Computational Photography

Computational Photography

Mastering New Techniques for Lenses, Lighting, and Sensors

Ramesh Raskar
Jack Tumblin

A K Peters, Ltd.
Wellesley, Massachusetts

Editorial, Sales, and Customer Service Office

A K Peters, Ltd.
888 Worcester Street, Suite 230
Wellesley, MA 02482
www.akpeters.com

Printed in the United States of America
12 11 10 09 08 10 9 8 7 6 5 4 3 2 1

Contents

1 Introduction

1.1 What is Computational Photography? . . . .

1.2 Elements of Computational Photography . . . .

1.3 Sampling the Dimensions of Imaging . . . .

1.3.1 Past: Film-Like Digital Photography . . . .

1.3.2 Present: Epsilon Photography . . . .

1.3.3 Future: Coded Photography . . . .

2 Camera Fundamentals

2.1 Understanding Film-Like Digital Photography . . . .

2.1.1 Lenses, Apertures and Aberrations . . . .

2.1.2 Sensors and Noise . . . .

2.1.3 Lighting . . . .

2.2 Image Formation Models . . . .

2.2.1 Generalized Projections . . . .

2.2.2 Generalized Linear Cameras . . . .

2.2.3 Ray-based Concepts and Light Fields . . . .

3 Extending Film-Like Digital Photography

3.1 Understanding Limitations . . . .

3.2 Fusion of Multiple Images . . . .

3.2.1 Sort-First versus Sort-Last Capture . . . .

3.2.2 Time and Space Multiplexed Capture . . . .

3.2.3 Hybrid Space-Time Multiplexed Systems . . . . .

3.3 Improving Dynamic Range . . . .
3.3.1 Capturing High Dynamic Range . . . .
3.3.2 Tone Mapping . . . .
3.3.3 Compression and Display . . . .
3.4 Extended Depth of Field . . . .
3.5 Beyond Tri-Color Sensing . . . .
3.6 Wider Field of View . . . .
3.6.1 Panorama via Image Stitching . . . .
3.6.2 Extreme Zoom . . . .
3.7 Higher Frame Rate . . . .
3.8 Improving Resolution . . . .
3.8.1 Camera or Sensor Translation . . . .
3.8.2 Apposition Eyes . . . .
3.9 Suppression of Glare . . . .

4 Illumination
4.1 Exploiting Duration and Brightness . . . .
4.1.1 Stroboscope for Freezing High Speed Motion . . .
4.1.2 Sequential Multi-Flash Stroboscopy . . . .
4.2 Presence or Absence of Flash . . . .
4.2.1 Flash/No-Flash Pair for Denoising . . . .
4.2.2 Flash, Exposure and Dynamic Range . . . .
4.2.3 Removing Flash Artifacts . . . .
4.2.4 Flash-Based Matting . . . .
4.3 Modifying Color and Wavelength . . . .
4.4 Position and Orientations of Lighting . . . .
4.4.1 Shape and Detail Enhancement using Multi-Position Flashes . . . .
4.4.2 Relighting using Domes and Light Waving . . . .
4.4.3 Towards Reflectance Fields Capture in 4D, 6D, and 8D . . . .
4.5 Modulation in Space and Time . . . .
4.5.1 Projector for Structured Light . . . .
4.5.2 High Spatial Frequency Patterns . . . .
4.5.3 Modulation in Time . . . .

4.6 Exploiting Natural Illumination Variations . . . .
4.6.1 Intrinsic Images . . . .
4.6.2 Context Enhancement of Night-Time Photos . . .

5 Optics
5.1 Introduction . . . .
5.1.1 Classification of Animal Eyes . . . .
5.1.2 Traditional Optics in Cameras . . . .
5.2 Grouping of Rays in Incident Light Field . . . .
5.2.1 Sort First: Before Lens . . . .
5.2.2 Sort Middle: Inside the Lens . . . .
5.2.3 Sort Last: Near the Sensor . . . .
5.3 Apertures and Masks . . . .
5.3.1 Masks in Non-Imaging Devices . . . .
5.3.2 Lensless Imaging . . . .
5.3.3 Slits and Coded Masks in Aperture . . . .
5.3.4 Non-aperture Masks and Gratings in Object Space
5.3.5 Synthetic Aperture . . . .
5.3.6 Wavefront Coding . . . .
5.4 Focus . . . .
5.4.1 Autofocus Mechanisms . . . .
5.4.2 Modern Methods for Extending Depth of Field . .
5.4.3 Confocal Imaging for Narrow Depth of Field . . .
5.4.4 Defocus for Estimating Scene Geometry . . . . . .
5.5 Color and Wavelength . . . .
5.5.1 Spectrometers . . . .
5.5.2 Diffractive Optics for Reduced Chromatic Aberration . . . .
5.6 Non-standard View and Perspective . . . .
5.6.1 Catadioptric Imaging . . . .
5.6.2 Increasing the Field of View . . . .
5.6.3 Displaced Centers of Projection . . . .
5.6.4 Orthographic View . . . .
5.7 Polarization . . . .
5.7.1 Dehazing in Foggy Conditions . . . .
5.7.2 Underwater Imaging . . . .

6 Modern Image Sensors
6.1 Introduction . . . .
6.1.1 Historic Perspective . . . .
6.1.2 Noise and Resolution Limits . . . .
6.2 Extended Dynamic Range . . . .
6.2.1 Multi-sensor Pixels . . . .
6.2.2 Logarithmic Sensing . . . .
6.2.3 Gradient Camera . . . .
6.3 Color Sensing . . . .
6.4 Three-dimensional (Range) Measurement . . . .
6.4.1 Time of Flight Techniques . . . .
6.4.2 Triangulation-Based Techniques . . . .
6.5 Encoding Identifiable Information . . . .
6.5.1 Identity of Active Beacons . . . .
6.6 Handling Object and Camera Motion . . . .
6.6.1 Line Scan Cameras . . . .
6.6.2 Methods for Image Stabilization . . . .
6.6.3 Performance Capture via Markers . . . .
6.6.4 Hybrid Imaging . . . .
6.6.5 Coded Exposure via Fluttered Shutter . . . .

7 Processing and Reconstruction
7.1 Filtering and Detection . . . .
7.1.1 Noise Reduction . . . .
7.1.2 Colorization and Color to Gray Conversion . . . .
7.1.3 Motion and Focus Deblurring . . . .
7.2 Geometric Operations . . . .
7.2.1 Image Warping . . . .
7.2.2 Smart Image Resizing . . . .
7.2.3 3D Analysis . . . .
7.3 Segmentation and Tracking . . . .
7.3.1 Matching . . . .
7.3.2 Segmentation . . . .
7.3.3 Smart Region Selection . . . .
7.4 Data-Driven Techniques . . . .
7.4.1 Image Collections . . . .
7.4.2 On-Line Photo Collections . . . .

7.4.3 Probabilistic and Inferential Methods . . . .
7.4.4 Indexing and Search . . . .
7.5 Image Sequences . . . .
7.5.1 Time Lapse Imaging . . . .
7.5.2 Motion Depiction . . . .

8 Future Directions
8.1 Great Ideas from Scientific Imaging . . . .
8.1.1 Tomography imaging . . . .
8.1.2 Coded Aperture imaging . . . .
8.1.3 Negative Index of Refraction . . . .
8.1.4 Schlieren Photography . . . .
8.1.5 Phase Contrast Microscopy . . . .
8.2 Displays . . . .
8.2.1 High Dynamic Range Displays . . . .
8.2.2 Light-Sensitive Displays . . . .
8.2.3 3D, Volumetric, and View Dependent Displays . .
8.2.4 Digital Photoframes . . . .
8.2.5 Printers . . . .
8.3 Fantasy Imaging Configurations . . . .
8.4 Intriguing Concepts . . . .
8.5 Photo Sharing and Community . . . .
8.6 Challenges . . . .

A Appendix
A.1 Gradient Manipulation . . . .
A.2 Bilateral Filtering . . . .
A.3 Deconvolution . . . .
A.4 Graph Cuts . . . .

Chapter 1
Introduction

Photography, literally, drawing with light,' is the process of making pictures by recording the visually meaningful changes in the light reflected by a scene. This goal was envisioned and realized for plate and film photography somewhat over 150 years ago by pioneers Joseph Nicphore Nipce (View from the Window at Gras, 1826), Louis-Jacques-Mand Daguerre (see http://www.hrc.utexas.edu/exhibitions/permanent/wfp/), and William Fox Talbot, whose invention of the negative led to reproducible photography. Though revolutionary in many ways, modern *digital photography* is essentially electronically implemented film photography, except that the film or plate is replaced by an electronic sensor. The goals of the classic film camera, which are at once enabled and limited by chemistry, optics, and mechanical shutters, are pretty much the same as the goals of the current digital camera. Both work to copy the image formed by a lens, without imposing judgement, understanding, or interpretive manipulations: both film and digital cameras are faithful but mindless copiers. For the sake of simplicity and clarity, let's call photography accomplished with today's digital cameras film-like, since both work only to copy the image formed on the sensor. Like conventional film and plate photography, film-like photography presumes (and often requires) 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.

This book will explore a progression away from film and film-like methods to a more comprehensive technology that exploits plentiful low-cost computing and memory with sensors, optics, probes, smart lighting and communication.

1.1 What is Computational Photography?

Computational photography (CP) is an emerging field. We cannot know where the path will lead, nor can we yet give the field a precise, complete definition or its components a reliably comprehensive classification. But here is the scope of what researchers are currently exploring:

- Computational photography attempts to record a richer, even a multilayered visual experience, captures information beyond just a simple set of pixels, and renders the recorded representation of the scene far more machine-readable.
- It exploits computing, memory, interaction and communications to overcome inherent limitations of photographic film and camera mechanics that have persisted in film-like digital photography, such as constraints on dynamic range, limitations of depth of field, field of view, resolution and the extent of subject motion during exposure.
- It enables new classes of recording the visual signal such as the "moment", shape boundaries for non-photorealistic depiction, foreground versus background mattes, estimates of 3D structure, "relightable" photos, and interactive displays that permit users to change lighting viewpoint, focus, and more, capturing some useful, meaningful fraction of the "light-field" of a scene, a 4D set of viewing rays.
- It enables synthesis of impossible photos that could not have been captured with a single exposure in a single camera, such as wraparound views ("multiple-center-of-projection" images), fusion of time-lapsed events, the motion-microscope (motion magnification), video textures and panoramas.It supports seemingly impossible camera movements such as the "bullet time" sequences as in *The Matrix* made with multiple cameras using staggered exposure times and "free-viewpoint television" (FTV) recordings.
- It encompasses previously exotic forms of imaging and data-gathering techniques in astronomy, microscopy, tomography, and other scientific fields.

1.2 Elements of Computational Photography

Traditional film-like digital photography involves a lens, a 2D planar sensor, and a processor that converts sensed values into an image. In addi-

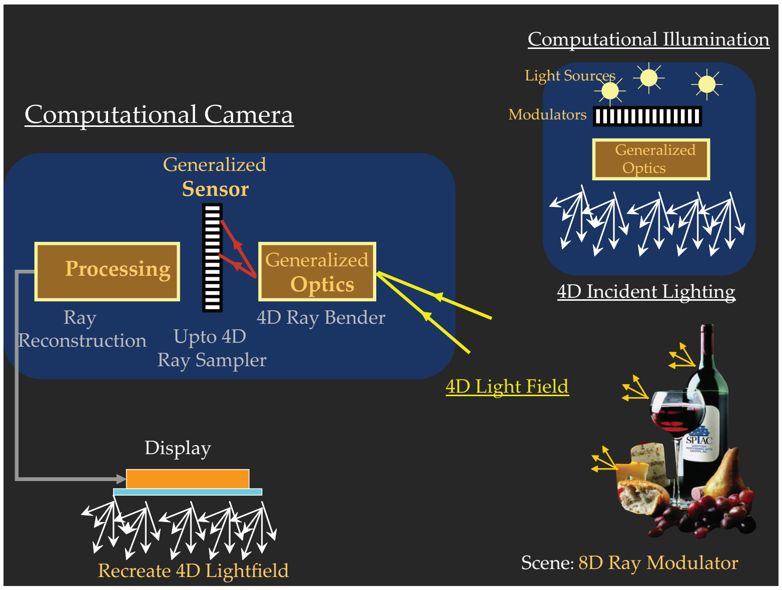


Figure 1.1. Elements of computational photography.

tion, such photography may entail external illumination from point sources (e.g., flash units) and area sources, for example, studio lights.

We like to categorize and generalize computational photography into the following four elements. Our categorization is influenced by Shree Nayar's original presentation [Nayar 05]. We refine it by considering the external illumination and the geometric dimensionality of the involved quantities.

(a) Generalized optics. Each optical element is treated as a 4D ray-bender that modifies a light-field. The incident 4D light-field[1] for a given wavelength is transformed into a new 4D light-field. The optics may involve more than one optical axis [Georgiev et al. 06]. In some cases, perspective foreshortening of objects based on distance may be modified [Popescu 05], or depth of field extended computationally by wavefront coded optics [Dowski and Cathey 95].

[1]4D refers here to the parameters (in this case 4) necessary to select one light ray. The light-field, discussed in the next chapter, is a function that describes the light traveling in every direction through every point in three-dimensional space. This function is alternately called "the photic field," the 4D light-field," or the "Lumigraph."

In some imaging methods [Zomet and Nayar 06], and in coded-aperture imaging [Zand 96] used for gamma-ray and X-ray astronomy, the traditional lens is absent entirely. In other cases optical elements such as mirrors outside the camera adjust the linear combinations of ray bundles reaching the sensor pixel to adapt the sensor to the imaged scene [Nayar et al 04].

(b) Generalized sensors. All light sensors measure some combined fraction of the 4D light-field impinging on it, but traditional sensors capture only a 2D projection of this light-field. Computational photography attempts to capture more—a 3D or 4D ray representation using planar, non-planar, or even volumetric sensor assemblies. For example, a traditional out-of-focus 2D image is the result of a capture-time decision: each detector pixel gathers light from its own bundle of rays that do not converge on the focused object. A *plenoptic camera*, however, [Adelson and Wang 92, Ren et al. 05] subdivides these bundles into separate measurements. Computing a weighted sum of rays that converge on the objects in the target scene creates a digitally refocused image, and even permits multiple focusing distances within a single computed image. Generalizing sensors can extend both their dynamic range [Tumblin et al. 05] and their wavelength selectivity [Mohan 08]. While traditional sensors trade spatial resolution for color measurement (wavelengths) using a Bayer grid or red, green, or blue filters on individual pixels, some modern sensor designs determine photon wavelength by sensor penetration, permitting several spectral estimates at a single pixel location [Foveon 04].

(c) Generalized reconstruction. Conversion of raw sensor outputs into picture values can be much more sophisticated. 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. Reconstruction might combine disparate measurements in novel ways by considering the camera intrinsic parameters used during capture. For example, the processing might construct a high dynamic range image out of multiple photographs from coaxial lenses [McGuire et al. 05], from sensed gradients [Tumblin et al 05], or compute sharp images of a fast moving object from a single image taken by a camera with a “fluttering” shutter [Raskar et al 06]. Closed-loop control during photographic capture itself can be extended, exploiting the exposure control, image stabilizing, and focus of traditional cameras as opportunities for modulating the scene's optical signal for later decoding.

(d) Computational illumination. Photographic lighting has changed very little since the 1950s. With digital video projectors, servos, and device-to-device communication, we have new opportunities for controlling the sources of light with as much sophistication as that with which we control our digital sensors. What sorts of spatio-temporal modulations of lighting might better reveal the visually important contents of a scene? Harold Edgerton showed that high-speed strobes offer tremendous new appearance-capturing capabilities; how many new advantages can we realize by replacing "dumb" flash units, static spot lights, and reflectors with actively controlled spatio-temporal modulators and optics? We are already able to capture occluding edges with multiple flashes [Raskar 04], exchange cameras and projectors by Helmholz reciprocity [Sen et al. 05], gather relightable actor's performances with light stages [Wagner et al. 05] and see through muddy water with coded-mask illumination [Levoy et al. 04]. In every case, better lighting control during capture allows for richer representations of photographed scenes.

1.3 Sampling the Dimensions of Imaging

1.3.1 Past: Film-Like Digital Photography

Even though photographic equipment has undergone continual refinement, the basic approach remains unchanged: a lens admits light into an otherwise dark box, and forms an image on a surface inside. This "camera obscura" idea has been explored for over a thousand years,but became photography only when combined with light-sensitive materials to fix the incident light for later reproduction. Early lenses, boxes, and photosensitive materials were crude in nearly every sense—in 1826, Niepce made an 8-hour exposure to capture a sunlit farmhouse through a simple lens onto chemically altered asphalt-like bitumen resulting in a coarse, barely discernible image. Within a few decades, other capture strategies based on the light-sensitive properties of sensitized silver and silver salts had reduced that time to minutes, and by the 1850s were displaced by wet-plate collodion emulsions prepared on a glass plate just prior to exposure.

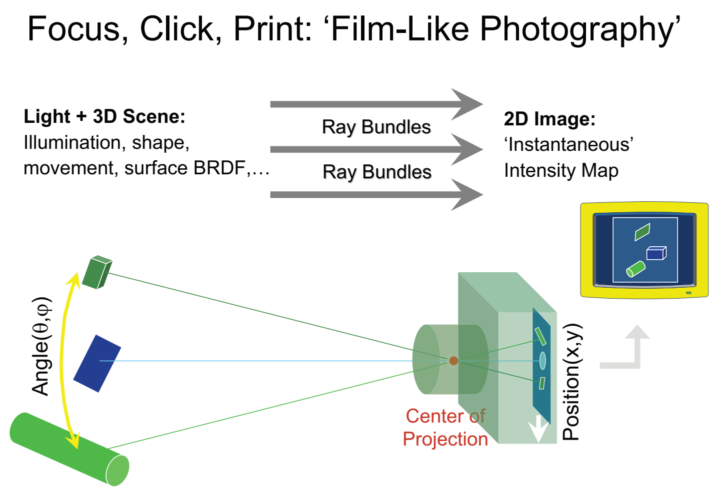


Figure 1.2. Ideal film-like photography uses a lens to form an image on a light-sensitive surface, then records that image instantly with light-sensitive materials. Practical limits such as lens light-gathering efficiency, sensitivity, and exposure time necessitate tradeoffs.

Though messy, complex and noxious to prepare, wet plates could produce larger, more subtle photos, and were fast enough to record human portraits. By the late 1870s, pre-manufactured gelatine dry plates were replacing the cumbersome collodion wet plates, and these in turn yielded to flexible film, introduced by George Eastman in 1884. Continual advances in thin-film chemistry have led to today's complex multi-layer film emulsions that offer widely varied choices in image capture. These are complemented by parallel camera development of complex multi-element lenses, shutters, aperture mechanisms, as well as of sophisticated lighting devices.[2]

With each set of improvements, photographers have gained an ever-expanding range of choices that affect the appearance of the captured image. The earliest cameras had neither shutters nor aperture mechanisms. Photographers chose their lens, adjusted its focus on a ground-glass sheet, replaced the ground glass with a light-sensitive plate, uncapped the lens, and waited for the lens to gather enough light to record the image. As light-sensing materials improved, exposure time dropped from minutes to

[2]For an authoritative technical review, see *The Theory of The Photographic Process*, edited by T. H. James.

seconds to milliseconds; adjustable-time shutters replaced lens caps; and adjustable lens apertures permitted regulation of the amount of light passing through the lens during exposure. By the 1880s, the basic camera settings were well-defined, and digital cameras have extended them only slightly. They are:

⋆ Lens. Aperture, focusing distance, and focal length;

⋆ Shutter. Exposure time;

⋆ Sensor. Light sensitivity (film speed; ASA, ISO, or DIN) latitude (or tonal range or dynamic range), and color-sensing properties;

⋆ Camera. Location, orientation, and the moment of exposure;

⋆ Auxilliary lighting. Position, intensity, timing.

Most digital film-like cameras can automatically choose these settings. Once the shutter is tripped, these choices are fixed; the resultant image is one among many possible photographs. At the instant of the shutter-click, the camera has chosen the following settings:

(a) 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 picture; a long (telephoto) focal length gives a narrow one. Though the image may be cropped later (at a corresponding loss of resolution), it cannot be widened.

(b) Exposure and dynamic range. The chosen lens aperture, exposure time, the sensors' film speed (ISO, sensitivity), and its latitude together determine how amounts of light in the scene map to picture values between black and white. Larger aperture settings, longer exposure times, or higher sensitivities map dimly-lit scenes to acceptable pictures, while smaller apertures, shorter exposure times, and lower sensitivity will be chosen for brilliantly sun-lit scenes. Poor choices here may mean loss of visible details in too-bright areas of the image, in too-dark areas, or both. Within the sensitometric response curve of any sensor, the latitude of the film or the dynamic range of the sensor (the intensity ratio between the darkest and lightest details) is not usually adjustable, and falls typically between 200:1 to 1000:1.

(c) Depth of field. The lens aperture focal length and sensor size together determine how wide a range of distances will appear in focus. A small aperture and short (wide) focal length gives the greatest depth of field, while large apertures with long focal lengths yield narrow ranges of focus.[3] (Note that increased depth of field normally require a smaller aperture, which may entail increased exposure time or sensor sensitivity (which in turn increases noise).)

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

(e) Spatial resolution. For a well-focused image, the sensor itself sets the spatial resolution. It may be artificially blurred, but no sharpening can recover more detail than that already recorded by the camera. Note that increased resolution reduces depth of focus and often increases visible noise due to reduced sensor pixel size.

(f) Wavelength resolution. Color-balance and saturation settings on the camera set sensitivity to color. Current film-like cameras sense color by measuring three primaries (usually R,G,B) with fixed, overlapping spectral response curves. While different sensors (especially black-and-white film stocks) offer varying spectral curves, none is adjustable.

In every case, film-like photography forces us to choose, to make tradeoffs among interdependent parameters, and to lock in those choices in a single photo at the moment we click the shutter. If we choose a long exposure time to gather enough light, movement in the scene may blur the picture, while too short an exposure time in order to freeze motion may make the picture too dark. We can keep the exposure time short if we increase the aperture size, but then we lose depth of focus, and foreground or background objects are no longer sharp. We can increase the depth of focus again if we shorten (widen) the focal length and move closer to the subject, but then we alter the foreshortening of the image. The basic

[3]Some portraits (e.g., Matthew Brady's close-up photos of Abraham Lincoln show eyes in sharp focus but employ soft focus in other planes to hide blemishes elsewhere on the face.

"camera obscura" design of film-like photography forces these tradeoffs; they are inescapable due to the hard limits of simple image formation and the measurement of light. 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, even a very tiny one in dim light; and we would like unbounded resolution that would allow capture of a very wide field of view. Unfortunately, this ideal camera's infinitesimal aperture and zero-length exposure time would gather no photons at all!

New methods of computational photography, however, offer a steadily growing number of ways to escape the bind of these tradeoffs, and gain new capabilities. Existing film-like camera designs are already excellent; we have economical cameras that offer a tremendous adjustment range for each of these parameters; We are increasingly confident of finding computational strategies to untangle them.

1.3.2 Present: Epsilon Photography

Think of film cameras at their best as defining a "box" in the multi-dimensional space of imaging parameters. The first, most obvious thing we can do to improve digital cameras is to expand this box in every conceivable dimension. The goal would be to build a super-camera that has enhanced performance in terms of the traditional parameters, such as dynamic range, field of view or depth of field. In this project of a super-camera, computational photography becomes "epsilon photography," in which the scene is recorded via multiple images that vary at least one of the camera parameters by some small amount or epsilon. For example, successive images (or neighboring pixels) may have different settings for parameters such as exposure, focus, aperture, view, illumination, or timing of the instant of capture. Each setting allows recording of partial information about the scene, 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, i.e., multiple film-style photos computationally merged to make a more complete photo or scene description. While the merged photo is superior, each of the individual photos is still useful and comprehensible independently. The merged photo contains the best features from of the group. Thus epsilon photography corresponds to the low-level vision: estimating pixels and pixel features with the best signal-to-noise ratio.

(a) Field of view. A wide field of view panorama is achieved by stitching and mosaicking pictures taken by panning a camera around a common center of projection or by translating a camera over a near-planar scene.

(b) Dynamic range. A high dynamic range image is captured by merging photos at a series of exposure values.and Picard 93, Debevec and Malik 97, Kang et al. 03].

(c) Depth of field. An image entirely in focus, foreground to background, is reconstructed from images taken by successively changing the plane of focus [Agrawala et al 05].

(d) Spatial resolution. Higher resolution is achieved by tiling multiple cameras (and mosaicing individual images) [Wilburn et al. 05] or by jittering a single camera [Landolt et al. 01].

(e) Wavelength resolution. Conventional cameras sample only three basis colors. But multi-spectral imaging (from multiple colors in the visible spectrum) or hyper-spectral imaging (from wavelengths beyond the visible spectrum) are accomplished by successively changing color filters in front of the camera during exposure, using tunable wavelength filters or diffraction gratings [Mohan et al. 08].

(f) 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 non-overlapping [Wilburn et al. 05] or overlaping [Shechtman et al 02].

Photographing multiple images under 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 using “assorted pixels” where each pixel is tuned to a different value for a given parameter [Nayar and Narsimhan 2002]. Just as some early digital cameras captured scanlines sequentially, including those that scanned a single one-dimensional detector array across the image plane, detectors are conceivable that intentionally randomize each pixel’s exposure time to trade off motion-blur and resolution, previously explored for interactive computer graphics rendering [Dayal 05]. Simultaneous capture of multiple samples can also be recorded 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 [Wilburn et al. 05] and single-axis multiple parameter (co-axial) cameras [Mcguire et al. 05].

1.3.3 Future: Coded Photography

But we wish to go far beyond the best possible film camera. Instead of high quality pixels, the goal is to capture and convey the mid-level cues: shapes, boundaries, materials, and organization. Coded photography reversibly encodes information about the scene in a single photograph (or a very few photographs) so that the corresponding decoding allows powerful decomposition of the image into light fields, motion-resolved images, global/direct illumination components, or distinction between geometric versus material discontinuities.

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 towards a single lens, and virtualize it further if we gather each ray with its own wavelength spectrum.

Coded photography is a notion of 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, a wrap-around view might be built from multiple images taken from a ring or a sphere of camera positions around the object, but the view takes only a few pixels from each input image for the final result; could we find a better, less wasteful way to gather the pixels we need? Coded aperture techniques, inspired by work in astronomical imaging, try to preserve the high spatial frequencies of light that passes through the lens so that out-of-focus blurred images can be digitally re-focused [Veeraraghavan 07] or resolved in depth [Levin07]. By coding illumination, it is possible to decompose radiance in a scene into direct and global components [Nayar06]. Using 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 photo. The fluttered shutter encodes the motion that conventionally appears blurred in a reversible way; we can compute a moving but un-blurred image. Other examples include confocal synthetic aperture imaging [Levoy 04] that lets us see through murky water, and techniques to recover glare by capturing selected rays through a calibrated grid [Talvala 07]. What other novel abilities might be possible by combining computation with sensing novel combinations of scene appearance?

In fact,the next phase of computational photography will go beyond the radiometric quantities and challenge the notion that a synthesized photo should appear to come from a device that mimics a single-chambered human eye. Instead of recovering physical parameters, the goal will be to capture the visual essence of the scene and scrutinize the perceptually critical components. This *essence photography* may loosely resemble depiction of the world after high-level vision processing. In addition to photons, additional elements will sense location coordinates, identities, and gestures via novel probes and actuators. With sophisticated algorithms, we will exploit priors based on natural image statistics and online community photo collections [Snavely 06, Hays 07]. Essence photography will spawn new forms of visual artistic expression and communication.

We may be converging on a new, much more capable box of parameters in computational photography that we can’t yet fully recognize; there is quite a bit of innovation yet to come!

Chapter 3

Extending Film-Like Digital Photography

As a thought experiment, suppose we accept our existing film-like concepts of photography, just as they have stood for well over a century. For the space of this chapter, let's continue to think of any and all photographs, whether captured digitally or on film, as a fixed and static record of a viewed scene, a straightforward copy of the 2D image formed on a plane behind a lens. How might we improve the results from these traditional cameras and the photographs they produce if we could apply unlimited computing, storage, and communication to them? The past few years have yielded a wealth of new opportunities, as miniaturization allows lightweight battery-powered devices such as mobile phones to rival the computing power of the desktop machines of only a few years ago, and as manufacturers can produce millions of low-cost, low-power and compact digital image sensors, high-precision motorized lens systems, bright, full-color displays, and even palm-sized projectors, integrated into virtually any form as low-priced products. How can these computing opportunities improve conventional forms of photography?

Currently, adjustments and tradeoffs dominate film-like photography, and most decisions are locked in once we press the camera's shutter release. Excellent photos are often the result of meticulous and artful adjustments, and the sheer number of adjustments has grown as digital camera electronics have replaced film chemistry, and now include ASA settings, tone scales, flash control, complex multi-zone light metering, color balance, and color saturation. Yet we make all these adjustments before we take the picture, and even our hastiest decisions are usually irreversible. Poor choices lead to poor photos, and an excellent photo may be possible only for an exquisitely narrow combination of settings taken with a shutter-click at just the right moment. Can we elude these tradeoffs? Can we defer choosing the camera's settings somehow, or change our minds and re-adjust them 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?

3.1 Understanding Limitations

This is a single-strategy chapter. As 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 and may 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* already familiar to most photographers. Bracketing lets photographers avoid uncertainty about critical camera settings such as focus or exposure; instead of taking just one photo at what we think are the correct 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 correct choice, the bracketed set of photos almost always contains a better one. The methods in this chapter are analogous, but often use a larger set of photos as multiple settings may

[1] For example, HDRShop from Paul Debevec's research group at USC-ICT (http://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 (http://www.cs.ubc.ca/~mbrown/autostitch/autostitch.html) and AutoPano-SIFT (http://user.cs.tu-berlin.de/~nowozin/autopano-sift/) let users construct cylindrical or spherical panoramas from overlapped images; and HD View from Microsoft Research (http://research.microsoft.com/ivm/hdview.htm) 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).

be changed, and we may digitally merge desirable features from multiple images in the set rather than simply select just one single best photo.

We need to broaden our thinking about photography to avoid missing some 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

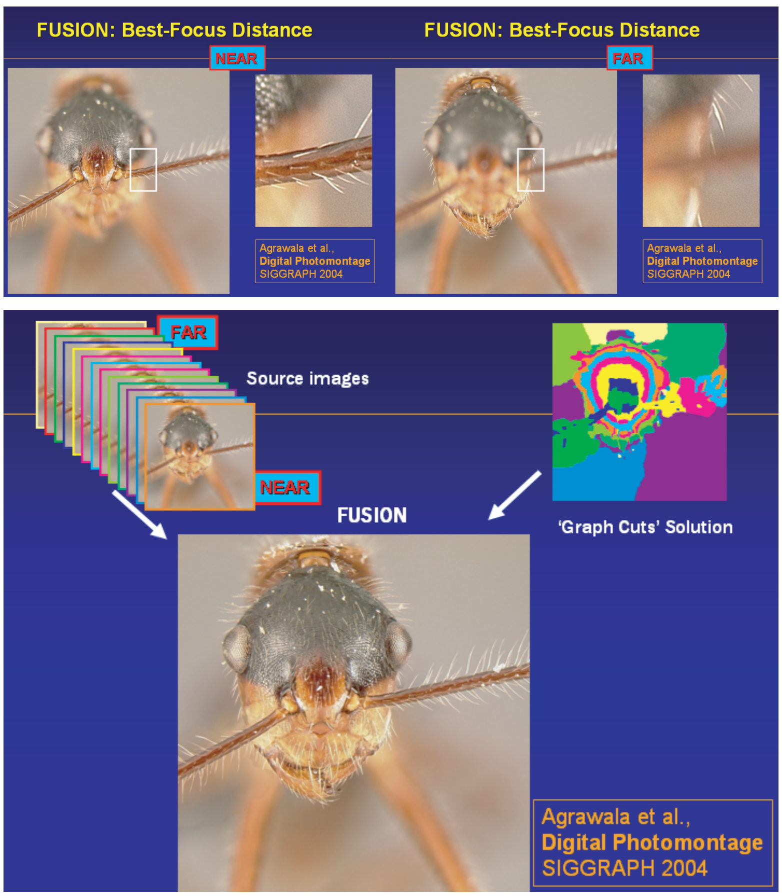


Figure 3.1. No physically realizable lens could achieve enough depth-of-field for this insect closeup: we can focus on antennae or thorax or in-between (top). However, with a large enough series of bracketed focus images and the right forms of optimization, we can assemble an "all-focus" image (bottom) from the best parts of each photo. (Image Credits: Digital Photomontage [Agrawal04]).

inescapable, a direct consequence of the laws of physics, image formation and light transport. For example, Chapter 2 reviewed how the depth-of-focus of an image formed behind a lens is a direct consequence of the thin-lens law. While true for a single image, merged multiple images let us construct an "all focus" image (as in Figure 3.1), or vary focus and depth-of-focus arbitrarily throughout the image.

In film-like photography, we cannot adjust a single knob to change the depth-of-focus: instead we must choose several interdependent settings that each impose different trade-offs. We can use a lens with a shorter focal length, but this will make the field-of-view wider; we can compensate for the wider field-of-view by moving the camera closer to the subject, but then we will change foreshortening in the scene.

We can keep the same image size for a photographed subject if we move the camera closer and zoom out to a wider-angle lens, or if we move the camera further away and zoom in to a narrow-angle telephoto lens, but the appearance of that subject may change dramatically due to foreshortening. Foreshortening is the subjective name for a mathematically simple rule for planar projection: the image size of an object is proportional to its depth, its distance to the camera. In a telephoto image of a face, the distance to the camera is much greater than the small depth differences

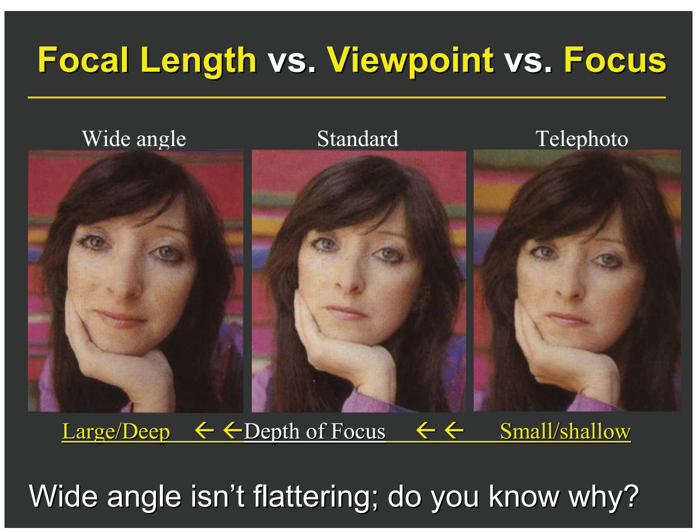


Figure 3.2. Foreshortening effects.

between forehead, nose, and chin, and all appear properly proportioned in the image. In a wide-angle close-up photo, the nose distance might be half the chin and forehead distance, exaggerating its image size in a very unflattering way.

We can compensate by cropping the sensed image using a smaller portion of the image sensor, but this reduces resolution and may make some lens flaws more noticeable, such as focus imperfections, chromatic aberration and coma, and other lens flaws less apparent, such as radial distortion. We can leave the lens unchanged but reduce the size of its limiting aperture, but this decreases the light falling on the image sensor and may increase visible noise. Compensating for the decreased intensity by increasing exposure time increases the chance of motion-blur or camera-blur. Increasing the sensor's light sensitivity further increases image noise.

What strategy should we choose to extend film-like photography in the most useful ways? Usually, no one answer is best; instead, we confront a host of interrelated tradeoffs that depend on scene, equipment, the photographer's intentions, and the ultimate display of the photograph.

What are our assumptions as photographers? Do they remain valid for bracketing, for merged sets of photographs? How might we transcend them by combining, controlling, and processing results from multiple cameras, lights, and photographs using computing methods?

We are misled by our strong beliefs. Surely every photo-making process has to employ a high-quality optical system for high-quality results. Surely any good camera must require focusing, adjusting zoom level, choosing the field 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 tradeoffs among sensitivity, 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 film (or digital sensor) 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, adjust the lights well, choose a good viewpoint, pose, and adjust the subject for its 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.

3.2 Strategies: Fusion of Multiple Images

Tradeoffs in film-like photography improve one measurable aspect of a photograph at the expense of another. While we can capture a series of photographs with different settings for each, we can also vary setting within the digital sensors themselves:

3.2.1 Sort First versus Sort Last Capture

With the sort-first method, we capture a sequence of photographs with one or more cameras. 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—if we photograph at high, moderate, and low exposure times, we sort the results by selecting the best whole-photo result; we don't need any further untangling of measured data to create the best photograph. For example, in 1909–1912, and 1915, commissioned and equipped by Tsar Nicholas II, Sergei Mikhailovich Prokudin-Gorskii (1863–1944) surveyed the Russian Empire in a set of beautiful

Figure 3.3. Prokudin-Gorskii captured scenes of Tsarist Russia with a custom-built sort-first, time-multiplexed camera that captured three color images in rapid succession on a tall, single-plate negative. (Prokudin-Gorskii, Sergei Mikhailovich, 1863–1944, photographer. The Bukhara Emir Prints and Photographs Division, Library of Congress. Reproduction number: LC-P87-8086A-2).

color photographs gathered by his own sort-first method for color photography (see http://www.loc.gov/exhibits/empire/gorskii.html). 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 3.3).

The sort-last method mixes together several different settings within each photographic image we take. After photography we must sort the contents of the photos, rearrange and recombine them somehow to construct a suitable output image. Such multiple simultaneous measurements in each image make sort-last methods less susceptible to scene variations over time, reducing the chance that a transient scene value will escape successful measurement. For example, suppose we photograph a scene as clouds cover or reveal the sun during sort-first exposure bracketing; our first high-exposure photo, taken before the sun went behind clouds appears overly bright, but our subsequent mid- and low-exposure photos are darker than they should be due to falling light levels, yielding no usable photos at all. Another example of sort-first difficulties appeared in merging Prokudin-

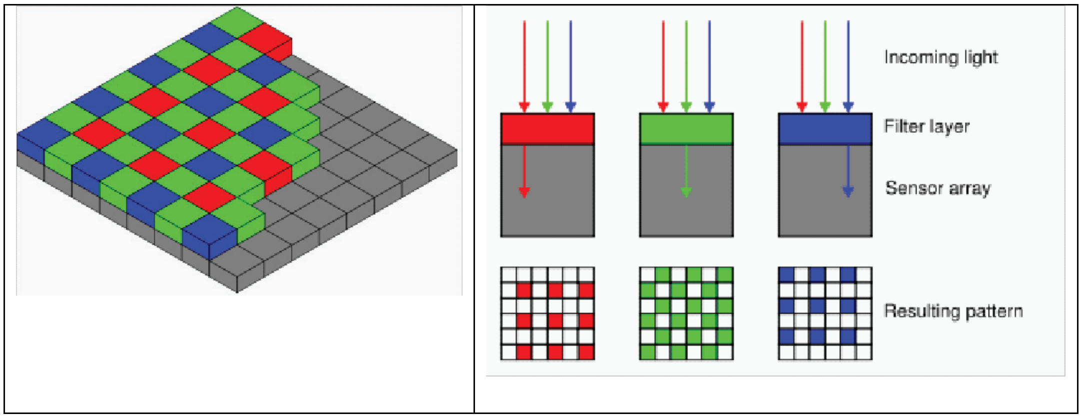


Figure 3.4. The Bayer mosaic in many modern digital cameras employs sort-last color sensing. De-mosaicking methods employ edge-following, estimation, and interpolation methods to approximate a full-resolution color image from these measurements. Alternatively, three-chip video cameras follow the sort-first method, and sense three complete, independent color images simultaneously.

Gorskii's color plates; unpredictable motions from wind-blown trees and swirling river water during photography caused mismatches among the three negatives, resulting in color fringing and rainbow-like artifacts, as in "Pinkhus Karlinskii. Eighty-four years [old]", viewable online at http://www.loc.gov/exhibits/empire/images/p87-5006.jpg.

The *Bayer* color mosaic pattern found on nearly all single-chip digital cameras is perhaps the most widely used form of sort-last measurement, while three-chip digital cameras follow the sort-first method. Three-chip cameras (more common for high-quality video applications than 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. In the patented Bayer mosaic method, individual, pixel-sized color filters cover adjacent pixels on this sensor, forming a red, green, and blue filter pattern as shown. 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 (de-mosaicking) to give the impression of a full-resolution image with all colors measured for every pixel.

3.2.2 Time- and Space-Multiplexed Capture

In addition to sort-first and sort-last, we can also classify multi-image gathering methods into time-multiplexed and space-multiplexed forms, which are more consistent with the 4D ray-space descriptions we encourage in this book. Time-multiplexed methods use one or more cameras to gather photos whose settings vary in a time-sequence: camera settings may change, the photographed scene may change, or both. Space-multiplexed methods are their complement, gathering a series of photos at the same time, but with camera settings that differ among cameras or within cameras (e.g., sort first, sort last).

Like sort-first methods, time-multiplexed capture can introduce inconsistencies from changing scenes. For example, suppose we wish to capture photographs for assembly into a panoramic image showing a 360-degree view from a single viewpoint. For a time-multiplexed sequence, we could 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 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, and any conventional panorama-making software will produce good results. 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 one, but alignment is not impossible. Tools in Photoshop CS3 are suitable for manually resolving modest mismatches. Video panoramas have proven capable of resolving more challenging scene changes that include flowing water, trees waving in the wind, and lighting changes [Agarwal et al. 05].

A space-multiplexed sequence neatly avoids these time-dependent mismatches. To capture a changing panorama, we can either construct a ring of cameras with aligned or slightly overlapping fields-of-view to capture all views simultaneously (e.g., Kodak's “Circle-Vision360” panoramic motion-picture attraction at Disney theme parks), or resort to specialized catadioptric (lenses-and-mirrors) optics to map the entire panorama onto a single image sensor [NayarCata 97, Benosman 01].

3.2.3 Hybrid Space-Time Multiplexed Systems

Hybrid systems of video or still cameras enable capture of each step of a complicated event over time in order to understand it 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. Even before the first motion pictures, in 1877–1879, Edweard Muybridge devised just such a hybrid by constructing an elaborate multi-camera system of wet-plate (collodion) cameras to take single short-exposure-time photos in rapid-fire sequences. Muybridge devised a clever electromagnetic shutter-release mechanism triggered by trip-threads to capture action photos of galloping horses. He also refined the system with electromagnetic shutter releases triggered by pressure switches or elapsed time to record walking human figures, dancers, and acrobatic performances (see http://www.kingston.gov.uk/browse/leisure/museum/museum_exhibitions/muybridge.htm). His sequences of short-exposure freeze-frame images allowed the first careful examination of the subtleties of motion that are too fleeting or complex for our eyes to absorb as they are happening—a fore-runner of slow-motion movies or video. Instead of selecting just one perfect instant for a single photograph, these event-triggered image sequences contain valuable visual information that

Figure 3.5. Edgerton's rapid multi-exposure sequence of images.

stretches across time and across a sequence of camera positions and is suitable for several different kinds of computational merging.

Perhaps the simplest *computational merging* of time-multiplexed images occurs within the camera itself. In his seminal work on fast, high-powered electronic (Xenon) strobe lights, Harold Edgerton showed that a rapid multiple-exposure sequence can be as revealing as a high-speed motion-picture sequence (see Figure 3.5).

In addition to its visual interest, photos lit by a precisely timed strobe sequence like this permit easy frame-to-frame measurements. For example, Figure 3.5 confirms that baseballs follow elastic collision dynamics.[2]

In some of Muybridge's pioneering efforts, two or more cameras were triggered at once to capture multiple views simultaneously. Modern work by Bregler and others on motion-capture from video merged these early multi-view image sequences computationally to infer the 3D shapes and the movements that caused them. By finding image regions undergoing movements consistent with rigid jointed 3D shapes in each image set, Bregler et al. could compute detailed estimates of the 3D position of each body segment in each frame and re-render the image sets as short movies at any frame rate viewed from any desired viewpoint [Bregler et al. 98].

In another ambitious experiment, at Stanford University, more than one hundred years after Muybridge's work, Marc Levoy and colleagues constructed an adaptable array of 128 individual film-like digital video cameras that perform both time-multiplexed and space-multiplexed image capture simultaneously [Wilburn 05]. The reconfigurable array enabled a wide range of computational photography experiments. Built on lessons from earlier arrays (e.g., [Kanade 97, Yang 02, Matusik 04, Zhang 04]), the system's interchangeable lenses, custom control hardware, and refined mounting system permitted adjustment of camera optics, positioning, aiming, arrangement, and spacing between cameras. One configuration kept the cameras packed together, just one 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.

[2]Similarly, you can try your own version of Edgerton's well-known milk-drop photo sequences (with a digital flash camera, an eye dropper, and a bowl of milk.

However, at high frame rates these differences were quite small, causing noisy results we wouldn't find acceptable as a conventional high-speed video camera.Instead, the team simultaneously computed three low-noise video streams with different tradeoffs using synthetic-aperture techniques [Levoy04]. 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

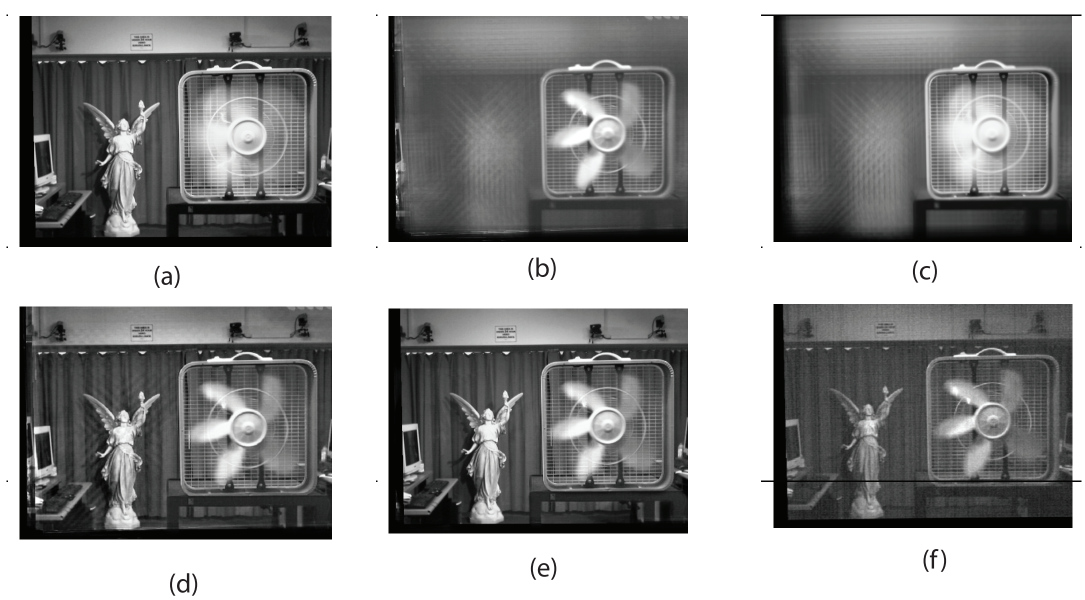


Figure 3.6. Staggered video frame times permit construction of a virtual high-speed video signal with a much higher frame-rate via hybrid synthetic aperture photography [Wilburn 05]. Hybrid synthetic aperture photography for combining high depth of field and low motion blur. (a-c) Images captured of a scene simultaneously through three different apertures: a single camera with a long exposure time (a), a large synthetic aperture with short exposure time (b), and 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 (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.

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 $I_s + I_t I_w$ exhibited reduced noise, sharp stationary features, and modest motion blur, as shown in Figure 3.6.

3.3 Improving Dynamic Range

Like any sensor, digital cameras have a limited input range: too much light dazzles the sensor, ruining the image with a featureless white glare, while too little light makes image features indistinguishable from perfect darkness. How can that range be improved, allowing our cameras to see details in the darkest shadows and brightest highlights?

Film-like cameras provide several mechanisms to match the camera's overall light sensitivity to the amount of light in a viewed scene, and digital cameras can adjust most of them automatically. These include adjusting the aperture size to limit the light admitted through the lens (though this alters the depth-of-field), adjusting exposure time (though this may allow motion blur), placing "neutral density" filters in the light-path (though this might accidentally displace the camera), or adjusting the sensitivity of the sensor itself—using a film with a different ASA rating (which changes film-grain size), or changing the gain-equivalent settings on a digital camera (which changes the amount of noise). Despite their tradeoffs, these mechanisms combine to give modern camera sensitivity an astoundingly wide sensitivity range, one that can rival or exceed that of the human eye, which adapts to sense light over 16 decades of intensity from the absolute threshold of vision at about $10^{-6}\mathrm{cd/m}^2$ up to the threshold of light-induced eye damage near $10^{8}\mathrm{cd/m}^2$.

However, sensitivity adjustment alone isn't enough to enable cameras to match our eye's ability to sense the variations in intensity of every possible scene. Many scenes with plenty of light are still quite difficult to photograph well because their contrasts are too high; the intensity ratio between their brightest and darkest regions overwhelms the camera, so that it cannot capture a detailed record of the darkest blacks and the brightest whites simultaneously. Troublesome high-contrast scenes often include large visible light sources aimed directly at the camera, strong back-lighting and deep shadows, reflections, and specular highlights such as in Figure 3.7.

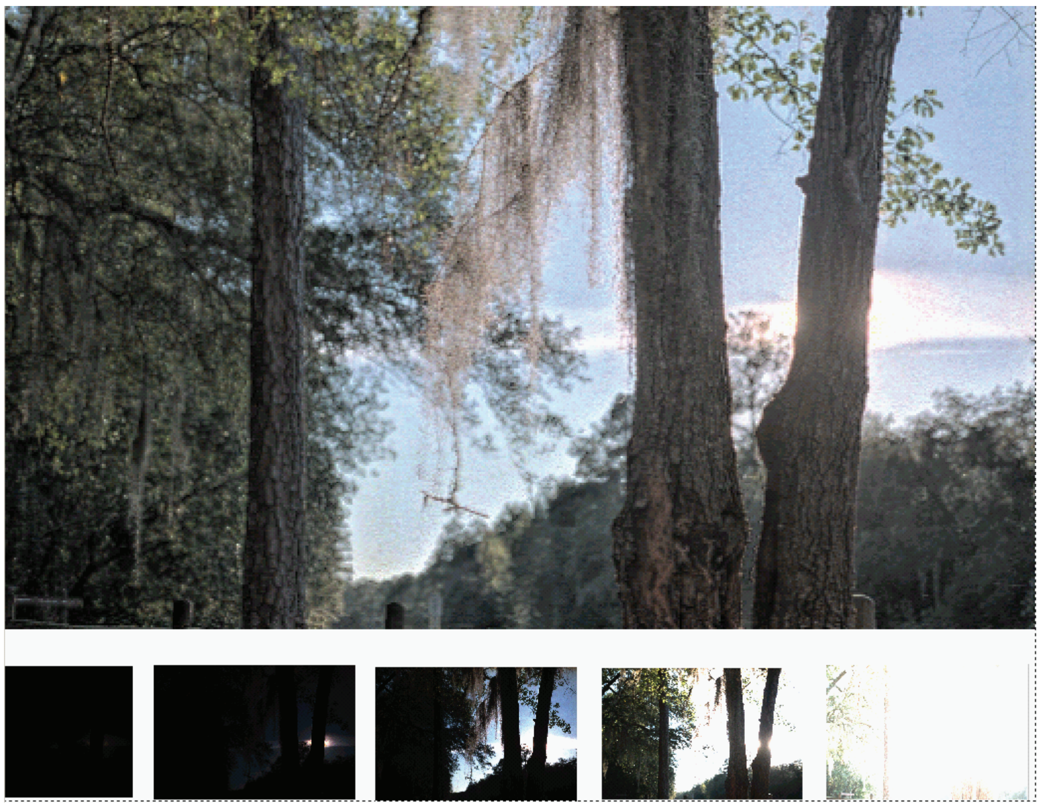


Figure 3.7. Tone-mapped HDR (high dynamic range) image from [Choud 03]. Many back-lit scenes such as this one can easily exceed the dynamic range of most cameras. Bottom row shows the original scene intensities scaled by progressive factors of ten; note that scene intensities in the back-lit cloud regions at left are approximately 10,000 times higher than shadowed forest details, well beyond the 1000:1 dynamic range typical of conventional CMOS or CCD camera sensors.

Film-like photography offers us little recourse other than to add light to the shadowy regions with flash or fill-lighting; rather than adjust the camera to suit the scene, we adjust the scene to suit the camera!

Unlike its sensitivity, the camera's maximum contrast ratio, known as its *dynamic range* is not adjustable. Formally, it is the ratio between the brightest and darkest light intensities a camera can capture from a scene within a single image without losing its detail-sensing abilities—the maximum intensity ratio between the darkest detailed shadows and brightest textured brilliance, as shown in Figure 3.7. 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.

Lens and sensor together limit the camera's dynamic range. In a high-contrast scene, glare effects and unwanted light scattering within complex lens structures cause glare and flare effects that depend on the image itself and cause traces of light from bright parts of the scene to "leak" into dark image areas, washing out shadow details and limiting the maximum contrast the lens can form on the image on the sensor, typically between 100,000:1 to 10 million to 1 [McCann 07, Levoy 07]. The sensor's dynamic range (typically $< 1000 : 1$) imposes further limits. Device electronics (e.g., charge transfer rates) typically set the upper bound on the amount of sensed light, and the least amount of light distinguishable from darkness is set by both the sensor's sensitivity and its *noise floor,*, 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 dwarfs the contrast abilities of cameras and displays. When plotted on a logarithmic scale (where distance depicts ratios; each tic marks a factor-of-10 change), the range of human vision spans about 16 decades, but typical film-like cameras and displays span no more than 2–3 decades. For the daylight-to-dusk (photopic) intensities (upper 2/3rds of scale), humans can detect some contrasts 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 is barely adequate for cameras and displays whose dynamic range may exceed 2 decades (100:1); many use 10, 12, or 14-bit internal representations to avoid visible contouring artifacts.

3.3.1 Capturing High Dynamic Range

Film-like photography is frustrating for high-contrast scenes because even the most careful bracketing of camera-sensitivity settings will not allow us to capture the whole scene's visible contents in a single picture. Sensitivity set high enough to reveal the shadow details will cause severe overexposure for dark parts of the scene; sensitivity set low enough to capture visible details in the brightest scene portions are far too low to capture any visible features in the dark parts of the image. However, several practical methods are available that let us capture all the scene contents in a usable way.

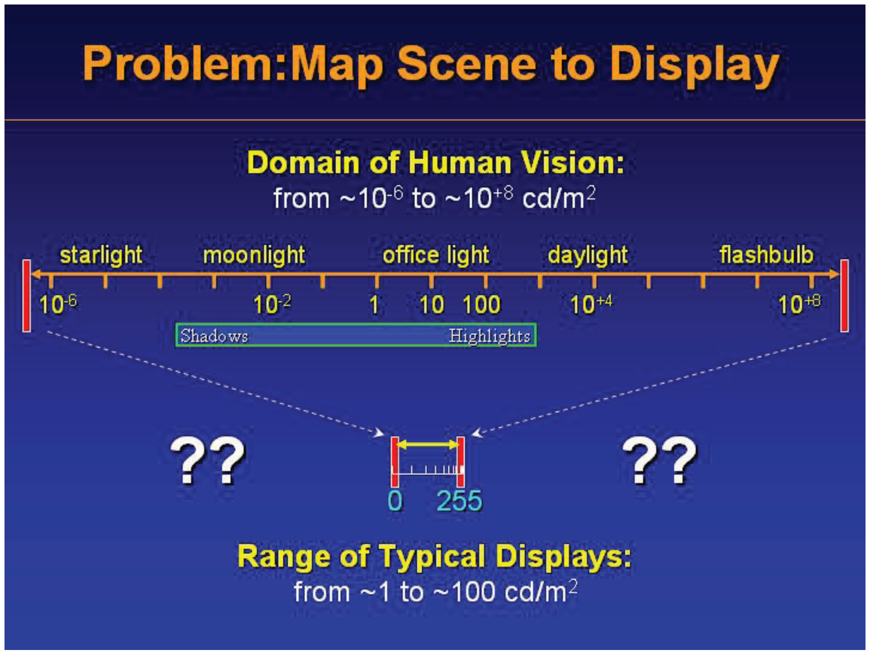


Figure 3.8. Visual dynamic range mismatch.

The resulting image covers a much wider dynamic range (see Figure 3.8 than conventional image file formats can express; storing only eight- or ten bits per color per pixel is inadequate to depict the much wider range of intensities in these high dynamic range (HDR) images. Many early file formats, using extravagant amounts of memory 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" devised by Greg Ward [Ward 95], and popularized by use in his photometrically accurate 3D renderer RADIANCE [Ward 98]. Later, a psychophysically well-motivated extension was proposed for the TIFF 6.0 image standard [logLUV 98], which formed the basis for the slightly simpler format used by HDRShop. Announced in 2003, the openEXR format developed by Industrial Light and Magic and independent partners provided a much simpler storage format, flexible bit-depth, and compression capabilities, backwards compatibility, suitability for motion-picture workflows, computing platform independence, and open-source licensing and has gained widespread acceptance.

3.3.2 HDR by Multiple Exposures

The sort-first approach is very suitable for capturing HDR images. To capture the finely-varied intensities in a high dynamic range scene, we can capture multiple images using a motionless camera that takes perfectly aligned images with different exposure settings and then merge these images. 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 pictures we took, ignoring badly over-and under-exposed images.

This simple form of merging is quick to compute and has found widespread early use as *exposure bracketing* [Morimura 93, Burt and Kolczynski 93, Madden 93, Tsai 94], but many methods assumed the linear camera response curves found on instrumentation cameras. Most digital cameras intended for photography introduce intentional nonlinearities in their light response, often mimicking the s-shaped response curves of film when plotted on log-log axes (H-D or Hurter-Driffield curves). These curves enable cameras to capture a wider usable range of intensities and provide a visually pleasing response to HDR scenes, retaining weak ability to capture intensity changes even at their extremes of over- and under-exposure, and varying among different cameras. Some authors have proposed the use of images acquired with different exposures to estimate the radiometric response function of an imaging device and use the estimated response function to process the images before merging them [Mann and Picard 95, Debevec and Malik 97, Mitsunaga and Nayar 99]. This approach has proven robust and is now widely available in both commercial software tools (Adobe Photoshop CS2 and later, CinePaint) and open-source projects (HDRShop (http://www.hdrshop.com/), PFStools (http://www.mpi-inf.mpg.de/resources/pfstools/), and others).

3.3.3 HDR by Exotic Image Sensors

While quite easy and popular for static scenes, exposure-time bracketing methods is not the only option available for capturing HDR scenes, and it is, moreover, 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 [Nayar and Narsihman 03], multiple sensor designs with beam-splitters, gradient-measuring sensors [Tumblin et al. 05]. In addition, we will explore techniques for dealing with high dynamic range scenes with video cameras [Kang et al. 05] or for capturing panoramas with panning cameras via attenuating ramp filters [Ahuja et al. 02, Nayar et al. 02].

3.4 Beyond Tri-Color Sensing

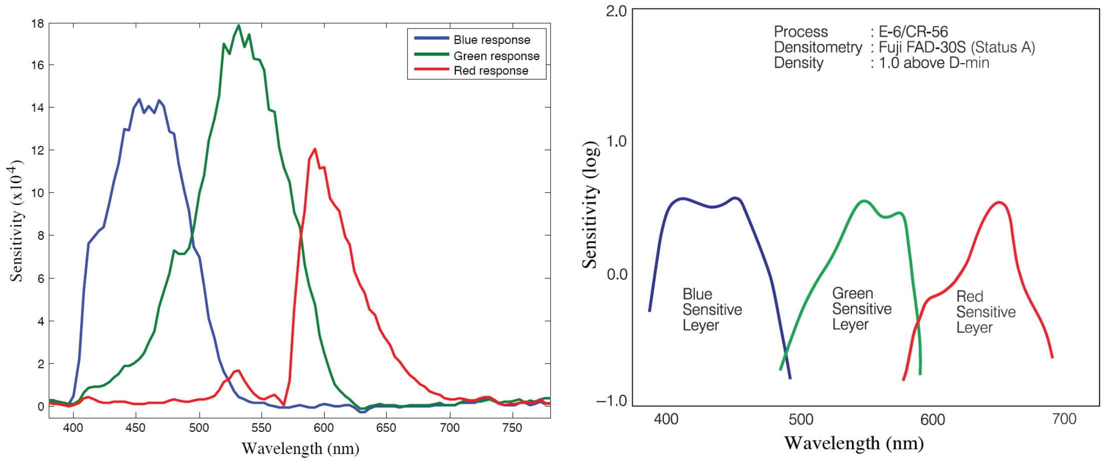
At first glance an increase in the spectral resolution of camera, lights, and projectors might not seem to offer any significant advantages in photography. Existing photographic methods quite sensibly rely on the well-established trichromatic response of human vision, and use three or more fixed color primaries such as red, green, and blue (RGB) to represent any color in the color gamut of the device.

Fixed-spectrum photography limits our ability to detect or depict several kinds of visually useful spectral differences. In the common phenomena of metamerism, 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 the device can accurately capture or reproduce. As demonstrated in the CIE 1931 color space chromaticity diagram, 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 Bayer grid of fixed, passive R, G, B filters overlaid on pixel detectors set the color primaries. Current DMD projectors use broad-band light sources passed through a spinning wheel that holds similar passive R, G, B filters. These filters compromise between narrow spectra that provide a large color gamut and broad spectra that provide greatest on-screen brightness.

3.4.1 Metamers and Contrast Enhancement

Photographers often use yellow, orange, red, and green filters for various effects in black and white photography. For example, white clouds and blue sky are often rendered as roughly the same intensity in a black and



(a) (b)

Figure 3.9. Comparison of the spectral response of a typical color film and digital camera sensor. (a) Spectral response of FujiFilm Velvia 50 color slide film (from:[Fuji 08] FUJIFILM. FUJICHROME Velvia for Professionals [RVP]. Data Sheet AF3-960E) (b) Spectral response of the Nikon D70 sensor [Moh 03].

white photograph. An orange or red filter placed in front of the lens makes the sky darker than the clouds, thus rendering them as a different shade on the black and white film. A red filter essentially attenuates the wavelength corresponding to blue and green colors in the scene, thus enabling the viewer to distinguish between the clouds and the sky in the resulting photograph. This is a classic case of effectively modifying the illumination to distinguish between metamers. In color photography, use of warming filters to enhance the contrast in a photograph is quite common.

Unfortunately, photographers can carry only a limited number of filters with them. Even these filters are often rather broad-band and useful for only very standard applications. We argue that a camera that allows arbitrary and instantaneous attenuation of specific wavelength ranges in a scene would give increased flexibility to the photographer. 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.

3.5 Wider Field of View

A key wonder of the human vision is its seemingly endless richness and detail; the more we look, the more we see. Most of us with normal- or corrected-to-normal vision 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 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 and soft, we do not confuse a tree trunk with the distant grassy field beyond it. Like any optical system, our eye's lens imperfections and photoreceptor array offers little or no resolving ability beyond 60–100 cycles per degree, yet we infer that the edge of a knife blade is discontinuous, it is disjoint from its background, and is not optically mixed with it on even the most minuscule scale. Of course we cannot see behind our heads, but we rarely have any sense of our limited field of view,[3] which stops abruptly approximately outside a cone spanning about $+/-80$ degrees away from our direction of gaze.This visual richness, and its tenuous connection to viewpoints, geometry, and sensed amounts of light can make convincing hand-drawn depictions of 3D scenes more difficult to achieve, as 2D marks on a page can seem ambiguous and contradictory (for an intriguing survey, see [Durand 02d]).

By comparison, camera placement, resolution, and framing are key governing attributes in many great film-like photographs. How might we achieve a more free-form visual record computationally? How might we

[3]Try this to map out the limits of your own peripheral vision; (some people have quite a bit more or less than others): gaze straight ahead at a fixed point in front of you, stretch out your arms back behind you, wiggle your fingers continually, and without bending your elbows, slowly bring your hands forward until you sense movement in your peripheral vision. Map it out from all directions; is it mostly circular? Is it different for each eye? Is it shaped by your facial features (nose, eyebrows, cheekbones, eye shape)? Your glasses or contact lenses? Do you include these fixed features in your conscious assessment of your surroundings?

construct a photograph to better achieve the impression of unlimited, unbounded field of view, limitless visual richness revealed with little more effort than an intent gaze?

We seldom find our impressions of our surroundings lacking in subtlety and richness; we seek out mountaintops, ocean vistas, and spectacular "big-sky" sunsets and dramatic weather effects in part because the more we look around, the more we see in these visually rich scenes. With close attention, we almost never exhaust our eye's abilities to discover interesting visual details, from the fine vein structure of a leaf to the slow boiling formation of a thunderstorm to the clouds in coffee to the magnificently complex composition of the luxurious fur on a hare.

A panorama is often created as a composite picture by stitching multiple photos of overlapping but distinct parts of the scene. Capturing a panorama requires the user to manually point the camera at interesting parts of the scene while ensuring there is adequate overlap in the various captured photos. The stitching process works best for scenes far away from the camera, thus making this very useful for capturing landscapes etc. Panoramas are popular because (a) ultra wide-angle lenses are expensive and usually not very good, and (b) the composite obtained by stitching has a much higher resolution than that of the digital sensor. Additionally, the photographer can select exactly the parts of the scene that are photographed and the parts that are skipped. The resulting panorama might not have a regular shape, but contains all the required "information." The main disadvantages of panoramas are that they require capture of multiple photos, and stitching photos may not give perfect results and might produce visible seams. As both resolution and field of view increase together, images are not only very large, but also become awkward to display and explore visually. Recently several efforts have led to progress in capturing giga-pixel resolution images via panoramic stitching and viewing those giga-pixel images using novel interfaces, such as HDview [Kopf et al 07]. The HDview system, for example, cleverly selects the image-browsing parameters by continuously varying the blend between spherical and planar projections.

Chapter 4
Illumination

Since adaptation from its precursor, the camera oscura, the photographic camera has evolved from a cumbersome view camera on a tripod to an easily portable, hand-held device. The technology of lighting the photographic subject, however, remains problematic—often bulky and awkward, not to say expensive. In view of the sophistication of modern consumer cameras, its arguable that today, only the use of elaborate auxiliary lighting distinguishes the amateur photographer from the professional. What can we learn from the expert? How can we create programmable lighting that minimizes critical human judgement at the time of capture?

Though the phrase had not been conceived at the time, in retrospect we can regard Harold Edgertons strobe photography at M.I.T. in the 1930s as an early instance of computational illumination. Instead of shortening the exposure of his cameras shutter, he employed a camera with a traditional shutter, but lighted his subjects with a novel strobe that emitted bursts of light of extremely short duration.

Every photographer knows how to capture a variety of subjects, under different lighting conditions, by manipulating the variables of the camerafocus, film speed, lens aperture, and shutter speed. These camera functions have been increasingly automated, or programmed. Similarly, the following parameters of auxiliary photographic lighting are programmable:

1. Presence or absence of auxiliary lighting;
2. Duration and intensity;
3. Color, wavelength, and polarization;

4. Position and orientation;

5. Modulation in space and time (strobing).

As we will see later, one can also exploit the change in natural lighting.

In earlier days of electro-chemical flashes, controlling the duration and intensity of flashes was quite challenging. But todays sources of illumination provide a high level of programmability. The advances in solid state lighting based on light emitting diodes or lasers, as well as sophisticated time modulation via strobes and space-modulation via spatial light modulators (SLMs) or video projectors allow for programmability. For ultimate programmability, researchers have developed domes in which hundreds of lights (or projectors) are distributed surrounding a subject.

4.1 Exploiting Duration and Intensity

4.1.1 Stroboscopic Freezing of High-Speed Motion

Harold Edgerton, along with Gjon Mili, in the 1930s pushed instantaneous photography to extremes by employing ultra-short strobes to illuminate transient phenomena, on the one hand, and ultra-short shutters to capture ultra-bright phenomena, such as his famous moving pictures of atomic detonations. These photos capture beautiful intricacy and the graceful flow of transient movement too rapid or complex for the eye to discern. Edgerton used the technique to capture dramatic images of balloons bursting, and of a bullet at the instant of its impact with an apple, for example. A key challenge was triggering the flash at the appropriate time. An audio trigger or laser-tripping trigger is commonly used for synchronization.

4.1.2 Sequential Multi-Flash Stroboscopy

A related technique was to employ a rapid sequence of strobe flashes to capture a time-sampled sequence of images onto a single photograph. The technique works well when the subject is photographed against a dark background and when subsequent frames have limited overlap. A good example is a golf swing performed in a plane perpendicular to the camera axis. The narrow golf club appears at distinct non-overlapping positions in successive frames. The results are less compelling when the scene is not contrasted against the background or the motion is towards or away from the camera.

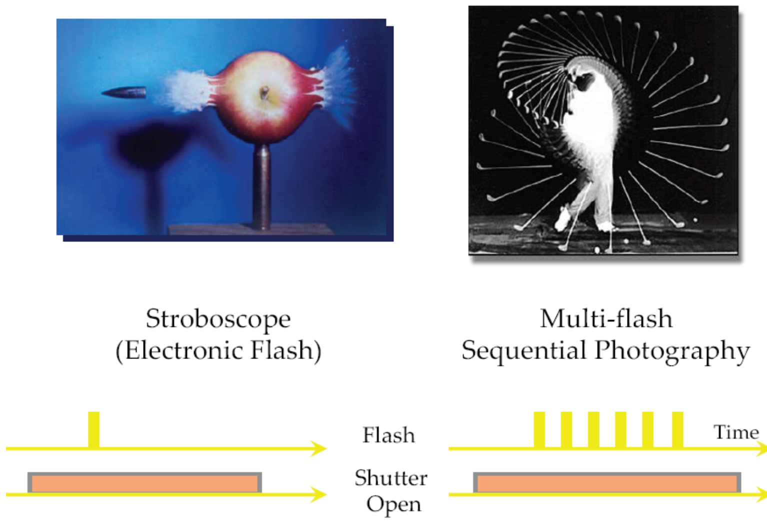


Figure 4.1. An early instance of computational illumination. By controlling flash duration, Edgerton captured an instant of frozen motion in one case, and the sequential components of a complex motion in the other.

4.1.3 Presence or Absence of Flash

The simplest form of computational illumination is perhaps the ubiquitous camera flash. Di Carlo et al. [01] first explored the idea of capturing a pair of images from the same camera position, one illuminated with ambient light only, the other using the cameras 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. [03] take multiple photos at different flash intensities, allowing the user to interpolate among them to simulate intermediate flash intensities.

4.1.4 Flash/No-Flash Pair for Noise Reduction

Petschnigg et al. [04] and Eisemann et al. [04] concurrently proposed similar strategies for combining information contained in the flash/no-flash image pair to generate a satisfactory single image. The photo without flash captures the large-scale illumination and overall ambiance of the scene. But in low light, the no-flash photo generally displays excessive noise. The flash photo, by contrast, shows much lower noise and greater high-frequency detail, but makes the image unnantural and fails to convey the

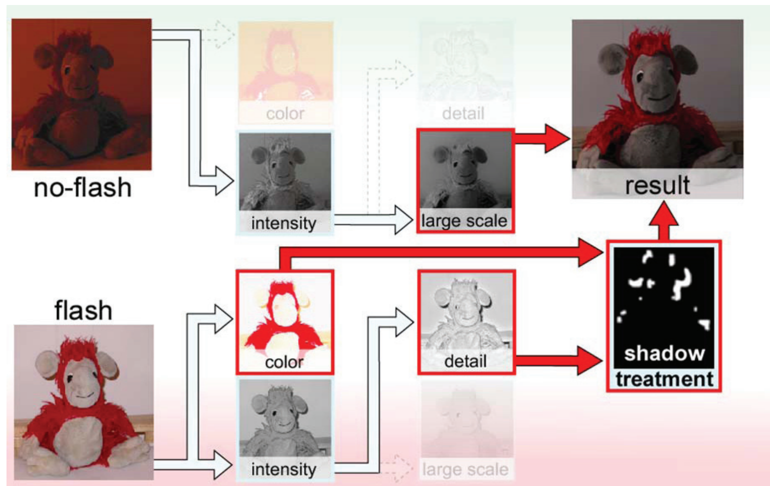


Figure 4.2. The process for reducing noise in a no-flash image. [Eisemann and Durand 04]

mood of the scene. The technique to combine the photos here is to decouple the high and low frequency components of each of the two photos and recombine them preserving the desired characteristics—detail and low noise from the flash photo, and overall ambiance from the photo taken without flash. 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 reduce noise in the no-flash image without excessive blurring of sharp features. Traditionally smoothing is performed on an image using information available in the same image. Smoothing of an image with a filter such as a Gaussian filter reduces high-frequency noise but it also blurs sharp edges. Using a traditional bilateral filter instead, the image filtering produces edge-preserving blur. The bilateral filter performs smoothing based on spatial extent as well as intensity similarity within the kernel filter. By exploiting the intensity similarity constraint, the traditional bilateral filter can create reduced noise while still preserving sharp details. Nevertheless, smoothing causes unnecessary suppression of weak (low-intensity high frequency) details along with noise.

With the joint bilateral filter, smoothing is influenced by detail in a companion image. For example, one can use a high-quality flash image

Figure 4.3. Merging a no-flash and a flash image. (Left) Top: Photograph taken in a dark environment; the image is noisy and/or blurry. Bottom: Flash photography yields a sharp but flat image with distracting shadows at the edges of objects. (Middle) Zoom, showing the noise of the available-light image. (Right) The technique merges the two images to transfer the ambiance of the available lighting. Note the shadow of the candle on the table. (Courtesy, Elmar Eisemann and Fredo Durand, 2004)

to denoise a no-flash image. This provides enhanced ability to find and preserve weak details (low confidence edges) in the presence of noise. The basic idea is to smooth the no-flash image while preserving all edges that exist in the flash image. The spatial kernel remains the same within the no-flash image, but the intensity similarity is measured with respect to the corresponding flash-image pixels. Since the flash photo displays lower noise, a better result is achieved and over- or under-blurring is avoided.

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 detail).

4.1.5 Flash, Exposure, and Dynamic Range

Present-day cameras use onboard sensors and algorithms to approximate the correct intensity of flash and proper exposure settings. But these estimates, based on aggregate measurements, lead often to under- or over-illumination. A single flash intensity value cannot illuminate distant or dark objects without simultaneously saturating or "blowing out" nearby or bright objects. Image quality is affected when a scene exceeds the dynamic range of the camera. For such situations, Agrawal et al. [Agrawal et al. 05]

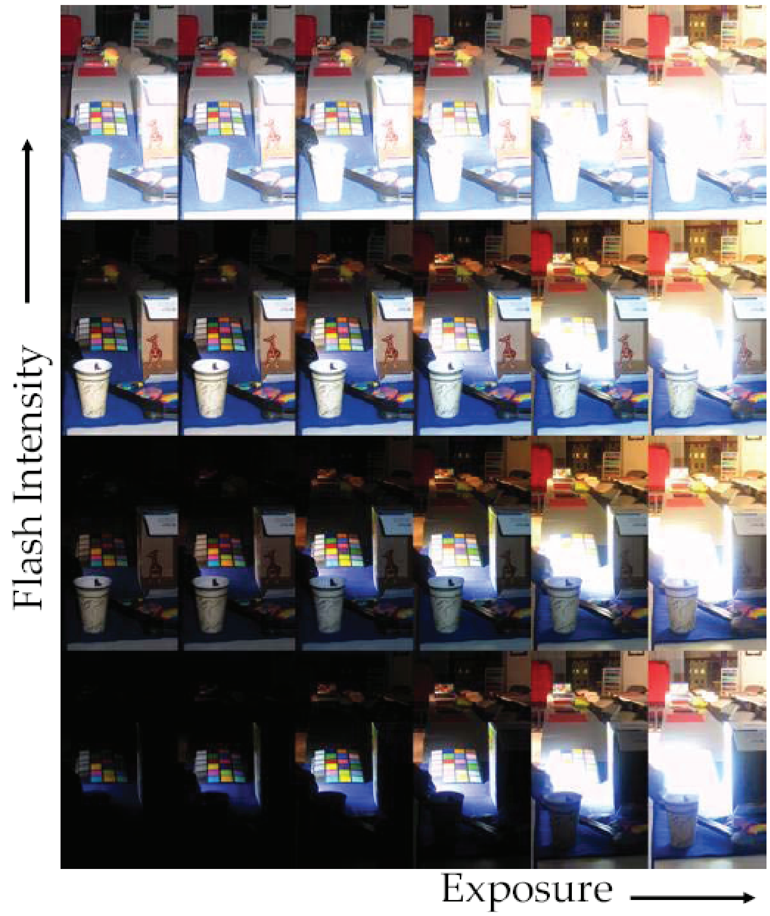


Figure 4.4. Flash exposure high dynamic range sampling [Agrawal et al. 05]
.

suggest merging multiple images captured with varying flash intensities and exposures to construct a correct HDR image.

Figure 4.4 shows an example of this strategy for exploring the impact on an image by varying flash intensity and exposure parameters. Given any scene in three dimensions, the requisite brightness of flash is a function of the depth of the scene, its natural illumination, and the surface reflectance of its visual elements. For example, a distant object with low reflectance will require a bright flash, whereas a nearby point or well lit area naturally will be over-exposed by a flash, even at lower flash intensity. The scene might extend to a far off depth that would not be illuminated by even an intensely bright flash; only a longer exposure time would properly capture it. To capture such a challenging scene with a mixed depth, reflectance, and natural lighting, one may need to make multiple exposures, each at a different setting along the exposure and flash-intensity axes. The example shows photos taken at six different exposure settings, without flash and at three different flash brightness settings—a total of 24 exposures. Many consumer cameras as well as professional cameras offer manual setting of flash intensity. Though making 24 captures to achieve a single image may be excessive, Agrawal et al. present a greedy approach: pixel values of each capture are analyzed for over- or under-exposure, suggesting optimal exposure and flash parameter setting for the subsequent capture [Agrawal et al. 05]. A greedy algorithm makes the locally optimal choice at each stage in order to calculate the global optimum. By adaptive sampling of the flash-exposure space, the number of captured images required for any given scene is minimized.

4.1.6 Removing Flash Artifacts

Flash images suffer notoriously from several problems: over-exposure, or blowing-out of nearby objects, poor illumination of distant objects, reflections from objects strongly lit by the flash and strong highlights, reflections of the flash itself, on glossy surfaces. One approach has been the merging of a flash and an ambient image pair to produce better flash images. Agrawal et al. [Agrawal et al. 05] use a technique based on image intensity gradients. The orientation of the gradient's vector at a pixel in a rasterized image is given by the direction with maximum rate of change of intensity. The magnitude is the rate of that change. For example, in an intensity edge, the gradient vector orientation is perpendicular to the edge and gradient vector magnitude is the strength of the edge. Agrawal et al. observe that the orientation of image gradients due to reflectance or geometry are illumination-invariant, whereas those due to changes in lighting are not. Hence, the gradient coherence model indicates that in the absence of artifacts, the gradient vector orientation in flash and ambient (no-flash) images should be the same. On the other hand, a change in gradient vector orientation indicates presence of an artifact. They propose a gradient projection scheme to decompose the illumination effects from the rest of the image. The gradient projection scheme is based on a gradient coherence model.

Given a set of gradient vectors at each pixel, i.e., a gradient vector field, it is possible to reconstruct an image that satisfies the gradient field. Several new techniques have emerged since 2002 based on such gradient vector manipulation and image reconstruction.

Figure 4.5 shows flash and ambient images of a painting. The ambient image includes distracting reflections of the photographer. The low-exposure flash image avoids these reflections, but shows a hot spot. Reflections in the ambient image are removed by subtracting the component of the ambient image gradients perpendicular to the flash image gradients. Reconstruction from the projected gradients produces a reflection-free image. Reconstruction from residual gradients recovers the reflection layer. However, the gradient orientation in not available when both images have co-located artifacts (photographer reflection as well as flash hot-spot). In addition, gradient orientation is unstable in homogeneous flat regions, so photographer reflection in such parts will be difficult to recover in such regions. In later works, Agrawal et al. have introduced a gradation projection tensor which is more robust compared to the simple gradient projec-

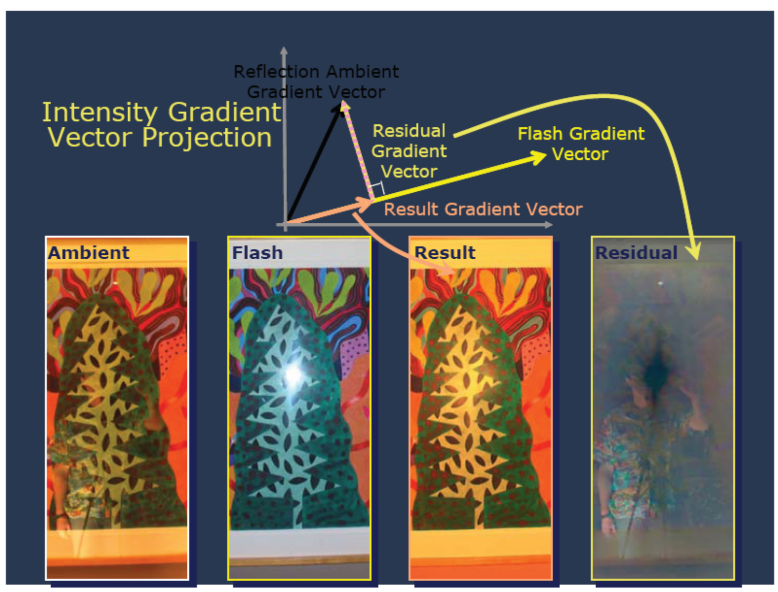


Figure 4.5. Removing flash artifacts with gradient vector projection. Undesirable artifacts in photography can be reduced by comparing image gradients at corresponding locations in a flash and ambient image pair [Agrawal et al. 05].

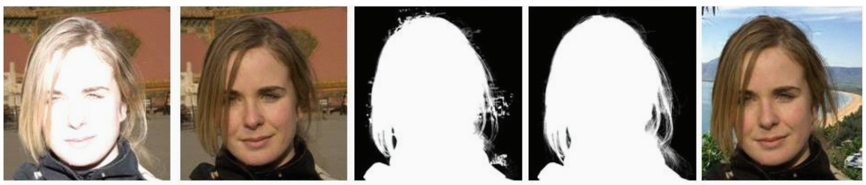


Figure 4.6. Flash matting allows extraction of foreground using an additional flash image [Jian et al. 06].

tion procedure. Their work also shows how to compensate for the falloff of flash intensity along the axis of depth of a scene by exploiting the ratio of the flash and ambient photos.

4.1.7 Flash-Based Matting

Matting, to extract a foreground subject from its background, can be made more precise by combining flash matting with more conventional Bayesian matting Jian et al. 06]. Inspired by the simple observation that the greatest difference between the flash and the ambient images of such a pair—provided the background is sufficiently distant—is the quality of illumination of the foreground subject, this approach is readily applicable to images produced by off-the-shelf flash-equipped cameras. For example, a pixel-wise ratio of flash image and no-flash image will be close to 1.0 for distant points (background) but significantly higher for nearer points (foreground). With a joint Bayesian matting, even foreground subjects with complex edges, such as fur or hair, can be more precisely extracted with an alpha-matt and placed into a new image context. Unlike traditional Bayesian matting, this 2-photo technique works even when foreground and background have similar colors. However, the technique fails when the flash image does not encode the distance related fall-off in the expected way. For example the background is not sufficiently far away, or when the object is rounded

4.2 Modifying Color and Wavelength

The scene radiance is a product of incident illumination and reflectance. By changing the wavelength profile (often simplified as color profile) of the incoming light or capturing specific wavelength channels, one can perform programmable color manipulations of images.

By changing the spectrum of illumination, it is possible to probe a scene and create multi-spectral photos or overcome confusion due to metamers (colors that 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. The scene is illuminated with higher frequency (lower wavelength) illumination which results in emission in lower frequency (higher wavelength). Thus, for example, subjects irradiated with ultraviolet may release, green, yellow, or pink light, and subjects irradiated with visible light may emit infrared fluorescence. Household fabrics are treated with fluorescent dyes to make them look whiter. When illuminated with ultraviolet light (in dimly lit discos, say), the clothes emit several lower frequencies and appear bright. In most fluorescence photography, a UV wavelength selective filter is placed at the light source. Another filter of a different (visible) wavelength selection 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 a scattering medium such as internal biological samples features in microscopy. By using a wavelength-rejecting optical filter in front of a camera, we can reject all scattered light that has the same wavelength. The induced fluorescence, however, has a different wavelength and can be imaged by this camera.

Let us look at an example where this wavelength manipulation is done in post-capture stage. Paul Haeberli 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 Haeberli 92]. He illustrates the technique with a scene lighted with two lamps, to the left and to the right of the subject (Figure 4.7), in addition to ambient lighting. Three exposures are made, one with ambient lighting only, one 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 lighted by the lamps. “This gives us an image that shows exactly what light is contributed by each light source ... Now we can simulate what the scene would look like if the lamp on the left was blue instead of white ... By applying a similar process to the lamp on

Figure 4.7. Programmable combination of colors [Haeberli 06].

the right, we can now synthetically illuminate the scene with multicolored lamps. The brightness and color of any number of lamps can be controlled in this way." This strategy allows also for negative lighting by subtracting light coming from a particular lamp.

4.3 Position and Orientation of Lighting

The placement and orientation of auxiliary lighting can also be altered, modifying the shading of subjects as well as shadows throughout an image. Changing the orientation of light with shaped output profile also changes its absolute intensity, but does not change the angle of incidence of light rays at any point in the scene.

4.3.1 Shape and Detail Enhancement using Multi-Position Flashes

A moving light source can be used to inspect and extract subtle surface detail and also distinguish silhouettes of objects. A traditional edge-detection filter in images can detect the reflectance-discontinuities but does a poor job in estimating edges due to shape discontinuities. Shape discontinuities occur due to depth difference (between a foreground and background patch) or due to sharp change in surface orientation (a ridge or a valley). By observation under a moving light source, and noting shading and shadows, one can highlight such discontinuity.

Raskar et. al. employed a camera equipped with multiple flashes to find the silhouettes in a scene and create stylized or cartoon-like images [Raskar et al. 04]. The multi-flash camera employs four strategically placed flashes to cast shadows along the depth discontinuities of a scene. Depth discontinues are edges in the scene due to shape boundaries or sil-

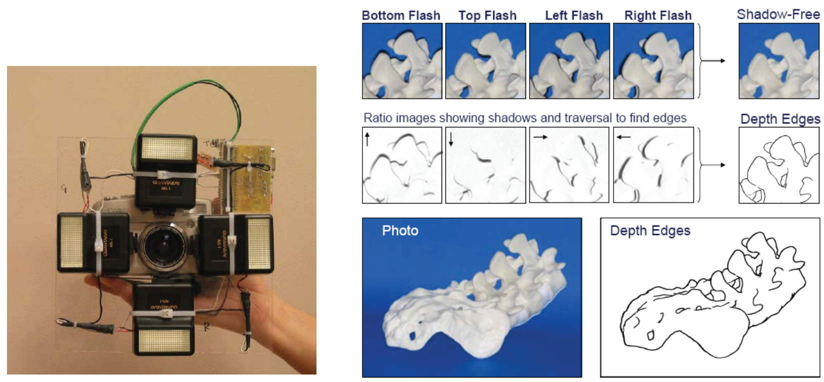


Figure 4.8. Multi-flash camera for depth edge detection. (Left) A camera with four flashes. (Right) Photos resulting from individual flash exposures, highlighted shadows and epipolar traversal to compute the single-pixel depth edges.

houettes, where the depth value of neighboring pixels is different. More precisely, depth discontinuity are *depth edges* due to C_0 discontinuity in the depth map with respect to the camera. The flashbulbs illuminate the scene during image capture creating thin slivers of shadow along the depth discontinuities. The position of shadows is of course determined by the position of the flash: when the flash is on the right, shadows 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 4.8, 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 of an image are detected by first computing a shadow-free image, approximated with the maximum 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 produce the depth edges.

The technique, however, fails to mark a depth edge when it is difficult to detect the shadow slivers attached to the image of the depth edge. The shadow detection fails, for example, when the background is too far away relative to the depth edge. If the foreground object is too narrow, (think of a nail), the shadow observed in the image is detached from the object. Since specularities from shiny surfaces can confuse the max-composite image, the authors described a method using an intrinsic image (described in Section 4.5 instead of the max-image. The detected silhouettes are then used to stylize the photograph and highlight important features. Raskar et al. demonstrated similar silhouette detection in video using a high-speed flash sequence [Raskar et al. 04].

Using a larger number of images captured with varying light positions around the photographic subject in a studio (or laboratory) setting, one can enhance the subtle surface features as observed in grazing angle illumination, shadows due to complex geometry, specularities, and subsurface scattering.

Akers et al. [Akers et al. 03] use spatially varying image weights on images acquired with a light stage similar to the work of Debevec [Debevec et al. 01]. 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. [Anrys et al. 04] and Mohan et al. [Mohan et al. 05] 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.

4.3.2 Relighting Using Domes and Light Waving

The goal of image-based relighting is to create arbitrary novel lighting in a photo in post-capture editing. Instead of building an accurate 3D model of the scene appearance, image-based relighting relies on the simple observation that light interacts linearly with material objects [Nimeroff, 94, Haeberli 92]. 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. Adding two photos, each taken with only one light, is equivalent to capturing a photo with two lights. More precisely, if a fixed camera makes 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 $I_{\text{out}} = \sum_i (w_i I_i)$. Adjusting weights allows us to create an output image from a linear combination of input images. However, due to linearity, the effective output image is the same as if the light sources had been modulated (turned brighter or dimmer). This achieves digital post-capture *relighting* of the image.

For accurate relighting of a scene to synthesize arbitrary virtual lighting conditions, ideally we need to photograph the scene by moving the light through every possible position of the lighting fixture—a challenging task. For example, let us say that we limit light positions within a square on a flat surface and take successive photos by moving the light source within that square. From this dataset, we can resynthesize photos only from virtual light sources lying within that square. To overcome this limitation and reduce the number of lighting variations required, we can exploit the fact that all incident light can be geometrically parameterized by a 5D plenoptic function, i.e., 5D ray-representation. Effectively, we need to make an exposure by turning on lighting just one ray at a time. If we limit ourselves to resynthesizing novel lights positioned only outside the convex hull of the object, however, the problem is slightly simplified. In this case, we can represent the incident light field (illumination field) using a 4D ray parameterization. To understand this, we need to consider the higher-dimensional incident light field (illumination field) and its impact on the resultant outgoing light field.

We discussed light fields earlier in Chapter 2. Light fields and lumigraphs reduced the more general 5D plenoptic function to a four-dimensional function, $L(u,v,s,t)$ that describes the presence of light in free space, ignoring the effects of wavelength and time [Adelson 91]. Here (u,v) and (s,t), respectively, are the parameters on two parallel planes that describe a ray of light in space. To represent the illumination field on an object, a slightly different parameterization can be used.

Imagine the object surrounded by a spherical dome with projectors aimed inwards. Parameter (θ_i, ϕ_i) describes the angular position of the projector on the unit sphere, and (u,v) the pixel position in the projected image from that 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 array of cameras on that spherical dome aimed inwards would capture the entire radiant light field of an object, $L_r(u,v,\theta,\phi)$. Debevec et al. introduced the 8D reflectance field that describes the relationship of the 4D incident light fileds plus the 4D radiant light fields of a scene [Debevec et al. 01]. An additional dimension of time is sometimes added to describe the changing interaction of light with an object over time.

For relighting, we are interested in a fixed viewpoint; hence from a 4D radiant field, we capture only a 2D photo. Along with the 4D incident illumination field, this becomes a problem of 6D reflectance-field estimation. 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 will 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. employed a light stage comprising a light mounted on a rotating robotic arm to acquire the non-local reflectance field of a human face [Debevec et al. 01]. The point-like light source can be thought of as a simplified projector of a single pixel. Thus, the incident light field is reduced to a 2D function. The reflectance field with a 2D incident light field plus a 2D photos is therefore only 4D. The generation of novel images under arbitrary lighting was demonstrated. This was accomplished simply by adjusting the weights w_i to match the desired intensity of illumination from various directions. Going beyond relighting, the authors added a small number of cameras all firing in parallel to capture images of the face from neighboring viewpoints. They were able to simulate small alterations of viewpoint using a simple model for skin reflectance. Hawkins et al. employed a similar configuration to digitize cultural artifacts by capturing reflectance fields. They argue for the use of the reflectance field in digital archiving, rather than geometric models and reflectance textures [Hawkins et al. 01]. Koudelka, et. al. 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 [Koudelka et al. 01].

Debevec et al. proposed an enhanced light stage comprising a large number (156) of inwardly oriented LEDs distributed over a spherical structure approximately two meters in diameter around the photographic subjectin this instance, an actor (Figure 4.9 (left)). Each light was set to an arbitrary color and intensity to simulate the effect of a real-world environment around the actor (Figure 4.9 (center)). The images gathered from the light stage, together with a mask of the actor captured with infrared sources and detector, were used to composite the actor seamlessly into a virtual set, while maintaining consistent illumination (Figure 4.9 (right)) [Debevec et al. 02]. Malzbender et al. employed 50 inwardly oriented flashes distributed over a hemispherical dome, together with a novel scheme for compressing and storing the 4D reflectance field called the *polynomial texture map* [Malzbender et al. 01]. They assumed that the color of a pixel changed smoothly as the light moved around the object

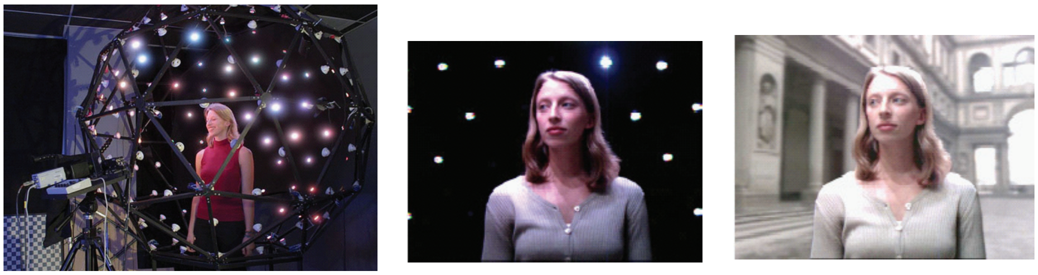


Figure 4.9. The light stage for lighting-aware matting [Debevec et al. 02].

and stored only the coefficients of the biquadratic polynomial that best modelled 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, although specular highlights are not modeled well by the polynomial model and result in visual artifacts.

To avoid the extensive light-stage, one can employ a more flexible set-up and use, say, a handheld light source freely moving around the photographic subject. Then the task is to estimate these light positions. The free-form light stage [Masselus 02] presented a strategy where 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 25–30 minutes. Winnemoller et al. used dimensionality reduction and a slightly constrained light scanning pattern to estimate the light source position without the need for additional fiducials in the scene [Winnemoeller et al. 05].

Mohan et al. argue that accurate calibration of light positions is unnecessary for the application of photographic relighting, and propose a novel reflector-based acquisition system [Mohan et al. 05]. They place 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 [Fuchs 05].

The disadvantage of these 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 problems can be addressed using motion compensation via image registration. But often the motion of any one of the elements creates two different relative motions. This makes it quite challenging to use these methods for traditional photography. Nevertheless, in many controllable setting these methods can be used.

4.3.3 Towards 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 [Debevec et al. 00]. The measurement of reflectance fields is an active area of research. The 8D reflectance field is a transport that 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. Note that the synthesized results demonstrate all global illumination effects such as diffuse inter-reflections, shadows, caustics and sub-surface scattering, without the need for an explicit physical simulation.

However, most of this research has focused on capturing meaningful lower-dimensional slices of the 8D reflectance field. Earlier, we saw examples 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 one obtains a 6D slice of an 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 projector pixels in each photo. Masselus et al. captured such data sets using a single moving projector positioned in the k positions [Masselus et al. 03]. Sen et al. exploited Helmholtz reciprocity to develop a dual photography approach [Sen et al. 05]. The Helmholtz reciprocity allows you 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 data sets. An earlier method for capturing the full 8D reflectance field exploited the data-sparseness of the 8D transport matrix to represent the transport matrix by local rank-1 approximations. Based on the sparsity observations, the authors developed a hierarchical parallelized acquisition technique that significantly sped up the reflectance field capture process [Garg et al. 06].

4.4 Modulation in Space and Time

The capacity to modulate the intensity of flash over space and time provides additional control of the resulting image. An intelligent flash would behave much like a projector. A projector allows modulation of ray intensities in each ray direction by changing pixel intensities and is an easily available programmable spatio-temporally modulated 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 us to change not only the overall brightness but also rapidly change the radiance of every ray emitted from the projector-flash, is a powerful alternative to conventional flash, as it provides full control over the 2D set of rays it emits via pixel intensity manipulation. Shree Nayar coined the term “CamPro” to designate a projector that is supporting the operation of a camera [Nayar 06]. A projector can project arbitrarily complex illumination patterns onto the scene, capture the corresponding images, and compute information regarding the scene that is impossible to obtain with traditional flash. Captured images are optically coded by the patterned illumination of the scene. In the future, then, the unwieldy projector may be replaced by smart lasers or light sources with highly programmable mask patterns.

4.4.1 Projector for Structured Light

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 light pattern and imaging the illuminated scene from one or more cameras to simplify the well known correspondence problem. Given a pair of images of the same 3D scene, captured from two different points of view, 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 representation in the two images generates the pair of corresponding pixels. In turn, given the pair of corresponding pixels in two images, by triangulaton one can compute the 3D location of that point. In the case of projected structured light, a single camera view can be used along with the projector view. The projected pattern is coded so that over time each projector pixel is assigned a unique binary. Thanks to this coding, correspondences between camera image points and points of the projector pattern can easily be decoded. By triangulating the decoded points, 3D information is recovered. In place of a passive camera, the projector actively encodes the space via illumination. Hence this is known as active stereo triangulation. For a good overview, see [Salvi et al. 02]. 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 Rusinkiewicz et al. 02].

4.4.2 High Spatial Frequency Patterns

Active illumination approaches have been used to analyze multi-path 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 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 4.10 shows a scene captured using a checkerboard illumination pattern. If the checkerboard patterns frequency 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 using the complement of this illumination pattern. In this case, the point does not have a direct component but still produces exactly half of the global component. This is 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 of Figure 4.10 shows separation results for a scene with peppers of different colors. The direct image includes mainly the specular reflections from the surfaces of

Figure 4.10. The projector can be used as a programmable camera flash. By projecting high frequency (checkerboard) patterns, one can separate direct and global illumination effects.

the peppers. The colors of the peppers come from subsurface scattering effects the global image captures. Altering the colors of the peppers in the global image and recombining it with the direct image yields a novel image, like the one shown on the right in Figure 4.10. In addition to subsurface scattering, this separation method is applicable to a variety of global illumination effects, including interreflections among opaque surfaces and volumetric scattering from participating media. Thus, one can distinguish the first bounce direct illumination effect from the multi-path scattering caused by global illumination.

Seitz et al. go beyond this partial inversion of light transport. They developed a mechanism to represent the impact of individual bounces. 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 [Seitz et al. 05].

4.4.3 Modulation in Time

The pattern of the flash in time can also be changed. We can use strobes to synchronize 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 scene activity and strobed flash have slightly different frequencies, the perceived periodic variationss rate is the difference between the two frequencies. For example, vocal folds moving at 1000 Hz can be viewed with a laryngoscope with auxiliary lighting. If the strobe is also at 1000 Hz, the vocal folds appear frozen if the person maintains a continuously pitched sound. If the strobed frequency is 999 Hz, 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.

4.5 Exploiting Natural Illumination Variations

Sometimes we cannot actively change the illumination for photography, but we can still exploit natural variations such as changes in sunlight throughout the day.

4.5.1 Intrinsic Images

In an intrinsic image decomposition, the goal is to decompose the input image I into a reflectance image and an illumination image.

An image is produced as a result 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 of the scene describes the pattern and ma-

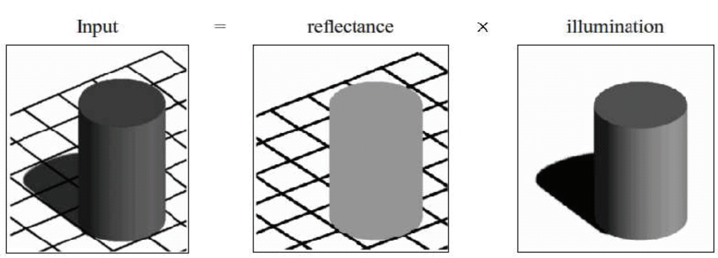


Figure 4.11. Intrinsic images. The goal is to decompose an image into its reflectance (intrinsic) and shading (illumination) layer.

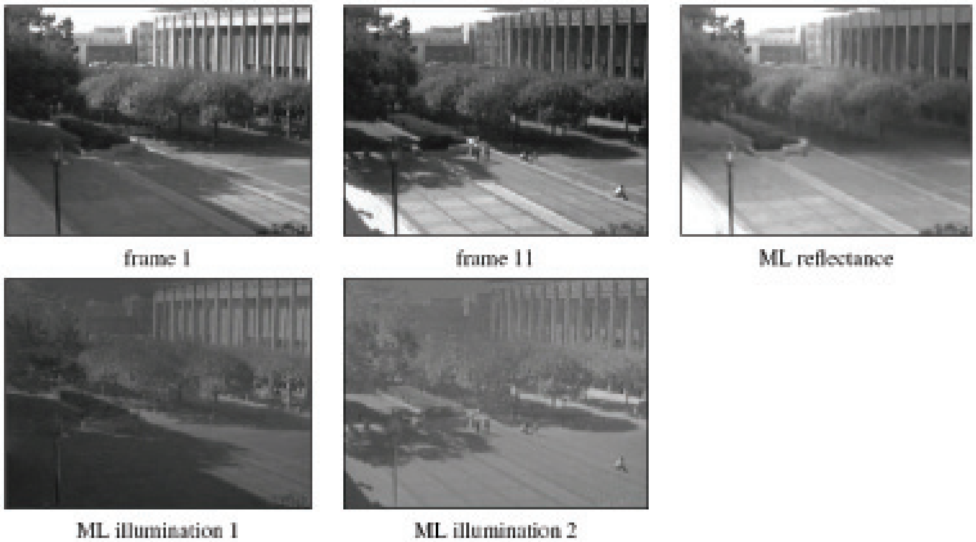


Figure 4.12. Intrinsic images from a web camera sequence. From 35 web camera images varying mostly in illumination, the maximum-likelyhood (ML) reflectance image free of cast shadows can be created. (Permission Yair Weiss)

terial and how each point reflects light. 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 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 [Weiss 01]. 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, one can do augmentation of real scene photos.

Figure 4.13. The night-time photo is context enhanced to the photo on the right using a prior daytime image [Raskar et al. 04].

4.5.2 Context Enhancement of Night-Time Photos

night-time images such as the one shown in Figure 4.13 (left) are difficult to understand because, due to poor illumination, they lack background context. 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, one can simply enhance the context of the low-quality image or video by fusing the appropriate pixels as shown in Figure 4.13 (right). Raskar et al. show that an image-fusion approach is based on a gradient domain technique that preserves important local perceptual cues while avoiding traditional problems such as aliasing, ghosting, and haloing [Raskar et al. 04]. 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 a linear combination 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. Notice that the dark regions of the night image in Figure 4.13 (left) are filled in by day-image pixels but with a smooth transition.

Bibliography

Fusion of Images Taken by Varying Camera and Scene Parameters

General

[Akers et al. 03] David Akers, Frank Losasso, Jeff Klingner, Maneesh Agrawala, John Rick, and Pat Hanrahan. "Conveying Shape and Features with Image-Based Relighting." In *Proceedings of the 14th IEEE Visualization 2003 (VIS'03)*, pp. 349–354. Washington, DC: IEEE Computer Society, 2003.

[Burt and Kolczynski 93] P. Burt and R. Kolczynski. "Enhanced Image Capture through Fusion." In *Proceedings of the Fourth International Conference on Computer Vision*, pp. 173–182. Los Alamos, CA: IEEE Press, 1993.

[Levin et al. 04] A. Levin, A. Zomet, S. Peleg, and Y. Weiss. "Seamless Image Stitching in the Gradient Domain." In *Computer Vision-ECCV 04: 8th European Conference on Computer Vision, Prague, Czech Republic, May 11–14, 2004, Proceedings, Lecture Notes in Computer Science 3024*, pp. 377–389. Berlin: Springer, 2004.

[Massey and Bender 96] M. Massey and W. Bender. "Salient Stills: Process and Practice." *IBM Systems Journal* 35:3&4 (1996), 557–574.

[Mutter and Krause 92] S. Mutter and M. Krause. *Surrational Images: Photomontages*. Champaign, IL: University of Illinois Press, 1992.

[Robinson 69] H. P. Robinson. *Pictorial Effect in Photography: Being Hints on Composition and Chiaroscuro for Photographers*. London: Piper & Carter. 1869.

[Muybridge 55] E. Muybridge. *The Human Figure in Motion*. New York: Dover Publications, Inc., 1955.

[Agarwala et al. 04] A. Agarwala, M. Dontcheva, M. Agrawala, S. Drucker, A. Colburn, B. Curless, D. Salesin and M. Cohen. "Interactive Digital Photomontage." In *ACM SIGGRAPH 2004 Papers*, pp. 294–302. New York: ACM Press, 2004.

[Nayar and Narasimhan 02] S. K. Nayar and S. G. Narasimhan. ""Assorted Pixels: Multi-Sampled Imaging with Structural Models." In *Computer Vision-ECCV 2002: 7th European Conference on Computer Vision, Copenhagen, Denmark, May 28-31, 2002. Proceedings, Part IV, Lecture Notes in Computer Science 2353*, pp. 636–652. Berlin: Springer, 2002.

Time

[Braun 92] M. Braun. *Picturing Time: The Work of Etienne-Jules Marey*. Chicago: The University of Chicago Press, 1992.

[Freeman and Zhang 03] W. T. Freeman and H. Zhang. "Shape-Time Photography." In *Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 151–157. Los Alamitos, CA: IEEE Press, 2003.

Exposure

[Fattal et al. 02] R. Fattal, D. Lischinski and M. Werman. "Gradient Domain High Dynamic Range Compression." *ACM Transactions on Graphics* 21:3 (2002), 249–256.

[Reinhard et al. 02] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda. "Photographic Tone Reproduction for Digital Images." *ACM Transactions on Graphics* 21:3 (2002), 267-276.

[Debevec and Malik 97] P. E. Debevec and J. Malik. "Recovering High Dynamic Range Radiance Maps from Photographs." In Proceedings of SIGGRAPH 97, Computer Graphics Proceedings, Annual Conference Series, edited by Turner Whitted, pp. 369-378. Reading, MA: Addison-Wesley, 1997.

[Durand and Dorsey 02] F. Durand and J. Dorsey. "Fast Bilateral Filtering for the Display of High-Dynamic-Range Images." ACM Transactions on Graphics 21:3 (2002), 257-266.

[Mann and Picard 95] S. Mann and R. W. Picard. "Being 'Undigital' with Digital Cameras: Extending Dynamic Range by Combining Differently Exposed Pictures." In Proceedings of the 47th Annular Conference of the Society for Imaging Science and Technology, pp. XX-XX. Springfield, VA: The Society of Imaging Science and Technology, 1995.

[Tumblin and Turk 99] Jack Tumblin and Greg Turk. "LCIS: A Boundary Hierarchy for Detail-Preserving Contrast Reduction." In Proceedings of SIGGRAPH 99, Computer Graphics Proceedings, Annual Conference Series, edited by Alyn Rockwood, pp. 83-90. Reading, MA: Addison Wesley Longman, 1999.

[DiCarlo and Wandell 00] J. DiCarlo and B. Wandell. "Rendering High Dynamic Range Images." In Sensors and Camera Systems for Scientific, Industrial, and Digital Photography Applications (Proceedings Volume), Proceedings of SPIE 3965, pp. 392-401. Bellingham, WA: SPIE, 2000.

[Kang et al. 03] S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski. "High Dynamic Range Video." ACM Transactions on Graphics 22:3 (2003), 319-325.

Focus

[Haeberli 94] P. Haeberli. Grafica Obscura: Collected Computer Graphics Hacks. Available at http://www.sgi.com/grafica/, 1994

[McGuire et al. 05] Morgan McGuire, Wojciech Matusik, Hanspeter Pfister, John F. Hughes, and Fredo Durand. "Defocus Video Matting." Proceedings of ACM SIGGRAPH 2005: ACM Transactions on Graphics 24:3 (2005), 567-576.

Illumination

[Anrys and Dutre 04] Frederik Philip. "Image Based Lighting Design." In Proceedings of the Fourth IASTED International Conference on Visualization, Imaging, and Image Processing, edited by Juan J. Villanieva, pp. 222- 230. Calgary: ACTA Press, 2004.

[Akers et al. 03] David Akers, Frank Losasso, Jeff Klingner, Maneesh Agrawala, John Rick, and Pat Hanrahan. "Conveying Shape and Features with Image-Based Relighting." In Proceedings of the 14th IEEE Visualization 2003 (VIS'03), pp. 349-354. Washington, DC: IEEE Computer Society, 2003.

[Agrawa et al. 05] Amit Agrawal, Ramesh Raskar, Shree K. Nayar, and Yuanzhen Li. "Removing Photography Artifacts Using Gradient Projection and Flash-Exposure Sampling." In Proceedings of ACM SIGGRAPH 2005: ACM Transactions on Graphics 24:3 (2005), 828-835.

[Debeve et al. 00] Paul Debevec, Tim Hawkins, Chris Tchou, Haarm-Pieter Duiker, Westley Sarokin, and Mark Sagar. "Acquiring the Reflectance Field of a Human Face." In Proceedings of SIGGRAPH 2000, Computer Graphics Proceedings, Annual Conference Series, edited by Kurt Akeley, pp. 145-156. Reading, MA: Addison- Wesley, 2000.

[Debevec et al. 02] Paul Debevec, AndreasWenger, Chris Tchou, Andrew Gardner, JamieWaese, and Tim Hawkins. "A Lighting Reproduction Approach to Live-Action Compositing." In Proceedings of ACM SIGGRAPH 2002: ACM Transactions on Graphics 21:3 (2002), 547-556.

Passive Illumination

[Raskar et al. 04] R. Raskar, A. Ilie, and J. Yu. "Image Fusion for Context Enhancement and Video Surrealism." In Proceedings of the 3rd International Symposium on Non-photorealistic Animation and Rendering, pp. 85-152. New York: ACM Press, 2004.

[Weiss 01] Y. Weiss. "Deriving Intrinsic Images from Image Sequences." In Proceedings of the Eighth IEEE International Conference on Computer Vision (ICCV 2001), pp. 68-75. Los Alamitos, CA: IEEE Press, 2001.

Polarization

[Schechner et al. 01] Y. Y. Schechner, S. G. Narasimhan, and S. K. Nayar. "Instant Dehazing of Images Using Polarization." In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Vol. I, pp. 325- 332. Los Alamitos, CA: IEEE Press, 2001.

[Nayar et al. 93] S. K. Nayar, X. Fang, and T. E. Boult. "Removal of Specularities Using Color and Polarization." In Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 583-590. Los Alamitos, Ca: IEEE Computer Society, 1993.

[Wu and Walsh 03] P. Wu and J. T. Walsh. "Tissue Polarization Imaging of Multiply-Scattered Reflected Light Lasers." in Lasers in Surgery and Medicine 23, Supplement 15 (2003), XX-XX.

Wavelength

[Socolinsky 00] D. A. Socolinsky. "Dynamic Range Constraints in Image Fusion and Realization." Proc. IASTED Int. Conf. Signal and Image Process, 349-354 (2000).

[Schechner and Nayar 04] Y. Y. Schechner and S. K. Nayar. "Uncontrolled Modulation Imaging." In Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 2, pp. 197-204. Los Alamitos, CA: IEEE Computer Society, 2004.

Location

[Agarwala et al. 06] Aseem Agarwala, Maneesh Agrawala, Michael Cohen, David Salesin, and Rick Szeliski. "Photographing Long Scenes with Multi-Viewpoint Panoramas." Proceedings of SIGGRAPH 2006: ACM Transactions on Graphics 25:3 (2006), 853-861.

[Wilburn et al. 05] B. Wilburn, N. Joshi, V. Vaish, E. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy. "High-Performance Imaging Using Large Camera Arrays." Proceedings of ACM SIGGRAPH 2005: ACM Transactions on Graphics 24:3 (2005), 765-776.

[Snavely et al. 06] Noah Snavely, Steven M. Seitz, and Richard Szeliski. "Photo Tourism: Exploring Photo Collections in 3D." In ACM SIGGRAPH 2006 Papers, pp. 835-846. New York: ACM Press, 2006.

[Rother et al. 06] Carsten Rother, Lucas Bordeaux, Youssef Hamadi, and Andrew Blake. "AutoCollage." In ACM SIGGRAPH 2006 Papers, pp. 847-852. New York: ACM Press, 2006.

Matting

[Chuang et al. 01] Y.-Y. Chuang, B. Curless, D. Salesin, and R. Szeliski. "A Bayesian Approach to Digital Matting." In Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. II, pp. 264-271. Los Alamitos, CA: IEEE Computer Society, 2001.

[Porter and Duff 84] T. Porter and T. Duff. "Compositing Digital Images." In Proceedings of ACM SIGGRAPH 84: Computer Graphics 18 (1984), 253-259.

[Smith and Blinn 96] A. R. Smith and J. F. Blinn. "Blue screen Matting." In Proceedings of SIGGRAPH 96, Computer Graphics Proceedings, Annual Conference Series, edited by Holly Rushmeier, pp. 259-268. Reading, MA: Addison Wesley, 1996.

[Sun et al. 04] Jian Sun, Jiaya Jia, Chi-Keung Tang, and Heung-Yeung Shum. "Poisson Matting." Proceedings of SIGGRAPH 2004: ACM Transactions on Graphics 23:3 (2004), 315-321.

Image Manipulation

[Jia et al. 06] Jiaya Jia, Jian Sun, Chi-Keung Tang, and Heung-Yeung Shum. "Drag-and-Drop Pasting." In ACM SIGGRAPH 2006 Papers, pp. 631-637. New York: ACM Press, 2006.

[Bae et al. 06] Soonmin Bae, Sylvain Paris, and Fredo Durand. "Two-Scale Tone Management for Photographic Look." In ACM SIGGRAPH 2006 Papers, pp. 637-645. New York: ACM Press, 2006.

[Lischinski et al. 06] Dani Lischinski, Zeev Farbman, Matt Uyttendaele, and Richard Szeliski. "Interactive Local Adjustment of Tonal Values." In ACM SIGGRAPH 2006 Papers, pp. 646-653. New York: ACM Press, 2006.

[Khan et al. 06] Erum Arif Khan, Erik Reinhard, Roland W. Fleming, and Heinrich H. Bulthoff. "Image-Based Material Editing." In ACM SIGGRAPH 2006 Papers, pp. 654-663. New York: ACM Press, 2006.

[Oliva et al. 06] Aude Oliva, Antonio Torralba, and Philippe G. Schyns. "Hybrid Images." In ACM SIGGRAPH 2006 Papers, pp. 527-532. New York: ACM Press, 2006.

Techniques

General

[Danielsson 80] P.-E. Danielsson. "Euclidean Distance Mapping." Computer Graphics and Image Processing 14 (1980), 227-248.

[Lucas and Kanade 81] B. D. Lucas and T. Kanade. "An Iterative Image Registration Technique with an Application to Stereo Vision. In Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI '81), pp. 674-679. San Francisco: Morgan Kaufmann, 1981.

[Mortensen and Barrett 95] E. N. Mortensen and W. A. Barrett. "Intelligent Scissors for Image Composition." In Proceedings of SIGGRAPH 95, Computer Graphics Proceedings, Annual Conference Series, edited by Robert Cook, pp. 191-198. Reading, MA: Addison Wesley, 1995.

Graph Cuts

[Boykov et al. 01] Y. Boykov, O. Veksler, and R. Zabih. "Fast Approximate Energy Minimization via Graph Cuts." IEEE Transactions on Pattern Analysis and Machine Intelligence 23:11 (2001), 1222-1239.

[Kwatra et al. 03] V. Kwatra, A. Schodl, I. Essa, G. Turk, and A. Bobick. "Graphcut Textures: Image and Video Synthesis Using Graph Cuts." ACM Transactions on Graphics 22:3 (2003), 277-286.

[Shi and Malik 97] J. Shi and J. Malik. "Normalized Cuts and Image Segmentation." In Proceedings of the 1997 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 731-737. Los Alamitos, CA: IEEE Computer Society, 1997.

Gradient Domain

[Perez et al. 03] P. Perez, M. Gangnet, and A. Blake. "Poisson Image Editing." ACM Transactions on Graphics 22:3 (2003), 313-318.

[Georgiev 06] T. Georgiev. "Covariant Derivatives and Vision." In Computer Vision-ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006, Proceedings, Part IV, Lecture Notes in Computer Science 3954, pp.56-69. Berlin: Springer, 2006.

Smoothing, Bilateral, and Trilateral Filters

[Tomasi and Manduchi 98] C. Tomasi and R. Manduchi. "Bilateral Filtering of Gray and Colored Images." In Proceedings of the Sixth IEEE International Conference on Computer Vision, pp. 836-846. Los Alamitos, CA: IEEE Computer Society, 1998.

[Choudhury and Tumblin 03] P. Choudhury and J. Tumblin. "The Trilateral Filter for High Contrast Images and Meshes." In Proceedings of the 14th Eurographics Workshop on Rendering, edited by Per. H. Christensen and Daniel Cohen, pp. 186-196. Aire-la-Ville, Switzerland: Eurographics Association, 2003.

Video Textures

[Schodl et al. 00] A. Schodl, R. Szeliski, D. H. Salesin, and I. Essa. "Video Textures." In Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques, pp. 489-498. New York: ACM Press/Addison-Wesley Publishing Co., 2000.

[Agarwala et al. 05] A. Agarwala, K. C. Zheng, C. Pal, M. Agrawala, M. Cohen, B. Curless, D. Salesin, and R. Szeliski. "Panoramic Video Textures." ACM Transaction on Graphics 24:3 (2005), 821-827.

Feature Extraction and Scene Understanding

Shape/Material/Illumination and Surface Normals

[Basri and Jacobs 01] R. Basri and D. Jacobs. "Photometric Stereo with General, Unknown Lighting." In Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. II, pp. 374-381. Los Alamitos, CA: IEEE Computer Society, 2001.

[Horn 70] B. K. P. Horn. "Shape from Shading: A Method for Obtaining the Shape of a Smooth Opaque Object from One View." MIT Project MAC Internal Report TR-79 and MIT AI Lab Technical Report 232, 1970.

[Zickler et al. 02] Todd Zickler, Peter N. Belhumeur, and David J. Kriegman, "Helmholtz Stereopsis: Exploiting Reciprocity for Surface Reconstruction." In Computer Vision-ECCV 2002: 7th European Conference on Computer Vision, Copenhagen, Denmark, May 28-31, 2002, Proceedings, Part III, Lecture Notes in Computer Science 2352, Part III, pp 869-884. Berlin, Springer, 2002.

[Wenger et al. 05] Andreas Wenger, Andrew Gardner, Chris Tchou, Jonas Unger, Tim Hawkins, and Paul. Debevec. "Performance Relighting and Reflectance Transformation with Time-Multiplexed Illumination." In ACM SIGGRAPH 2005 Papers, pp. 756-764. New York: ACM Press, 2005.

Depth Edges

[Raskar et al. 04] Ramesh Raskar, Karhan Tan, Rogerio Feris, Jingyi Yu, and Matthew Turk. "Non-photorealistic Camera: Depth Edge Detection and Stylized Rendering Using Multi-flash Imaging." In ACM SIGGRAPH 2004 Papers, pp. 679-688. New York, ACM Press, 2004.

Depth

[Nayar and Nakagawa 94] S. K. Nayar and Y. Nakagawa. "Shape from Focus." IEEE Transactions on Pattern Analysis and Machine Intelligence 16:8 (1994), 824-831.

Motion

[Ben-Ezra and Nayar 04] M. Ben-Ezra and S. K. Nayar, "Motion-Based Motion Deblurring." IEEE Transactions on Pattern Analysis and Machine Intelligence 26:6 (2004), 689-698.

[Raskar et al. 06] R. Raskar, A. Agrawal, and J. Tumblin. "Coded Exposure Photography: Motion Deblurring Using Fluttered Shutter." ACM Transactions on Graphics 25:3 (2006), 795-804.

[Fergus et al. 06] Rob Fergus, Barun Singh, Aaron Hertzmann, Sam T. Roweis, and William T. Freeman. "Removing Camera Shake from a Single Photograph." Proceedings of ACM SIGGRAPH 2006: ACM Transactions on Graphics 25:3 (2006), 787-794.

[Liu et al. 05] C. Liu, A. Torralba, W. T. Freeman, F. Durand, and E. H. Adelson. "Motion Magnification." ACM Transactions on Graphics 24:3 (2005), 519-526.

Decomposition of Scene Parameters

[Nayar et al. 06] 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 ACM SIGGRAPH 2006 Papers, pp. 935-944. New York: ACM Press, 2006.

Transfer and Denoising

Flash to No-Flash

[Eisemann and Durand 04] Elmar Eisemann and Fredo Durand. "Flash Photography Enhancement via Intrinsic Relighting." In ACM SIGGRAPH 2004 Papers, pp. 673-678. New York: ACM Press, 2004.

[Petschnigg et al. 04] Georg Petschnigg, Richard Szeliski, Maneesh Agrawala, Michael Cohen, Hugues Hoppe, and Kentaro Toyama. "Digital Photography with Flash and No-Flash Image Pairs." Proceedings of SIGGRAPH 2004: ACM Transactions on Graphics 23:3 (2004), 664-672.

[DiCarlo et al. 01] J. M. DiCarlo, F. Xiao, and B. A. Wandell. "Illuminating Illumination." In Proceedings of the Ninth Color Imaging Conference, pp. 27-34. Springfield, VA: The Society of Imaging Science and Technology, 2001.

Noise

[Meer et al. 90] P. Meer, J. Jolion, and A. Rosenfeld. "A Fast Parallel Algorithm for Blind Estimation of Noise Variance." IEEE Transactions on Pattern Analysis and Machine Intelligence, 12:2 (1990), 216-223.

[Bennett and McMillan 05] Eric P. Bennett and Leonard McMillan. "Video Enhancement Using Per-Pixel Virtual Exposures." In ACM SIGGRAPH 2005 Papers, pp. 845-852. New York: ACM Press, 2005.

Geometric Operations

Panorama

[Davis 98] J. Davis. "Mosaics of Scenes with Moving Objects." In Proceedings of the 1998 Conference on Computer Vision and Pattern Recognition, pp. 354-360. Los Alamitos, CA: IEEE Computer Society, 1998.

[Uyttendaele et al. 01] M. Uyttendaele, A. Eden, and R. Szeliski. "Eliminating Ghosting and Exposure Artifacts in Image Mosaics." In Proceedings of the 2001 Conference on Computer Vision and Pattern Recognition, Vol. II, pp. 509-516. Los Alamitos, CA: IEEE Computer Society, 2001.

[Szeliski and Shum 97] R. Szeliski and H.-Y. Shum. "Creating Full View Panoramic Mosaics and Environment Maps." In Proceedings of SIGGRAPH 97, Computer Graphics Proceedings, Annual Conference Series, edited by Turner Whitted, pp. 251-258. Reading, MA: Addison Wesley, 1997.

Synthetic Aperture and Multi-views

[Adelson and Wang 92] E. H. Adelson and J. Y.A. Wang. "Single Lens Stereo with a Plenoptic Camera." IEEE Transactions on Pattern Analysis and Machine Intelligence 14:2 (1992), 99-106.

[Levoy and Hanrahan 96] Marc Levoy and Pat Hanrahan. "Light Field Rendering." In Proceedings of SIGGRAPH 96, Computer Graphics Proceedings, Annual Conference Series, edited by Holly Rushmeier, pp. 31-42. Reading, MA: Addison Wesley, 1996.

[Gortler et al. 96] Steven J. Gortler, Radek Grzeszczuk, Richard Szeliski, and Michael F. Cohen. "The Lumigraph." In Proceedings of SIGGRAPH 96, Computer Graphics Proceedings, Annual Conference Series, edited by Holly Rushmeier, pp. 43-54. Reading, MA: Addison Wesley, 1996.

[Levoy et al. 06] Marc Levoy, Ren Ng, Andrew Adams, Matthew Footer, and Mark Horowitz. "Light Field Microscopy." In ACM SIGGRAPH 2006 Papers, pp.924-934. New York: ACM Press, 2006.

[Levoy et al. 04] Marc Levoy, Billy Chen, Vaibhav Vaish, Mark Horowitz, Ian McDowall, and Mark Bolas. "Synthetic Aperture Confocal Imaging." In ACM SIGGRAPH 2004 Papers, pp. 825-834. New York: ACM Press, 2004.

[Hg 05] Ren Ng. "Fourier Slice Photography." In ACM SIGGRAPH 2005 Papers, pp. 735-744. New York: ACM Press, 2005.

[Georgiev et al. 06] Todor Georgiev, Ke Colin Zheng, Brian Curless, David Salasin, Shree Nayar, and Chintan Intwala. "Spatio-Angular Resolution Tradeoff in Integral Photography." In Rendering Techniques 2006: Eurographics Symposium on Rendering, edited by T. Akenine-Moller and W. Heidrich, pp. 263-272. Aire-la-Ville, Switzerland: Eurographics Association, 2006.

[Stern and Javidi 03] A. Stern and B. Javidi. "3-D Computational Synthetic Aperture Integral Imaging (COMPSAII)." Optics Express 11 (2003), 2446-2451.

[Oliver and Quegan 98] C. Oliver and S. Quegan. Understanding Synthetic Aperture Radar Images. London: Artech House, 1998.

[Vaish et al. 05] 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 Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recongnition, Vol. 3, pp. 129. Los Alamitos, CA: IEEE Computer Society, 2005.

[Joshi et al. 06] Neel Joshi, Wojciech Matusik, and Shai Avidan. "Natural Video Matting Using Camera Arrays." In ACM SIGGRAPH 2006 Papers, pp. 779-786. New York: ACM Press, 2006.

[Kuthirummal and Nayar 06] Sujit Kuthirummal and Shree K. Nayar. "Multiview Radial Catadioptric Imaging for Scene Capture." In ACM SIGGRAPH 2006 Papers, pp. 916-923. New York: ACM Press, 2006.

Deblurring and Superresolution

[Ben-Ezra and Nayar 03] M. Ben-Ezra and S. K. Nayar. "Motion Deblurring Using Hybrid Imaging." In Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. I, pp. 657-664. Los Alamitos, CA: IEEE Computer Society, 2003.

[Lin and Shum 04] 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 (2004), 83-89.

[Landolt et al. 01] O. Landolt, A. Mitros, and C. Koch. "Visual Sensor with Resolution Enhancement by Mechanical Vibrations." In Proceedings of the

2001 Conference on Advanced Research in VLSI, pp. 249-264. Los Alamitos, CA: IEEE Computer Society, 2001.

[Zhang and Nayar 06] Li Zhang and Shree Nayar. "Projection Defocus Analysis for Scene Capture and Image Display." In ACM SIGGRAPH 2006 Papers, pp. 907-915. New York: ACM Press, 2006.

Dynamic Range

[Morimura 93] A. Morimura. "Imaging Method for a Wide Dynamic Range and an Imaging Device for a Wide Dynamic Range." U.S. Patent no. 5,455,621, 1993.

[Burt and Kolczynski 93] P. Burt and R. J. Kolczynski. "Enhanced Image Capture Through Fusion." In Proceedings of the Fourth International Conference on Computer Vision, pp. 173-182, 1993. Los Alamitos, CA: IEEE Press, 1993.

[Madden 93] B. Madden. "Extended Intensity Range Imaging." Technical Report MS-CIS-93-96, Grasp Laboratory, University of Pennsylvania, 1993.

[Tsai 94] Y. T. Tsai. "Method and Apparatus for Extending the Dynamic Range of an Electronic Imaging System." U.S. Patent no. 5,309,243, 1994.

[Mitsunaga and Nayar 99] T. Mitsunaga and S. K. Nayar. "Radiometric Self Calibration." In Proceedings of the 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 2, pp. 374-380. Los Alamitos, CA: IEEE Computer Society, 1999.

[Street 98] R. A. Street. "High Dynamic Range Segmented Pixel Sensor Array." U.S. Patent no. 5,789,737, 1998.

[Handy 86] R. J. Handy. "High Dynamic Range CCD Detector Imager." U.S. Patent no. 4,623,928, 1986.

[Wen 89] D. D. Wen. "High Dynamic Range Charge Coupled Device." U.S. Patent no. 4,873,561, 1989.

[Hamazaki 96] M. Hamazaki. "Driving Method for Solid-State Image Pickup Device." Japanese Patent no. 08- 331461, 1996.

[Knight 83] T. Knight. "Design of an Integrated Optical Sensor with On-Chip Pre-processing." Ph.D. diss., Massachusetts Institute of Technology, 1983.

[Sayag 91] M. Sayag, "Non-linear Photosite Response in CCD Imagers." U.S Patent no. 5,055,667, 1991.

[Decker et al. 98] S. J. Decker, R. D. McGrath, K. Brehmer, and C. G. Sodini, “A 256 ? 256 CMOS Imaging Array with Wide Dynamic Range Pixels and Column-Parallel Digital Output.” IEEE Journal of Solid-State Circuits 33 (1998), 2081-2091.

[Brajovic and Kanade 96] V. Brajovic and T. Kanade. “A Sorting Image Sensor: An Example of Massively Parallel Intensity-to-Time Processing for Low-Latency Computational Sensors.” In Proceedings of the 1996 IEEE International Conference on Robotics and Automation, Vol. 2, pp. 1638-1643. Los Alamitos, CA: IEEE Press, 1996.

[Scheffer et al. 00] D. Scheffer S. Kavadias, B. Dierickx, A. Alaerts, D. Uwaerts, and J. Bogaerts. “A Logarithmic Response CMOS Image Sensor with On-Chip Calibration.” IEEE Journal of Solid-State Circuits 35:8 (2000), 1146-1152.

[Shree and Mitsunaga 00] Shree K. Nayar and Tomoo Mitsunaga. “High Dynamic Range Imaging: Spatially Varying Pixel Exposures.” In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1472-1479. Los Alamitos, CA: IEEE Press, 2000:

Active Illumination

[Adelson and Bergen 91] Edward H. Adelson and James R. Bergen. “The Plenoptic Function and the Elements of Early Vision.” In Computational Models of Visual Processing, edited by M. Landy and J. A. Movshon, pp. 3-20. Cambridge, MA: MIT Press, 1991.

[Fuchs et al. 05] Martin Fuchs, Volker Blanz, and Hans-Peter Seidel. “Bayesian Relighting.” In Proceedings of the Eurographics Symposium on Rendering Techniques, pp. 157-164. Aire-la-Ville, Switzerland: Eurographics Association 2005.

[Haeberli 08] Paul Haeberli. Graphics Obscura. Available at http://www.graficaobscura.com/, 2008.

[Hawkins et al. 01] Tim Hawkins, Jonathan Cohen, and Paul Debevec. “A Photometric Approach to Digitizing Cultural Artifacts.” In Proceedings of the 2001 Conference on Virtual Reality, Archeology, and Cultural Heritage, pp. 333-342. New York: ACM Press, 2001.

[Hawkins et al. 05] Tim Hawkins, Per Einarsson, and Paul E. Debevec. “A Dual Light Stage.” In Proceedings of the Eurographics Symposium on Rendering Techniques, pp. 91-98. Aire-la-Ville, Switzerland: Eurographics Association, 2005.

[Hoppe and Toyama 03] Hugues Hoppe and Kentaro Toyama. "Continuous Flash." Technical Report 63, Microsoft Research, 2003.

[Jones et al. 05] Andrew Jones, Andrew Gardner, Mark Bolas, Ian McDowall, and Paul Debevec. "Performance Geometry Capture for Spatially Varying Relighting." In ACM SIGGRAPH 2005 Sketches, Article no. 74. New York: ACM Press, 2005.

[Xiao et al. 01] Feng Xiao, Jeffrey M. DiCarlo, and Brian A Wandell. "Illuminating Illumination." In The 9th Color Imaging Conference, pp. 27-34. Springfield, VA: The Society for Imaging Science and Technology, 2001.

[Koudelka et al. 01] Melissa L Koudelka, Peter N Belhumeur, Sebastian Magda, and David J. Kriegman. "Image- Based Modeling and Rendering of Surfaces with Arbitrary BRDFs." In Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. I, pp. 568-575. Los Alamitos, CA: IEEE Computer Society, 2001.

[Malzbender et al. 01] Tom Malzbender, Dan Gelb, and Hans Wolters. "Polynomial Texture Maps." In Proceedings of SIGGRAPH 2001, Computer Graphics Proceedings, Annual Conference Series, edited by E. Fiume, pp. 519-528. New York: ACM Press, 2001.

[Masselus 03] Vincent Masselus, Pieter Peers, Philip Dutre, and Yves D. Willems. "Relighting with 4D Incident Light Fields." In SIGGRAPH 2003: ACM Transactions on Graphics 22:3 (2003), 613-620.

[Mohan et al. 05] Ankit Mohan, Jack Tumblin, Bobby Bodenheimer, Cindy Grimm, and Reynold J. Bailey. "Table- Top Computed Lighting for Practical Digital Photography." In Proceedings of the Eurographics Symposium on Rendering Techniques, pp. 165-172. Aire-la-Ville, Switzerland: Eurographics Association, 2005.

[Nimeroff et al. 94] Jeffry S. Nimeroff, Eero Simoncelli, and Julie Dorsey. "Efficient Re-rendering of Naturally Illuminated Environments." In Proceedings of the Fifth Eurographics Workshop on Rendering, pp. 359-373. Aire- la-Ville, Switzerland: Eurographics Association, 1994.

[Wenger et al. 05] Andreas Wenger, Andrew Gardner, Chris Tchou, Jonas Unger, Tim Hawkins, and Paul Debevec. "Performance Relighting and Reflectance Transformation with Time-Multiplexed Illumination." In ACM SIGGRAPH 2005 Papers, pp. 756-764. New York: ACM Press, 2005.

[Winnemoeller et al. 05] Holger Winnemoeller, Ankit Mohan, Jack Tumblin, and Bruce Gooch. "Light Waving: Estimating Light Positions from Photographs Alone." Computer Graphics Forum 24:3 (2005), 433-438.

[Sun et al. 06] Jian Sun, Yin Li, Sing Bing Kang, and Heung-Yeung Shum. "Flash Matting." In ACM SIGGRAPH 2006 Papers, pp. 772-778. New York: ACM Press, 2006.

Smart, Unconventional Cameras

MEMS Technology

[Nayar et al. 04] S. K. Nayar, V. Branzoi, and T. Boult. "Programmable Imaging Using a Digital Micromirror Array." In Proceedings of 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. I, pp. 436-443. Los Alamitos, CA: IEEE Computer Society, 2004.

High Speed Imaging

[Wandell et al. 99] B. Wandell, P. Catrysse, J. DiCarlo, D. Yang, and A. El Gamal. "Multiple Capture Single Image Architecture with a CMOS Sensor." In Proceedings of the International Symposium on Multispectral Imaging and Color Reproduction for Digital Archives, pp. 11-17. Chiba, Japan: Society of Multispectral Imaging of Japan, 1999.

[Kleinfelder et al. 01] S. Kleinfelder, S. H. Lim, X. Q. Liu, and A. El Gamal. "A 10,000 Frames/s CMOS Digital Pixel Sensor." IEEE Journal of Solid-State Circuits 36:12 (2001), 2049-2059.

[Liu and Gamal 03] X. Q. Liu and Abbas El Gamal. "Synthesis of High Dynamic Range Motion Blur Free Image from Multiple Captures." IEEE Transactions on Circuits and Systems 50:4 (2003), 530-539.

[Raskar et al. 06] Ramesh Raskar, Amit AGrawal, and Jack Tumblin. "Coded Exposure Photography: Motion Deblurring Using Fluttered Shutter." In ACM SIGGRAPH 2006 Papers, pp. 795-804. New York: ACM Press, 2006.

[Shechtman et al. 02] E. Shechtman, Y. Caspi, and M. Irani. "Increasing Space-Time Resolution in Video." In Computer Vision-ECCV 2002: 7th European Conference on Computer Vision, Copenhagen, Denmark, May 28-31, 2002, Proceedings, Part I, Lecture Notes in Computer Science 2350, pp. 753-768. Berlin: Springer-Verlag, 2002.

[Johansson et al. 03] R. Johansson, L. Lindgren, J. Melander, and B. Moller. "A Multi-resolution 100-GOPS 4-Gpixels/s Programmable CMOS Image Sensor for Machine Vision." Paper presented at the 2003 IEEE Workshop on Charge-Coupled Devices and Advanced Image Sensors, 2003.

Advanced, Programmable, Demodulating Cameras, and Temporal Correlation

[Xiao et al. 01] F. Xiao, J. DiCarlo, P. Catrysse and B. Wandell, Image Analysis using Modulated Light Sources, In Proceedings of the SPIE Electronic Imaging '2001 conference, Vol. 4306, San Jose, CA, January 2001.

[Ando et al. 97] Ando, S., K. Nakamura, and T. Sakaguchi. Ultrafast Correlation Image Sensor: Concept, Design, and Applications,. in Proc. IEEE CCD/AIS Workshop. 1997. Bruges, Belgium: IEEE.

[Ando and Kimachi 99] Ando, S. and A. Kimachi. Time-Domain Correlation Image Sensor: First CMOS Realization of Demodulator Pixels Array. in Proc. '99 IEEE CCD/AIS Workshop. 1999.

Others

[Cohen 05] Michael Cohen. "Capturing the Moment", Symposium on Computational Photography and Video, Boston in May, 2005.

[Tumblin et al. 05] TUMBLIN, J., AGRAWAL, A. AND RASKAR, R. Why I want a Gradient Camera, IEEE CVPR 2005

Optics

[Stam 99] Jos Stam, "Diffraction Shaders", In SIGGRAPH 99 Conference Proceedings, Annual Conference Series, August 1999, 101-110.

[Elber 94] Elber, G. 1994. Low cost illumination computation using an approximation of light wavefronts. In Proceedings of the 21st Annual Conference on Computer Graphics and interactive Techniques SIGGRAPH '94. ACM Press, New York, NY, 335-342.

[Rosenfeld and Kak 87] Rosenfeld A. and Kak A.C., Digital Image Processing, New York; Academic Press, Cap 11, 1987.

[Goodman 68] J.W. Goodman, Introduction to Fourier Optics, McGraw-Hill Book Co.,New York, N.Y., (1968).

[Dowski and Cathey 95] Dowski, E.R.,Jr., Cathey, W.T., Extended depth of field through wave-front coding, Applied Optics, Vol. 34, No. 11, 10 April 1995, pp. 1859-1866.

[Zomet and Nayar 06] A. Zomet and S.K. Nayar, "Lensless Imaging with a Controllable Aperture," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jun, 2006.

[Zand 96] Zand, J., Coded aperture imaging in high energy astronomy, NASA Laboratory for High Energy Astrophysics (LHEA) at NASA's GSFC, 1996.

Light Field Optics

[Hecht and Zajac 79] E. Hecht and A. Zajac. Optics. Addison-Wesley, 1979.

[Gerrard and Burch 94] A. Gerrard and J. M. Burch. Introduction to Matrix Methods in Optics. Dover Publications, 1994.

[Guillemin and Sternberg] V. Guillemin and S. Sternberg. Symplectic techniques in physics. Cambridge University Press.

[Durand et al. 05] Fredo Durand, Nicolas Holzschuch, Cyril Soler, Eric Chan & Francois Sillion. A Frequency analysis of Light Transport. Proceedings of SIGGRAPH 2005.

[Georgiev and Intwala 06] Todor Georgiev, Chintan Intwala. Integral View Photography. Adobe Tech Report 2006:

[Georgiev et al. 06] Todor Georgiev, Colin Zheng, Shree Nayar, David Salesin, Brian Curless, Chintan Intwala. Spatio-angular Resolution Trade-Offs in Integral Photography. In proceedings, EGSR 2006. http://www.tgeorgiev.net/Spatioangular.pdf

Fourier Optics

[Goodman 68] J. Goodman, Introduction to Fourier Optics, McGraw-Hill, 1968.

[Born 64] M. Born, E. Wolf, Principles of Optics, Pergamon Press, 1964.