Computational Photography

Ramesh Raskar, Jack Tumblin

MERL and Northwestern University, USA

Abstract

Computational photography combines plentiful computing, digital sensors, modern optics, actuators, probes and smart lights to escape the limitations of traditional film cameras and enables novel imaging applications. Unbounded dynamic range, variable focus, resolution, and depth of field, hints about shape, reflectance, and lighting, and new interactive forms of photos that are partly snapshots and partly videos are just some of the new applications found in Computational Photography. The computational techniques encompass methods from modification of imaging parameters during capture to sophisticated reconstructions from indirect measurements. We provide a practical guide to topics in image capture and manipulation methods for generating compelling pictures for computer graphics and for extracting scene properties for computer vision, with several examples.

Many ideas in computational photography are still relatively new to digital artists and programmers and there is no up-to-date reference text. A larger problem is that a multi-disciplinary field that combines ideas from computational methods and modern digital photography involves a steep learning curve. For example, photographers are not always familiar with advanced algorithms now emerging to capture high dynamic range images, but image processing researchers face difficulty in understanding the capture and noise issues in digital cameras. These topics, however, can be easily learned without extensive background. The goal of this article is to present both aspects in a compact form.

The new capture methods include sophisticated sensors, electromechanical actuators and on-board processing. Examples include adaptation to sensed scene depth and illumination, taking multiple pictures by varying camera parameters or actively modifying the flash illumination parameters. A class of modern reconstruction methods is also emerging. The methods can achieve a 'photomontage' by optimally fusing information from multiple images, improve signal to noise ratio and extract scene features such as depth edges.

The participants learn about topics in image capture and manipulation methods for generating compelling pictures for computer graphics and for extracting scene properties for computer vision, with several examples. We hope to provide enough fundamentals to satisfy the technical specialist without intimidating the curious graphics researcher interested in recent advances in photography.

The intended audience is photographers, digital artists, image processing programmers and vision researchers using or building applications for digital cameras or images. They will learn about camera fundamentals and powerful computational tools, along with many real world examples.

The article briefly examines the topics commonly appearing at academic conferences so that the reader is familiar with the general problems in Computational Photography. Our focus is on the review of problems rather than an exhaustive review of the emerging solutions. We refer the reader to the Course Notes for Siggraph 2007 course available on our website for a comprehensive discussion.

1   Introduction

1.1 Film-like Photography

Photography is the process of making pictures by, literally, 'drawing with light' or recording the visually meaningful changes in the light leaving a scene. This goal was established for film photography about 150 years ago.

Currently, 'digital photography' is electronically implemented film photography, refined and polished to achieve the goals of the classic film camera which were governed by chemistry, optics, mechanical shutters. 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 labelling.

In this article we plan to explore a progression away from film and film-like methods to something more comprehensive that exploits plentiful low-cost computing and memory with sensors, optics, probes, smart lighting and communication.

1.2 What is Computational Photography?

Computational Photography (CP) is an emerging field, just getting started. We don't know where it will end up, we can't yet set its precise, complete definition, nor make a reliably comprehensive classification. But here is the scope of what researchers are currently exploring in this field.

- Computational photography attempts to record a richer visual experience, captures information beyond just a simple set of pixels and makes the recorded scene representation far more machine readable.

- It exploits computing, memory, interaction and communications to overcome long-standing limitations of photographic film and camera mechanics that have persisted in film-style digital photography, such as constraints on dynamic range, depth of field, field of view, resolution and the extent of scene motion during exposure.

- It enables new classes of recording the visual signal such as the 'moment' [Cohen 2005], shape boundaries for non-photorealistic depiction [Raskar et al 2004] , 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 4-D set of viewing rays.

- It enables synthesis of impossible photos that could not have been captured at a single instant with a single camera, such as wrap-around views ('multiple-center-of-projection' images [Rademacher and Bishop 1998]), fusion of time-lapsed events [Raskar et al 2004], the motion-microscope (motion magnification [Liu et al 2005]), video textures and panoramas [Agarwala et al 2005]. They also support seemly impossible camera movements such as the 'bullet time' (Matrix) sequence recorded with multiple cameras with staggered exposure times.

- It encompass previously exotic forms of scientific imaging and data gathering techniques e.g. from astronomy, microscopy, and tomography.
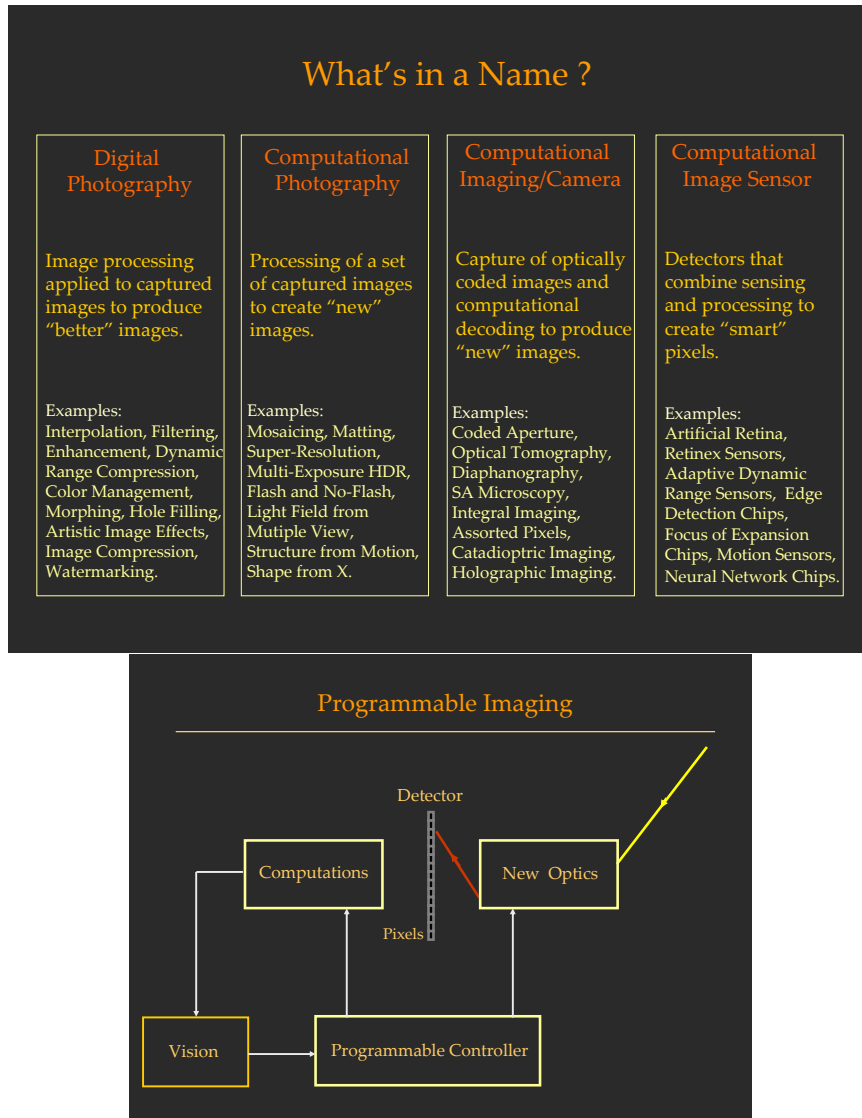
What's in a Name ?

Digital Photography

Image processing applied to captured images to produce "better" images.

Examples:
Interpolation, Filtering, Enhancement, Dynamic Range Compression, Color Management, Morphing, Hole Filling, Artistic Image Effects, Image Compression, Watermarking.

Computational Photography

Processing of a set of captured images to create "new" images.

Examples:
Mosaicing, Matting, Super-Resolution, Multi-Exposure HDR, Flash and No-Flash, Light Field from Mutiple View, Structure from Motion, Shape from X.

Computational Imaging/Camera

Capture of optically coded images and computational decoding to produce "new" images.

Examples:
Coded Aperture, Optical Tomography, Diaphanography, SA Microscopy, Integral Imaging, Assorted Pixels, Catadioptric Imaging, Holographic Imaging.

Computational Image Sensor

Detectors that combine sensing and processing to create "smart" pixels.

Examples:
Artificial Retina, Retinex Sensors, Adaptive Dynamic Range Sensors, Edge Detection Chips, Focus of Expansion Chips, Motion Sensors, Neural Network Chips.

Programmable Imaging

Detector

Computations

New Optics

Pixels

Vision

Programmable Controller

Figure 1 Shree Nayar's views on Computational Photography (Permission: nayar@cs.columbia.edu)

1.3 Elements of Computational Photography

Traditional film-like photography involves (a) a lens, (b) a 2D planar sensor and (c) a processor that converts sensed values into an image. In addition, the photography may involve (d) external illumination from point sources (e.g. flash units) and area sources (e.g. studio lights).

We like to categorize and generalize Computational Photography into these four elements. Our categorization is influenced by Shree Nayar's presentation during a symposium in May 2005. 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 for a given wavelength is transformed into a new 4D lightfield. The optics may involve more than one optical axis [Georgiev et al 2006]. In some cases the perspective foreshortening of objects based on distance may be modified using wavefront coded optics [Dowski and Cathey 1995]. In recent lensless imaging methods [Zomet and Nayar 2006] and coded-aperture imaging [Zand 1996] used for gamma-ray and X-ray astronomy, the traditional lens is missing entirely. In some

cases optical elements such as mirrors [Nayar et al 2004] outside the camera adjust the linear combinations of ray bundles that reach the sensor pixel to adapt the sensor to the viewed scene.

(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 lightfield. Computational photography attempts to capture more; a 3D or 4D ray representation using planar, non-planar or even volumentric 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. But a Plenoptic Camera [Adelson and Wang 1992, Ren et al 2005] subdivides these bundles into separate measurements. Computing a weighted sum of rays that converge on the objects in the scene creates a digitally refocused image, and even permits multiple focusing distances within a single computed image. Generalizing sensors can extend their dynamic range [Tumblin et al 2005] and wavelength selectivity as well. 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 2004].

(c) Generalized Reconstruction: Conversion of raw sensor outputs into picture values can be much more sophisticated. While existing digital cameras perform 'de-mosaicking,' (interpolate the Bayer grid), remove fixed-pattern noise, and hide 'dead' pixel sensors, recent work in computational photography can do more. 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 scene from multiple photographs from coaxial lenses, from sensed gradients [Tumblin et al 2005], or compute sharp images a fast moving object from a single image taken by a camera with a 'fluttering' shutter [Raskar et al 2006]. Closed-loop control during photography itself can also be extended, exploiting traditional cameras' exposure control, image stabilizing, and focus, as new opportunities for modulating the scene's optical signal for later decoding.

(d) Computational Illumination: Photographic lighting has changed very little since the 1950's: with digital video projectors, servos, and device-to-device communication, we have new opportunities to control the sources of light with as much sophistication as we use to control our digital sensors. What sorts of spatio-temporal modulations for light might better reveal the visually important contents of a scene? Harold Edgerton showed high-speed strobes offered tremendous new appearance-capturing capabilities; how many new advantages can we realize by replacing 'dumb' the flash units, static spot lights and reflectors with actively controlled spatio-temporal modulators and optics? Already we can capture occluding edges with multiple flashes [Raskar 2004], exchange cameras and projectors by Helmholz reciprocity [Sen et al 2005], gather relightable actor's performances with light stages [Wagner et al 2005] and see through muddy water with coded-mask illumination [Levoy et al 2004]. In every case, better lighting control during capture to builds richer representations of photographed scenes.
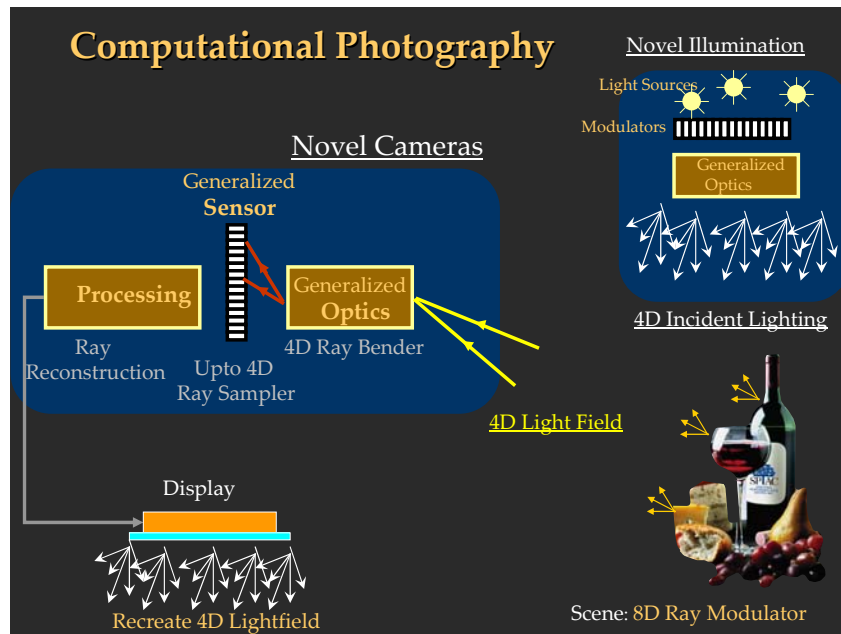
Figure 2 Our categorization of the elements of Computational Photography

2   Sampling Dimensions of Imaging

2.1 Epsilon Photography for Optimizing Film-like Camera

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. This effort reduces Computational Photography to 'Epsilon Photography', where the scene is recorded via multiple images, each captured by epsilon variation of the camera parameters. For example, successive images (or neighboring pixels) may have different settings for parameters such as exposure, focus, aperture, view, illumination, or the instant of capture. Each setting allows recording of partial information about the scene and the final image is reconstructed from these multiple observations. Epsilon photography is thus concatenation of many such boxes in parameter space; 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 on its own, without any of the others. The merged photo contains the best features from all of them.

(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 [Debevec and Malik 1997, Kang et al 2003]

(c) Depth of field: All-in-focus image is reconstructed from images taken by successively changing the plane of focus [Agrawala et al 2005].

(d) Spatial Resolution: Higher resolution is achieved by tiling multiple cameras (and mosaicing individual images) [Wilburn et al 2005] or by jittering a single camera [Landolt et al 2001].

(e) Wavelength resolution: Traditional cameras sample only 3 basis colors. But multi-spectral (multiple colors in the visible spectrum) or hyper-spectral (wavelengths beyond the visible spectrum) imaging is accomplished by taking pictures while successively changing color filters in front of the camera, using tunable wavelength filters or using diffraction gratings.

(f) Temporal resolution: High speed imaging is achieved by staggering the exposure time of multiple low-framerate cameras. The exposure durations of individual cameras can be non-overlapping) [Wilburn et al 2005] or overlaping [Shechtman et al 2002].

Taking multiple images under varying camera parameters can be achieved in several ways. The images can be taken with a single camera over time. The images can be captured simultaneously using 'assorted pixels' where each pixel is a tuned to a different value for a given parameter [Nayar and Narsimhan 2002]. 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 used for multi-camera solutions: a camera array [Wilburn et al 2005] and single-axis multiple parameter (co-axial) cameras [Mcguire et al 2005].

2.2 Coded Photography

But there is much more beyond the 'best possible film camera'. Instead of increasing the field of view by panning a camera, can we create a wrap-around view of an object ? Panning a camera allows us to concatenate and expand the 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 dimensions. We can virtualize the notion of the camera itself if we consider it as a device that collects bundles of rays, 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 are not comprehensible as 'images' without further decoding, re-binning or reconstruction. For example, a wrap around view is built from images taken with multiple centers of projection but by taking only a few pixels from each input image. Coded aperture techniques, inspired by work in astronomical imaging, try to preserve high spatial frequencies so that out of focus blurred images can be digitally re-focused [Veeraraghavan07]. By coding illumination, it is possible to decompose radiance in a scene into direct and global components [Nayar06]. Using a coded exposure technique, one can rapidly flutter open and close the shutter of a camera in a carefully chosen binary sequence, to capture a single photo. The fluttered shutter encoded the motion in the scene in the observed blur in a reversible way. Other examples include confocal images and techniques to recover glare in the images [Talvala07].

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

In the rest of the article, we survey recent techniques that exploit exposure, focus and active illumination.

3   High Dynamic Range

A camera sensor is limited in the range of highest and lowest intensities it can measure. To capture the high dynamic range, one can adaptively exposure the sensor so that the signal to noise ratio is high over the entire image, including in the the dark and brightly lit regions.

3.1 Multiple Exposures

One approach for faithfully recording the intensities in a high dynamic range scenes is to capture multiple images using different exposures, and then to merge these images. The basic idea is that when longer exposures are used, dark regions are well exposed but bright regions are saturated. On the other hand, when short exposures are used, dark regions are too dark but bright regions are well imaged. If exposure varies and multiple pictures are taken of the same scene, value of a pixel can be taken from those images where it's neither too dark nor saturated. This type of approach is often referred to as exposure bracketing, and has been widely adopted [Morimura 1993, Burt and Kolczynski 1993,Madden 1993,Tsai 1994]. Imaging devices usually contain nonlinearities, where pixel values are nonlinearly related to the brightness values in the scene. Some authors have proposed to use images acquired under 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 1995, Debevec and Malik 1997, Mitsunaga and Nayar 1999.]

Figure 3 Exposure Bracketing. Earlier examples by Mann and Picard 1995 (Courtesy, Steve Mann and Rosaling Picard 1995)

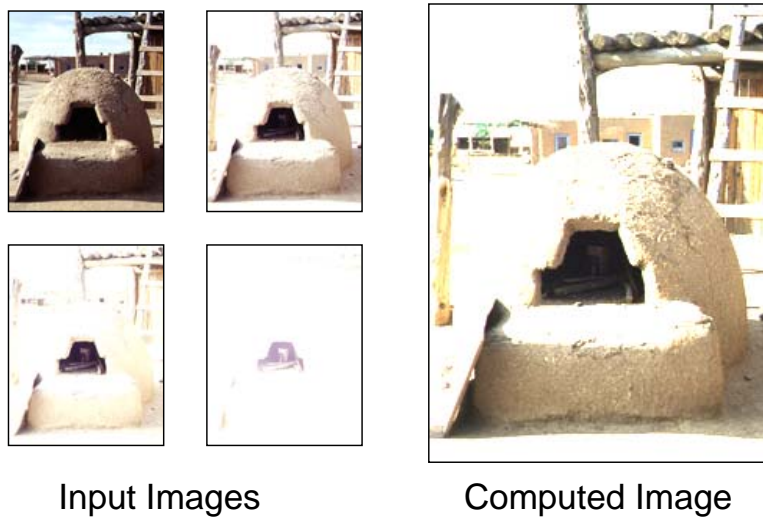Input Images   Computed Image

Figure 4 High Dynamic Range imaging using exposure bracketing (Courtesy, Nayar et al 1999)

3.2 Sensor Design

At the sensor level, various approaches have also been proposed for high dynamic range imaging. One type of approach is to use multiple sensing elements with different sensitivities within each cell [Street 1998, Handy 1986, Wen 1989, Hamazaki 1996]. Multiple measurements are made from the sensing elements, and they are combined on-chip before a high dynamic range image is read out from the chip. Spatial sampling rate is lowered in these sensing devices, and spatial resolution is sacrificed. Another type of approach is to adjust the well capacity of the sensing elements during photocurrent integration [Knight 1983, Sayag 1990, Decker 1998] but this gives higher noise. A different approach is proposed by [Brajovic and Kanade 1996], where the time it takes to reach saturation is measured, by a computation element attached to each sensing element. This time encodes high dynamic range information, as it is inversely proportional to the brightness at each pixel. Logarithmic sensors [Scheffer et al 2000] have also been proposed to increase the dynamic range. Brightside exploits the interline transfer of a charge coupled device (CCD) based camera to capture two exposures during a single mechanical shutter timing.

High dynamic range sensor design is in progress, but the implementation is usually costly. A rather novel and flexible approach is proposed by [Nayar and Mitsunaga 2000, Narasimhan and Nayar 2005], where exposures vary across space of the imager. A pattern with varying sensitivities is applied to the pixel array. It resembles the Bayer pattern in color imaging, but the sampling is made along the exposure instead of wavelength. The particular form of the sensitivity pattern, and the way of implementing it, are both quite flexible. One way of implementing it is to place a mask with cells of varying optical transparencies in front of the sensing array. Here, just as in Bayer mosaic, spatial resolution is sacrificed to some extent and aliasing can occur. Measurements under different exposures (sensitivities) are spatially interpolated, and combined into a high dynamic range image.

4   Aperture and Focus

Several concepts in exploiting focus and aperture parameters can be understood by considering the 4D lightfields transfer via lens and its 2D, 3D or 4D projection recorded on the image sensor.

Defocus Video Matting

Video matting is the process of recovering a high-quality alpha matte and foreground from a video sequence. Common approaches require either a known background (e.g., a blue screen) or extensive user interaction (e.g., to specify known foreground and background elements). The matting problem is generally under-constrained, unless additional information is recorded at the time of capture. McGuire et. al. have proposed a novel, fully autonomous method for pulling a matte using multiple synchronized video cameras that share the center of projection but differ in their plane of focus [McGuire et. al 2005]. The multi-camera data stream over-constrains the problem and the solution is obtained by directly minimizing the error in filter-based image formation equations. Their system solves the fully dynamic video matting problem without user assistance: both the foreground and background may be high frequency and have dynamic content, the foreground may resemble the background, and the scene may be lit by natural (as opposed to polarized or collimated) illumination. The authors capture 3 synchronized video streams using a 3 cameras and beam splitters. The first camera has a pinhole sensor has a small aperture that creates a large depth of field. The second and third cameras have large apertures, creating narrower depths of field focused on foreground and background, respectively. The foreground sensor produces sharp images for objects within about 0.5m of depth of the foreground object and defocuses objects farther away. The background sensor produces sharp images for objects from about 5m to infinity and defocuses the foreground object. Given the three video streams, at each frame the optical formation of each of the three images is expressed as the function of the unknowns background, foreground and alpha values.

Lightfield Camera

A detailed survey of the integral and lightfield photography is available at Roberts and Trebor Smith'2007. Here we describe more recent efforts. Ren et. al. have developed a camera that can capture the 4D light field incident on the image sensor in a single photographic exposure [Ren et al 2005]. This is achieved by inserting a microlens array between the sensor and main lens, creating a plenoptic camera. Each microlens measures not just the total amount of light deposited at that location, but how much light arrives along each ray. By re-sorting the measured rays of light to where they would have terminated in slightly different, synthetic cameras, one can compute sharp photographs focused at different depths. A linear increase in the resolution of images under each microlens results in a linear increase in the sharpness of the refocused photographs. This property allows one to extend the depth of field of the camera without reducing the aperture, enabling shorter exposures and lower image noise.
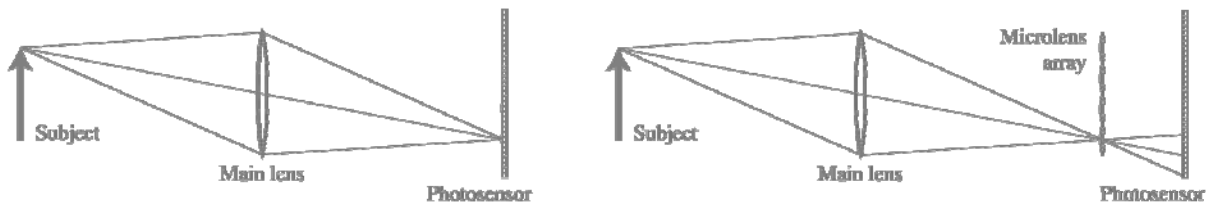

Figure 5 Light field photography using a handheld plenoptic camera. (Left) A scene point in sharp focus on the image photosensor. (Right) A microlens in front of the image photosensor. The scene point is in sharp focus on the lenslet and the image sensor captures how much light arrives along each ray. (Courtesy, Ren Ng et al 2005) (Permission renng@graphics.stanford.edu)

To the photographer, the plenoptic camera operates exactly like an ordinary hand-held camera. The ability to digitally refocus and extend the depth of field is ideal of portraits, high-speed action and macro close-ups. In a related paper, the authors have derived a Fourier representation of photographic imaging. The Fourier representation is conceptually and computationally simpler than the spatial domain representation. The theory enables one to compute photographs focused at different depths more quickly from the 4D light field data.

It is also possible to create a plenoptic camera using a patterned mask instead of a lenslet array. The geometric configurations remains nearly identical [Veeraraghavan2007]. The method is known as 'spatial optical heterodyning'. Instead of remapping of rays in 4D using microlens array so that they can be captured on a 2D sensor, spatial optical heterodyning remaps frequency components of the 4D lightfield so that the frequency components can be recovered from Fourier transform of the captured 2D image. In microlens array based design, each pixel effectively records light along a

single ray bundle. With patterned masks, each pixel records a linear combination multiple ray-bundles. By carefully coding the linear combination, the coded heterodyning method can reconstruct the values of individual ray-bundles.
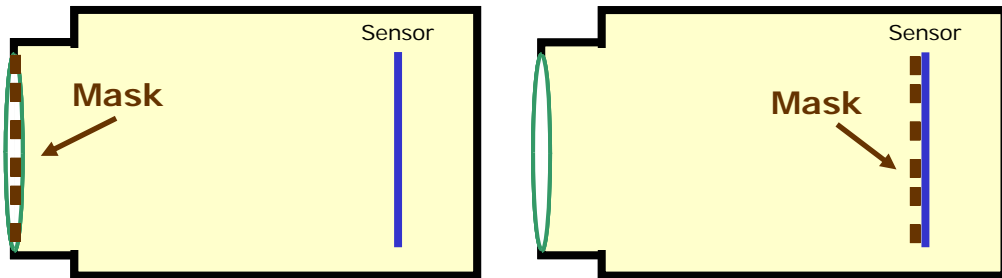

Figure 6 Mask based encoding and capture of Light Field. (Left) Encoded Blur Camera: A mask with a broad-band pattern placed at the lens preserves high spatial frequencies in out of focus parts and allows digital refocusing at full sensor resolution for images of layered Lambertian scenes. (Right) Heterodyne Lightfield Camera: A mask with narrow-band pattern performs reversible modulation of 4D light field by via optical heterodyning in space and angle. (Veeraraghavan et. al. 2007)
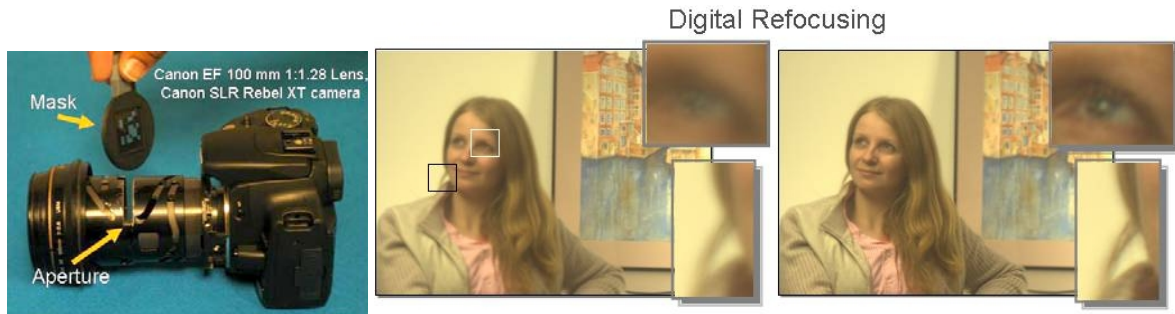
Figure 7 Encoded Blur Camera, i.e. with mask in the aperture, can preserve high spatial images frequencies in the defocus blur. Notice the glint in the eye. In the misfocused photo, on the left, the bright spot appears blurred with the bokeh of the chosen aperture (shown in the inset). In the deblurred result, on the right, the details on the eye are correctly recovered. (Veeraraghavan et. al. 2007)

Synthetic Aperture Imaging

Synthetic aperture focusing consists of warping and adding together the images in a 4D light field so that objects lying on a specified surface are aligned and thus in focus, while objects lying off this surface are misaligned and hence blurred. This provides the ability to see through partial occluders such as foliage and crowds, making it a potentially powerful tool for surveillance [Vaish et al 2004].
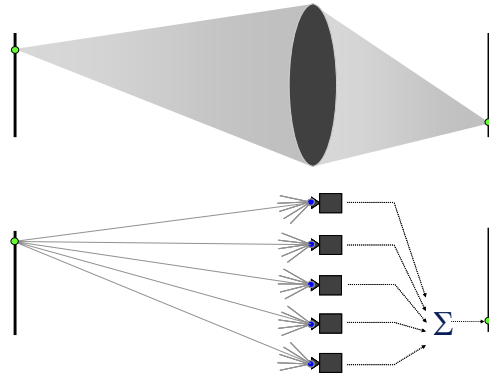
Figure 8 Synthetic aperture imaging. (Top) A large physical lens is ideal for creating a very narrow depth of field. (Bottom) One can simulate a large physical aperture by rebinning rays captured from multiple smaller aperture cameras. (Courtesy, Marc Levoy) (Permission levoy@graphics.stanford.edu)

Confocal microscopy is a family of imaging techniques that employ focused patterned illumination and synchronized imaging to create cross-sectional views of 3D biological specimens. Levoy et. al. have adapted confocal imaging to large-scale scenes by replacing the optical apertures used in microscopy with arrays of real or virtual video projectors and cameras [Levoy 2004]. A dense array of projectors allows to simulate a wide aperture (Synthetic Aperture Illumination) projector which can produce a real image with small depth of field. By projecting coded patterns and combining the resulting views using an array of virtual projectors, one can selectively image any plane in a partially occluded environment. These ideas were demonstrated on enhancing visibility in weakly scattering environments, such as murky water, to compute cross-sectional images and to see through partially occluded environments, such as foliage.

5   Motion Blur

Motion blur due to camera motion can significantly degrade the quality of an image. Since the path of the camera motion can be arbitrary, deblurring of motion blurred images is a hard problem. Previous methods to deal with this problem have included blind restoration of motion blurred images, optical correction using stabilized lenses, and special CMOS sensors that limit the exposure time in the presence of motion.

Motion Deblurring using Hybrid Imaging

Ben-Ezra et. al. exploit the fundamental trade off between spatial resolution and temporal resolution to construct a hybrid camera that can measure its own motion during image integration [Ben-Ezra and Nayar 2005]. The acquired motion information is used to compute a point spread function (PSF) that represents the path of the camera during integration. This PSF is then used to deblur the image. Results were shown on several indoor and outdoor scenes using long exposure and complex camera motion paths.

The hybrid imaging system proposed by the author consists of a high resolution primary detector and a low resolution secondary detector. The secondary detector is used to compute the motion information and the PSF. The motion between successive frames is limited to a global rigid transformation model which is computed using a multi-resolution iterative algorithm that minimizes the optical flow based error function. The resulting continuous PSF is then used for motion deblurring using the Richardson-Lucy algorithm. The authors used a 3M pixel Nikon still camera as the primary detector and a Sony DV camcoder as the secondary detector. The two detectors were calibrated offline. Results on several real sequences with exposure time ranging from 0.5 seconds to 4 seconds and the blur ranging up to 130 pixels were shown.

Recently, Fergus et al have shown that, in case of camera shake, the point spread function can be estimated from a single image. They exploit the natural image statistics on image gradients and then use the probably blur function to deblur the image [Fergus et al 2006].

Blur due to camera shake is different from blur due to object motion. And so far, there appears to be no good techniques for estimating object motion blur function.

Coded Exposure

In a conventional single-exposure photograph, moving objects or moving cameras cause motion blur. The exposure time defines a temporal box filter that smears the moving object across the image by convolution. This box filter destroys important high-frequency spatial details so that deblurring via deconvolution becomes an ill-posed problem. Raskar et. al. have proposed to flutter the camera's shutter open and closed during the chosen exposure time with a binary pseudo-random sequence, instead of leaving it open as in a traditional camera [Raskar et al 2006]. The flutter changes the box filter to a broad-band filter that preserves high-frequency spatial details in the blurred image and the corresponding deconvolution becomes a well-posed problem.

Results on several challenging cases of motion-blur removal including outdoor scenes, extremely large motions, textured backgrounds and partial occluders were presented. However, the authors assume that PSF is given or is obtained by simple user interaction. Since changing the integration time of conventional CCD cameras is not feasible, an external ferro-electric shutter is placed in front of the lens to code the exposure. The shutter is driven opaque and transparent according to the binary signals generated from PIC using the pseudo-random binary sequence.

Coded Exposure   **Coded Aperture**

Temporal 1-D
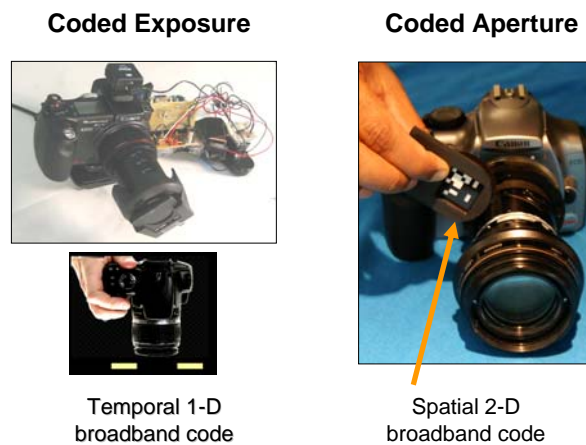broadband code

Spatial 2-D
broadband code

Figure 9 Coded Photography. (Left) Using a 1-D code in time to block and unblock light over time, a coded exposure photo can reversibly encode motion blur (Raskar et al 2006). (Right) Using a 2-D code in space to block parts of the light via a masked aperture, a coded aperture photo can reversibly encode defocus blur (Veeraraghavan et al 2007).

Figure 10 Preserving High Spatial Frequencies using a Flutter Shutter Camera. (Top) Photo of a fast moving taxi taken with coded exposure camera. Notice the subtle patterns in the blur which appear as multiple replicas rather than a continuous smear. (Bottom) The deblurred image. (Raskar et. al. 2006)

6   Computational Illumination

One of the oldest examples of computational illumination is strobe photography of Edgerton at MIT in the 1930's. Instead of a short exposure shutter, he used a traditional camera but with a novel strobe that has a very short duration. By varying the parameters of the camera, one can capture a variety of challenging scenes. The parameters include presence/absence, brightness, duration, position and wavelength of light. In addition, several new techniques employ spatial or temporal modulation of light using a projector.

6.1 Flash-no flash

The simplest form of computational illumination is perhaps the ubiquitous camera flash. [DiCarlo et al 2001] first explored the idea of capturing a pair of images for the same camera position - one illuminated with ambient light only, and the other using the camera flash as an additional light source. They use this image pair to estimate object reflectance functions, an the spectral distribution of the ambient ilumination. [Hoppe et al.2003] acquire multiple photos under different flash intensities, and allow the user to interpolate between them to simulate intermediate flash intensities.

Concurrent work by [Petschnigg et al. 2004] and [Eisemann et al.2004] proposed very similar techniques of combining the information contained in the flash and no-flash image pair to generate a single nice image. The no-flash photo captures the large-scale illumination effects such as the ambiance of the scene. However, in a low-light situation, the no-flash photo generally has excessive noise. The flash photo in contrast has much lower noise and more high frequency details, but fails to preserve the mood of the scene. The basic idea here is to decouple the high and low frequency components of the images, and then recombine to preserve the desired characteristics (detail from the flash photo, and large scale ambiance from the no-flash photo). This decoupling is done using a modified bilateral filter called joint bilateral filter,
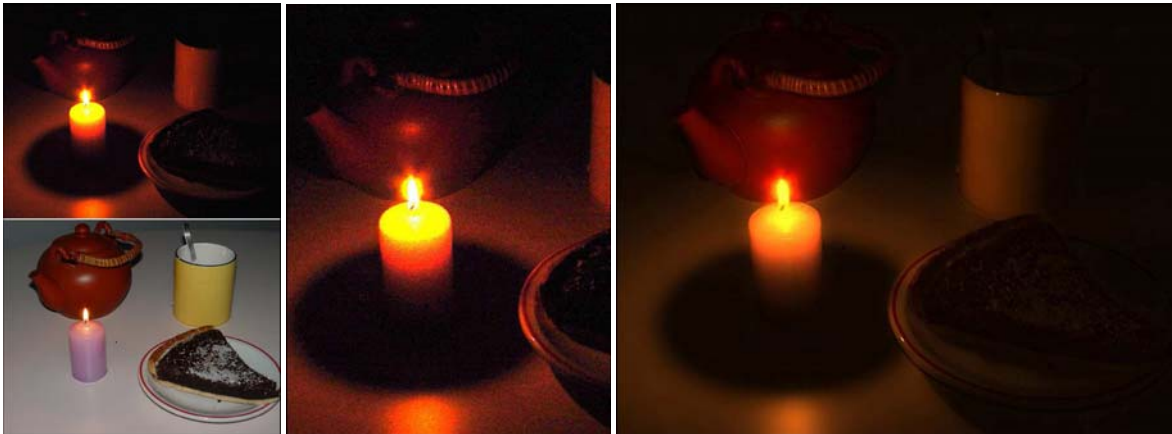
Figure 11 Combining a no-flash and flash image. (Left) Top: Photograph taken in a dark environment, the image is noisy and/or blurry. Bottom: Flash photography provides a sharp but flat image with distracting shadows at the silhouette of objects. (Middle) Zoom showing the noise of the available-light 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) (Permission fredo@mit.edu)

The bilateral filter is basically an edge-preserving blur that gives the low frequency component of the photo. In the joint bilateral filter, the intensity difference in the flash photo is used. Since the flash photo has lower noise, this gives a better results and avoids over or under blurring.

Agrawal et al. [Agrawal et al 20005] use the flash no-flash photo pair to remove reflections and hotspots from flash photos. They rely on the observation that the orientation of image gradients due to reflectance geometry are illumination invariant, while those due to changes in illumination are not. They propose a gradient projection scheme to decompose the illumination effects from the rest of the image. Based on the ratio of the flash and no-flash photos, they compensate for flash intensity falloff due to depth. Finally, they also propose a unified flash-exposure space that contains photos taken by varying the flash intensity and the shutter speed, and a method for adaptively sampling this space to capture a flash-exposure high dynamic range image.

Raskar et al.[Raskar et al 2004] used a multi-flash camera to find the silhouettes in a scene. They take four photos of an object with four different light positions (above, below, left and right of the lens). They detect shadows cast along the depth discontinuities are use them to detect depth discontinuities in the scene. The detected silhouettes are then used for stylizing the photograph and highlighting important features. They also demonstrate silhouette detection in a video using a repeated fast sequence of flashes.
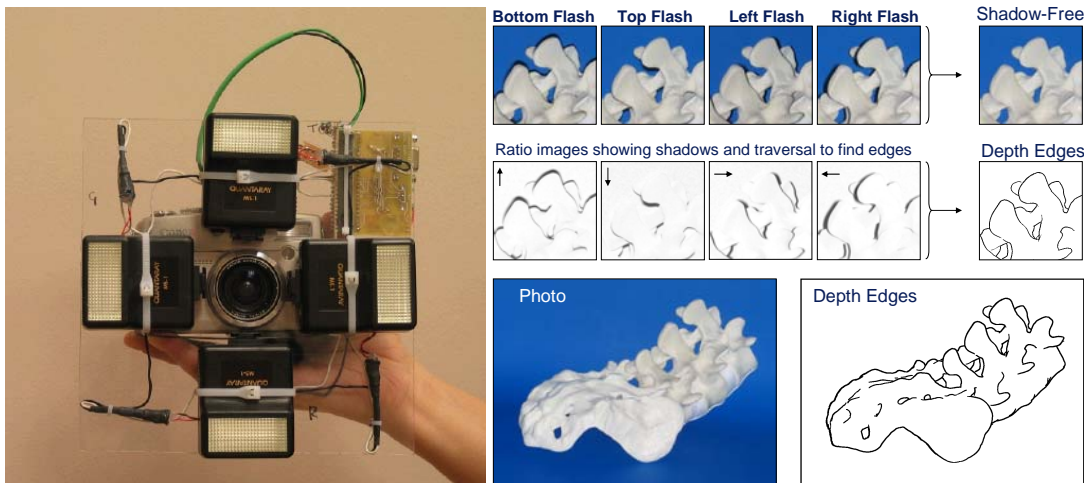

Figure 12 Multi-flash Camera for Depth Edge Detection. (Left) A camera with four flashes. (Right) Photos due to individual flashes, highlighted shadows and epipolar traversal to compute the single pixel depth edges.

6.2 4D acquisition

Light fields [Levoy 1996] and Lumigraph [Gortler 1996} reduced the more general plenoptic function [Adelson 1991] to a four dimensional function, $L(u,v,s,t)$ that describes the the presence of light in free space, ignoring the effect of wavelength and time. Here (u,v) and (s,t) are the parameters on two parallel planes respectively that describe a ray of light in space. A slightly different parameterization can be used to decribe the incident light field on an object. If we think of the object surrounded by a while sphere of imaginary projectors looking inwards, $(theta_i, phi_i)$ describes the angular position of the projector on the unit sphere, and (u,v) the pixel position on that projector. Thus, the function $L_i(u,v,theta,phi)$ gives complete control over the incident light on an object in free space. Similarly a sphere of inward looking cameras would capture the entire radiant light field of an object, $L_r(u,v,theta,phi)$. Debevec et al.[Debevec et al 2001] introduced the 8D reflectance field that describes relationship of the incident and radiant light fields of a scene. An additional dimension of time is sometimes added to describe light interaction with an object that changes over time.

While the reflectance field gives a complete description of how light interacts with a scene, acquiring this complete function would require enormous amounts of time and storage. Significant work has been done in trying to acquire lower dimensional subsets of this function, and using it for restricted re-lighting and rendering.

Most image based relighting work relies on the simple observation that light interacts linearly with materials [Nimeroff 1994, Haeberli 1992]. 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 make an image $I_{out}=sum_i (w_iI_i)$. Adjusting weights lets us ``relight'' the image, as if the weights modulate the lights rather than the images.

Debevec et al.[Debevec et al 2001] used a light stage comprising of a light mounted on a rotating robotic arm to acquire the non-local reflectance field of a human face. The point-like light source can be thought of as a simplified projector with a single pixel. Thus the incident light field is reduced to a 2D function. They acquired images of the face using a small number of cameras with densely sampled lighting directions. They demonstrated generation of novel images from the original viewpoints under arbitrary illumination. This is done by simply adjusting the weights w_i to match the desired illumination intensity from different directions. They also are also able to simulate small changes in the viewpoint using a simple model for the skin reflectance. Hawkins et al.[Hawkins et al 2001] used a similar setup and used it for digitizing cultural artifacts. They argue for the use reflectance field in digital archiving instead of geometric models and reflectance textures. Koudelka et al.[Koudelka et al 2001} acquire a set of images from a single viewpoint as a point light source moved around the object, and estimate the surface geometry by using two set of basis images. They then estimate the apparent BRDF for each pixel in the images, and use this to render the object under arbitrary illumination.

Debevec et al.[Debevec ey al 2002} proposed an enhanced light stage comprising of a large number (156) of inward pointing LEDs distributed on a spherical structure, about two meters in diameter, around the actor. They set each light to an arbitrary color and intensity to simulate the effect of a real world environment around the actor. The images gathered by the light stage, together with a mask of the actor captured using infrared sources and detector, were used to seamlessly composite the actor into a virtual set while maintaining consistent illumination. Malzblender et al. [Malzbender et al 2001] used 50 inward looking flashes placed on a hemispherical dome and a novel scheme for compressing and storing the 4D reflectance field, called the Polynomial Texture Map. They assumed that the color of a pixel changed smoothly as the light moved around the object, and store only the coefficients of a biquadratic polynomial that best models this change for each pixel. This highly compact representation allows for real time rendering of the scene with arbitrary illumination, and works fairly well for diffuse objects; specular highlights are not modeled very nicely by the polynomial model and result in visual artifacts.

The free-form light stage [Masselus 2002] presented a way to acquire a 4D slice of the reflectance field without the use of an extensive light-stage. Instead, they used a handheld, free-moving light source around the object. The light position was estimated automatically from four diffuse spheres placed near the object in the field of view of the camera. The data acquisition time was reported as 25-30 minutes. Winnemoller et al. [Winnemoeller et al 2005] used dimensionality reduction and a slightly constrained light scanning pattern to estimate approximate light source position without the need for any additional fiducials in the scene.

Akers et al. [Akers et al 2003] use spatially varying image weights on images acquired with a light stage similar to [Debevec et al 2001]. They use a painting interface allow an artist to locally modify the relit image as desired. While the spatially varying mask gives greater flexibility, it might also gives results that are not physically realizable and look unrealistic. [Anrys et al.2004] and [Mohan et al.2005] used a similar painting interface to help a novice user in lighting design for photography. The users sketch a target image, and the system finds optimal weights for each basis image to get a physically realizable result that is closest to the target. [Mohan et al.2005] argue that accurate calibration is not necessary for the

application photographic relighting, and propose a novel reflector based acquisition system. They place a moving-head gimbaled disco light inside a diffuse enclosure, together with the object to be photographed. The spot from the light on the enclosure acts as an area light source that illuminates the object. The light source is moved by simply rotating the light and capturing images for various light positions. The idea of area light sources was also used in bayesian relighting [Fuchs 2005].

7   Future Directions

7.1 Smart Sensors

Digital camera sensors typically use a color mosaic or a Bayer pattern of R,G, B filters to sense 3 different spectral bands, forming a basis for color reproduction. So-called 'demosaicing' methods, though widely varied and often proprietary, convert raw, interleaved color sensor values from the Bayer grid into R,G,B estimates for each pixel with as many luminance details and as few chrominance artifacts as possible, but the task itself forces tradeoffs and continued innovation.. Sony's four color CCD uses 'emerald' pixels which allow for correcting for defects in the rendition of red tones at certain frequencies. The Foveon sensor found in some Sigma digital cameras avoids the Bayer filter entirely, and instead detects wavelength bands for color according to photon penetration depths in a novel silicon detector design that stacks three layers of photodetectors, one below the other. This eliminates all the potential errors and artifacts of demosaicking, and reduces post-processing requirements substantially.

By sensing different between neighboring pixels instead of actual intensities, Tumblin et al [Tumblin et al 2005] have shown that a 'Gradient Camera' can record large global variations in intensity. Rather than measure absolute intensity values at each pixel, this proposed sensor measures only forward differences between them, which remain small even for extremely high-dynamic range scenes, and reconstructs the sensed image from these differences using Poisson solver methods. This approach offers several advantages: the sensor is nearly impossible to over- or under-expose, yet offers extremely fine quantization, even with very modest A/D convertors (e.g. 8 bits). The thermal and quantization noise occurs in the gradient domain, and appears as low frequency 'cloudy' noise in the reconstruction, rather than uncorrelated high-frequency noise that might obscure the exact position of scene edges.

Several companies now offer '3D cameras' that estimate depth for each pixel of the images they gather. Systems by Canesta and Zcam operate by precise measurement of the 'time-of-flight' (TOF) required for modulated infrared illumination to leave the camera, reflect from the scene and return to fast camera sensors. Several earlier, laser-based TOF systems, e.g. Cyberware, used 'flying spot' scanning to estimate depth sequentially. Without scanning these newer systems apply incoherent light (e.g. IR LEDs) and electronic gating to build whole-frame depth estimates at video rates. Canesta systems integrate the emitters in the same chip substrate as the detector, enabling a compact single-chip sensor unit; the Zcam device augments professional television camera units (ENG) to provide real-time depth keying and 3D reprojection.

Line Scan cameras. Several systems for critically-timed sports (e.g. sprints, horse racing) high-speed narrow-view or line-scan cameras hold more opportunities for capturing visual appearance. The 'FinishLynx' Lynx System Developers Inc. camera views a race finish-line through a narrow vertical slit, and assembles and image whose horizontal axis measures time instead of position. Despite occasionally strange distortions, the camera reliably depicts the first racer's body part to cross the finish line as the right-most feature in the time-space image.

7.2 Smart Optics

Wavefront coded imaging. Geometric aberrations in lenses cause image distortions, but these distortions can be modeled, computed, and in some cases robustly reversed. In 1995, Dowski and Cathey introduced a 'wavefront coded' optical element that forms intentionally distorted images with small, low cost optics [Dowski and Cathey 1995]. These seemingly out-of-focus images are computationally reversible, and allow reconstruction of an image with extended depth of focus, forming images with a focusing range up to 10X the abilities of conventional lenses. What other optical distortions might prove similarly advantageous?

Plenoptic Camera. As early as 1992 [Adelson and Wang 1992] several researchers have recognized the value of sensing the direction of incident light at each point on the focal plane behind a lens. Adelson's 1992 camera system combined a large front lens and a field of micro-lenses behind it, gathering what is now known as a 4D light field estimate, and he used it for single-lens stereo reconstruction. More recently, [Ng et al 2005] refined the idea further with an elegant hand-held digital camera for light-field capture that permits digital re-focussing and slight changes of viewpoint computationally.

Recently [Georgiev et al 2006] modeled the optics of these cameras using ray-matrix formulation, and showed an intriguing alternative. Instead of adding many tiny microlenses directly on top of delicate camera sensors, he builds a bundle large lenses and prisms attached to externally to the camera. The resulting light-field captured allows much larger computational changes in viewpoint in exchange for coarser digital re-focussing.

As these examples indicate, we have scarcely begun to explore the possibilities offered by combining computation, 4D modeling of light transport, and novel optical systems. Nor have such explorations been limited to photography and computer graphics. Computer vision, microscopy, tomography, astronomy and other optically driven fields already contain some ready-to-use solutions to borrow and extend. For example, N. Ahuja has explored cameras with spinning dispersion plates to allow a single camera to gather images from many virtual viewpoints for robust stereo reconstruction. How might other spinning optical elements help with appearance capture?

Tools for Optics

Until recently, ray-based models of light transport have been entirely adequate for computer graphics and computer vision, sometimes extended with special-case models for diffraction [Stam 1999]. Some early excursions into **wave optics** models by Gershon Elber [Elber 1994] proved computationally intense, and pinhole-camera and ideal-thin-lens models of optics have been entirely adequate for computer graphics use. As computational photography considers more complex lens systems, ray-only models of light transport begin to fail; to adequately model the spatial frequency response and wavelength dependence of optical systems we can move first to ray-matrix formulations, commonly used for optical fiber models and single-axis multi-lens systems, or move to **Fourier Optics** models to more accurately model the diffraction effects that predict the spatial frequency response of lens systems with adjustable apertures, and include accurate modeling of coherent light as well, including holography. The classic text by Goodman [Goodman 1968] is an elegant introduction to this topic. The computational requirements for Fourier analysis of optics is no longer formidable, especially with GPU assistance, and recent work by [Ng 2005] has already tied lightfields to images by showing it follows the 4D projection-slice theorem [Rosenfeld and Kak, 1987] that became a fundamental tenet of medical tomography.

Beyond Fourier Optics we can resort to specialized lens-design descriptors such as Zernicke polynomials and remain within the realm of practical computation. Further refinement by resorting to full electromagnetic simulation can model polarization and optical effects due to structures smaller than the wavelength of light. These models can directly predict the optical behavior of superlattice structures such as iridescent butterfly wings, the transparency of finely-fibred structures such as the lens and cornea of the eye, and strange retro-reflectance properties of some classes diseased cell bodies. While medical researchers and others are actively pursuing such simulations, the computational requirements are still daunting, and appear out of reach for current experiments in computational photography.

7.3 Other Dimensions

As noted in the 'Assorted pixels' paper [Nayar2003], photographic capture gathers optical data along many dimensions, and few are fully exploited. In 4-dimensional ray space we sense and measure more than simple intensity (or more formally, radiance), but also visually assess wavelength, time, materials, illumination direction and more. **Polarization** is also sometimes revealing, and the mapping from polarization direction of the illuminant to the polarization of reflected light is not a simple one: for some biological materials, the mappings are nonlinear and unexplored [Wu et al 2003]. Extended exploration of wavelength dependence is already well advanced. Hyperspectral imaging has already gathered a rich and growing literature for a broad range of applications from astronomy to archival imaging of museum treasures.

Film-style photography relies on an 'instantaneous' ideal: we attempt 'stop time' by capturing any photographed scene quickly enough to ignore any movement that happens during the measurement process. Even 'motion pictures' commit serial attempts at instantaneous capture, rather than direct sensing of the motions themselves. Harold Edgerton pushed the instantaneous ideal to extremes by using ultra-short strobes to illuminate transient phenomena, and ultra-short shutters to measure ultra-bright phenomena quickly, such as his famous high-speed movies of atomic bomb explosions.

Digital sensors offer new opportunities for more direct sensing, and digital displays permit interactive display of the movements we capture. Accordingly, Michael Cohen has proposed that the film-rooted distinction between 'still' cameras and 'video' cameras should gradually disappear. He proposed that we need an intermediate digital entity he calls a '**moment'**; one visually meaningful action we wish to remember—a child's fleeting expression of delighted surprise, a whisper of wind that sways the trees, etc., and it might fit in short video clips [Cohen 2005]. Motion sensing and deblurring itself can improve in the future [BenEzra 2004, Raskar 2006]. Movement also causes difficulties for constructing panoramas. However, if the movement is statistically consistent, it is possible to combine conventional image stitching operations with so-called 'video texturing' [Schödl 2000] methods to create consistent, seamless movement that captures the 'moment' of the panorama quite well. It can be further extended to capture video texture panoramas [Agarwala 2005]

7.4 Scientific Imaging

Scene measurement and representation in 4-D and beyond encompasses previously isolated "Islands" of Ingenious Scientific Imaging & Measuring. What can we learn from them? Can we extend their methods? Particularly promising fields include the following.

(i) Tomography: For any penetrating measurements, attenuation along straight-line paths can be used to construct 3D images of internal structures This is currently used measuring sound transmission to electrical capacitance, from seismographic disturbances to ultrasonics to X-rays.

(ii) Spectrographic methods: complex interdependencies between wavelengths, reflectance, and transmissions are used for image forming, and broad classes of statistical measurements help decipher or identify useful features for land management, pollution studies, atmospheric patterns, wildlife migration, and geological and mineral features..

(iii) Confocal Methods and Synthetic Aperture methods: As described above, one can achieve very narrow depth-of-field image by collecting a widely divergent rays from each imaged point and these methods can extend to macroscopic scales via multiple cameras and multiple video projectors.

(iv) Fluorescence Methods: Some materials respond to absorbed photons by re-emitting other photons at different wavelengths, a phenomena known as fluorescence While very few materials fluoresce in the narrow range (< 1 octave!) of visible wavelengths, hyperspectral imaging reveals instructive fluorescence phenomena occur over much wider bands of wavelengths. Many organic chemicals have strongly varied fluorescent responses to ultraviolet light, and some living tissues can be chemically or genetically tagged with fluorescent markers that reveal important biological processes. Accordingly, hyperspectral imaging and illuminants can directly reveal chemical or biological features that may be further improved by 4D methods.

7.5 Fantasy Configurations

Beyond what we can do now, what would we like to achieve in computational photography? Freed from practical limits, a few fantasy devices come to mind. If the goal of photography is to capture the visual essence of an object in front of us, then perhaps the ideal photography studio is not a room full of lights and box-like cameras at all, but a flexible cloth we can rub gently over the surface of the object itself. The cloth would hold microscopic, interleaved video projectors and video cameras. It would emit hyperspectrally colorful patterns of light in all possible directions from all possible points on the cloth (a flexible 4D light source), while simultaneously making coordinated hyperspectral measurements in all possible directions from all possible points on the cloth (a flexible 4D camera). Wiping the cloth over a surface would illuminate and photograph inside even the tiniest crack or vent hole of the object, banishing occlusion from the data set; a quick wipe would characterize any rigid object thoroughly.

Suppose we wish to capture the appearance of a soft object, without touching it? Then perhaps a notebook-like device made of two plates hinged together would help. Each panel would consist of interleaved cameras and projectors in a sheet-like arrangement; simply placing it around the object would provide sufficient optical coupling between the embedded 4D illuminators and 4D cameras to assess the object thoroughly.

Yet even these are not the whole answer. If the goal of photography is to capture, reproduce, and manipulate a meaningful visual experience, then the 'camera cloth' is not sufficient to capture even the most rudimentary birthday party. The human experience and our personal viewpoint is missing. Ted Adelson suggested 'camera wallpaper' or the 'balloon camera', ubiquitous sensors that would enable us to compute arbitrary viewpoints at arbitrary times. Thad Starner and other 'cybernauts' who began personally instrumenting themselves in the 1990s have experimented with 'always-on' video cameras, and projects at Microsoft and the MIT Media Lab have explored gathering 'video memories' of every waking moment. So called 'smart dust' sensors and other unstructured ubiquitous sensors might gather views, sounds, and appearance from anywhere in a large city. What makes these moments special? What parts of this video will become keepsakes or evidence? How do we find what we care about in this flood of video? Computational Photography can supply us with visual experiences, but can't decide which one's matter most to humans.

The article briefly examined the topics commonly appearing at academic conferences so that the reader is familiar with the general problems in Computational Photography. Our focus is on the review of problems rather than an exhaustive review of the emerging solutions. We refer the reader to the Course Notes for Siggraph 2007 course available on our website for a comprehensive discussion. http://www.merl.com/people/raskar/photo/

Acknowledgements

Bibliography

Fusion of Images Taken by Varying Camera and Scene Parameters

General

AKERS, D., LOSASSO, F., KLINGNER, J., AGRAWALA, M., RICK, J., AND HANRAHAN, P. 2003. Conveying shape and features with image-based relighting. In IEEE Visualization, 349–354.

BURT, P., AND KOLCZYNSKI, R. 1993. Enhanced image capture through fusion. In International Conference on Computer Vision (ICCV 93), 173–182.

LEVIN, A., ZOMET, A., PELEG, S., AND WEISS, Y. 2004. Seamless image stitching in the gradient domain. In European Conference on Computer Vision (ECCV 04).

MASSEY, M., AND BENDER, W. 1996. Salient stills: Process and practice. IBM Systems Journal 35, 3&4, 557–574.

MUTTER, S., AND KRAUSE, M. 1992. Surrational Images: Photomontages. University of Illinois Press.

ROBINSON, H. P. 1869. Pictorial Effect in Photography: Being Hints on Composition and Chiaroscuro for Photographers. Piper & Carter.

MUYBRIDGE, E. 1955. The human figure in motion. Dover Publications, Inc.

AGARWALA, A., DONTCHEVA, AGRAWALA, M., DRUCKER, COLBURN, CURLESS, SALESIN AND COHEN, M. Interactive Digital Photomontage. ACM Transactions on Graphics (Proceedings of SIGGRAPH 2004), 2004.

S.K. Nayar and S.G. Narasimhan , "Assorted Pixels: Multi-Sampled Imaging With Structural Models," Europian Conference on Computer Vision (ECCV), Vol.IV, pp.636-652, May, 2002.

Time

BRAUN, M. 1992. Picturing Time: The Work of Etienne-Jules Marey. The University of Chicago Press.

FREEMAN, W. T., AND ZHANG, H. 2003. Shape-time photography. In Conference on Computer Vision and Pattern Recognition (CVPR 03), 151–157.

Exposure

FATTAL, R., LISCHINSKI, D., AND WERMAN, M. 2002. Gradient domain high dynamic range compression. ACM Transactions on Graphics 21, 3, 249–256.

REINHARD, E., STARK, M., SHIRLEY, P., AND FERWERDA, J. 2002. Photographic tone reproduction for digital images. ACM Transactions on Graphics 21, 3, 267– 276.

DEBEVEC, AND MALIK. 1997. Recovering high dynamic range radiance maps from photographs. In Proc. SIGGRAPH.

DURAND, AND DORSEY. 2002. Fast bilateral filtering for the display of high-dynamic-range images. ACMTrans. on Graphics 21, 3.

MANN, AND PICARD. 1995. Being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures. In Proc. IS&T 46th ann. conference.

TUMBLIN, AND TURK. 1999. LCIS: A boundary hierarchy for detail-preserving contrast reduction. In Proc. SIGGRAPH.

DICARLO, J., AND WANDELL, B. 2000. Rendering high dynamic range images. Proc. SPIE: Image Sensors 3965, 392–401.

Kang, S. B., Uyttendaele, M., Winder, S., and Szeliski, R. 2003. High dynamic range video. ACM Trans. Graph. 22, 3 (Jul. 2003), 319-325.

Focus

HAEBERLI, P. 1994. Grafica Obscura web site. http://www.sgi.com/grafica/.

MORGAN MCGUIRE, MATUSIK, PFISTER, HUGHES, AND DURAND, Defocus Video Matting, ACM Transactions on Graphics, Vol 24, No 3, July 2005, (Proceedings of ACM SIGGRAPH 2005).

Illumination

Frederik Anrys and Philip Dutr´e. Image based lighting design. In The 4th IASTED International Conference on Visualization, Imaging, and Image Processing, 2004.

David Akers, Frank Losasso, Jeff Klingner, Maneesh Agrawala, John Rick, and Pat Hanrahan. Conveying shape and features with image-based relighting. In IEEE Visualization, 2003.

Amit Agrawal, Ramesh Raskar, Shree K. Nayar, and Yuanzhen Li. Removing photography artifacts using gradient projection and flash-exposure sampling. In SIGGRAPH, pages 828–835, 2005.

Paul Debevec, Tim Hawkins, Chris Tchou, Haarm-Pieter Duiker, Westley Sarokin, and Mark Sagar. Acquiring the reflectance field of a human face. In SIGGRAPH, pages 145–156, 2000.

Paul Debevec, AndreasWenger, Chris Tchou, Andrew Gardner, JamieWaese, and Tim Hawkins. A lighting reproduction approach to live-action compositing. In SIGGRAPH,pages 547–556, 2002.

Passive Illumination

RASKAR, R., ILIE, A., AND YU, J. 2004. Image fusion for context enhancement and video surrealism. In NPAR 2004: Third International Symposium on Non- Photorealistic Rendering.

WEISS, Y. 2001. Deriving intrinsic images from image sequences. In International Conference On Computer Vision (ICCV 01), 68–75.

Polorization

Y. Y. SCHECHNER, S. G. NARASIMHAN and S. K. NAYAR, Instant Dehazing of Images Using Polarization, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Hawaii, December 2001.

S. K. NAYAR, X. FANG, and T. E. BOULT, Removal of Specularities using Color and Polarization, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition.

Wu, P. Walsh, J.T. "Tissue Polarization Imaging of Multiply-Scattered Reflected Light Lasers in Surgery and Medicine, Supplement 15, 2003.

Wavelength

D. A. SOCOLINSKY, "Dynamic range constraints in image fusion and realization." Proc. IASTED Int. Conf. Signal and
Image Process, 349-354 (2000).

Y. Y. SCHECHNER and S. K. NAYAR , Uncontrolled Modulation Imaging, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Washington DC, June 2004.

Location

Aseem Agarwala, Maneesh Agrawala, Michael Cohen, David Salesin, Rick Szeliski. ``Photographing long scenes with multi-viewpoint panoramas,'' ACM Transactions on Graphics (Proceedings of SIGGRAPH 2006), To appear, 2006.

B. Wilburn, N. Joshi, V. Vaish, E. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, M. Levoy, High-Performance Imaging Using Large Camera Arrays.. ACM Transactions on Graphics, Vol 24, No 3, July 2005, pp 765-776 (Proceedings of ACM SIGGRAPH 2005).

Matting

CHUANG, Y.-Y., CURLESS, B., SALESIN, D., AND SZELISKI, R. 2001. A Bayesian approach to digital matting. In Proceedings of Computer Vision and Pattern Recognition (CVPR 2001), vol. II, 264 – 271.

PORTER, T., AND DUFF, T. 1984. Compositing digital images. In Computer Graphics (Proceedings of ACM SIGGRAPH 84), vol. 18, 253–259.

SMITH, A. R., AND BLINN, J. F. 1996. Blue screen matting. In Proceedings of ACM SIGGRAPH 96, 259–268.

Jian SUN, Jiaya JIA, Chi-Keung TANG and Heung-Yeung SHUM, Poisson Matting, ACM Transactions on Graphics, also in SIGGRAPH 2004, vol. 23, no. 3, July 2004, pages 315-321.

Techniques

General

DANIELSSON, P.-E. 1980. Euclidean distance mapping. Computer Graphics and Image Processing 14, 227–248.

LUCAS, B. D., AND KANADE, T. 1981. An iterative image registration technique with an application to stereo vision. In Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI '81), 674–679.

MORTENSEN, E. N., AND BARRETT, W. A. 1995. Intelligent scissors for image composition. In Proceedings of SIGGRAPH 95, Computer Graphics Proceedings, Annual Conference Series, 191–198.

Graph Cuts

BOYKOV, Y., VEKSLER, O., AND ZABIH, R. 2001. Fast approximate energy minimization via graph cuts. IEEE Transactions on Pattern Analysis and Machine Intelligence 23, 11, 1222–1239.

KWATRA, V., SCH ¨ODL, A., ESSA, I., TURK, G., AND BOBICK, A. 2003. Graphcut textures: Image and video synthesis using graph cuts. ACM Transactions on Graphics 22, 3, 277–286.

SHI, J., AND MALIK, J. Normalized Cuts and Image Segmentation.   IEEE Conf. Computer Vision and Pattern Recognition (CVPR), June 1997, Puerto Rico

Gradient Domain

PEREZ, P., GANGNET, M., AND BLAKE, A. 2003. Poisson image editing. ACM Transactions on Graphics 22, 3, 313–318.

T. Georgiev. Covariant Derivatives and Vision. Proceedings of ECCV 2006.

Smoothing, Bilateral and Trilateral Filter

C. TOMASI, AND R. MANDUCHI, Bilateral Filtering of gray and colored images, Proc. IEEE Intl. Conference on
Computer Vision, pp. 836-846, 1998.

CHOUDHURY, P., TUMBLIN, J., "The Trilateral Filter for High Contrast Images and Meshes", Proc. of the Eurographics Symposium on Rendering, Per. H. Christensen and Daniel Cohen eds., pp. 186-196, 2003

Video Textures

Schödl, A., Szeliski, R., Salesin, D. H., and Essa, I. 2000. Video textures. In Proceedings of the 27th Annual Conference on Computer Graphics and interactive Techniques International Conference on Computer Graphics and Interactive Techniques. ACM Press/Addison-Wesley Publishing Co., New York, NY, 489-498.

Agarwala, A., Zheng, K. C., Pal, C., Agrawala, M., Cohen, M., Curless, B., Salesin, D., and Szeliski, R. 2005. Panoramic video textures. ACM Trans. Graph. 24, 3 (Jul. 2005), 821-827.

Feature Extraction and Scene Understanding

Shape/Material/Illumination, Surface normals

BASRI, R. JACOBS, D. Photometric stereo with general, unknown lighting, Computer Vision and Pattern Recognition, 2001

B. K. P. HORN, "Shape from shading: A method for obtaining the shape of a smooth opaque object from one view," MIT Project MAC Int. Rep. TR-79 and MIT AI Lab, Tech. Rep. 232, Nov. 1970.

TODD ZICKLER, PETER N. BELHUMEUR, AND DAVID J. KRIEGMAN, "Helmholtz Stereopsis: Exploiting Reciprocity for Surface Reconstruction." Proc. 7th European Conference on Computer Vision, May 2002. Vol. III, pp 869-884.

ANDREAS WENGER, A GARDNER, CHRIS TCHOU, J UNGER, T HAWKINS, P DEBEVEC, Postproduction Relighting and Reflectance Transformation With Time-Multiplexed Illumination, SIGGRAPH 2005

Depth edges

Ramesh RASKAR , Karhan TAN, Rogerio FERIS, Jingyi YU, Matthew TURK, Non-photorealistic Camera: Depth Edge Detection and Stylized Rendering Using a Multi-Flash Camera, SIGGRAPH 2004

Depth

S. K. NAYAR and Y. NAKAGAWA,, Shape from Focus, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 16, No. 8, pp. 824-831,

Motion

M. Ben-Ezra and S.K. Nayar, "Motion-based Motion Deblurring,", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.26, No.6, pp.689-698, Jun, 2004.

Raskar, R., Agrawal, A., and Tumblin, J. 2006. Coded exposure photography: motion deblurring using fluttered shutter. ACM Trans. Graph. 25, 3 (Jul. 2006), 795-804.

Rob Fergus, Barun Singh, Aaron Hertzmann, Sam T. Roweis, William T. Freeman.,"Removing Camera Shake from a Single Photograph", ACM Trans. on Graphics (Proc. SIGGRAPH 2006).

Liu, C., Torralba, A., Freeman, W. T., Durand, F., and Adelson, E. H. 2005. Motion magnification. ACM Trans. Graph. 24, 3 (Jul. 2005), 519-526.

Transfer and denoising

Flash to no-flash

Elmar EISEMANN and Frédo DURAND, Flash Photography Enhancement Via Intrinsic Relighting, SIGGRAPH 2004

Georg PETSCHNIGG, Maneesh AGRAWALA, Hugues HOPPE, Richard SZELISKI, Michael COHEN, Kentaro TOYAMA. Digital Photography with Flash and No-Flash Image Pairs.ACM Transactions on Graphics (Proceedings of SIGGRAPH 2004), 2004.

DICARLO, J. M., XIAO, F., AND WANDELL, B. A. 2001. Illuminating illumination. In 9th Color Imaging Conference, 27–34.

Noise

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, vol. 12, no. 2, pp. 216-223, 1990.

Eric P. BENNETT and Leonard McMILLAN "Video Enhancement Using Per-Pixel Virtual Exposures" SIGGRAPH 2005

Geometric Operations

Panorama

DAVIS, J. 1998. Mosaics of scenes with moving objects. In Computer Vision and Pattern Recognition (CVPR 98), 354–360.

UYTTENDAELE, M., EDEN, A., AND SZELISKI, R. 2001. Eliminating ghosting and exposure artifacts in image mosaics. In Conference on Computer Vision and Pattern
Recognition (CVPR 01), 509–516.

SZELISKI, R., AND SHUM, H.-Y. 1997. Creating full view panoramic mosaics and environment maps. In Proceedings of SIGGRAPH 97, Computer Graphics Proceedings,
Annual Conference Series, 251–258.

Synthetic Aperture and Multi-views

E. H. Adelson and J. Y.A. Wang, "Single Lens Stereo with a Plenoptic Camera" in IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 14, No. 2, Feb 1992.

Marc LEVOY, Billy CHEN, Vaibhav VAISH, Mark HOROWITZ, Ian MCDOWALL, Mark BOLAS, Synthetic Aperture Confocal Imaging. ACM SIGGRAPH 2004.

REN NG, Fourier Slice Photography, SIGGRAPH 2005

Georgiev, T., Zheng, C., Nayar, Sh., Curless, B., Salasin, D., Intwala, Ch., Spatio-angular Resolution Trade-offs in Integral Photography, EGSR 2006

STERN and B. JAVIDI, "3-D computational synthetic aperture integral imaging (COMPSAII)," Opt. Express 11, 2446-2451 (2003),

OLIVER and S. QUEGAN, Understanding Synthetic Aperture Radar Images. London: Artech House, 1998.

Vaibhav Vaish, Gaurav Garg, Eino-Ville (Eddy) Talvala, Emilio Antunez, Bennett Wilburn, Mark Horowitz, Marc Levoy, "Synthetic Aperture Focusing using a Shear-Warp Factorization of the Viewing Transform", Proc. Workshop on Advanced 3D Imaging for Safety and Security (A3DISS) 2005 (in conjunction with CVPR 2005)

Deblurring and Superresolution

M. BEN-EZRA AND S. K. NAYAR , Motion Deblurring using Hybrid Imaging, In Proc. IEEE Computer Vision and Pattern Recognition (CVPR), Wisconsin, June 2003.

ZHOUCHEN LIN, HEUNG-YEUNG SHUM Fundamental Limits of Reconstruction-Based Superresolution Algorithms under Local Translation PAMI, January 2004 - (Vol. 26, No. 1) pp. 83-9

LANDOLT, A. MITROS, AND C. KOCH, "Visual Sensor with Resolution Enhancement by Mechanical Vibrations," Proc. 2001 Conf. Advanced Research in VLSI, pp. 249-264, 2001.

Dynamic Range

Morimura. Imaging method for a wide dynamic range and an imaging device for a wide dynamic range. U.S. Patent 5455621, October 1993.

P. Burt and R. J. Kolczynski. Enhanced Image Capture Through Fusion. Proc. Of International Conference on Computer Vision (ICCV), pages 173-182, 1993.

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

Y. T. Tsai. Method and apparatus for extending the dynamic range of an electronic imaging system. U.S. Patent 5309243, May 1994.

T. Mitsunaga and S. K. Nayar. Radiometric self calibration. In Proc CVPR, volume 2, pages 374--380, June 1999.

R. A. Street. High dynamic range segmented pixel sensor array. U.S. Patent 5789737, August 1998.

R. J. Handy. High dynamic range ccd detector imager. U.S. Patent 4623928, November 1986.

D. D.Wen. High dynamic range charge coupled device. U.S. Patent 4873561, October 1989.

M. Hamazaki. Driving method for solid-state image pickup device. Japanese Patent 08-331461, December 1996.

Knight, T. (1983) "Design of an Integrated Optical Sensor with On-Chip Pre-Processing," Ph.D. Thesis, Department of Electrical Engineering and Computer Science, MIT.

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

S. J. Decker, R. D. McGrath, K. Brehmer, and C. G. Sodini, "A 256x256 CMOS imaging array with wide dynamic range pixels and column-parallel digital output." IEEE J. of Solid State Circuits, Vol. 33, pp. 2081-2091, Dec. 1998

V. Brajovic and T. Kanade. A Sorting Image Sensor: An Example of Massively Parallel Intensity-to-Time Processing for Low-Latency Computational Sensors. Proc. of IEEE Conference on Robotics and Automation, pages 1638-1643, April 1996.

D. Scheffer S. Kavadias, B. Dierickx, A. Alaerts, D. Uwaerts, and J. Bogaerts. A Logarithmic Response CMOS Image Sensor with On-Chip Calibration. IEEE JSSC, 35(8):1146--52, August 2000

Shree K. Nayar, Tomoo Mitsunaga: High Dynamic Range Imaging: Spatially Varying Pixel Exposures. CVPR 2000: 1472-1479

Plenoptic Function and Light Field

David Roberts and Trebor Smith, 'Lens Array Print Techniques: The History of Integral Print Methods', Ihttp://www.integralresource.org/Integral_History.pdf

Edward H. Adelson and James R. Bergen. The plenoptic function and the elements of early vision. M. Landy and J. A. Movshon, (eds) Computational Models of Visual Processing, 1991.

Marc Levoy and Pat Hanrahan. Light field rendering. In SIGGRAPH, pages 31–42, 1996.

Steven J. Gortler, Radek Grzeszczuk, Richard Szeliski, and Michael F. Cohen. The lumigraph. In SIGGRAPH, pages 43–54, 1996.

Active Illumination

Paul Haeberli. Graphics obscura.

Tim Hawkins, Jonathan Cohen, and Paul Debevec. A photometric approach to digitizing cultural artifacts. In Proceedings of conference on Virtual reality, archeology, and cultural heritage, pages 333–342, 2001.

Tim Hawkins, Per Einarsson, and Paul E. Debevec. A dual light stage. In Rendering Techniques, pages 91–98, 2005.

Hugues Hoppe and Kentaro Toyama. Continuous flash. Technical Report 63, Microsoft Research, 2003.

Andrew Jones, Andrew Gardner, Mark Bolas, Ian McDowall, and Paul Debevec. Performance geometry capture for spatially varying relighting. SIGGRAPH Sketch, 2005.

Feng Xiao Jeffrey M. DiCarlo and Brian A Wandell. Illuminating illumination. In Ninth Color Imaging Conference, pages 27–34, 2001.

Melissa L Koudelka, Peter N Belhumeur, Sebastian Magda, and David J. Kriegman. Image-based modeling and rendering of surfaces with arbitrary brdfs. In IEEE CVPR, pages 568–575, 2001.

Martin Fuchs, Volker Blanz, and Hans-Peter Seidel. Bayesian relighting. In Rendering Techniques, pages 157–164, 2005.

Tom Malzbender, Dan Gelb, and Hans Wolters. Polynomial texture maps. In SIGGRAPH, pages 519–528. ACM Press, 2001.

Vincent Masselus, Pieter Peers, Philip Dutr´e, and Yves D. Willems. Relighting with 4d incident light fields. In SIGGRAPH, volume 22, pages 613–620, 2003.

Ankit Mohan, Jack Tumblin, Bobby Bodenheimer, Cindy Grimm, and Reynold J. Bailey. Table-top computed lighting for practical digital photography. In Rendering Techniques, pages 165–172, 2005.

Jeffry S. Nimeroff, Eero Simoncelli, and Julie Dorsey. Efficient re-rendering of naturally illuminated environments. In Proceedings of the Fifth Eurographics Workshop on Rendering, pages 359–373, 1994.

AndreasWenger, Andrew Gardner, Chris Tchou, Jonas Unger, Tim Hawkins, and Paul Debevec. Performance relighting and reflectance transformation with time-multiplexed illumination. In SIGGRAPH, volume 24, pages 756–764, 2005.

Holger Winnemöeller, Ankit Mohan, Jack Tumblin, and Bruce Gooch. Light waving: Estimating light positions from photographs alone. Computer Graphics Forum, 24(3):to appear, 2005.

Smart, Unconventional Cameras

MEMS Technology

S. K. NAYAR, V. BRANZOI, AND T. BOULT. Programmable Imaging using a Digital Micromirror Array, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Washington DC, June 2004.

High Speed Imaging

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, October 21-22 1999. (Society of Multispectral Imaging of Japan.)

S. KLEINFELDER, S.H. LIM, X.Q. LIU AND A. EL GAMAL A 10,000 Frames/s CMOS Digital Pixel Sensor, In IEEE Journal of Solid State Circuits, Vol.36, No.12, Pages 2049-2059, December 2001

X.Q. LIU AND ABBAS EL GAMAL, Synthesis of High Dynamic Range Motion Blur Free Image From Multiple Captures, In IEEE Transactions on circuits and systems (TCASI), VOL. 50, NO. 4, pp 530-539, APRIL 2003

Ramesh RASKAR, Amit AGRAWAL, Jack TUMBLIN, Coded Exposure Photography: Motion Deblurring using Fluttered Shutter, ACM SIGGRAPH 2006.

SHECHTMAN, E., CASPI, Y., AND IRANI, M. 2002. Increasing space-time resolution in video. In ECCV, Springer-Verlag, London, UK, 753–768.

Programmable SIMD

JOHANSSON, R., LINDGREN, L., MELANDER, J., AND MOLLER, B. 2003. A multi-resolution 100 gops 4 gpixels/s programmable cmos image sensor for machine vision. In Proceedings of the 2003 IEEE Workshop on Charge-Coupled Devices and Advanced Image Sensors, IEEE.

Advanced, Programmable, Demodulating Cameras and Temporal Correlation

CANESTA Inc, 2004

PIXIM Inc, 2004

FOVEON Inc, 2004

JENOPTIK Inc, 2004

IVP Inc, Ranger Camera, 2004

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, 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, 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

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

Pilu, Maurizio, "Casual Capture Project at HP labs", http://www.hpl.hp.com/news/2004/jan-mar/casualcapture.html

TUMBLIN, J., AGRAWAL, A. AND RASKAR, R. Why I want a Gradient Camera, IEEE CVPR 2005

Optics

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

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

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

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

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

Light field optics

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

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

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

Frédo Durand, Nicolas Holzschuch, Cyril Soler, Eric Chan & François Sillion. A Frequency analysis of Light Transport. Proceedings of SIGGRAPH 2005.

Todor Georgiev, Chintan Intwala. Integral View Photography. Adobe Tech Report 2006:

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

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

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