Less Is More: Coded Computational Photography

Ramesh Raskar

Mitsubishi Electric Research Labs (MERL), Cambridge, MA, USA

Abstract. Computational photography combines plentiful computing, digital sensors, modern optics, actuators, and smart lights to escape the limitations of traditional cameras, enables novel imaging applications and simplifies many computer vision tasks. However, a majority of current Computational Photography methods involve taking multiple sequential photos by changing scene parameters and fusing the photos to create a richer representation. The goal of Coded Computational Photography is to modify the optics, illumination or sensors at the time of capture so that the scene properties are encoded in a single (or a few) photographs. We describe several applications of coding exposure, aperture, illumination and sensing and describe emerging techniques to recover scene parameters from coded photographs.

1 Introduction

Computational photography combines plentiful computing, digital sensors, modern optics, actuators, and smart lights to escape the limitations of traditional cameras, enables novel imaging applications and simplifies many computer vision tasks. 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.

In this paper, we discuss Coded Photography which involves encoding of the photographic signal and post-capture decoding for improved scene analysis. With film-like photography, the captured image is a 2D projection of the scene. Due to limited capabilities of the camera, the recorded image is a partial representation of the view. Nevertheless, the captured image is ready for human consumption: what you see is what you almost get in the photo. In Coded Photography, the goal is to achieve a potentially richer representation of the scene during the encoding process. In some cases, Computational Photography reduces 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 a different exposure, focus, aperture, view, illumination, or instant of capture. Each setting allows recording of partial information about the scene and the final image is reconstructed from these multiple observations. In Coded Computational Photography, the recorded image may appear distorted or random to a human observer. But the corresponding decoding recovers valuable information about the scene.

'Less is more' in Coded Photography. By blocking light over time or space, we can preserve more details about the scene in the recorded single photograph. In this paper we look at four specific examples.

(a) Coded Exposure: By blocking light in time, by fluttering the shutter open and closed in a carefully chosen binary sequence, we can preserve high spatial frequencies of fast moving objects to support high quality motion deblurring.

(b) Coded Aperture Optical Heterodyning: By blocking light near the sensor with a sinusoidal grating mask, we can record 4D light field on a 2D sensor. And by blocking light with a mask at the aperture, we can extend the depth of field and achieve full resolution digital refocussing.

(c) Coded Illumination: By observing blocked light at silhouettes, a multi-flash camera can locate depth discontinuities in challenging scenes without depth recovery.

(d) Coded Sensing: By sensing intensities with lateral inhibition, a gradient sensing camera can record large as well as subtle changes in intensity to recover a high-dynamic range image.

We describe several applications of coding exposure, aperture, illumination and sensing and describe emerging techniques to recover scene parameters from coded photographs.

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.

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).
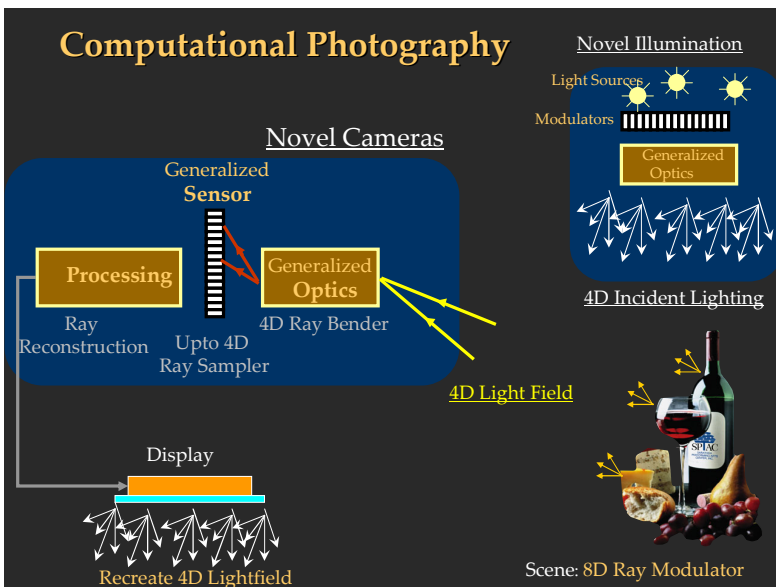

Fig. 1. Elements of Computational Photography

Computational Photography generalizes these four elements.

(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 commun-ication, 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.

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].

Coded Exposure   **Coded Aperture**

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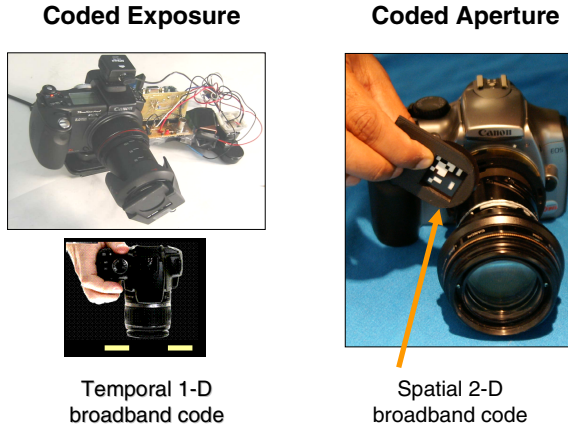
Fig. 2. Blocking light to achieve Coded Photography. (Left) Using a 1-D code in time to block and unblock light over time, a coded exposure photo can reversibly encode motion blur (Raskar et al 2006). (Right) Using a 2-D code in space to block parts of the light via a masked aperture, a coded aperture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

2.2 Coded Photography

But there is much more beyond the 'best possible film camera'. 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 and exposure duration.

Coded Photography is a notion of an 'out-of-the-box' photographic method, in which individual (ray) samples or data sets may or may not be comprehensible as 'images' without further decoding, re-binning or reconstruction. 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, active illumination and sensors.
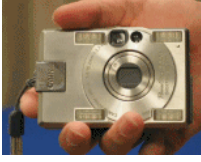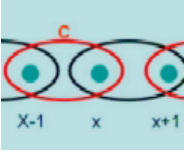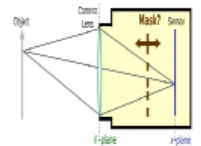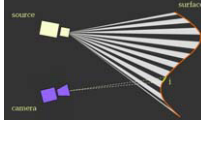
Coding in Time	Coding in Space	Coded Illumination	Coded Sensing
Exposure	Aperture	Inter-View	Gradient Sensor (Differential Encoding)
[Raskar et al 2006]	[Veeraraghavan et al 07]	[Raskar et al 2004]	[Tumblin et al 2005]
	Mask, Optical Heterodyning	Intra-view	
	[Veeraraghavan et al 07]	[Nayar et al 2006]	

Fig. 3. An overview of projects. Coding in time or space, coding the incident active illumination and coding the sensing pattern.

3   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. We 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.

Fig. 4. The flutter shutter camera. The coded exposure is achieved by fluttering the shutter open and closed. Instead of a mechanical movement of the shutter, we used a ferro-electric LCD in front of the lens. It is driven opaque and transparent according to the desired binary sequence.

4   Coded Aperture and Optical Heterodyning

Can we capture additional information about a scene by inserting a patterned mask inside a conventional camera? We use a patterned attenuating mask to encode the light field entering the camera. Depending on where we put the mask, we can effect desired frequency domain modulation of the light field. If we put the mask near the lens aperture, we can achieve full resolution digital refocussing. If we put the mask near the sensor, we can recover a 4D light field without any additional lenslet array.
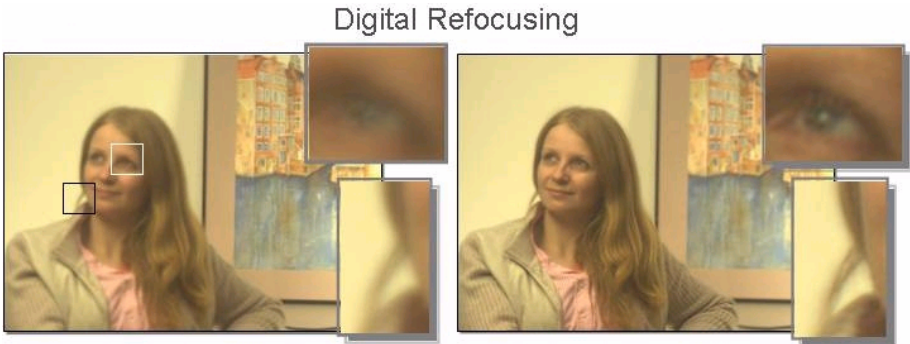

Fig. 5. Encoded Blur Camera, i.e. with mask in the aperture, can preserve high spatial images frequencies in the defocus blur. Notice the glint in the eye. In the misfocused photo, on the left, the bright spot appears blurred with the bokeh of the chosen aperture (shown in the inset). In the deblurred result, on the right, the details on the eye are correctly recovered.

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.

Our group has shown that 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.

This is reversible modulation of 4D light field by inserting a patterned planar mask in the optical path of a lens based camera. We can reconstruct the 4D light field from a 2D camera image. The patterned mask attenuates light rays inside the camera instead of bending them, and the attenuation recoverably encodes the ray on the 2D sensor. Our mask-equipped camera focuses just as a traditional camera might to capture conventional 2D photos at full sensor resolution, but the raw pixel values also hold a modulated 4D light field. The light field can be recovered by rearranging the tiles of the 2D Fourier transform of sensor values into 4D planes, and computing the inverse Fourier transform.
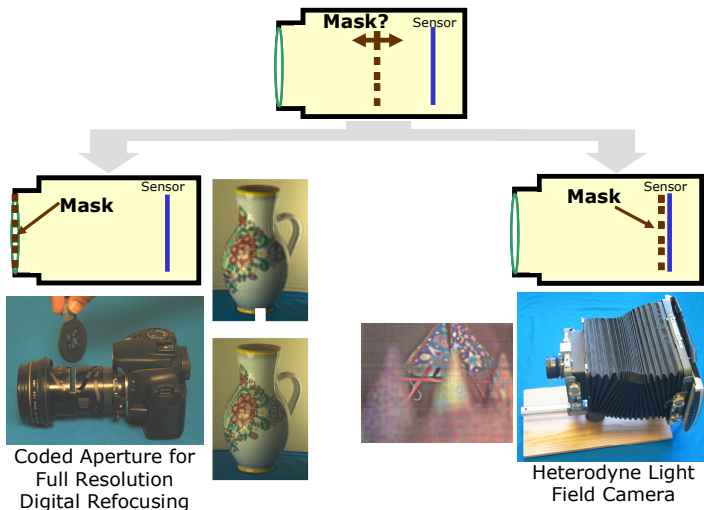

Fig. 6. Coding Light Field entering a camera via a mask

5 Coded Illumination

By observing blocked light at silhouettes, a multi-flash camera can locate depth discontinuities in challenging scenes without depth recovery. We used a multi-flash camera to find the silhouettes in a scene [Raskar et al 2004]. We take four photos of an object with four different light positions (above, below, left and right of the lens). We 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. We also demonstrate silhouette detection in a video using a repeated fast sequence of flashes.
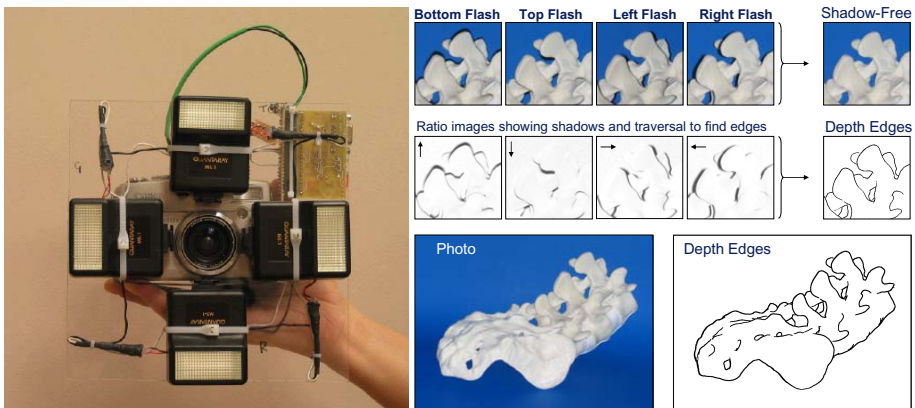

Fig. 7. Multi-flash Camera for Depth Edge Detection. (Left) A camera with four flashes. (Right) Photos due to individual flashes, highlighted shadows and epipolar traversal to compute the single pixel depth edges.

6 High Dynamic Range Using a Gradient Camera

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. 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.

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.

By sensing intensities with lateral inhibition, a gradient sensing camera can record large as well as subtle changes in intensity to recover a high-dynamic range image. By sensing different between neighboring pixels instead of actual intensities, our group has shown that a 'Gradient Camera' can record large global variations in intensity [Tumblin et al 2005]. 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.

7 Conclusion

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 or computer vision. Microscopy, tomography, astronomy and other optically driven fields already contain some ready-to-use solutions to borrow and extend. If the goal of photography is to capture, reproduce, and manipulate a meaningful visual experience, then the camera is not sufficient to capture even the most rudimentary birthday party. The human experience and our personal viewpoint is missing. Computational Photography can supply us with visual experiences, but can't decide which one's matter most to humans. Beyond coding the first order parameters like exposure, focus, illumination and sensing, maybe the ultimate goal of Computational Photography is to encode the human experience in the captured single photo.

Acknowledgements

References

Raskar, R., Tan, K., Feris, R., Yu, J., Turk, M.: Non-photorealistic Camera: Depth Edge Detection and Stylized Rendering Using a Multi-Flash Camera. SIGGRAPH 2004 (2004)

T umblin, J., Agrawal, A., Raskar, R.: Why I want a Gradient Camera. In: CVPR 2005, IEEE, Los Alamitos (2005)

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

Veeraraghavan, A., Raskar, R., Agrawal, A., Mohan, A., Tumblin, J.: Dappled Photography: Mask-Enhanced Cameras for Heterodyned Light Fields and Coded Aperture Refocusing. ACM Siggraph (2007)

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

Debevec, Malik.: Recovering high dynamic range radiance maps from photographs. In: Proc. SIGGRAPH (1997)

Mann, Picard.: Being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures. In: Proc. IS&T 46th ann. conference (1995)

McGuire, M., Matusik, Pfister, Hughes, Durand.: Defocus Video Matting, ACM Transactions on Graphics. Proceedings of ACM SIGGRAPH 2005 24(3) (2005)

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

Ng, R.: Fourier Slice Photography, SIGGRAPH (2005)

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

Levoy, M., Hanrahan, P.: Light field rendering. In: SIGGRAPH, pp. 31–42 (1996)

Dowski Jr., E.R., Cathey, W.T.: Extended depth of field through wave-front coding. Applied Optics 34(11), 1859–1866 (1995)

Georgiev, T., Zheng, C., Nayar, S., Salesin, D., Curless, B., Intwala, C.: Spatio-angular Resolution Trade-Offs in Integral Photography. In: proceedings, EGSR 2006 (2006)