View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by DSpace@MIT

DOI: 10.1111/j.1467-8659.2009.01648.x

EUROGRAPHICS 2010 / T. Akenine-Möller and M. Zwicker (Guest Editors)

*Volume 29 (2010), Number 2*

# Reinterpretable Imager: Towards Variable Post-Capture Space, Angle and Time Resolution in Photography

Amit Agrawal[1], Ashok Veeraraghavan[1] and Ramesh Raskar[2]

[1]Mitsubishi Electric Research Labs (MERL), Cambridge, MA, USA
[2]MIT Media Lab, Cambridge, MA, USA

**Abstract**

*We describe a novel multiplexing approach to achieve tradeoffs in space, angle and time resolution in photography. We explore the problem of mapping useful subsets of time-varying 4D lightfields in a single snapshot. Our design is based on using a dynamic mask in the aperture and a static mask close to the sensor. The key idea is to exploit scene-specific redundancy along spatial, angular and temporal dimensions and to provide a programmable or variable resolution tradeoff among these dimensions. This allows a user to reinterpret the single captured photo as either a high spatial resolution image, a refocusable image stack or a video for different parts of the scene in post-processing.*

*A lightfield camera or a video camera forces a-priori choice in space-angle-time resolution. We demonstrate a single prototype which provides flexible post-capture abilities not possible using either a single-shot lightfield camera or a multi-frame video camera. We show several novel results including digital refocusing on objects moving in depth and capturing multiple facial expressions in a single photo.*

Categories and Subject Descriptors (according to ACM CCS): I.4.1 [Computer Graphics]: Digitization and Image Capture—Sampling

## 1. Introduction

Multiplexing techniques allow cameras to go beyond capturing a 2D photo and capture additional dimensions or information, leading to post-processing outputs not possible with traditional photography. These techniques usually trade-off one image parameter for another, e.g., spatial resolution for angular resolution in lightfield cameras to support digital refocusing [NLB*05, GZN*06] and pupil plane multiplexing to capture wavelength and polarization information by reducing spatial resolution [HEAL09]. Similarly, high speed cameras tradeoff spatial resolution for temporal resolution. In this paper, we describe a novel multiplexing technique which also allows capturing temporal information along with angular information in a single shot. Unlike traditional multiplexing techniques, the resolution tradeoff is not fixed, but is scene dependent. We show that this leads to two novel post-processing outputs: (a) digital refocusing on an object moving in depth, and (b) low spatial resolution video from a single photo.

Mapping angular variations in rays to spatial intensity variations is well-known for lightfield capture. This has been done by inserting a micro-lens array [NLB*05] as well as a high frequency mask [VRA*07] close to the sensor. We use a time-varying mask in the aperture to control angular variations and a static mask near the sensor (similar to [VRA*07]) that enables capture of those angular variations. Simultaneously modifying lens aperture and sensor-optics has been used for encoding color. Kodachrome films used a rainbow filter to map wavelength variations to angular variations, and then a lenticular-pattern on sensor to record colors to separate pixels. To the best of our knowledge, mask pattern manipulation for mapping temporal variations to angular variations and encoding a video clip in a single photo have been unexplored.

We show that we can encode angular as well as temporal variations of a scene in a single photo. We modulate the mask in the aperture *within* a single exposure to encode the angular and temporal ray variations. A important character-

© 2010 The Author(s)
Journal compilation © 2010 The Eurographics Association and Blackwell Publishing Ltd.
Published by Blackwell Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA.

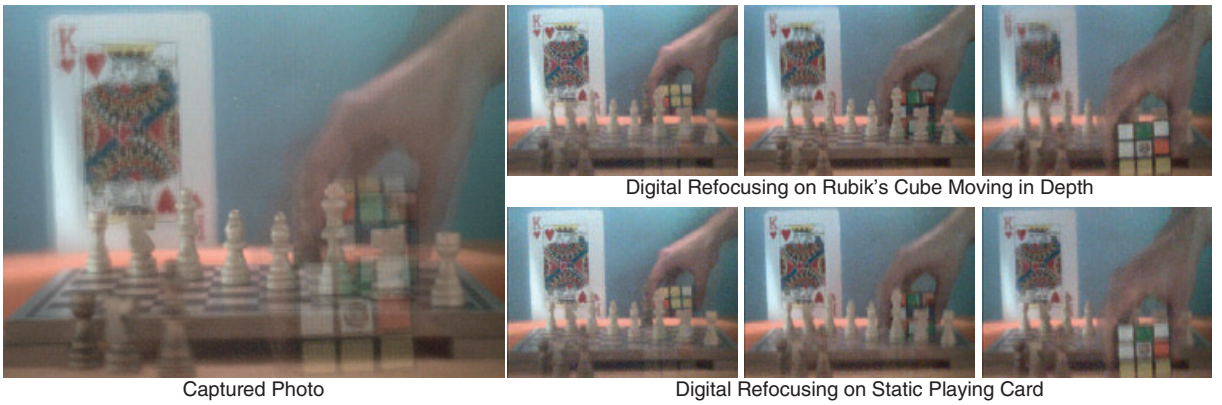


**Figure 1:** *We show how to achieve digital refocusing on both static and moving objects in the scene. (Left) Captured photo. (Right) Low spatial resolution digitally refocused images. Top row shows that the playing card and chess pieces go in and out of focus as the Rubik's cube moving in depth is digitally refocused. Note the correct occlusions between the Rubik's cube and the static objects. Bottom row shows digital refocusing on the static playing card in the back. Notice that the moving cube is focus blurred, but not motion blurred.*

istic of our design is that it does not waste samples if the scene does not have information along specific dimensions. Thus, it allows *scene dependent* variable resolution tradeoffs. For example, if the scene is static, we automatically obtain a 4D lightfield of the scene as would have captured by other lightfield cameras. If the scene is in-focus but is changing over time, the captured photo can be converted into a low spatial resolution video. If the scene is static and also within the depth of field of the lens, the captured photo gives the full spatial resolution 2D image of the scene. Thus, we are able to *reinterpret* the pixels in multiple ways among spatial, angular and temporal dimensions depending on the scene.

This differentiates our design from previous lightfield cameras and a traditional video camera, where a fixed resolution tradeoff is assumed at the capture time. While temporal variations in scene could be better captured using multiple sequential images or using a video camera, a video camera does not allow digital refocusing. Similarly, previous lightfield cameras allow digital refocusing, but cannot handle dynamic scenes. Our design provides the flexibility to capture both temporal and angular variations in the scene, not supported by any existing cameras. In addition, it also allows *variable* resolution tradeoffs in spatial, angular and temporal dimensions depending on the scene, while previous approaches allowed fixed scene independent resolution tradeoffs. In this paper, we conceptualize that such a tradeoff is possible in a single shot, propose an optical design to achieve it and demonstrate it by building a prototype.

We show that the simplest time-varying mask to achieve such a modulation consist of moving a finite size pinhole across the aperture within the exposure time. This scheme maps temporal variations in the scene to angular variations in the aperture, which are subsequently captured by the static mask near the sensor. This allows lightfield capture for static scene, video for in-focus dynamic scene (lightfield 'views' now correspond to low spatial resolution temporal frames) and 1-D refocusing on objects moving in depth. We also show that one can exploit the redundancy in Lambertian scenes to capture the temporal variations along the horizontal aperture dimension, and angular variations along the vertical aperture dimension by moving a vertical slit in the aperture. This allows 1-D refocusing on moving objects.

### 1.1. Contributions

Our contributions are as follows:

- We conceptualize the notion of simultaneously capturing both angular and temporal variations in a scene in a single shot.
- We propose a mask based optical design to achieve spatio-angular-temporal tradeoffs using a time-varying aperture mask and a static mask close to the sensor. Our design allows variable resolution tradeoff depending on the scene.
- We develop a prototype camera (reinterpretable imager) that can provide one of the three outputs from a single photo: video, lightfield or high resolution image. Further, different outputs can be obtained for different parts of the scene.
- Our design provides a unique mechanism for taking linear combinations of video frames optically in a single device.
- We demonstrate two novel post-processing outputs: (a) 1D refocusing on an object moving in depth and (b) single-shot video capture, not realizable by existing lightfield or video cameras.

### 1.2. Related work

**Lightfield capture:** To measure the directional intensity of rays, integral photography was proposed almost a cen-

© 2010 The Author(s)
Journal compilation © 2010 The Eurographics Association and Blackwell Publishing Ltd.

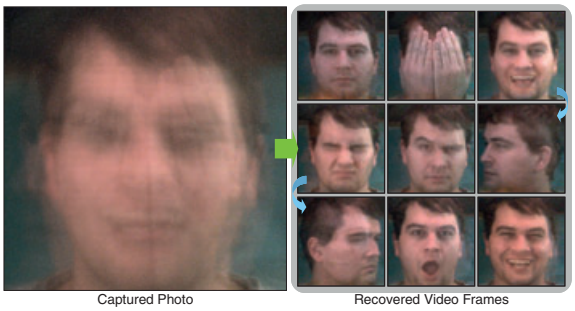


**Figure 2:** *Capturing multiple facial expressions in a single shot. (Left) Photo of a person showing different facial expressions within the exposure time of the camera. (Right) The* $3 \times 3$ *'views' of the recovered lightfield directly correspond to the* 9 *video frames.*

tury ago [Lip08, Ive28]. The concept of the 4D lightfield as a representation of all rays of light in free-space was introduced by Levoy and Hanrahan [LH96] and Gortler et al. [GGSC96]. In the pioneering work of Ng et al. [NLB*05], a focused micro-lens array was placed on top of the sensor. Each micro-lens samples the angular variations in the aperture at its spatial location, thereby capturing a low spatial resolution lightfield. Georgiev et al. [GZN*06] and Okano et al. [OAHY99] instead placed a combination of prisms and lenses in front of a main lens for juxtaposed sampling. Frequency domain modulation of lightfields was described in [VRA*07]. The modulated lightfield was captured by placing a sum of cosines mask close to the sensor. Our approach is inspired by these *single-shot* capture methods which lose spatial resolution to capture extra dimensions of the lightfield. A multi-image lightfield capture using dynamic masks in the aperture was shown in [LLW*08]. However, all these approaches are targeted towards 4D lightfields for static scenes and cannot handle dynamic scenes.

**Coding and multiplexing:** Multiplexed sensing has been used to increase the SNR during image capture. Schechner et al. [SNB03] proposed illumination multiplexing for increasing capture SNR using Hadamard codes. Improved codes that take into account sensor noise and saturation were described in [RS07]. Liang et al. [LLW*08] also used similar codes in aperture for multi-image lightfield acquisition. Coded aperture techniques use MURA codes [FC78] to improve capture SNR in non-visible imaging, invertible codes for out-of-focus deblurring for photography [VRA*07] and special codes for depth estimation [LFDF07]. Wavefront coding extends the depth of field (DOF) using cubic phase plates [DC95, CD02] in the aperture. Zomet and Nayar [ZN06] used an array of attenuating layers in a lensless setting for novel imaging applications such as split field of view, which cannot be achieved with a single lens. In [NM00, NN05], an optical mask with spatially varying transmittance was placed close to the sensor for high dynamic range imaging. Other imaging modulators include digital micro-mirror arrays [NBB04], holograms [SB05], and mirrors [FTF06].

**Motion photography:** Push-broom cameras and slit-scan photography [Dav] are used for finish-line photos and satellite imaging to avoid motion blur and to capture interesting motion distortions. A high speed camera can capture complete motion information, but is expensive, requires high bandwidth and does not allow digital refocusing on moving objects. Unlike techniques based on motion deblurring for removing blur, recovery of video frames in our approach does not require deblurring or knowledge of motion PSF allowing us to capture arbitrary scene changes within the exposure time. In [WJV*05], a dense array of low frame rate cameras were used for high speed motion photography. Our approach can also be viewed as a single camera that works as a low spatial resolution camera array to capture video in a single-shot for in-focus scene.

**Mapping methods:** Information in a non-geometric dimension can be captured by mapping it to a geometric dimension. Bayer filter mosaics [Bay76] map wavelength information directly to sensor pixels by losing spatial resolution. By using a rainbow in the aperture, wavelength (color) can be mapped to angular dimensions, which can be captured on a 2D image. Pupil-plane multiplexing [HEAL09] has been used for capturing polarization as well as color information. Our approach shows how multiplexing in aperture can be used to map temporal information to the spatial dimension using a lightfield camera.

## 2. Plenoptic function and mappings

The plenoptic function [AB91] describes the complete holographic representation of the visual world as the information available to an observer at any point in space and time. Ignoring wavelength and polarization effects, it can be described by *time-varying 4D lightfields* (TVLF) in free-space. Using the two-plane parametrization, let $(x,y)$ denote the sensor plane, $(\theta_x,\theta_y)$ denote the aperture plane and $L_0(x,y,\theta_x,\theta_y,t)$ denote the TVLF (Figure 3). Familiar structures of the visual world lead to redundancies in TVLF and we exploit this to capture useful *subsets* of TVLF for interesting applications.

Common optical devices essentially sample subsets of TVLF with underlying assumptions about the scene. For example, a traditional camera makes the inherent assumption that the scene is in-focus and static during the exposure time. Thus, it assumes absence of angular and temporal variations in TVLF and provides an adequate and accurate characterization of the resulting 2D subset under these assumptions. A video camera assumes that the scene is in-focus but changing over time. By assuming lack of angular variations, it provides an adequate characterization of the resulting 3D subset of the TVLF. A lightfield camera assumes absence of temporal variations and captures the 4D subset of the TVLF. When

© 2010 The Author(s)
Journal compilation © 2010 The Eurographics Association and Blackwell Publishing Ltd.

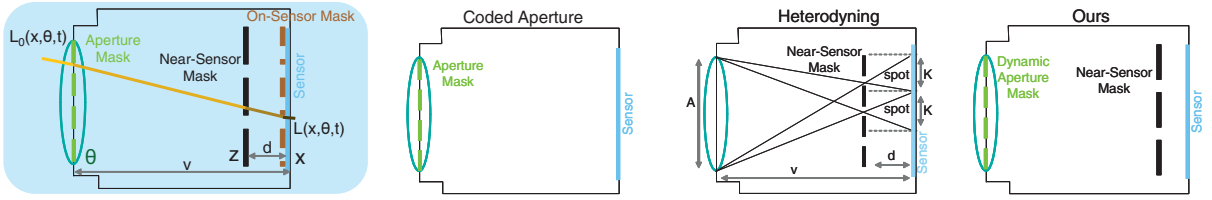


**Figure 3:** *General modulation of incoming light ray $L_0(x,\theta,t)$ can be achieved by placing a mask in the aperture, on-sensor and/or in the near-sensor plane. Coded aperture techniques use a mask in the aperture to control angular variations and achieve defocus PSF manipulation. Heterodyning employs a mask near-sensor to capture the incoming angular variations but does not control them in aperture. Our design uses a dynamic mask in the aperture to control the angular variations along with a static mask near-sensor to capture them, allowing variable tradeoff in spatial, angular and temporal resolution.*

the capture-time assumptions about the scene are not met, the acquired photos from these devices exhibit interesting and artistic artifacts such as focus blur, motion blur, highlights, specularities etc. In addition, the resolution tradeoff is decided at the capture time and cannot be modified depending on the scene. For example, if the scene was static, a video camera will continue capturing redundant frames at the same spatial resolution. It cannot provide a higher spatial resolution photo. Similarly, if the scene was dynamic, the output of a lightfield camera will be meaningless.

We show that one can have a single device that can act as a traditional camera, lightfield camera or video camera depending on the scene. The recovered resolution along different dimensions in each of these cases would be different, but the product of spatial, angular and temporal resolution is equal to the number of sensor pixels. The resolution tradeoff can be scene dependent and can vary across the image, i.e., different parts of the same photo can have different spatio-temporal-angular resolutions. The advantage is that with the same optical setup, we can tradeoff spatial, angular and temporal resolution as required by the scene properties. We believe that ours is the first system that allows such flexibility and show how to achieve it using a mask based design. Note that we capture up to 4D subsets of the TVLF and our design cannot capture the complete 5D information in TVLF.

### 2.1. Mapping methods

Any design to capture the information in TVLF onto a 2D sensor (single shot) must map the variations in angular and temporal dimensions into spatial intensity variations on the sensor. This can be achieved in following ways.

**Mapping angle to space:** The angular variations in rays can be captured by mapping it to spatial dimensions. This is well known, by placing lenslets or masks close to the sensor. A lenslet based design [NLB*05] maps individual rays to sensor pixels, thereby capturing the angular variations in the lightfield. A juxtaposed mapping can be achieved by placing an array of lenses outside the main lens [GZN*06]. The heterodyning design samples linear combination of rays at each sensor pixel, which can be inverted in frequency domain [VRA*07].

**Mapping time to space (direct):** Temporal variations can be mapped directly to the sensor by having controllable integration for each individual pixel within the exposure time. In order to capture $N$ low resolution frames in a single exposure time $T$, every $N^{th}$ pixel is allowed to integrate light only for $T/N$ time period. This is similar to Bayer mosaic filter, that maps wavelength to space. However, current sensor technology only allows controllable integration for *all* pixels simultaneously (IEEE DCAM Trigger mode 5).

**Mapping time to space (indirect):** To achieve time to space mapping, one can map temporal variations in rays to angles, in conjunction with mapping angle to space. Our design is based on this idea using a dynamic aperture mask and a static near-sensor mask.

## 3. Reinterpreting pixels for variable resolution tradeoffs

In this section, we describe our optical design using masks, which is shown in Figure 3. It consist of a mask in the aperture which is modulated within the exposure time and a mask close to the sensor. Effectively, this design *rebins* the rays on to pixels and the captured radiance is then interpreted as spatial, angular or temporal samples of TVLF. It differs from previous designs in the following ways. While a mask based heterodyning camera [VRA*07] captures the angular variations in the rays at the sensor plane, it does not modulate them in the aperture. It outputs lightfield for static scene, but in presence of motion, the output is unusable. On the other hand, [LLW*08] capture multiple images for lightfield reconstruction by changing the aperture mask for each image, without any mask close to the sensor. Such a design cannot handle dynamic scenes. In contrast, we modulate the aperture mask within a single exposure time as well as capture those variations using a static mask near the sensor.

We now describe our design using finite size pinhole masks. Note that for implementation, the static pinhole mask at the sensor can be replaced with a sum-of-cosines mask [VRA*07] or a tiled-broadband mask [LRAT08] to gain more light.

© 2010 The Author(s)
Journal compilation © 2010 The Eurographics Association and Blackwell Publishing Ltd.

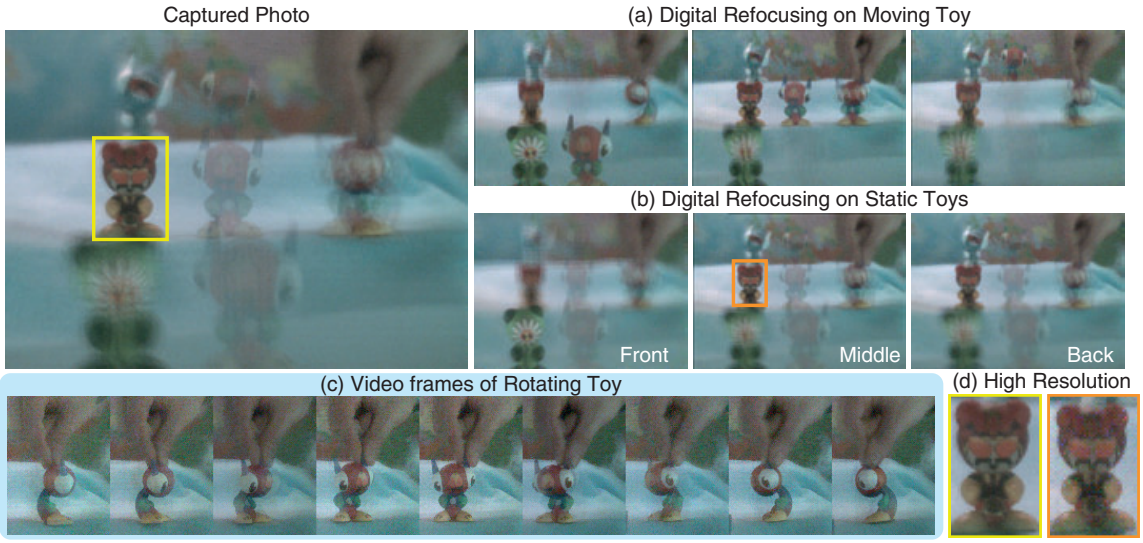


**Figure 4:** *From a single captured photo of a scene consisting of static and dynamic objects in and out-of-focus, we generate (a) 1D refocusing (vertical) on object moving towards the camera, (b) digital refocusing on static scene parts, (c) 9 frames of video for the in-focus rotating object, and (d) high spatial resolution image for the in-focus static object.*

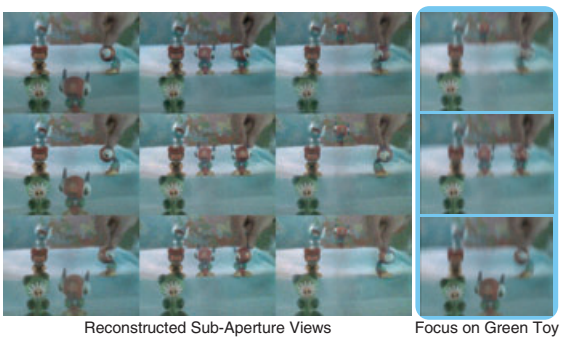


**Figure 5:** *Reconstructed* $3 \times 3$ *sub-aperture views from the captured photo shown in Figure 4. Notice that both static and dynamic objects are sharp in all sub-aperture views, without any defocus or motion blur. Right column shows novel rendering where focus is maintained on the static green toy, while the moving toy is brought in and out-of-focus without any motion blur.*

### 3.1. Optical design

Consider the heterodyning design shown in Figure 3, consisting of a static pinhole array placed at a distance $d$ from the sensor, such that the individual pinhole images (referred to as spots) do not overlap with each other (F-number matching). Each pinhole captures the angular variations across the aperture on $K \times K$ pixels. If the sensor resolution is $P \times P$, this arrangement allows capturing a lightfield with spatial resolution of $\frac{P}{K} \times \frac{P}{K}$ and angular resolution of $K \times K$. To map temporal variations to angular variations, the exposure time $T$ of the camera is sub-divided into $K^2$ slots of duration $T_\delta = \frac{T}{K^2}$ each, and the aperture is divided into a $K \times K$ grid. In each of the $K^2$ time slots, one of the $K^2$ grid location in the aperture is open, while others are closed. This modification of the heterodyning design with moving pinholes in the aperture achieves the objectives of post-capture flexibility about scene characteristics.

Figure 4 shows a visually rich scene consisting of static objects in and out of focus on the left, an object moving towards the camera in the center and an object rotating in the focus plane on the right. We will use this single captured photo to describe how different resolution tradeoffs can be made for different parts of the scene. For this example, $K = 3$.

### 3.2. Static scenes

It is easy to see that for single shot capture of a static scene, the dynamic aperture mask does not play any role except that of losing light. For static scene, since there are no temporal variations, moving the pinhole does not affect the angular variations of the rays over time. Each pinhole position captures a subset of the rays in the aperture and as long as the moving pinhole covers the entire aperture, all angular variations across the aperture are captured, albeit for lower time durations. Thus, the photo taken by moving a pinhole in the aperture is equivalent to the photo taken by keeping all the pinholes open for a reduced exposure time of $T_\delta$. In comparison with [VRA*07], the light gets attenuated by a factor of $K^2$ and the captured photo can be used to recover a lightfield with angular resolution $K \times K$ and spatial reso-

© 2010 The Author(s)
Journal compilation © 2010 The Eurographics Association and Blackwell Publishing Ltd.

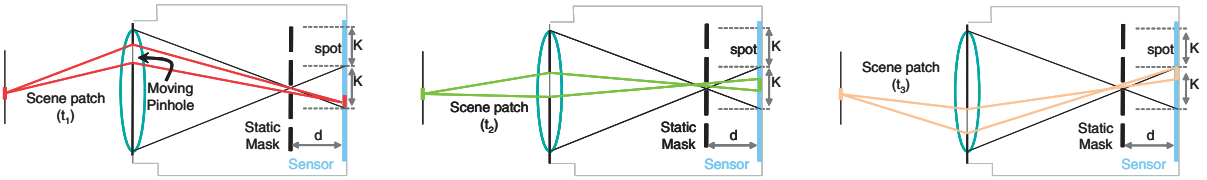


**Figure 6:** *Indirect mapping of time to space via angle. By dynamically changing the aperture mask, temporal variations in rays can be mapped to angular variations of the lightfield, which are captured using a mask near the sensor. The same scene patch changes color in three different time instants, which are mapped to different locations in the corresponding spot.*

lution $\frac{P}{K} \times \frac{P}{K}$. In Figure 4, we show that digital refocusing can be performed on static parts of the scene *independent* of the moving objects, which results in artifacts on the moving objects. The out of focus blur on static objects corresponds to the shape of the aperture. Figure 5 shows that all static objects are in-focus in the recovered sub-aperture views.

#### 3.2.1. In-focus static scene

Similar to previous mask based designs [VRA*07, RAWV08], we recover high resolution 2D image for in-focus static parts of the scene. As shown by these designs, a near-sensor mask simply attenuates the image of in-focus scene parts by a spatially varying attenuation pattern. This can be compensated by normalizing with a calibration photo of a uniform intensity Lambertian plane.

Now consider the additional effect of the moving pinhole mask in the aperture. For each in-focus scene point, the cone of rays focuses perfectly on a sensor pixel. Since the scene is static, this cone of rays does not change during the exposure time. The moving pinhole in the aperture simply allows different parts of the cone to enter the camera at different time slots. Moreover, since the scene is in-focus, all rays within the cone have the same radiance. Thus, the effect of moving pinhole is to attenuate the intensity by an additional factor of $K^2$. Figure 4 shows that the spatial resolution on the static object in focus (red toy on left) can be increased beyond the resolution in the refocused image obtained using the lightfield.

### 3.3. In-focus dynamic scenes

Now let us consider a more interesting case of a scene that is in-focus but changing arbitrarily over time. We assume that the scene changes at a rate comparable to $T_\delta$, and hence it remains static during each of the time slot. Note that this assumption is true for any video camera which assumes scene to be static within the time frame. Since the scene is in focus, the cone of rays at the aperture have the *same* radiance at any given time instant. By capturing *any* subset of this cone, we can record the scene radiance at that time instant, albeit at lower intensity. This fact can be utilized to capture different subsets of rays at different time instants to capture a dynamic scene, and the moving pinhole exactly achieves it. The effect of the moving pinhole mask at the aperture is to map rays at different time instants to different angles (indirect time to angle mapping) as shown in Figure 6. The 'views' of the captured lightfield (Figure 5) now automatically correspond to different frames of a video (temporal variations) at lower spatial resolution for the rotating object. This is further evident in the bottom row of Figure 4, which shows the cropped toy region. Thus, one can convert the captured $P \times P$ photo into $K^2$ temporal frames, each with a spatial resolution of $\frac{P}{K} \times \frac{P}{K}$ pixels.

### 3.4. Out of focus dynamic Lambertian scenes

Now consider the Lambertian object in the middle of Figure 4, which is moving towards the camera. Its motion results in both focus blur and motion blur in the captured photo. Recently, great progress has been made in tackling the problem of object motion blur [RAT06], camera shake [FSH*06] and focus blur [VRA*07, LFDF07]. Nevertheless, none of these approaches can handle motion blur and focus blur simultaneously, and they require PSF estimation and deblurring to recover the sharp image of the object. If parts of a dynamic scene are out-of-focus, we will not be able to differentiate between the temporal and angular variations since both of them will result in angular variations in the rays. Since we map temporal variations to angles, we cannot capture *both* temporal and angular variations *simultaneously*. However, we utilize redundancy in the TVLF to capture both these variations as follows.

In general, the lightfield has two angular dimensions. But for Lambertian scenes, since the apparent radiance of a scene point is same in all directions, the angular information is redundant. If we capture the angular information using only one dimension (1D parallax) by placing a *slit* in the aperture, the resulting 3D lightfield captures the angular information for Lambertian scenes. This 3D lightfield enables refocusing just as a 4D lightfield. But since the out-of-focus blur depends on the aperture shape, it will be 1D instead of being 2D for regular full aperture lightfields. The key idea then is to map the temporal variations in the scene to the 'extra' angular dimension available.

By moving a vertical slit horizontally in the aperture, we can map the temporal variations to the horizontal dimension

© 2010 The Author(s)
Journal compilation © 2010 The Eurographics Association and Blackwell Publishing Ltd.

| Coding Scheme | Captured Dimensions (space, angle, time) | Captured Resolution (space, angle, time) | Output |
|---|---|---|---|
| Static/Dynamic Aperture Mask | 2, 0, 0 | P*P, 0, 0 | 2D Photo |
| Static Near-Sensor Mask | 2, 0, 0 | P*P, 0, 0 | 2D Photo |
| | 2, 2, 0 | P/K*P/K, K*K, 0 | 4D Light Field |
| Static Near-Sensor Mask + Dynamic Aperture Mask | 2, 0, 0 | P*P, 0, 0 | 2D Photo |
| | 2, 2, 0 | P/K*P/K, K*K, 0 | 4D Light Field |
| | 2, 0, 1 | P/K*P/K, 0, $K^2$ | Video |
| | 2, 1, 1 | P/K*P/K, K, K | 1D Parallax + Motion |

**Figure 7:** *Comparison of various designs for single-shot capture on a 2D sensor having $P \times P$ pixel resolution.*

and angular variations to the vertical dimensions of the captured lightfield. Notice that for in-focus dynamic scene, temporal variations are mapped to both the angular dimensions of the lightfield (horizontal and vertical) by moving the pinhole as described in Section 3.3. Thus, the captured $P \times P$ photo again results in $K^2$ images of spatial resolution $\frac{P}{K} \times \frac{P}{K}$. But these $K^2$ images correspond to refocusing using $K$ angular samples *independently* for $K$ different instants in time. This allows digital refocusing on moving objects, as shown in Figure 4 for the object in center. Notice that the out of focus blur for static objects is now only in the vertical direction, since the vertical direction of the aperture was used to capture the angular variations. In comparison, digital refocusing on the static objects using the entire lightfield results in two dimensional defocus blur as shown in Figure 4. However, compared to the previous case, the temporal resolution is reduced to $K$ from $K^2$.

Thus, we showed how different parts of the same captured photo can have different resolutions in spatial, angular and temporal dimensions. In general, we capture up to 4D subsets of TVLF using our design. Figure 7 provides a summary of the captured resolution and dimensions for each of the above cases.

## 4. Applications

Now we show several novel results using our design assuming $K = 3$.

**Lightfield for static scene:** Figure 8 shows digital refocusing on a static scene. For static scenes, the mask in the aperture does not play any role except that of losing light. The lightfield encoding shows that higher spatial resolution is preserved for in-focus scene parts as compared to out-of-focus scene parts. Figure 9 shows recovery of high resolution image for in-focus scene parts from the captured photo along with the upsampled refocused image for comparison.

**Region adaptive output:** Figure 10 shows a scene with several static objects along with a rotating doll in the right. The $3 \times 3$ views recovered from the captured photo correspond to the 9 video frames for the rotating doll and $3 \times 3$ angular samples for the rest of the scene. Notice that the sharp features of the doll as well as the hand rotating it are recovered. For static scene, the angular samples allow digital refocusing on the front doll and the flower in the background as

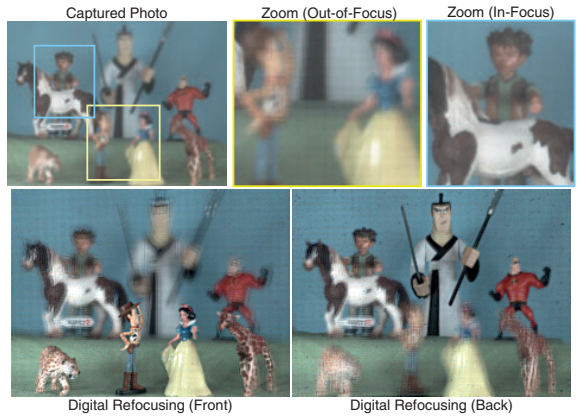


**Figure 8:** *Our design does not waste samples if the scene was static and provides a 4D lightfield. The zoom-in on the in-focus regions shows that high spatial resolution information is preserved for static in-focus parts using masks. Bottom row shows digital refocusing on the front and the back.*

shown. Thus, different outputs can be obtained for different parts of the scene.

**Digital refocusing on moving object:** Digital refocusing using lightfields have been demonstrated only for static scenes. Figure 1 show refocusing on an object moving in depth, by capturing 1D parallax+motion information. Notice the correct occlusions and dis-occlusions between the moving and static objects in Figure 1. This is a challenging example; object moving across the image is much easier to refocus on than object moving across depth due to change of focus. Notice that the resulting bokeh is 1D, since the 1D angular information is captured.

**Novel effects:** We can generate novel effects such as keeping a static object in sharp focus while bringing the moving object out of focus *without* any motion blur as shown in Figure 1 and Figure 5.

**Capturing facial expressions:** Family photographs as well as portraits are challenging to capture as the photographer may not take the snap at the *right* moment. The moment camera [CS06] continuously captures frames in a buffer to avoid taking a snapshot. Our technique can capture multiple facial expressions in a single photo as shown in Figure 2. The recovered frames can be combined using software techniques such as digital photomontage [ADA*04] for generating novel images.

## 5. Implementation and analysis

Our prototype is shown in Figure 11. We use a 22 megapixel medium-format Mamiya 645ZD digital back and place a sum-of-cosines mask [VRA*07] directly on top of the protective sensor glass. This setup allows a maximum aperture

© 2010 The Author(s)
Journal compilation © 2010 The Eurographics Association and Blackwell Publishing Ltd.

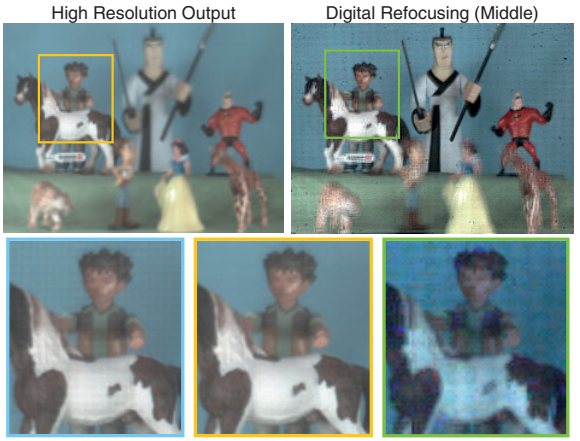


**Figure 9:** *By normalizing with a calibration photo, higher spatial resolution can be recovered for static in-focus scene parts. For comparison, the corresponding region from the refocused image obtained using the lightfield is also shown.*

of $f/8$ for $11 \times 11$ angular resolution and $242 \times 182$ color spatial resolution lightfield capture, similar to [RAWV08].

To implement time-varying aperture mask, a servo motor is used to drive a circular plastic wheel with appropriate patterns printed on it. The wheel is placed adjacent to a 100 mm focal length achromatic lens from Edmund Optics. Note that conventional camera lenses have their aperture plane *inside* the lens body which is difficult to access. Although [LLW*08] used a 50 mm conventional lens for coding in aperture, the constraint of $f/8$ as the maximum aperture size would lead to small aperture for us. More importantly, placing a mask on either side of the lens body typically leads to spatially varying defocus blur and vignetting, and would lead to spatially varying encoding of rays within each spot unsuitable for our application. Thus, we avoid using a conventional camera lens. An external shutter in front is synchronized with the rotating wheel to block light during its motion.

In our experiments, we use $K = 3$. The circular plastic wheel have $K^2 = 9$ patterns, each being a pinhole of size 3 mm$^2$. In each pattern, the location of the pinhole is changed to cover the entire aperture. The pinhole locations have spacing between them as shown in Figure 11 to avoid blurring between neighboring pinhole images due to diffraction. For Figure 1, we use a vertical slit as the aperture mask. Thus, $K = 3$ slit patterns were used within the exposure time as shown in Figure 11. We use up to 8 seconds exposure time for indoor scenes. We capture RAW images and use dcraw to get 16 bit linear Bayer pattern. Instead of color interpolation, we simply pick RGB in each $2 \times 2$ block. We did not observe any color issues with nearby pixels. For each experiment, we first recover the $3 \times 3$ angular resolution and $242 \times 182$ spatial resolution lightfield using frequency domain technique [VRA*07]. The recovered lightfield 'views' correspond to either $3 \times 3$ angular samples for static scene, 9 temporal frames for dynamic in-focus scene, or 3 angular samples with each having 3 temporal frames for dynamic out-of-focus scene.

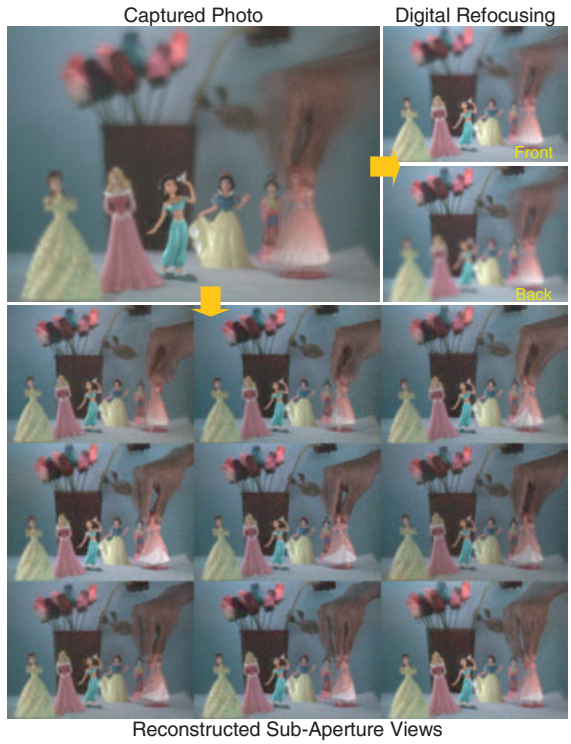


**Figure 10:** *Region adaptive output can be obtained from the captured photo. The reconstructed sub-aperture views (bottom) correspond to angular samples for the static scene and temporal samples for the rotating doll.*

**Dataset capture:** As noted above, in our implementation $K = 3$. Since the temporal resolution is low for our implementation, it is impossible to show any continuous motion. Figures 1, 2, 4 and 10 show discontinuous motion: objects were kept static during sub-aperture exposure times and were moved rapidly otherwise. However, note that this is not a fundamental restriction of the design, but rather a limitation of our implementation.

**Failure cases and artifacts:** Objects moving faster than our temporal sampling rate results in motion blur in decoded video frames as shown in Figure 12. Brightly lit moving objects could leave 'ghosts' on dark backgrounds during lightfield reconstruction due to low SNR on dark regions. Misalignment of the masks on the wheel could cause additional blurring/ghosting. Non-Lambertian, transparent and translucent objects would cause additional angular variations in the rays and would lead to artifacts in reconstructed sub-aperture lightfield views. For in-focus scene, the viewpoint of the re-

© 2010 The Author(s)
Journal compilation © 2010 The Eurographics Association and Blackwell Publishing Ltd.

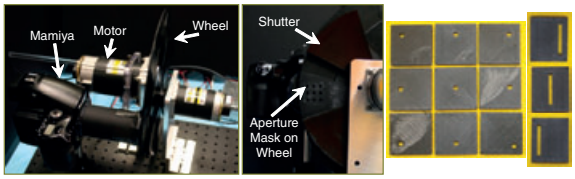


**Figure 11:** *Our prototype uses a motor driven wheel to dynamically change the aperture mask in a stop and go fashion, along with a heterodyne mask placed on the Mamiya digital back. Shown are* 9 *pinhole aperture-masks and* 3 *slit aperture masks.*

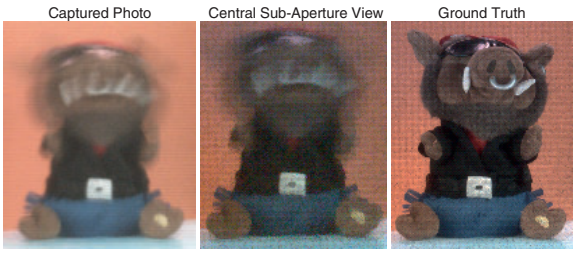


**Figure 12:** *Objects moving faster than the temporal sampling rate results in motion blur in the recovered frames, similar to a video camera.*

covered video frames are slightly shifted due to the moving pinhole in the aperture, which could be accounted for based on focal plane distance and aperture size.

## 6. Discussions

We now discuss the benefits and limitations of our design. Let $T$ be the total exposure time of the captured photo.

**Temporal resolution:** The maximum achievable temporal resolution is limited by the angular resolution provided by the mask close to the sensor. Thus, if the angular resolution of the design is $K \times K$, maximum number of distinct frames that can be captured for in-focus dynamic scene is $K^2$. Using the mechanical setup, the effective integration time of each sub-aperture view is $T/K^2$, and is thus *coupled* with the number of frames. The effective integration time decreases with increasing number of frames for same $T$. In contrast, for a video camera, increasing the frame rate (not the number of frames) reduces the effective integration time.

**Light loss:** In comparison with a single-shot lightfield camera having the same exposure time and angular resolution, each sub-aperture view integrates light for a duration of $T/K^2$ as opposed to $T$. For a video camera capturing $K^2$ frames within $T$, the exposure duration of each frame is $T/K^2$, same as ours. However, the effective aperture size for each frame in our design is reduced by $K^2$. Thus, the light loss in both cases is a factor of $K^2$, and increases with the temporal/angular resolution. Conversely, to capture the same amount of light, $K^2$ times longer exposure duration is required. This will increase dark noise and enhance temperature dependent signal degradations in the sensor. However, a lightfield camera cannot handle dynamic scenes and video cameras do not support digital refocusing. Unlike a video camera or a lightfield camera, in absence of variations along temporal or angular dimensions, resolution is not wasted in our design (at the expense of light loss). While a video camera is best suited for capturing temporal variations at fixed spatial resolution, our design provides the user with more flexibility in post-capture decisions for computational photography.

**Video lightfield camera:** A brute force way to capture time-varying lightfields would be to design a video lightfield camera (burst-mode SLR with lenslet/mask close to the sensor), which would offer fixed resolution tradeoff and would require large bandwidth. Our design offers an alternative using a fast modulation device in the aperture in contrast to a high bandwidth sensor. The benefit of video lightfield camera is that full 5D information can be captured, whereas our approach can capture only 4D subsets of TVLF. Dynamic scenes with non-Lambertian objects and lens interreflections [RAWV08] will cause artifacts in capturing temporal variations in our design.

Our design uses the available $K \times K$ angular samples either for $K \times K$ angular resolution, $K^2$ temporal resolution or $K$ angular with $K$ temporal resolution, depending on the scene properties. In contrast, to provide $K^2$ temporal resolution, a video lightfield camera would require $K^2$ times more bandwidth. Thus, our design could be useful in limited bandwidth scenarios. If bandwidth is not an issue, a video lightfield camera will provide greater light benefit than ours. Camera arrays [WJV*05] can also capture video lightfields with reduced bandwidth, but require extensive hardware compared to our design.

**LCD's for modulation:** Though our in-house prototype has a low temporal resolution, a LCD in the aperture plane can be used to increase temporal resolution. Commercially available color LCD's and monochrome LCD's used in projectors lead to diffraction when used in the aperture plane due to the RGB filters and/or small pixel size. Off the shelf monochrome LCD's have a larger pixel size, but have lower contrast ratio which significantly reduces the capture SNR [ZN06]. However, low-cost 'single' pixel FLC shutters (DisplayTech) can provide switching rate of 1000Hz and contrast ratio of 1:1000, and application specific custom LCD solutions can be designed.

**Conclusions:** In this paper, we have taken an initial step in the direction of providing variable resolution tradeoffs along spatial, angular and temporal dimensions in post-processing. We conceptualize that such tradeoffs are possible in a single shot and analyzed these tradeoffs in capturing different 4D subsets of time-varying lightfield. We utilize the redundancy in Lambertian scenes for capturing simultaneous angular and temporal ray variations, required to handle motion

© 2010 The Author(s)
Journal compilation © 2010 The Eurographics Association and Blackwell Publishing Ltd.

blur along with focus blur. Using dynamic masks in the camera, we demonstrated how we can obtain a video, 4D light-field or high resolution image from a single captured photo depending on the scene, without prior scene knowledge at the capture time. In medical and scientific microscopy, the ability to refocus on moving objects will be beneficial. Our current masks are simple attenuators but in the future, angle dependent holographic optical elements may support a full capture of the 5D plenoptic function. We hope our approach will lead to further research in innovative devices for capturing the visual experience and will inspire a range of new software tools to creatively unravel the captured photo.

**Acknowledgements** We thank the anonymous reviewers and several members of MERL for their suggestions. We also thank Brandon Taylor, John Barnwell, Jay Thornton, Keisuke Kojima, Joseph Katz, along with Haruhisa Okuda and Kazuhiko Sumi, Mitsubishi Electric, Japan for their help and support.

## References

[AB91] ADELSON E., BERGEN J.: The plenoptic function and the elements of early vision. *Computational Models of Visual Processing, MIT Press* (1991), 3–20.

[ADA*04] AGARWALA A., DONTCHEVA M., AGRAWALA M., DRUCKER S., COLBURN A., CURLESS B., SALESIN D., COHEN M.: Interactive digital photomontage. *ACM Trans. Graph. 23*, 3 (2004), 294–302.

[Bay76] BAYER B. E.: Color imaging array. US Patent 3,971,065, July 1976.

[CD02] CATHEY W. T., DOWSKI E. R.: A new paradigm for imaging systems. *Appl. Optics 41* (2002), 6080–6092.

[CS06] COHEN M., SZELISKI R.: The moment camera. *Computer 39* (Aug. 2006), 40–45.

[Dav] DAVIDHAZY A.: Slit-scan photography. http://people.rit.edu/andpph/text-slit-scan.html.

[DC95] DOWSKI E. R., CATHEY W.: Extended depth of field through wavefront coding. *Appl. Optics 34*, 11 (Apr. 1995), 1859–1866.

[FC78] FENIMORE E., CANNON T.: Coded aperture imaging with uniformly redundant arrays. *Appl. Optics 17* (1978), 337–347.

[FSH*06] FERGUS R., SINGH B., HERTZMANN A., ROWEIS S. T., FREEMAN W. T.: Removing camera shake from a single photograph. *ACM Trans. Graph. 25*, 3 (2006), 787–794.

[FTF06] FERGUS R., TORRALBA A., FREEMAN W.: *Random lens imaging*. Tech. rep., MIT, 2006.

[GGSC96] GORTLER S., GRZESZCZUK R., SZELISKI R., COHEN M.: The lumigraph. In *SIGGRAPH* (1996), pp. 43–54.

[GZN*06] GEORGIEV T., ZHENG C., NAYAR S., CURLESS B., SALASIN D., INTWALA C.: Spatio-angular resolution trade-offs in integral photography. In *EGSR* (2006), pp. 263–272.

[HEAL09] HORSTMEYER R., EULISS G., ATHALE R., LEVOY M.: Flexible multimodal camera using a light field architecture. In *ICCP* (Apr. 2009).

[Ive28] IVES H.: Camera for making parallax panoramagrams. *J. Opt. Soc. of America 17* (1928), 435–439.

[LFDF07] LEVIN A., FERGUS R., DURAND F., FREEMAN W. T.: Image and depth from a conventional camera with a coded aperture. *ACM Trans. Graph. 26*, 3 (2007), 70.

[LH96] LEVOY M., HANRAHAN P.: Light field rendering. In *SIGGRAPH 96* (1996), pp. 31–42.

[Lip08] LIPPMANN G.: Epreuves reversible donnant la sensation du relief. *J. Phys 7* (1908), 821–825.

[LLW*08] LIANG C.-K., LIN T.-H., WONG B.-Y., LIU C., CHEN H.: Programmable aperture photography: Multiplexed light field acquisition. *ACM Trans. Graphics 27*, 3 (2008), 55:1–55:10.

[LRAT08] LANMAN D., RASKAR R., AGRAWAL A., TAUBIN G.: Shield fields: modeling and capturing 3d occluders. *ACM Trans. Graph. 27*, 5 (2008), 1–10.

[NBB04] NAYAR S. K., BRANZOI V., BOULT T.: Programmable imaging using a digital micromirror array. In *CVPR* (2004), vol. 1, pp. 436–443.

[NLB*05] NG R., LEVOY M., BR DIF M., DUVAL G., HOROWITZ M., HANRAHAN P.: *Light Field Photography with a Hand-held Plenoptic Camera*. Tech. rep., Stanford Univ., 2005.

[NM00] NAYAR S., MITSUNAGA T.: High dynamic range imaging: spatially varying pixel exposures. In *CVPR* (2000), vol. 1, pp. 472–479.

[NN05] NARASIMHAN S., NAYAR S.: Enhancing Resolution along Multiple Imaging Dimensions using Assorted Pixels. *IEEE Trans. Pattern Anal. Machine Intell. 27*, 4 (Apr 2005), 518–530.

[OAHY99] OKANO F., ARAI J., HOSHINO H., YUYAMA I.: Three dimensional video system based on integral photography. *Optical Engineering 38* (1999), 1072–1077.

[RAT06] RASKAR R., AGRAWAL A., TUMBLIN J.: Coded exposure photography: motion deblurring using fluttered shutter. *ACM Trans. Graph. 25*, 3 (2006), 795–804.

[RAWV08] RASKAR R., AGRAWAL A., WILSON C. A., VEERARAGHAVAN A.: Glare aware photography: 4d ray sampling for reducing glare effects of camera lenses. *ACM Trans. Graph. 27*, 3 (2008), 1–10.

[RS07] RATNER N., SCHECHNER Y. Y.: Illumination multiplexing within fundamental limits. In *CVPR* (June 2007).

[SB05] SUN W., BARBASTATHIS G.: Rainbow volume holographic imaging. *Optics Letters 30* (2005), 976–978.

[SNB03] SCHECHNER Y. Y., NAYAR S. K., BELHUMEUR P. N.: A theory of multiplexed illumination. In *ICCV* (2003), vol. 2, pp. 808–815.

[VRA*07] VEERARAGHAVAN A., RASKAR R., AGRAWAL A., MOHAN A., TUMBLIN J.: Dappled photography: Mask enhanced cameras for heterodyned light fields and coded aperture refocusing. *ACM Trans. Graph. 26*, 3 (2007), 69.

[WJV*05] WILBURN B., JOSHI N., VAISH V., TALVALA E.-V., ANTUNEZ E., BARTH A., ADAMS A., HOROWITZ M., LEVOY M.: High performance imaging using large camera arrays. *ACM Trans. Graph. 24*, 3 (2005), 765–776.

[ZN06] ZOMET A., NAYAR S.: Lensless imaging with a controllable aperture. In *CVPR* (2006), pp. 339–346.

© 2010 The Author(s)
Journal compilation © 2010 The Eurographics Association and Blackwell Publishing Ltd.