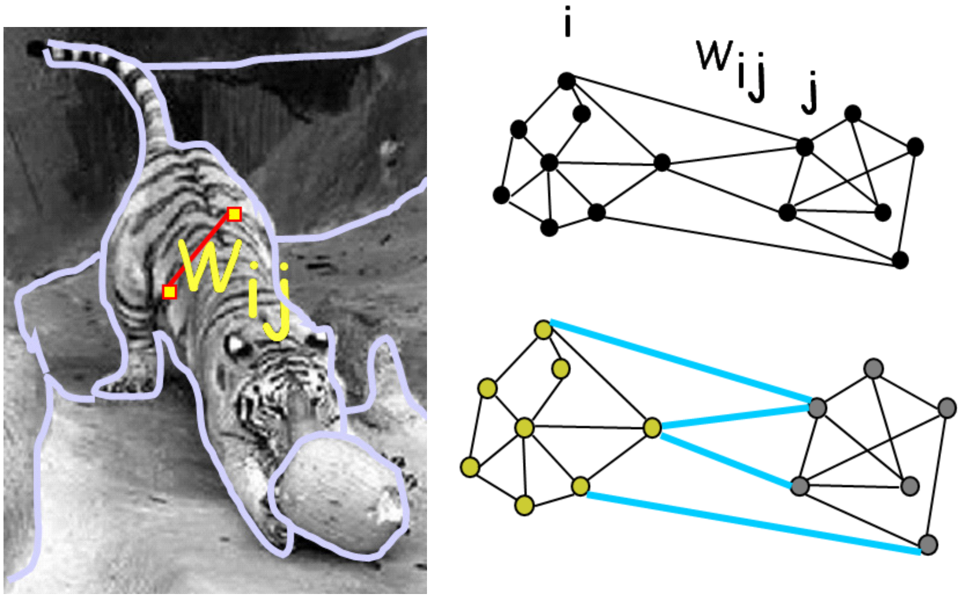


Figure 9.4: Segmentation as a graph cut problem. Graph is created by assigning pixel similarity weight w_{ij} between pixel pairs i and j. The minimum cost cut, shown in blue, partitions the graph i.e. segments the image by labeling pixels as belonging to one group.

For computational photography applications, a common application is as follows: Given a set of images, we wish to obtain a new image which combines relevant and useful information from all of them. This can be solved using graph cuts by treating the given set of images as different labels. In the final output, each pixel is thus assigned a label from this set.

- Define a cost function per pixel for assigning a particular label to that pixel
- Define a smoothness cost between pixels to have same labels.
- Minimize the total cost function over all image pixels using graph cut technique and obtain the labels.
- Generate the final output by choosing pixels from given set of images according to the labels.

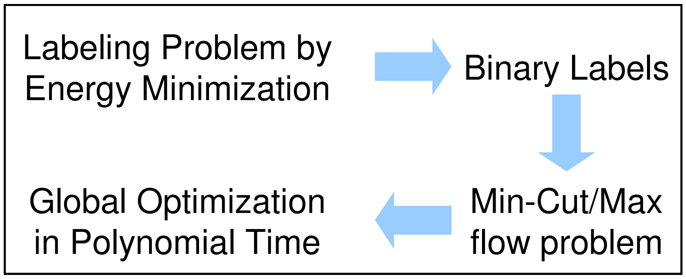

Labeling Problem by Energy Minimization
Binary Labels
Min-Cut/Max flow problem
Global Optimization in Polynomial Time

Figure 9.5: Labeling problems can be efficiently solved using graph cuts which provide global optimization in polynomial time.

Matlab Code for graph cut optimization: www.csd.uwo.ca/~olga/code.html
Web resources: www.cs.cornell.edu/~rdz/graphcuts.html

Bibliography

[1] Website. http://en.wikipedia.org/wiki/Electromagnetic_spectrum.

[2] Website. http://en.wikipedia.org/wiki/Double-slit_experiment.

[3] Website. The Astro 7000 is available at http://www.searchlightsandleds.com.

[4] Website. http://en.wikipedia.org/wiki/Wien\%27s_displacement_law.

[5] Website. http://4colorvision.com/files/photopiceffic.htm.

[6] Website. http://www.cie.co.at/publ/abst/datatables15_2004/y2.txt.

[7] Website. http://www.helios32.com/Measuring\%20Light.pdf.

[8]

[9] Website. http://www.helios32.com/resources.htm.

[10] Website. http://www.siggraph.org/publications/newsletter/volume-43-number-2/a-survey-of-brdf-models-for-computer-graphics.

[11]

[12] Website. http://www.instructables.com/id/How_To_Make_A_Pinhole_Camera/.

[13] Website. http://www.kodak.com/global/en/consumer/education/lessonPlans/pinholeCamera/.

[14] Website. http://www.instructables.com/id/build_a_digital_pin_hole_camera/.

[15] Website. http://hugin.sourceforge.net/.

[16] Website. Interactive Java tutorial on biconvex lenses, http://micro.magnet.fsu.edu/primer/java/lens/biconvex.html.

[17] Website. Grepstad, J., Pinhole Photography,http://photo.net/learn/pinhole/pinhole.

[18]

[19] optical bench simulator (chapter 2) to be added.

[20] *to be added in chapter 6*.

[21] Extremely low-frequency antenna, November 1965.

[22] *Bilateral Filtering for Gray and Color Images*, Washington, DC, USA, 1998. IEEE Computer Society.

[23] *Panoramic vision: sensors, theory, and applications*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2001.

[24] Hawk-i digital system, motion analysis corporation, 2006.

[25] Visualeyez vz 4000, pti inc, 2006.

[26] Nimeroff 1994. light interacts linearly with material objects. 1994.

[27] Masselus 2002. free form light stage. 2002.

[28] Agrawal Raskar Nayar Li SIGGRAPH 2005. In *SIGGRAPH*, 2005.

[29] Fuchs 2005. Bayesian relighting. 2005.

[30] Phase Space 2007. expensive camera-based system from phase space 2007. 2007.

[31] Andrew Adams and Marc Levoy. General linear cameras with finite aperture. In *EGSR*, 2007.

[32] Ansel Adams. *The Negative*, volume 2 of *The Ansel Adams Photography Series.* Little, Brown and Company, 1981.

[33] Jim Adams, Ken Parulski, and Kevin Spaulding. Color processing in digital cameras. *IEEE Micro*, 18(6):20–30, 1998.

[34] Edward H. Adelson and James R. Bergen. The plenoptic function and the elements of early vision. *Computational Models of Visual Processing*, pages 3–20, 1991.

[35] Edward H. Adelson and John Y. A. Wang. Single lens stereo with a plenoptic camera. *Transactions on Pattern Analysis and Machine Intelligence*, 14(2):99–106, 1992.

[36] A. Agarwala, K.C. Zheng, C. Pal, M. Agrawala, M. Cohen, B. Curless, D. Salesin, and R. Szeliski. Panoramic video textures. In *ACM Transactions on Graphics (TOG)*, volume 24, pages 821–827. ACM, 2005.

[37] Aseem Agarwala, Mira Dontcheva, Maneesh Agrawala, Steven Drucker, Alex Colburn, Brian Curless, David Salesin, and Michael Cohen. Interactive digital photomontage. In *SIGGRAPH*, volume 23, pages 294–302, 2004.

[38] Manoj Aggarwal and Narendra Ahuja. Split aperture imaging for high dynamic range. *International Journal of Computer Vision*, 58(1):7–17, 2004.

[39] Agrawal. to be added. 2005.

[40] Agrawal and Raskar. a technique for obtaining superresolution from a single photograph (chapter 3) to be add.

[41] A. Agrawal and R. Raskar. Optimal single image capture for motion deblurring. *CVPR*, pages 2560–2567, 2009.

[42] A. Agrawal, R. Raskar, and R. Chellappa. What is the range of surface reconstructions from a gradient field? *Computer Vision–ECCV 2006*, pages 578–591, 2006.

[43] Marc Alexa, Daniel Cohen-Or, and David Levin. As-rigid-as-possible shape interpolation. In *SIGGRAPH '00: Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 157–164, New York, NY, USA, 2000. ACM Press/Addison-Wesley Publishing Co.

[44] I. Ashdown. *Radiosity: a programmer's perspective.* John Wiley & Sons, Inc., 1994.

[45] I. Ashdown. Photometry and radiometry–a tour guide for computer graphics enthusiasts. Technical report, Technical report, Ledalite Architectural Products, Inc, 1996.

[46] Michael Ashikhmin. Synthesizing natural textures. In *I3D '01: Proceedings of the 2001 symposium on Interactive 3D graphics*, pages 217–226, New York, NY, USA, 2001. ACM.

[47] Jackie Assa, Yaron Caspi, and Daniel Cohen-Or. Action synopsis: Pose selection and illustration. ACM Press, 2005.

[48] Volker Aurich and Jörg Weule. Non-linear gaussian filters performing edge preserving diffusion. In *Mustererkennung 1995, 17. DAGM-Symposium*, pages 538–545, London, UK, 1995. Springer-Verlag.

[49] Shai Avidan and Ariel Shamir. Seam carving for content-aware image resizing. In *SIGGRAPH '07: ACM SIGGRAPH 2007 papers*, page 10, New York, NY, USA, 2007. ACM.

[50] Bae. real-time bilateral filtering–based tone-mapping framework applied it to interesting photographic manipulations—to be added. 2006.

[51] Soonmin Bae and Frédo Durand. Defocus magnification. In *Eurographics*, volume 26, pages 571–579, 2007.

[52] Simon Baker and Takeo Kanade. Limits on super-resolution and how to break them. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(9):1167–1183, 2002.

[53] Banterle. Hdr star. 2009.

[54] Barbastathis. to be added. 2003.

[55] Bryce E. Bayer. Color imaging array. US Patent 3971065, 1976.

[56] Benjamin B. Bederson, Ben Shneiderman, and Martin Wattenberg. Ordered and quantum treemaps: Making effective use of 2d space to display hierarchies. *ACM Trans. Graph.*, 21(4):833–854, 2002.

[57] Nayar Ben-Ezra. Ben-ezra and nayar 2005; a hybrid camera that can measure its own motion. 2005.

[58] Eric P. Bennett and Leonard McMillan. Video enhancement using per-pixel virtual exposures. In *ACM SIGGRAPH*, volume 24, pages 845–852, 2005.

[59] Eric P. Bennett and Leonard McMillan. Video enhancement using per-pixel virtual exposures. In *SIGGRAPH '05: ACM SIGGRAPH 2005 Papers*, pages 845–852, New York, NY, USA, 2005. ACM.

[60] Eric P. Bennett and Leonard McMillan. Computational time-lapse video. In *SIGGRAPH '07: ACM SIGGRAPH 2007 papers*, page 102, New York, NY, USA, 2007. ACM.

[61] Marcelo Bertalmio, Guillermo Sapiro, Vincent Caselles, and Coloma Ballester. Image inpainting. In *SIGGRAPH '00: Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 417–424, New York, NY, USA, 2000. ACM Press/Addison-Wesley Publishing Co.

[62] Yan Betremieux, Timothy A. Cook, Daniel M. Cotton, and Supriya Chakrabarti. Spinr: two-dimensional spectral imaging through tomographic reconstruction. *Optical Engineering*, 32(12):3133–3138, 1993.

[63] Michael J. Black and P. Anandan. The robust estimation of multiple motions: parametric and piecewise-smooth flow fields. *Comput. Vis. Image Underst.*, 63(1):75–104, 1996.

[64] Sean Borman and Robert Stevenson. Spatial resolution enhancement of low-resolution image sequences - a comprehensive review with directions for future research. Technical report, Laboratory for Image and Signal Analysis (LISA), University of Notre Dame, July 1998.

[65] Ronald N. Bracewell. *The Fourier Transform and Its Applications.* McGraw-Hill, 1999.

[66] D.J. Brady and N. Hagen. Multiscale lens design. *Optics Express*, 17(13):10659–10674, 2009.

[67] V. Brajovic and T. Kanade. A sorting image sensor: An example of massively parallel intensity-to-time processing for low-latency computational sensors. In *Robotics and Automation, 1996. Proceedings., 1996 IEEE International Conference on*, volume 2, pages 1638–1643. IEEE, 1996.

[68] C. Bregler and J. Malik. Tracking people with twists and exponential maps. In *Computer Vision and Pattern Recognition, 1998. Proceedings. 1998 IEEE Computer Society Conference on*, pages 8–15. IEEE, 1998.

[69] S. W. Brown, Joseph P. Rice, Jorge E. Neira, B. C. Johnson, and J. D. Jackson. Spectrally tunable sources for advanced radiometric applications. *Journal of Research of the National Institute of Standards and Technology*, 111:401–410, 2006.

[70] A. Buades, B. Coll, and J. M. Morel. A review of image denoising algorithms, with a new one. *Simul*, 4:490–530, 2005.

[71] Theodor V. Bulygin and Gennady N. Vishnyakov. Spectrotomography: a new method of obtaining spectrograms of two-dimensional objects. In *SPIE, Analytical Methods for Optical Tomography*, volume 1843, pages 315–322, 1992.

[72] Burt and Kolczynski. exposure bracketing–to be added. 1993.

[73] Peter J. Burt and Edward H. Adelson. A multiresolution spline with application to image mosaics. *ACM Trans. Graph.*, 2(4):217–236, 1983.

[74] Canon. Canon diffractive optics lenses reduce chromatic aberration (chapter 1) to be added. Technical report.

[75] Thomas W. Cathey and Edward R. Dowski. wavefront coding modifies the defocus blur. 1995.

[76] Thomas W. Cathey and Edward R. Dowski. New paradigm for imaging systems. *Applied Optics*, 41(29):6080–6092, 2002.

[77] Jin-Xiang Chai, Shing-Chow Chan, Heung-Yeung Shum, and Xin Tong. Plenoptic sampling. In *SIGGRAPH*, pages 307–318, 2000.

[78] Shing-Chown Chan and Heung-Yeung Shum. A spectral analysis for light field rendering. In *IEEE ICIP*, pages 25–28, 2000.

[79] Ltd. Charnwood Dynamics. Codamotion. www.charndyn.com, 2007.

[80] Ed. Chaudhuri, S. *Super-Resolution Imaging*. Kluwer Academic, 2001.

[81] Shenchang Eric Chen. Quicktime vr: an image-based approach to virtual environment navigation. In *SIGGRAPH '95: Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pages 29–38, New York, NY, USA, 1995. ACM.

[82] P. Choudhury and J. Tumblin. The trilateral filter for high contrast images and meshes. In *Proceedings of the 14th Eurographics workshop on Rendering*, pages 186–196. Eurographics Association, 2003.

[83] Yung-Yu Chuang, Brian Curless, David H. Salesin, and Richard Szeliski. A bayesian approach to digital matting. In *Proceedings of IEEE CVPR 2001*, volume 2, pages 264–271. IEEE Computer Society, December 2001.

[84] Y.Y. Chuang, B. Curless, D.H. Salesin, and R. Szeliski. A bayesian approach to digital matting. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 2, pages 264–271. IEEE, 2001.

[85] Brian Coe. *Colour Photography: the first hundred years 1840-1940*. Ash & Grant, 1978.

[86] M. Cohen. Capturing the moment. In *Symposium on Computational Photography and Video, Boston*, May 2005.

[87] CSR2009. compressive sensing review(?). 2009.

[88] Dana. The btf is a 6d function (chapter 4) to be added.

[89] James Davis. James davis papers on time of flight for depth detection (chapter 1) to be added.

[90] Dayal. randomized trade-offs between motion blur and resolution–to be added.

[91] Debevec and Malik. use the estimated response function to process the images before merging them–to be added. 1997.

[92] Paul E. Debevec, Camillo J. Taylor, and Jitendra Malik. Modeling and rendering architecture from photographs: a hybrid geometry- and image-based approach. In *SIGGRAPH '96: Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pages 11–20, New York, NY, USA, 1996. ACM.

[93] Tchou C. Duiker H.-P. Sarokin W. Sagar M. Debevec P., Hawkins T. Acquiring the reflectance field of a human face. 2000.

[94] DeCarlo. to be added. 2001.

[95] Decker. Decker article in 1998. 1998.

[96] James C. Demro, Richard Hartshorne, Loren M. Woody, Peter A. Levine, and John R. Tower. Design of a multispectral, wedge filter, remote-sensing instrument incorporating a multiport, thinned, ccd area array. In *Proceedings of SPIE*, volume 2480, pages 280–286, 1995.

[97] Michael R. Descour and Eustace L. Dereniak. Computed-tomography imaging spectrometer: experimental calibration and reconstruction results. *Applied Optics*, 34(22):4817–4826, 1995.

[98] Michael R. Descour, Curtis E. Volin, Eustace L. Dereniak, Kurtis J. Thome, A. B. Schumacher, Daniel W. Wilson, and Paul D. Maker. Demonstration of a high-speed nonscanning imaging spectrometer. *Optics Letters*, 22(16):1271–1273, 1997.

[99] David Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, pages 1289–1306, April 2006.

[100] Edward R. Dowski and Thomas W. Cathey. Extended depth of field through wavefront coding. *Applied Optics*, 34(11):1859–1866, April 1995.

[101] Durand. Bilateral filtering methods proposed by durand et al.–to be added. 2002.

[102] P. Dutre, K. Bala, P. Bekaert, and P. Shirley. *Advanced global illumination*, volume 2. AK Peters New York, 2006.

[103] Alexei A. Efros and William T. Freeman. Image quilting for texture synthesis and transfer. *Proceedings of SIGGRAPH 2001*, pages 341–346, August 2001.

[104] Alexei A. Efros and Thomas K. Leung. Texture synthesis by non-parametric sampling. In *IEEE International Conference on Computer Vision*, pages 1033–1038, Corfu, Greece, September 1999.

[105] Eisemann and Duran. In *SIGGRAPH*, 2004.

[106] M. Elad and A. Feuer. Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images. *IEEE Transactions on Image Processing*, 6:1646–1658, December 1997.

[107] Ahuja et al. capturing panoramas with panning cameras via attenuating ramp filters–to be added. 2002.

[108] Eisemann et al. to be added. 2004.

[109] Kang et al. techniques for dealing with high dynamic range scenes with video cameras–to be added. 2005.

[110] Mantiuk et al. display adaptive tone-mapping operator–to be added. 2008.

[111] Nayar et al. capturing panoramas with panning cameras via attenuating ramp filters–to be added. 2002.

[112] Petschnigg et al. In *SIGGRAPH*, 2004.

[113] Debevec et al. 2001. images acquired with a light stage. 2001.

[114] Koudelka et al. 2001. estimate surface geometry. 2001.

[115] Debevec et al. 2002. enhanced light stages. 2002.

[116] Akers et al. 2003. 2003.

[117] Anrys et al. 2004. painting interface for photographic lighting design. 2004.

[118] Feris et al. 2004. 2004.

[119] Tan et al. 2004. 2004.

[120] Agrawal et al. 2005. merging multiple images captured under varying flash intensities (chapter 5) to be added. 2005.

[121] Mohan et al. 2005. painting interface for photographic lighting design. 2005.

[122] Crispell et al. 2006. 2006.

[123] Chuang et al. CVPR 2001. In *CVPR*, 2001.

[124] Kim et al. LNCS 2005. Alternate methods for active detection of depth discontinuities using structured illumination. In *LNCS*, 2005.

[125] Casio EXILIM Pro EX-F1. http://www.exilim.com/intl/ex_f1.

[126] Hui Fang and John C. Hart. Detail preserving shape deformation in image editing. In *SIGGRAPH '07: ACM SIGGRAPH 2007 papers*, page 12, New York, NY, USA, 2007. ACM.

[127] Hany Farid and Eero P. Simoncelli. Range estimation by optical differentiation. *Journal of the Optical Society of America*, 15(7):1777–1786, 1998.

[128] Ivar Farup, Jan Henrik Woldy, Thorstein Seimy, and Torkjel Søndrol. Generating lights with specified spectral power distributions. *Applied Optics*, 46:2411–2422, 2007.

[129] Fattal. Gradient domain hdr compression–to be included? 2002.

[130] Fattal. Gradient domain manipulation methods–to be added. 2002.

[131] M. Agrawala Fattal, R. and S. Rusinkiewicz. Multiscale shape and detail enhancement from multilight image collections. *ACM Transactions on Graphics*, 2007.

[132] Rob Fergus, Barun Singh, Aaron Hertzmann, Sam T. Roweis, and William T. Freeman. Removing camera shake from a single photograph. In *SIGGRAPH '06: ACM SIGGRAPH 2006 Papers*, pages 787–794, New York, NY, USA, 2006. ACM.

[133] Rob Fergus, Antonio Torralba, and William T. Freeman. Random lens imaging. Technical report, MIT CSAIL, 2006.

[134] Russell D. Fernald. Casting a genetic light on the evolution of eyes. *Science*, 313(5795):1914–1918, September 2006.

[135] R. Feynman. *QED: The strange theory of light and matter*. Princeton University Press, 1985.

[136] K. Fife, A. El Gamal, and H.-S. P. Wong. A 3mpixel multi-aperture image sensor with 0.7um pixels in 0.11um cmos. *IEEE ISSCC Digest of Technical Papers*, 2008.

[137] first camera phone. the first camera phone (chapter 8)—to be added.

[138] Foveon. to be added. 2004.

[139] G.R. Fowles. *Introduction to modern optics,2nd edition.* Dover Pubns, 1989.

[140] W.T. Freeman and Hao Zhang. Shape-time photography. volume 2, pages II–151–II–157 vol.2, June 2003.

[141] Joseph Solomon Friedman. *History of Color Photography.* Style Press, 1968.

[142] Irena Fryc, Steven W. Brown, George P. Eppeldauer, and Yoshi Ohno. A spectrally tunable solid-state source for radiometric, photometric and colorimetric applications. In *SPIE*, volume 5530, pages 150–159, 2004.

[143] M. Fuchs, R. Raskar, H.P. Seidel, and H. Lensch. Towards passive 6d reflectance field displays. In *ACM Transactions on Graphics (TOG)*, volume 27, page 58. ACM, 2008.

[144] Fuji. en.wikipedia.org/wiki/Super_CCD.

[145] Fuji. Fujichrome velvia for professionals (rvp); data sheet af3-960e. 2008.

[146] Chunyu Gao and Narendra Ahuja. A refractive camera for acquiring stereo and super-resolution images. In *IEEE CVPR*, volume 2, pages 2316–2323, 2006.

[147] Chunyu Gao, Narendra Ahuja, and Hong Hua. Active aperture control and sensor modulation for flexible imaging. In *IEEE CVPR*, 2007.

[148] Garg Eino-Ville Talvala Marc Levoy-Hendrik P. A. Lensch Garg, Gaurav. Symmetric photography: exploiting data-sparseness in reflectance fields. In *Proc. Eurographics Symposium on Rendering*, 2006.

[149] Nahum Gat. Imaging spectroscopy using tunable filters: A review. In *SPIE Wavelet Applications VII*, volume 4056, pages 50–64, 2000.

[150] Nahum Gat, Gordon Scriven, John Garman, Ming De Li, and Jingyi Zhang. Development of four-dimensional imaging spectrometers (4d-is). In *SPIE Imaging Spectrometry XI*, volume 6302, page 63020M, 2006.

[151] Michael E. Gehm and David J. Brady. High-throughput hyperspectral microscopy. In *SPIE*, volume 6090, page 609007, 2006.

[152] Michael E. Gehm, Scott T. McCain, Nikos P. Pitsianis, David J. Brady, Prasant Potuluri, and Michael E. Sullivan. Static two-dimensional aperture coding for multimodal, multiplex spectroscopy. *Applied Optics*, 45(13):2965–2974, May 2006.

[153] Joe Geigel and F. Kenton Musgrave. A model for simulating the photographic development process on digital images. In *SIGGRAPH*, pages 135–142, 1997.

[154] T. Georgiev. Covariant derivatives and vision. In *9th European conference on Computer Vision (ECCV)*, pages 56–69, May 2006.

[155] Todor Georgiev, Chintan Intwala, and Derin Babacan. Light-field capture by multiplexing in the frequency domain. Technical report, Adobe Systems Incorporated, 2007.

[156] Todor Georgiev, Ke Colin Zheng, Brian Curless, David Salesin, Shree Nayar, and Chintan Intwala. Spatio-angular resolution tradeoffs in integral photography. In *Rendering Techniques*, pages 263–272, jun 2006.

[157] Dan B Goldman, Brian Curless, David Salesin, and Steven M. Seitz. Schematic storyboarding for video visualization and editing. In *SIGGRAPH '06: ACM SIGGRAPH 2006 Papers*, pages 862–871, New York, NY, USA, 2006. ACM.

[158] Evan Golub. Photocropr: A first step towards computer-supported automatic generation of photographically interesting cropping suggestions, 2007.

[159] Amy A. Gooch, Sven C. Olsen, Jack Tumblin, and Bruce Gooch. Color2gray: salience-preserving color removal. In *SIGGRAPH*, volume 24, pages 634–639. ACM Press, 2005.

[160] Amy A. Gooch, Sven C. Olsen, Jack Tumblin, and Bruce Gooch. Color2gray: salience-preserving color removal. In *SIGGRAPH '05: ACM SIGGRAPH 2005 Papers*, pages 634–639, New York, NY, USA, 2005. ACM.

[161] Joseph V. Goodman. *Introduction to Fourier Optics*. Roberts and Company Publishers, 2004.

[162] Steven J. Gortler, Radek Grzeszczuk, Richard Szeliski, and Michael F. Cohen. The lumigraph. In *SIGGRAPH*, pages 43–54, 1996.

[163] Stephen R. Gottesman and E. E. Fenimore. New family of binary arrays for coded aperture imaging. *Applied Optics*, 28(20):4344–4352, Oct 1989.

[164] Paul Green, Wenyang Sun, Wojciech Matusik, and Frédo Durand. Multi-aperture photography. In *SIGGRAPH*, volume 26, 2007.

[165] Rajiv Gupta and Richard I. Hartley. Linear pushbroom cameras. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(9):963–975, 1997.

[166] Paul Haeberli. *Synthetic Lighting for Photography*. 1992.

[167] Paul Haeberli. *Synthetic Lighting for Photography*. 2006.

[168] Hamazaki. *Hamazaki book from 1996*. 1996.

[169] Han and Perlin. used a tapered kaleidoscope with a single camera to view the same surface simultaneously from many directions (in chapter 4) to be added.

[170] J. Hanc and S. Tuleja. The feynman quantum mechanics with the help of java apllets and physlets in slovakia. In *10th Workshop on multimedia in physics teaching and larning*, 2005.

[171] Handy. *Handy book from 1986*. 1986.

[172] A Harvey, J Beale, A Greenaway, and T Hanlon. Technology options for imaging spectrometry. In *Proceedings of SPIE*, volume 4132, pages 13–24, 2000.

[173] Martin Harwit and N.J.A. Sloane. *Hadamard Transform Optics*. Academic Press, 1979.

[174] Samuel W. Hasinoff and Kiriakos N. Kutulakos. Confocal stereo. In *ECCV*, volume 1, pages 620–634, 2006.

[175] Samuel W. Hasinoff and Kiriakos N. Kutulakos. A layer-based restoration framework for variable-aperture photography-aperture photography. In *IEEE ICCV*, pages 1–8, 2007.

[176] Samuel W. Hasinoff and Kiriakos N. Kutulakos. Light-efficient photography. In *10th European Conference on Computer Vision (ECCV)*, pages 45–59, 2008.

[177] G. Häusler. A method to increase the depth of focus by two step image processing. *Optics Comm.*, 6(1):38–42, 1972.

[178] James Hays and Alexei A Efros. Scene completion using millions of photographs. *ACM Transactions on Graphics (SIGGRAPH 2007)*, 26(3), 2007.

[179] James Hays and Alexei A. Efros. im2gps: estimating geographic information from a single image. In *Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2008.

[180] X.D. He, K.E. Torrance, F.X. Sillion, and D.P. Greenberg. A comprehensive physical model for light reflection. In *ACM SIGGRAPH Computer Graphics*, volume 25, pages 175–186. ACM, 1991.

[181] O.S. Heavens. *Optical properties of thin solid films*. Dover Publications, 1991.

[182] S. Hecht, S. Shlaer, and M.H. Pirenne. Energy, quanta, and vision. *The Journal of general physiology*, 25(6):819–840, 1942.

[183] Roger D. Hersch, Philipp Donzé, and Sylvain Chosson. Color images visible under uv light. In *SIGGRAPH*, volume 26, 2007.

[184] Aaron Hertzmann, Charles E. Jacobs, Nuria Oliver, Brian Curless, and David H. Salesin. Image analogies. In *SIGGRAPH '01: Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 327–340, New York, NY, USA, 2001. ACM.

[185] Shinsaku Hiura and Takashi Matsuyama. Depth measurement by the multi-focus camera. In *IEEE CVPR*, pages 953–961, 1998.

[186] Derek Hoiem, Alexei A. Efros, and Martial Hebert. Automatic photo pop-up. In *SIGGRAPH '05: ACM SIGGRAPH 2005 Papers*, pages 577–584, New York, NY, USA, 2005. ACM.

[187] et al. Hoppe. to be added. 2003.

[188] Youichi Horry, Ken-Ichi Anjyo, and Kiyoshi Arai. Tour into the picture: using a spidery mesh interface to make animation from a single image. In *SIGGRAPH '97: Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, pages 225–232, New York, NY, USA, 1997. ACM Press/Addison-Wesley Publishing Co.

[189] Roarke Horstmeyer, Gary Euliss, Ravindra Athale, and Marc Levoy. Flexible multimodal camera using a light field architecture. In *IEEE International Conference on Computational Photography*, 2009.

[190] R.S. Hunter and R.W. Harold. *The measurement of appearance*. Wiley-Interscience, 1987.

[191] Takeo Igarashi, Tomer Moscovich, and John F. Hughes. As-rigid-as-possible shape manipulation. In *SIGGRAPH '05: ACM SIGGRAPH 2005 Papers*, pages 1134–1141, New York, NY, USA, 2005. ACM.

[192] Jean in 't Zand. Coded aperture imaging in high-energy astronomy.

[193] M. Irani and S. Peleg. Super resolution from image sequences. *International Conference on Pattern Recognition*, pages 115–120, 1990.

[194] Michal Irani and Shmuel Peleg. Improving resolution by image registration. *Graphical Models and Image Processing*, 53(3):231–239, 1991.

[195] Aaron Isaksen, Leonard McMillan, and Steven J. Gortler. Dynamically reparameterized light fields. In *SIGGRAPH*, 2000.

[196] Herbert E. Ives. A camera for making parallax panoramagrams. *JOSA*, 17(6):435–439, 1928.

[197] Herbert E. Ives. A camera for making parallax panoramagrams. *JOSA*, pages 332–342, 1930.

[198] Nathan Jacobs, Nathaniel Roman, and Robert Pless. Toward fully automatic geo-location and geo-orientation of static outdoor cameras. In *Proc. IEEE Workshop on Video/Image Sensor Networks*, 2008.

[199] John James. *Spectrograph Design Fundamentals*. Cambridge University Press, 2007.

[200] B. Javidi. *Javidi book 2002*. 2002.

[201] B. Javidi and F. Okano. Three-dimensional video and display: Devices and systems. In *Proceedings of a conference held*, volume 5, page 6, 2000.

[202] H.W. Jensen, S.R. Marschner, M. Levoy, and P. Hanrahan. A practical model for subsurface light transport. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 511–518. ACM, 2001.

[203] William R. Johnson, Daniel W. Wilson, and Greg Bearman. Spatial-spectral modulating snapshot hyperspectral imager. *Applied Optics*, 45(9):1898–1908, 2006.

[204] C.S. Johnson Jr. Science for the curious photographer: Appendix b. 2009.

[205] C.S. Johnson Jr. Science for the curious photographer: Appendix c. 2009.

[206] O. JOHNSTON and F. THOMAS. The illusion of life: Disney animation. Disney Editions, 1995.

[207] S. Johnston. *A history of light and colour measurement: science in the shadows*. Taylor & Francis, 2001.

[208] Andrew Jones, Ian McDowall, Hideshi Yamada, Mark Bolas, and Paul Debevec. Rendering for an interactive 360\$°\$*lightfielddisplay. ACMTrans.Graph.*, 26(3) : 40, 2007.

[209] Manjunath V. Joshi, Subhasis Chaudhuri, and Rajkiran Panuganti. Super-resolution imaging: Use of zoom as a cue. *Image and Vision Computing*, 22(14):1185–1196, 2004.

[210] Neel Joshi, Wojciech Matusik, and Shai Avidan. Natural video matting using camera arrays. In *SIGGRAPH*, volume 25, pages 779–786, 2006.

[211] T. Kanade, P. Rander, and PJ Narayanan. Virtualized reality: Constructing virtual worlds from real scenes. *Multimedia, IEEE*, 4(1):34–47, 1997.

[212] D. Keren, S. Peleg, and R. Brada. Image sequence enhancement using sub-pixel displacements. *CVPR*, pages 742–746, 1988.

[213] Bose N. Kim, S. and H. Valenzuela. Recursive reconstruction of high resolution image from noisy undersampled multiframes. *IEEE Trans. Acoustics, Speech, and Signal Processing*, 38:1013–1027, 1990.

[214] et al. Kim. Kim et al. 2008 (chapter 8)—to be added. 2008.

[215] S.P. Kim and W.-Y. Su. Recursive high-resolution reconstruction of blurred multiframe images. *IEEE Transactions on Image Processing*, 2:534–539, October 1993.

[216] Kirmani. Kirmani 2009 (chapter 8)—to be added. 2009.

[217] Knight. Knight article. 1983.

[218] Koike and Naemura. Koike and naemura 2007 (chapter 8)—to be added. 2007.

[219] Takashi Komatsu, Toru Igarashi, Koyoharu Aizawa, and Takahiro Saito. Very high resolution imaging scheme with multiple different-aperture cameras. *Signal Processing: Image Communication.*, 5(5-6):511–526, 1993.

[220] Johannes Kopf, Matt Uyttendaele, Oliver Deussen, and Michael F. Cohen. Capturing and viewing gigapixel images. In *SIGGRAPH '07: ACM SIGGRAPH 2007 papers*, page 93, New York, NY, USA, 2007. ACM.

[221] Vladislav Kraevoy, Alla Sheffer, Ariel Shamir, and Daniel Cohen-Or. Non-homogeneous resizing of complex models. In *SIGGRAPH Asia '08: ACM SIGGRAPH Asia 2008 papers*, pages 1–9, New York, NY, USA, 2008. ACM.

[222] Kuthirummal and Nayar. a class of imaging systems, called radial imaging systems (in chapter 4) to be added.

[223] E.P.F. Lafortune, S.C. Foo, K.E. Torrance, and D.P. Greenberg. Non-linear approximation of reflectance functions. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, pages 117–126. ACM Press/Addison-Wesley Publishing Co., 1997.

[224] Jean-François Lalonde, Derek Hoiem, Alexei A. Efros, Carsten Rother, John Winn, and Antonio Criminisi. Photo clip art. *ACM Transactions on Graphics (SIGGRAPH 2007)*, 26(3), August 2007.

[225] A. Mitros Landolt, O. and C. Koch. Visual sensor with resolution enhancement by mechanical vibrations. In *Proceedings of the Conference on Advanced Research in VLSI*. IEEE Computer Society, 249–264 2001.

[226] Gregory Ward Larson. LogLuv encoding for full-gamut, high-dynamic range images. *Journal of Graphics Tools: JGT*, 3(1):15–31, 1998.

[227] John Lasseter. Principles of traditional animation applied to 3d computer animation. In *SIGGRAPH '87: Proceedings of the 14th annual conference on Computer graphics and interactive techniques*, pages 35–44, New York, NY, USA, 1987. ACM.

[228] Ledda. compare various tone mapping operators using hdr displays. 2005.

[229] A. Levin, S.W. Hasinoff, P. Green, F. Durand, and W.T. Freeman. 4d frequency analysis of computational cameras for depth of field extension. In *ACM Transactions on Graphics (TOG)*, volume 28, page 97. ACM, 2009.

[230] A. Levin, Y. Weiss, F. Durand, and W.T. Freeman. Understanding and evaluating blind deconvolution algorithms. 2009.

[231] Anat Levin, Rob Fergus, Frédo Durand, and William T. Freeman. Image and depth from a conventional camera with a coded aperture. In *SIGGRAPH*, volume 26, 2007.

[232] Anat Levin, William T. Freeman, and Frédo Durand. Understanding camera tradeoffs through a bayesian analysis of light field projections. In *ECCV*, page to appear, 2008.

[233] Anat Levin, Dani Lischinski, and Yair Weiss. Colorization using optimization. In *SIGGRAPH '04: ACM SIGGRAPH 2004 Papers*, pages 689–694, New York, NY, USA, 2004. ACM.

[234] Anat Levin, Alex Rav-Acha, and Dani Lischinski. Spectral matting. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(10):1699–1712, 2008.

[235] Anat Levin, Peter Sand, Taeg Sang Cho, Frédo Durand, and William T. Freeman. Motion-invariant photography. In *SIGGRAPH '08: ACM SIGGRAPH 2008 papers*, pages 1–9, New York, NY, USA, 2008. ACM.

[236] G. G. Levin and G. N. Vishnyakov. On the possibilities of chronotomography of high-speed processes. *Optics Communications*, 56:231–234, 1985.

[237] Levoy. Digital michelangelo. In *SIGGRAPH*, 2004.

[238] M. Levoy, B. Chen, V. Vaish, M. Horowitz, I. McDowall, and M. Bolas. Synthetic aperture confocal imaging. In *ACM Transactions on Graphics (TOG)*, volume 23, pages 825–834. ACM, 2004.

[239] Marc Levoy, Billy Chen, Vaibhav Vaish, Mark Horowitz, Ian McDowall, and Mark Bolas. Synthetic aperture confocal imaging. In *SIGGRAPH*, volume 23, pages 825–834, 2004.

[240] Marc Levoy and Pat Hanrahan. Light field rendering. In *SIGGRAPH*, pages 31–42, 1996.

[241] Marc Levoy, Ren Ng, Andrew Adams, Matthew Footer, and Mark Horowitz. Light field microscopy. In *SIGGRAPH*, pages 924–934, 2006.

[242] Leyvand. data-driven enhancement of visual features–to be added.

[243] Wenchong Li and Chunhua Ma. Imaging spectroscope with an optical recombination system. In *SPIE Three-Dimensional Bioimaging Systems and Lasers in the Neurosciences*, volume 1428, pages 242–248, 1991.

[244] Chia-Kai Liang, Tai-Hsu Lin, Bing-Yi Wong, Chi Liu, and Homer Chen. Programmable aperture photography: Multiplexed light field acquisition. In *SIGGRAPH*, volume 27, pages 1–10, 2008.

[245] Chia-Kai Liang, Gene Liu, and Homer H. Chen. Light field acquisition using programmable aperture camera. In *IEEE International Conference on Image Processing*, 2007.

[246] Zhouchen Lin and Heung-Yeung Shum. Fundamental limits of reconstruction-based superresolution algorithms under local translation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(1):83–97, 2004.

[247] Linguatec. Useful information on shoot & translate. http://www.linguatec.net/products/mtr/information/tech_basics, 2009.

[248] M. G. Lippmann. Epreuves reversible donnant la sensation du relief. *Journal of Physics*, 7:821–825, 1908.

[249] Lischinski. Constraint propagation approaches proposed by lischinski et al.—to be added. 2006.

[250] Ce Liu, Antonio Torralba, William T. Freeman, Frédo Durand, and Edward H. Adelson. Motion magnification. In *SIGGRAPH '05: ACM SIGGRAPH 2005 Papers*, pages 519–526, New York, NY, USA, 2005. ACM.

[251] Nicholas MacKinnon, Ulrich Stange, Pierre Lane, Calum MacAulay, and Mathieu Quatrevalet. Spectrally programmable light engine for in vitro or in vivo molecular imaging and spectroscopy. *Applied Optics*, 44:2033–2040, 2005.

[252] Madden. exposure bracketing–to be added. 1993.

[253] Magnor. to be added. 2003.

[254] Dan Gelb Malzbender, Tom and Hans Wolters. Polynomial texture maps. In *Proceedings, SIGGRAPH 2001*, 2001.

[255] Picard Mann. use the estimated response function to process the images before merging them–to be added. 1995.

[256] Mantiuk. Gradient domain manipulation methods–to be added. 2006.

[257] Emil Martinec. Noise, dynamic range and bit depth in digital slrs. http://theory.uchicago.edu/~ejm/pix/20d/tests/noise/.

[258] Marwick and Andreou 2006. Marwick and andreou 2006 (chapter 8)—to be added.

[259] Masia. a method for evaluating the reverse tone mapping algorithms based on varying exposure conditions. 2009.

[260] V. Masselus, P. Peers, P. Dutré, and Y.D. Willems. Relighting with 4d incident light fields. In *ACM Transactions on Graphics (TOG)*, volume 22, pages 613–620. ACM, 2003.

[261] D. Hihara-T. Ushiro S. Yoshimura-J. Rekimoto Matsushita, N. and Y. Yamamoti. A smart camera for scene capturing and id recognition. In *International Symposium on Mixed and Augmented Reality*, 2003.

[262] W. Matusik and H. Pfister. 3d tv: a scalable system for real-time acquisition, transmission, and autostereoscopic display of dynamic scenes. In *ACM Transactions on Graphics (TOG)*, volume 23, pages 814–824. ACM, 2004.

[263] W. Matusik, H. Pfister, M. Brand, and L. McMillan. A data-driven reflectance model. In *Proceedings International Conference on Computer Graphics and Interactive Techniques, ACM SIGGRAPH*, pages 27–31, 2003.

[264] J.J. McCann and A. Rizzi. Camera and visual veiling glare in hdr images. *J. Soc. Information Display*, 15(9):721–730, 2007.

[265] Morgan McGuire, Wojciech Matusik, Billy Chen, John F. Hughes, Hanspeter Pfister, and Shree Nayar. Optical splitting trees for high-precision monocular imaging. *IEEE Computer Graphics and Applications*, 27(2):32–42, 2007.

[266] Morgan McGuire, Wojciech Matusik, Hanspeter Pfister, Billy Chen, John F. Hughes, and Shree K. Nayar. Optical splitting trees for high-precision monocular imaging. *IEEE Computer Graphics and Applications*, 27(2):32–42, 2007.

[267] Morgan McGuire, Wojciech Matusik, Hanspeter Pfister, John F. Hughes, and Frédo Durand. Defocus video matting. In *SIGGRAPH*, volume 24, pages 567–576, 2005.

[268] Leonard McMillan and Gary Bishop. Plenoptic modeling: An image-based rendering system. In *SIGGRAPH*, pages 39–46, 1995.

[269] O.G. Memis, A. Katsnelson, S.C. Kong, H. Mohseni, M. Yan, S. Zhang, T. Hossain, N. Jin, and I. Adesida. A photon detector with very high gain at low bias and at room temperature. *Applied Physics Letters*, 91(17):171112–171112, 2007.

[270] Klaus D. Mielenz. On the diffraction limit for lensless imaging. *Jour nal of Research of the National Institute of Standards and Technology*, 104(5):479–485, 1999.

[271] Peter J. Miller and Clifford C. Hoyt. Spectral imaging system. US Patent 6373568, 2002.

[272] Marvin Minsky. Memoir on inventing the confocal scanning microscope. *Scanning*, 10:128–138, 1988.

[273] James Miskin and David J. C. Mackay. Ensemble learning for blind image separation and.

[274] D.P. Mitchell and A.N. Netravali. Reconstruction filters in computer-graphics. In *ACM Siggraph Computer Graphics*, volume 22, pages 221–228. ACM, 1988.

[275] Mitsunaga and Nayar. use the estimated response function to process the images before merging them–to be added. 1999.

[276] Joan Moh, Hon Mun Low, and Greg Wientjes. Characterization of the nikon d-70 digital camera, 2003.

[277] Ankit Mohan, Xiang Huang, Ramesh Raskar, and Jack Tumblin. Sensing increased image resolution using aperture masks. In *IEEE CVPR*, 2008.

[278] Ankit Mohan, Douglas Lanman, Shinsaku Hiura, and Ramesh Raskar. Image destabilization: Programmable defocus using lens and sensor motion. In *IEEE International Conference on Computational Photography*, 2009.

[279] Ankit Mohan, Ramesh Raskar, and Jack Tumblin. Agile spectrum imaging: Programmable wavelength modulation for cameras and projectors. In *Eurographics 2008*, volume 27, pages 709–717, 2008.

[280] Jonathan M. Mooney, Virgil E. Vickers, Myoung An, and Andrzej K. Brodzik. High-throughput hyperspectral infrared camera. *JOSA A*, 14(11):2951–2961, 1997.

[281] Matusik Pfister-Hughes Durand Morgan, McGuire. Defocus video matting. *ACM Transactions on Graphics*, 24, No. 3:835–846, July 2005.

[282] Morimura. exposure bracketing–to be added. 1993.

[283] Eric N. Mortensen and William A. Barrett. Intelligent scissors for image composition. In *SIGGRAPH '95: Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pages 191–198, New York, NY, USA, 1995. ACM.

[284] Hajime Nagahara, Sujit Kuthirummal, Changyin Zhou, and Shree K. Nayar. Flexible depth of field photography. In *ECCV*, 2008.

[285] S. Nakajima, K. Nakamura, K. Masamune, I. Sakuma, and T. Dohi. Three-dimensional medical imaging display with computer-generated integral photography. *Computerized medical imaging and graphics*, 25(3):235–241, 2001.

[286] Srinivasa G. Narasimhan and Shree K. Nayar. Enhancing resolution along multiple imaging dimensions using assorted pixels. *IEEE PAMI*, 27(4):518–530, 2005.

[287] Nayar and Narsihman. pixels with assorted attenuation–to be added. 2003.

[288] Shree Nayar. Computational camera and programmable imaging. *Symposium on Computational Photography and Video*, May, 2005.

[289] Shree K. Nayar, Peter N. Belhumeur, and Terry E. Boult. Lighting sensitive display. *ACM Trans. Graph.*, 23(4):963–979, 2004.

[290] Shree K. Nayar and Vlad Branzoi. Adaptive dynamic range imaging: Optical control of pixel exposures over space and time. In *IEEE ICCV*, volume 2, pages 1168–1175, Oct 2003.

[291] Shree K. Nayar, Vlad Branzoi, and Terry E. Boult. Programmable imaging: Towards a flexible camera. In *IJCV*, volume 70, pages 7–22, 2006.

[292] Shree K. Nayar, Gurunandan Krishnan, Michael D. Grossberg, and Ramesh Raskar. Fast separation of direct and global components of a scene using high frequency illumination. In *SIGGRAPH*, volume 25, pages 935–944, 2006.

[293] Shree K. Nayar and Tomoo Mitsunaga. High dynamic range imaging: Spatially varying pixel exposures. In *IEEE CVPR*, volume 1, pages 472—479, 2000.

[294] S.K. Nayar. Computational cameras: Redefining the image. *Computer*, 39(8), Aug. 2006.

[295] S.K. Nayar, P.N. Belhumeur, and T.E. Boult. Lighting sensitive display. *ACM Transactions on Graphics (TOG)*, 23(4):963–979, 2004.

[296] V. Branzoi Nayar, S.K. and T. Boult. Programmable imaging using a digital micromirror array. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2004.

[297] Neel Neel Joshi. *Joshi thesis (chapter 8)—to be added*. PhD thesis.

[298] Neil R. Nelson, Paul Bryant, and Rob Sundberg. Development of a hyperspectral scene generator. In *SPIE Earth Observing Systems VIII*, volume 5151, pages 480–484, 2003.

[299] Isaac Newton. *Opticks*. Royal Society, 1704.

[300] Ng. to be added. 2005.

[301] Ren Ng. Fourier slice photography. In *SIGGRAPH*, volume 24, pages 735–744, 2005.

[302] Ren Ng and Pat Hanrahan. Digital correction of lens aberrations in light field photography. In *SPIE International Optical Design Conference*, volume 6342, page 63421E, 2006.

[303] Ren Ng, Marc Levoy, Mathieu Brédif, Gene Duval, Mark Horowitz, and Pat Hanrahan. Light field photography with a hand-held plenoptic camera. Technical Report CTSR 2005-02, Stanford University, 2005.

[304] Y. Nomura, L. Zhang, and S.K. Nayar. Scene Collages and Flexible Camera Arrays. In *Proceedings of Eurographics Symposium on Rendering*, Jun 2007.

[305] S.B. Oh, G. Barbastathis, and R. Raskar. Augmenting light field and ray optics to model wave optics effects. Technical report, MIT, 2009.

[306] Takayuki Okamoto and Ichirou Yamaguchi. Simultaneous acquisition of spectral image information. *Optics Letters*, 16(16):1277–1279, 1991.

[307] Takanori Okoshi. *Three-dimensional Imaging Techniques*. Academic Press Inc., 1976.

[308] James Olson, Robert K. Jungquist, and Zoran Ninkov. Tunable multispectral imaging system technology for airborne applications. In *Proceedings of SPIE*, volume 2480, pages 268–279, 1995.

[309] OPTOTRAK. Ndi optotrak certus spatial measurement, 2007.

[310] Tomáš Pajdla. Stereo with oblique cameras. *IJCV*, 47(123):161–170, 2002.

[311] Paris. Durand et al. then implemented a real-time bilateral filtering–based tone-mapping framework–to be added. 2006.

[312] Sylvain Paris, Pierre Kornprobst, Jack Tumblin, and Frédo Durand. A gentle introduction to bilateral filtering and its applications. ACM SIGGRAPH 2008, 2008.

[313] Jong-Il Park, Moon-Hyun Lee, Michael D. Grossberg, and Shree K. Nayar. Multi-spectral imaging using multiplexed illumination. In *ICCV*, pages 1–8, 2007.

[314] Sung C. Park, Min K. Park, and Moon G. Kang. Super-resolution image reconstruction: a technical overview. *Signal Processing Magazine, IEEE*, 20(3):21–36, 2003.

[315] P. Peers, D.K. Mahajan, B. Lamond, A. Ghosh, W. Matusik, R. Ramamoorthi, and P. Debevec. Compressive light transport sensing. *ACM Transactions on Graphics (TOG)*, 28(1):3, 2009.

[316] Alex P. Pentland. A new sense for depth of field. *IEEE PAMI*, pages 523–531, 1987.

[317] Patrick Pérez, Michel Gangnet, and Andrew Blake. Poisson image editing. *ACM Trans. Graph.*, 22(3):313–318, 2003.

[318] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 12(7):629–639, 1990.

[319] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 12(7):629–639, Jul 1990.

[320] C. Pethick and H. Smith. *Bose-Einstein condensation in dilute gases*. Cambridge Univ Pr, 2002.

[321] L.P. Pitaevskii and S. Stringari. *Bose-Einstein Condensation*, volume 116. Oxford University Press, USA, 2003.

[322] Thomas Porter and Tom Duff. Compositing digital images. In *SIGGRAPH '84: Proceedings of the 11th annual conference on Computer graphics and interactive techniques*, pages 253–259, New York, NY, USA, 1984. ACM.

[323] JL Posdamer and MD Altschuler. Surface measurement by space-encoded projected beam systems. *Computer graphics and image processing*, 18(1):1–17, 1982.

[324] Joanna L. Power, Brad S. West, Eric J. Stollnitz, and David H. Salesin. Reproducing color images as duotone. In *SIGGRAPH*, pages 237–248. ACM Press, 1996.

[325] Digital Light Processing. http://en.wikipedia.org/wiki/DLP.

[326] Prokudin-Gorskii. *Sergei Mikhailovich, 1863–1944, photographer*. Library of Congress, Prints and Photographs Division. Reproduction number: LC-P87-8086A-2.

[327] D Rajan, S Chaudhuri, and M.V. Joshi. Multi-objective super resolution: concepts and examples. *Signal Processing Magazine, IEEE*, 20(3):49–61, 2003.

[328] Karl Rasche, Robert Geist, and James Westall. Re-coloring images for gamuts of lower dimension. *Comput. Graph. Forum*, 24(3):423–432, 2005.

[329] Raskar. occluding edges with multiple flashes to be added.

[330] A. Ilie Raskar, R. and J. Yu. Image fusion for context enhancement and video surrealism. In *Proceedings of the 3rd International Symposium on Non-photorealistic Animation and Rendering*, pages 85–152, 2004.

[331] Ramesh Raskar, Amit Agrawal, and Jack Tumblin. Coded exposure photography: motion deblurring using fluttered shutter. In *SIGGRAPH '06*, pages 795–804, 2006.

[332] Ramesh Raskar, Amit Agrawal, Cyrus Wilson, and Ashok Veeraraghavan. Glare aware photography: 4d ray sampling for reducing glare effects of camera lenses. In *SIGGRAPH*, pages 1–10, 2008.

[333] Ramesh Raskar, Jack Tumblin, Ankit Mohan, Amit Agrawal, and Yuanzen Li. Eurographics 2006 star state of the art report computational photography. 2008.

[334] Rav-Acha. unwrapped mosaics–to be added.

[335] algorithms RBA. reconstruction-base superresolution algorithms (rba).

[336] Eric Reinhard. *HDR*. 2005.

[337] Rempel. a robust algorithm for converting legacy ldr video and photographs to hdr versions in real time. 2007.

[338] Vision Research. http://www.visionresearch.com/.

[339] J. P. Rice, S. W. Brown, and J. E. Neira. Development of hyperspectral image projectors. In *SPIE Infrared Spaceborne Remote Sensing XIV*, volume 6297, page 629701, 2006.

[340] Joseph P. Rice, Steven W. Brown, Jorge E. Neira, and Robert R. Bousquet. Hyperspectral image projector for hyperspectral imagers. In *SPIE*, volume 6565, page 65650C, 2007.

[341] J.P. Rice, S.W. Brown, B.C. Johnson, and J.E. Neira. Hyperspectral image projectors for radiometric applications. *Metrologia*, 43:S61–S65, 2006.

[342] David E. Roberts and Trebor Smith. The history of integral print methods. http://www.integralresource.org/integral_history.html.

[343] B. Robertson. Big moves. *Computer Graphics World*, 29(11), 2006.

[344] V. Ronchi and E. Rosen. *Optics: the science of vision*. Dover Publications, 1991.

[345] A. Rose. *Vision: Human and Electronic.* (Kluwer Academic, 1973.

[346] Eric Rosenthal, Richard Jay Solomon, and Clark Johnson. Full spectrum color projector. US Patent 6985294, 2006.

[347] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake. "grabcut": interactive foreground extraction using iterated graph cuts. In *SIGGRAPH '04: ACM SIGGRAPH 2004 Papers*, pages 309–314, New York, NY, USA, 2004. ACM.

[348] Michael Rubinstein, Ariel Shamir, and Shai Avidan. Improved seam carving for video retargeting. In *SIGGRAPH '08: ACM SIGGRAPH 2008 papers*, pages 1–9, New York, NY, USA, 2008. ACM.

[349] Olaf Hall-Holt Rusinkiewicz, Szymon and Marc Levoy. Real-time 3d model acquisition. 2002.

[350] D. Salesin. to be added. *Symposium on Computational Geometry.*

[351] Jordi Pagès Salvi, j. and Joan Batlle. Pattern codification strategies in structured light systems. In *Pattern Recognition*, 2004.

[352] Peter Sand and Seth Teller. Video matching. *ACM Trans. Graph.*, 23(3):592–599, 2004.

[353] Sayag. Sayag article in 1990. 1990.

[354] Scharstein. Scharstein 1996 (chapter 8)—to be added. 1996.

[355] Yoav Y Schechner, Srinivasa G Narasimhan, , and Shree K Nayar. Instant dehazing of images using polarization. *IEEE CVPR*, pages 325–332, 2001.

[356] Yoav Y. Schechner and Shree K. Nayar. Generalized mosaicing. In *IEEE ICCV*, volume 1, pages 17–24, 2001.

[357] Yoav Y. Schechner and Shree K. Nayar. Generalized mosaicing: Wide field of view multispectral imaging. *Pattern Analysis and Machine Intelligence*, 24(10):1334–1348, 2002.

[358] Yoav Y. Schechner and Shree K. Nayar. Uncontrolled modulation imaging. In *IEEE CVPR*, volume 2, pages 197—204, 2004.

[359] Scheffer. Scheffer article in 2000. 2000.

[360] Helge Seetzen, Wolfgang Heidrich, Wolfgang Stuerzlinger, Greg Ward, Lorne Whitehead, Matthew Trentacoste, Abhijeet Ghosh, and Andrejs Vorozcovs. High dynamic range display systems. In *SIGGRAPH*, 2004.

[361] S.M. Seitz, Y. Matsushita, and K.N. Kutulakos. A theory of inverse light transport. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, volume 2, pages 1440–1447. IEEE, 2005.

[362] P. Sen, B. Chen, G. Garg, S.R. Marschner, M. Horowitz, M. Levoy, and H. Lensch. Dual photography. In *ACM Transactions on Graphics (TOG)*, volume 24, pages 745–755. ACM, 2005.

[363] Pradeep Sen and Soheil Darabi. Compressive Dual Photography. *Computer Graphics Forum*, 28(2):609 – 618, 2009.

[364] Foveon X3 sensor. http://en.wikipedia.org/wiki/Foveon_X3_sensor.

[365] Vidya Setlur, Saeko Takagi, Ramesh Raskar, Michael Gleicher, and Bruce Gooch. Automatic image retargeting. In *MUM '05: Proceedings of the 4th international conference on Mobile and ubiquitous multimedia*, pages 59–68, New York, NY, USA, 2005. ACM.

[366] Ariel Shamir and Shai Avidan. Seam carving for media retargeting. *Commun. ACM*, 52(1):77–85, 2009.

[367] Shechtman. to be added.

[368] Eli Shechtman, Yaron Caspi, and Michal Irani. Space-time super-resolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(4):531–545, 2005.

[369] I. Simon, N. Snavely, and S.M. Seitz. Scene summarization for online image collections. pages 1–8, Oct. 2007.

[370] G. K. Skinner. X-ray imaging with coded masks. *Scientific American*, 259:84, August 1988.

[371] Alvy Ray Smith and James F. Blinn. Blue screen matting. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pages 259–268, New York, NY, USA, 1996. ACM.

[372] A.R. Smith. A pixel is not a little square, a pixel is not a little square, a pixel is not a little square! Technical report, Technical report, Microsoft, Inc., 1995. ftp://ftp. alvyray. com/Acrobat/6_Pixel. pdf, 1995.

[373] Kaleigh Smith, Pierre-Edouard Landes, Jöelle Thollot, and Karol Myszkowski. Apparent greyscale: A simple and fast conversion to perceptually accurate images and video. In Roberto Scopigno and Eduard Gröller, editors, *Computer Graphics Forum (Proc. EUROGRAPHICS)*, volume 27(2), pages 193–200, Crete, Greece, 2008. European Association for Computer Graphics (Eurographics), Blackwell.

[374] Stephen M. Smith and J. Michael Brady. Susan—a new approach to low level image processing. *Int. J. Comput. Vision*, 23(1):45–78, 1997.

[375] Noah Snavely, Rahul Garg, Steven M. Seitz, and Richard Szeliski. Finding paths through the world's photos. *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2008)*, 27(3):11–21, 2008.

[376] Noah Snavely, Steven M. Seitz, and Richard Szeliski. Photo tourism: exploring photo collections in 3d. In *SIGGRAPH '06: ACM SIGGRAPH 2006 Papers*, pages 835–846, New York, NY, USA, 2006. ACM.

[377] solarsails. Solar sails on spacecraft (chapter 2) to be added. 2008.

[378] R.L. Solso. *Cognition and the visual arts.* The MIT press, 1996.

[379] Sony. Evis chip, developed by sony and the sony kihara research center.

[380] Eric J. Stollnitz, Victor Ostromoukhov, and David H. Salesin. Reproducing color images using custom inks. In *SIGGRAPH*, pages 267–274. ACM Press, 1998.

[381] Street. *Street book from 1998.* 1998.

[382] Hiroaki Sugiura, Hideyuki Kaneko, Shuichi Kagawa, Jun Someya, and Hideki Tanizoe. Six-primary-color lcd monitor using six-color leds with an accurate calibration system. In Gabriel G. Marcu Reiner Eschbach, editor, *SPIE Color Imaging XI: Processing, Hardcopy, and Applications*, volume 6058, 2006.

[383] Sun. to be added.

[384] Sun. to be added. In *SIGGRAPH*, 2006.

[385] Jian Sun, Jiaya Jia, Chi-Keung Tang, and Heung-Yeung Shum. Poisson matting. In *SIGGRAPH '04: ACM SIGGRAPH 2004 Papers*, pages 315–321, New York, NY, USA, 2004. ACM.

[386] Kalyan Sunkavalli, Wojciech Matusik, Hanspeter Pfister, and Szymon Rusinkiewicz. Factored time-lapse video. In *SIGGRAPH '07: ACM SIGGRAPH 2007 papers*, page 101, New York, NY, USA, 2007. ACM.

[387] Daniel Sýkora, Jan Buriánek, and Jiří Žára. Unsupervised colorization of black-and-white cartoons. In *NPAR '04: Proceedings of the 3rd international symposium on Non-photorealistic animation and rendering*, pages 121–127, New York, NY, USA, 2004. ACM.

[388] Dharmpal Takhar, Jason N. Laska, Michael B. Wakin, Marco F. Duarte, Dror Baron, Shriram Sarvotham, Kevin F. Kelly, and Richard G. Baraniuk. A new compressive imaging camera architecture using optical-domain compression. In *Computational Imaging IV at SPIE Electronic Imaging*, 2006.

[389] Eino-Ville Talvala, Andrew Adams, Mark Horowitz, and Marc Levoy. Veiling glare in high dynamic range imaging. In *SIGGRAPH*, volume 26, 2007.

[390] Eino-Ville Talvala, Andrew Adams, Mark Horowitz, and Marc Levoy. Veiling glare in high dynamic range imaging. In *SIGGRAPH '07: ACM SIGGRAPH 2007 papers*, page 37, New York, NY, USA, 2007. ACM.

[391] E.F. Taylor, S. Vokos, J.M. OMeara, and N.S. Thornber. Teaching feynmans sum-over-paths quantum theory. *Computers in Physics*, 12:190, 1998.

[392] F. Tejada, A.G. Andreou, and P.O. Pouliquen. Stacked, standing wave detectors in 3d soi-cmos. In *Circuits and Systems, 2006. ISCAS 2006. Proceedings. 2006 IEEE International Symposium on*, pages 4–pp. IEEE, 2006.

[393] Laura Teodosio and Walter Bender. Salient video stills: content and context preserved. In *MULTIMEDIA '93: Proceedings of the first ACM international conference on Multimedia*, pages 39–46, New York, NY, USA, 1993. ACM.

[394] the Dolby Brightside. 2004dolbybrightside (chapter 8)—to be added.

[395] the Irvine Sensors. Irvine sensors (chapter8)—to be added.

[396] the MOSAIC initiative. Mosaic initiative (chapter 8)—to be added.

[397] the Optotune Corporation. Optotune inc. (chapter 8)—to be added.

[398] the Variotopic Corporation. Varioptic inc. (chapter 8) to be added.

[399] Tomasi and Manduchi. In *ICCV*, 1998.

[400] A. Torralba, K.P. Murphy, W.T. Freeman, and M.A. Rubin. Context-based vision system for place and object recognition. pages 273–280 vol.1, Oct. 2003.

[401] Tali Treibitz and Yoav Y. Schechner. Active polarization descattering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009.

[402] R. A. Stack R. L. Morrison Tremblay, E. J. and J. E. Ford. The origami lens to create ultrathin cameras using annular folded optics. *Applied Optics*, 2007.

[403] Trifonov. tomography. 2006.

[404] Tsai. exposure bracketing–to be added. 1994.

[405] R. Y. Tsai and T. S. Huang. Multiframe image restoration and registration. *Advances in Computer Vision and Image Processing*, pages 317–339, 1984.

[406] J. Tumblin, A. Agrawal, and R. Raskar. Why i want a gradient camera. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 103–110. IEEE, 2005.

[407] J. Tumblin and Rushmeier. to be included? 1993.

[408] J. Tumblin and Turk. Lcis by tumblin and turk. 1999.

[409] Vaibhav Vaish, Gaurav Garg, Eino-Ville Talvala, Emilio Antunez, Bennett Wilburn, Mark Horowitz, and Marc Levoy. Synthetic aperture focusing using a shear-warp factorization of the viewing transform. In *A3DISS (at CVPR)*, 2005.

[410] Vaibhav Vaish, Bennett Wilburn, Neel Joshi, and Marc Levoy. Using plane + parallax for calibrating dense camera arrays. In *IEEE CVPR*, 2004.

[411] Patrick Vandewalle, Sabine Süsstrunk, and Martin Vetterli. A frequency domain approach to registration of aliased images with application to super-resolution. *EURASIP Journal on Applied Signal Processing*, 2006.

[412] Ashok Veeraraghavan, Amit Agrawal, Ramesh Raskar, Ankit Mohan, and Jack Tumblin. Non-refractive modulators for encoding and capturing scene appearance and depth. In *IEEE CVPR*, 2008.

[413] Ashok Veeraraghavan, Ramesh Raskar, Amit Agrawal, Ankit Mohan, and Jack Tumblin. Dappled photography: mask enhanced cameras for heterodyned light fields and coded aperture refocusing. In *SIGGRAPH '07: ACM SIGGRAPH 2007 papers*, page 69, New York, NY, USA, 2007. ACM.

[414] Ashok Veeraraghavan, Ramesh Raskar, Amit Agrawal, Ankit Mohan, and Jack Tumblin. Dappled photography: Mask enhanced cameras for heterodyned light fields and coded aperture refocusing. In *SIGGRAPH*, volume 26, pages 69:1–69:12. ACM, 2007.

[415] Vicon. Viconpeak, camera mx 40. www.vicon.com/products/mx40.html, 2006.

[416] Wagner. to be added. 2005.

[417] Christine F. Wall, Andrew R. Hanson, and Julie A. F. Taylor. Construction of a programmable light source for use as a display calibration artefact. In *SPIE*, volume 4295, pages 259–266, 2001.

[418] Jue Wang and Michael F. Cohen. Image and video matting: a survey. *Found. Trends. Comput. Graph. Vis.*, 3(2):97–175, 2007.

[419] Jue Wang, Steven M. Drucker, Maneesh Agrawala, and Michael F. Cohen. The cartoon animation filter. In *SIGGRAPH '06: ACM SIGGRAPH 2006 Papers*, pages 1169–1173, New York, NY, USA, 2006. ACM.

[420] G. Ward and R. Shakespeare. Rendering with radiance. *The Art*, 1998.

[421] Greg Ward. a compact, easy-to-read "rgbe"–to be added.

[422] Li-Yi Wei and Marc Levoy. Fast texture synthesis using tree-structured vector quantization. In *SIGGRAPH '00: Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 479–488, New York, NY, USA, 2000. ACM Press/Addison-Wesley Publishing Co.

[423] Yair Weiss. Deriving intrinsic images from image sequences. In *Proceedings International Conference on Computer Vision*, 2001.

[424] Wen. *Wen book from 1989.* 1989.

[425] B. Wilburn, N. Joshi, V. Vaish, E.V. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy. High performance imaging using large camera arrays. *ACM Transactions on Graphics*, 24(3):765–776, 2005.

[426] Bennett Wilburn, Neel Joshi, Vaibhav Vaish, Marc Levoy, and Mark Horowitz. High speed videography using a dense camera array. In *IEEE CVPR*, volume 2, pages 294–301, 2004.

[427] Bennett Wilburn, Neel Joshi, Vaibhav Vaish, Eino-Ville Talvala, Emilio Antunez, Adam Barth, Andrew Adams, Mark Horowitz, and Marc Levoy. High performance imaging using large camera arrays. In *SIGGRAPH*, volume 24, pages 765–776, 2005.

[428] Alexander Wilkie, Andrea Weidlich, Caroline Larboulette, and Werner Purgathofer. A reflectance model for diffuse fluorescent surfaces. In *Graphite*, pages 321–328, 11 2006.

[429] Rebecca Willett, Michael E. Gehm, and David J. Brady. Multiscale reconstruction for computational spectral imaging. In *SPIE*, volume 6498, page 64980L, 2007.

[430] Lance Williams. to be added. 2008.

[431] Ankit Mohan Jack Tumblin Winnemöeller, Holger and Bruce Gooch. Light waving: Estimating light positions from photographs alone. *Computer Graphics Forum*, 24(3):433–438, 2005.

[432] Holger Winnemöller, Sven C. Olsen, and Bruce Gooch. Real-time video abstraction. In *SIGGRAPH '06: ACM SIGGRAPH 2006 Papers*, pages 1221–1226, New York, NY, USA, 2006. ACM.

[433] Jason C. Yang, Matthew Everett, Chris Buehler, and Leonard McMillan. A real-time distributed light field camera. In *EGRW02: Proceedings of the 13th Eurographics workshop on Rendering*, pages 77–86, Aire-la-Ville, Switzerland, Switzerland, 2002. Eurographics Association.

[434] Jason C. Yang, Matthew Everett, Chris Buehler, and Leonard McMillan. A real-time distributed light field camera. In *13th Eurographics workshop on rendering*, pages 77–86, 2002.

[435] Jingyi Yu and Leonard McMillan. A framework for multiperspective rendering. In *Eurographics Symposium on Rendering*, 2004.

[436] Jingyi Yu and Leonard McMillan. General linear cameras. In *ECCV*, 2004.

[437] Jingyi Yu, Leonard McMillan, and Peter Sturm. Multiperspective modeling, rendering, and imaging. 2008.

[438] Lu Yuan, Jian Sun, Long Quan, and Heung-Yeung Shum. Image deblurring with blurred/noisy image pairs. In *SIGGRAPH '07: ACM SIGGRAPH 2007 papers*, page 1, New York, NY, USA, 2007. ACM.

[439] C. Zhang and T. Chen. A self-reconfigurable camera array. In *Eurographics Symposium on Rendering*, volume 4, page 6, 2004.

[440] Cha Zhang and Tsuhan Chen. Light field capturing with lensless cameras. In *IEEE ICIP*, volume 3, pages 792–795, 2005.

[441] Cha Zhang and Tsuhan Chen. *Light Field Sampling.* Morgan & Claypool, 2006.

[442] T. Zickler. Reciprocal image features for uncalibrated helmholtz stereopsis. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 2, pages 1801–1808. IEEE, 2006.

[443] T.E. Zickler, J. Ho, D.J. Kriegman, J. Ponce, and P.N. Belhumeur. Binocular helmholtz stereopsis. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 1411–1417. IEEE, 2003.

[444] Assaf Zomet, Doron Feldman, Shmuel Peleg, and Daphna Weinshall. Non-perspective imaging and rendering with the crossed-slits projection. Technical Report 2002-41, Leibnitz Center, Hebrew University of Jerusalem, 2002.

[445] Assaf Zomet and Shree K. Nayar. Lensless imaging with a controllable aperture. In *IEEE CVPR*, volume 1, pages 339–346, 2006.