

Programmable Aperture Photography: Multiplexed Light Field Acquisition

Chia-Kai Liang*

Tai-Hsu Lin

Bing-Yi Wong

Chi Liu

Homer H. Chen†

National Taiwan University

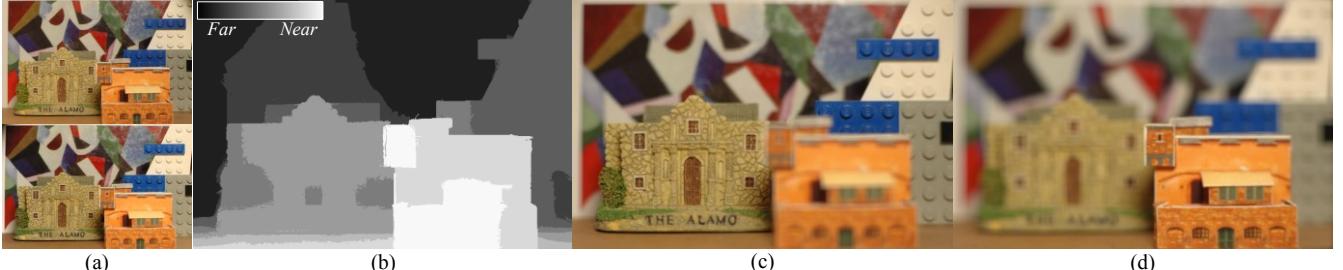


Figure 1: (a) Two demultiplexed light field images generated by the proposed system. The full 4D resolution is $4 \times 4 \times 3039 \times 1404$. (b) The estimated depth map of the top image of (a). (c-d) Post-exposure refocused images generated from the light field and the depth maps.

Abstract

In this paper, we present a system including a novel component called programmable aperture and two associated post-processing algorithms for high-quality light field acquisition. The shape of the programmable aperture can be adjusted and used to capture light field at full sensor resolution through multiple exposures without any additional optics and without moving the camera. High acquisition efficiency is achieved by employing an optimal multiplexing scheme, and quality data is obtained by using the two post-processing algorithms designed for self calibration of photometric distortion and for multi-view depth estimation. View-dependent depth maps thus generated help boost the angular resolution of light field. Various post-exposure photographic effects are given to demonstrate the effectiveness of the system and the quality of the captured light field.

1 Introduction

Computational photography is changing the way we capture images. While traditional photography simply captures 2D projection of 3D world, computational photography captures additional information by using generalized optics. The captured image may not be visually attractive, but together with the additional information, it enables novel post-processings that can deliver quality images and, more importantly, generate data such as scene geometry that were unobtainable in the past.

Light field acquisition is of fundamental importance among all aspects of computational photography. A complete 4D light field contains most visual information of a scene and allows various photographic effects to be generated in a physically correct way. However, existing light field cameras [Ng et al. 2005; Georgiev et al.

2006; Georgiev et al. 2007; Veeraraghavan et al. 2007] manipulate the light rays by means of lens arrays or attenuating masks that trade the spatial resolution for the angular resolution. Even with the latest sensor technology, one can hardly generate a light field with megapixel spatial resolution.

An alternative approach called programmable aperture places a programmable non-refractive mask at the aperture [Liang et al. 2007]. It exploits the fast multiple-exposure feature of digital sensors without trading off the sensor resolution to capture the light field sequentially, which, in turn, enables the multiplexing of light rays. However, the original implementation [Liang et al. 2007] is at best a proof of concept and is not compact and accurate enough for practical purposes. Besides, the traditional Hadamard-code based multiplexing fails in the presence of shot noise that appears in most digital sensors, resulting in poor data quality.

In this paper, we address these issues by employing an optimal multiplexing scheme and by designing two new prototypes of the programmable aperture using pattern scroll and liquid crystal array. The new system has the following advantages (Sec. 4):

- Better acquisition efficiency and higher flexibility angular resolution due to the optimal multiplexing scheme.
- Adjustable angular resolution and pre-filter kernel size. When the angular resolution is set to one, the light field camera becomes a conventional camera.
- Compact and economic realization with precise control. The prototypes can be placed in, and nicely integrated with, a conventional camera. Each prototype costs less than 5 US dollars.

To complete the design, a light transport analysis of the light field camera is also provided.

Moreover, we find that the light field captured by the cameras developed in the past, including ours, appear to have a common photometric distortion and aliasing that, if not properly managed, may render the data useless. To address this issue that has been somehow ignored, we propose two algorithms (Sec. 5):

- A calibration algorithm to remove the photometric distortion unique to light field without using any reference object. The distortion is directly estimated from the captured light field.
- A depth estimation algorithm utilizing multi-view property of light field and visibility reasoning to generate view-dependent depth maps for view interpolation.

These device and algorithms constitute a complete system for high quality light field acquisition. In comparison with other light field cameras, the spatial resolution of our camera is increased by

*liangck@gmail.com

†homer@cc.ee.ntu.edu.tw

orders of magnitude, and the angular resolution can be easily adjusted during operation or post-processing. The photometric calibration enables more consistent rendering and more accurate depth estimation. The multi-view depth estimation effectively increases the angular resolution for smoother transitions between views and makes depth-aware image editing possible. The output of the system is illustrated in Figure 1, which includes a high resolution light field, an estimated depth map, and two refocused images.

2 Related Work

Our work is inspired by previous research in light field acquisition, computational photography, and illumination multiplexing. This section reviews the remarkable progress in these areas. The related post-processing work is briefly reviewed in Sec. 5.

Light Field Acquisition: 4D light field representation of the ray space was first proposed for image-based rendering and then applied to various fields [Levoy and Hanrahan 1996; Gortler et al. 1996]. There are several ways to capture the light field. The most simple method uses a single moving camera whose position for each exposure is located by a camera gantry [Levoy and Hanrahan 1996] or estimated by a structure-from-motion algorithm [Gortler et al. 1996]. This method is slow and only works in a controlled environment. Another method simultaneously captures the full 4D dataset by using a camera array [Yang et al. 2002; Wilburn et al. 2005], which is cumbersome and expensive.

The third method, which is most related to our approach, inserts additional optical elements or masks in the camera to avoid the angular integration of the light field. The idea dates back nearly a century ago, called integral photography or parallax panoramograms and realized using a fly-eye lens array or a slit plate [Lippmann 1908; Ive 1930; Okoshi 1976]. Compact implementations and theoretical analysis of this method have been recently developed. In a *plenoptic camera*, for example, a microlens array is placed at the original image plane inside the camera [Adelson and Wang 1992; Ng et al. 2005]. The resulting image behind each microlens records an angular distribution of the light rays. Alternatively, one can place a positive lens array in front of the camera [Georgiev et al. 2006]. Along the same line, Veeraraghavan et al. replace the slit plate with a cosine mask to improve efficiency [2007]. In summary, these devices manipulate the 4D light field spectrum by modulation or reparameterization to make it fit in a 2D sensor slice [Georgiev et al. 2007].

All these devices share the following drawbacks. First, the spatial resolution, or spectrum bandwidth, is traded for the angular resolution. Although high resolution sensors can be made, capturing a light field with high spatial and high angular resolutions is still difficult. Second, inserting masks or optical elements in a camera automatically imposes a fixed sampling pattern. These components are usually permanently installed and cannot be easily removed from the camera to capture regular pictures.

Computational Photography: Two popular techniques, coded aperture and multiple capturing, are closely related to our work. The former treats the aperture (or shutter) as an optical modulator to preserve the high-frequency components of motion-blurred images [Raskar et al. 2006], to provide high-dynamic-range or multispectral imaging [Nayar and Branzoi 2003; Schechner and Nayar 2004], to split the field of view [Zomet and Nayar 2006] or to capture stereoscopic images [Farid and Simoncelli 1998]. One particularly relevant work uses a coded aperture to estimate the depth of a near-Lambertian scene from a single image [Levin et al. 2007]. In contrast, our method directly captures the 4D light field and estimates the depth from it when possible.

The multiple capturing technique captures the scene many times sequentially, or simultaneously by using beam splitters and camera arrays. At each exposure the imaging parameters, such as lighting [Raskar et al. 2004], exposure time, focus, viewpoints [Joshi et al.

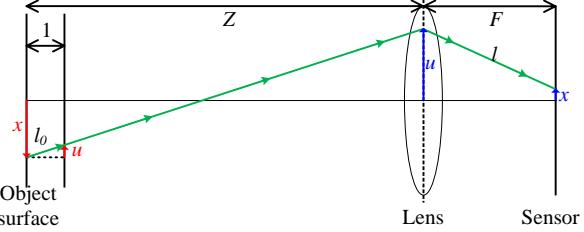


Figure 2: Light field and light transport. A light ray emitting from a point on an object surface at Z can be represented by $l_0([x \ u]^T)$ (red) or $l([x \ u]^T)$ (blue) after refraction by lens. These two representations only differ by a linear transformation (Equation 1).

2006], or spectral sensitivity [Zomet and Nayar 2006], are made different. Then a quality image or additional information (e.g., alpha matte) is obtained by computation. This technique can be easily implemented in digital cameras since the integration duration of the sensor can be electronically controlled. For example, Senkichi et al. [2003] split a given exposure time into a number of time steps and samples one image in each time step. The resulting images are then registered for correcting hand-shaking.

Illumination Multiplexing: Capturing the appearances of an object under different lightings is critical for image-based relighting and object recognition. Since the dimensionality of the signal (a 4D incident light field) is higher than that of the sensor (a 2D photon sensor array), the signal must be captured sequentially, one subset of the signal at a time. Multiplexing can be used to reduce the acquisition time and improve the signal-to-noise ratio by turning on multiple light sources at each exposure and recovering the signal corresponding to a single light source by computation [Schechner et al. 2003; Wenger et al. 2005].

We exploit the coded aperture and multiple capturing techniques in our proposed system. More specifically, we use multiple exposures to avoid the loss of spatial resolution and coded aperture to perform multiplexing for quality improvement. Although our method requires sequential multiple exposures, capturing a clear light field dataset takes the same amount of time as capturing a clear image with a conventional camera.

3 Light Transport in Photography

This section gives a brief review of the light field representation and the light transport theory of the photography process. For simplicity, only 2D geometry is considered here. But the result can be easily extended to 3D.

A light field can be represented as a function that maps the geometric entities of a light ray in free space to the radiance along the light ray. Each light ray is specified by the intersections of two planes with the light ray [Levoy and Hanrahan 1996; Gortler et al. 1996]. There are several ways to define the two planes, see Figure 2. For example, one plane can be located on the object surface of interest, and the other at unit distance from, and parallel to, the first one. The coordinates of the intersection of a light ray with the second plane is defined with respect to the intersection of the light ray with the first plane (red ones in Figure 2) [Durand et al. 2005]. Another common representation places the two planes at the lens and the film (sensor) of a camera and defines independent coordinate systems for these two planes [Ng 2005] (blue ones in Figure 2).

Suppose there is a light ray emitting from an object surface point and denote its radiance by the light field $l_0([x \ u]^T)$, where x and u are the intersections of the light ray with the two coordinate planes. The light ray first traverses the space to the lens of the camera at distance Z from the emitting point, as illustrated in Figure 2. According to the light transport theory, this causes a shearing to the light field [Durand et al. 2005]. Then the light ray changes its di-

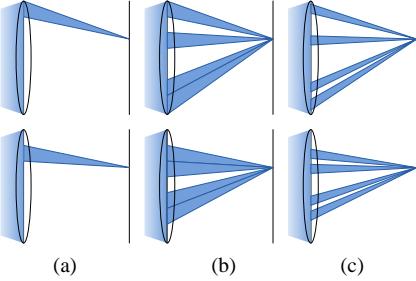


Figure 3: Configurations of the programmable aperture. (a) Capturing one single sample at a time. (b) Aggregating several samples at each exposure for quality improvement. (c) Adjusting the pre-filter kernel without affecting the sampling rate (Sec. 4.4).

rection after it leaves the lens. According to the matrix optics, this makes another shearing to the light field [Georgiev et al. 2007]. As the light ray traverses to the image plane at distance F from the lens plane, one more shearing is resulted. Finally, the light field is reparameterized into the coordinate system used in the camera. Since the shearings and the reparameterization are all linear transformations, they can be concatenated into a single linear transformation. Hence, the transformed light field $l([x u]^T)$ can be represented by

$$\begin{aligned} l([x u]^T) &= l_0(M[x u]^T) \\ &= l_0 \left(\begin{bmatrix} -\frac{Z}{F} & Z\Delta \\ \frac{1}{F} & \frac{1}{f} - \frac{1}{F} \end{bmatrix} \begin{bmatrix} x \\ u \end{bmatrix} \right), \end{aligned} \quad (1)$$

where f is the focal length of the lens and $\Delta = 1/Z + 1/F - 1/f$. This transformation, plus modulation due to the blocking of the aperture [Veeraraghavan et al. 2007], describes various photographic effects such as focusing [Georgiev et al. 2007].

In traditional photography, a sensor integrates the radiances along rays from all directions into an irradiance sample and thus loses all angular information of the light rays. The goal of this work is to capture the transformed light field $l([x u]^T)$ that contains both the spatial and the angular information. In other words, we want to avoid the integration step in the traditional photography.

4 Programmable Aperture Camera

In this section we show how a programmable aperture camera captures the light field and how multiplexing improves the acquisition efficiency. Then we describe the prototypes of the camera.

4.1 Sequential Light Field Acquisition

In a traditional camera we can only adjust the size of the aperture to change the depth of field. The captured image is always a 2D projection of the 3D scene, and the angular information is irrecoverably lost. However, if we modify the *shape* of the aperture so that only the light rays arriving in a small specified region of the aperture can pass through the aperture, we can avoid the angular integration. More specifically, if the aperture blocks all light rays but those around u , the resulting image is a subset of the light field. Denote such a *light field image* by I_u :

$$I_u(x) = l([x u]^T). \quad (2)$$

By capturing images with different aperture shapes (Figure 3 (a)), we construct a complete light field. However, unlike previous devices that manipulate the light rays after they enter the camera [Ng et al. 2005; Georgiev et al. 2007; Veeraraghavan et al. 2007], our method blocks the undesirable light rays and captures one subset of the data at a time. Thus the spatial resolution of the light field is the same as the sensor resolution. For the method to take effect, a *programmable aperture* is needed. Its transmittance has to be spatially variant and controllable.

An intuitive approach to such a programmable aperture is to replace the lens module with a volumetric light attenuator [Zomet and Nayar 2006]. However, according to the following frequency analysis of light transport, we find that the lens module should be preserved for efficient sampling. Let $L_0([f_x f_u]^T)$ and $L([f_x f_u]^T)$ denote the Fourier transform of $l_0([x u]^T)$ and $l([x u]^T)$, respectively. By Equation 1 and the Fourier linear transformation theory, L_0 and L are related by:

$$\begin{aligned} L([f_x f_u]^T) &= |\det(M)|^{-1} L_0(M^{-T}[f_x f_u]^T) \\ &= \frac{1}{|\det(M)|} L_0 \left(\begin{bmatrix} 1 - \frac{F}{f} & 1 \\ FZ\Delta & Z \end{bmatrix} \begin{bmatrix} f_x \\ f_u \end{bmatrix} \right). \end{aligned} \quad (3)$$

Consider the case where the scene is a Lambertian plane perpendicular to the optical path at $Z = 3010$, $f = 50$, and the camera is focused at $Z = 3000$ (so $F = 50.8475$). If the lens module is removed, $f \rightarrow \infty$, and we have to increase the sampling rate along the f_u axis by a factor of 18059 to capture the same signal content. As a result, we have to capture millions of images for a single dataset, which is practically infeasible. Thus the lens module must be preserved. The light rays are bent inwards at the lens due to refraction and consequently the spectrum of the transformed light field is compressed. With the lens module and by carefully selecting the in-focus plane, we can properly reshape the spectrum to reduce aliasing. A similar analysis is developed for multi-view displays [Zwicker et al. 2006].

4.2 Light Field Multiplexing

A light field with angular resolution N requires N exposures, one for each u . Compared with traditional photography, the light collection efficiency of this straightforward acquisition is decreased because only a small aperture is open at each exposure and each exposure time is only $1/N$ of the total acquisition time. As a result, given the same acquisition time, the captured images are noisier than those captured by conventional cameras.

To solve this problem, we *multiplex* the light field images at each exposure. Specifically, because the radiances of the light rays are additive, we aggregate multiple light field samples at each exposure by opening multiple regions of the aperture and recover individual signals afterwards.

At each exposure, the captured image M_u is a linear combination of N light field images (Figure 3 (b)):

$$M_u(x) = \sum_{k=0}^{N-1} w_{uk} I_k(x). \quad (4)$$

The weights $w_{uk} \in [0, 1]$ of the light field images can be represented by a vector $\mathbf{w}_u = [w_{u0} \ w_{u1} \ \dots \ w_{u(N-1)}]$ and is referred to as a *multiplexing pattern* since \mathbf{w}_u is physically realized as a spatial-variant mask on the aperture. After N captures with N different multiplexing patterns, we can recover the light field images by *demultiplexing* the captured images.

Intuitively one should open as many regions as possible (i.e., maximize $\|\mathbf{w}_u\|$) to allow the sensor to gather as much light as possible. In practice, however, noise always involves in the acquisition and complicates the design of the multiplexing patterns. In the case where the noise is independent and identically-distributed (i.i.d.), Hadamard-code based patterns are best in terms of the quality of the demultiplexed data [Harwit and Sloane 1979; Schechner et al. 2003; Liang et al. 2007]. However, digital sensor noises are often correlated with the input signal [HP components group 1998; Tsin et al. 2001]. For example, the variance of the shot noise grows linearly with the number of incoming photons. In this case, using the Hadamard-code based patterns actually degrades the data quality [Wenger et al. 2005]. Another drawback of the Hadamard-code based patterns is that they only exist for certain sizes.

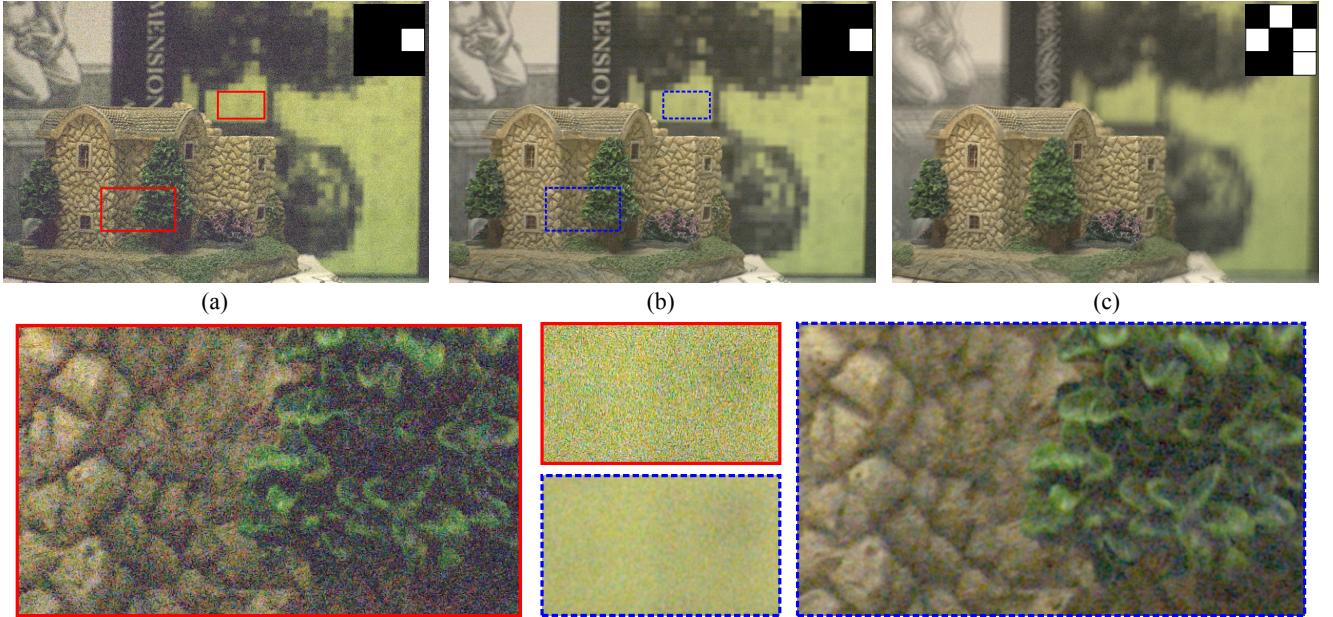


Figure 4: Performance improvement by multiplexing. (a) Light field image captured without multiplexing. (b) Demultiplexed light field image. (c) Image captured with multiplexing ($M_u(x)$ in Equation 4). The insets in (a-c) show the corresponding multiplexing patterns. (Bottom row) Close-up of (a) and (b).

Instead, we apply an optimization process to obtain multiplexing patterns. Given the noise characteristics of the device and the true signal value, the mean square error of the demultiplexed signal is proportional to a function $E(\mathbf{W})$:

$$E(\mathbf{W}) = \text{Trace}((\mathbf{W}^T \mathbf{W})^{-1}), \quad (5)$$

where \mathbf{W} is an $N \times N$ matrix and each row of \mathbf{W} is a multiplexing pattern w_{uk} . Finding a matrix \mathbf{W}^* that minimizes $E(\mathbf{W})$ can be formulated as a constrained convex optimization problem and solved by the projected gradient method [Ratner and Schechner 2007]. Because most entities (w_{uk}) of the \mathbf{W}^* thus obtained are either ones or zeros and because binary masks can be made more accurately in practice, we enforce all the entities of \mathbf{W}^* to be binary. This only slightly affects the performance. A result of multiplexing is given in Figure 4 where we can see that the demultiplexed image is much more clear than the one captured without multiplexing.

4.3 Prototypes

We implement two prototypes of the programmable aperture camera shown in Figure 5 using a regular Nikon D70 DSLR camera and a 50mm f/1.4D lens module. For simplicity, we dismount the lens module from the Nikon camera and insert the programmable aperture in between them. Hence the distance (F in Figure 1) between the lens and the sensor is lengthened and the focus range is shortened as compared to the original camera.

The optimization of the multiplexing patterns requires the noise characteristics of the camera and the scene intensity. The former is obtained by calibration and the latter is assumed to be one half of the saturation level. Both prototypes can capture the light field with or without multiplexing. The maximal spatial resolution of the light field is 3039×2014 and the angular resolution is adjustable.

In the first prototype, the programmable aperture is made up of a *pattern scroll*, which is an opaqued slit of paper used for film protection. The aperture patterns are manually cut and scrolled across the optical path. The pattern scroll is long enough to include tens of multiplexing patterns and the traditional aperture shapes. This quick and dirty method is simple and performs well except one mi-

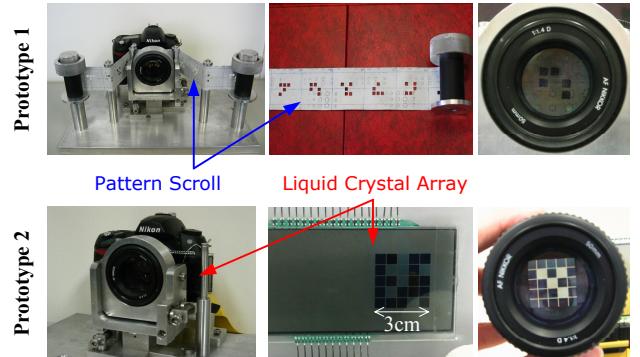


Figure 5: Prototypes of the programmable aperture cameras with aperture patterns (First row) on an opaque slip of paper and (Second row) on an electronically controlled liquid crystal array.

nor issue: The blocking cell ($w_{uk} = 0$) cannot stay on the pattern scroll if it loses support. We solve it by leaving a gap between cells.

In the second prototype, the programmable aperture is made up of a liquid crystal array (LCA) controlled by a Holtek HT49R30A-1 micro control unit that supports C language. Two different resolutions, 5×5 and 7×7 , of the LCA are made. The LCA is easier to program and mount than the pattern scroll, and the multiplexing pattern is no longer limited to binary. However, the light rays can leak from the gaps (used for routing) in between the liquid crystal cells and from the cells that cannot be completely turned off. We compensate for the leakage by capturing an extra image with all liquid crystal cells turned off and subtracting it from other images.

4.4 Summary

The proposed light field acquisition scheme does not require a high resolution sensor. Therefore, it can even be implemented on web cameras, cell-phone cameras, surveillance cameras, etc. The image captured by previous light field cameras must be decoded before visualization. In contrast, the image captured by our device can be directly displayed. Even when the multiplexing is applied, the in-focus regions remain sharp (Figure 4 (c)).

Another advantage of the proposed device is that the sampling grid and the pre-filtering kernel are decoupled. Therefore, the aperture size can be chosen regardless of the sampling rate, Figure 3 (c). We can choose a small pre-filter to preserve the details and remove aliasing by view interpolation. Also the sampling lattice on the lens plane in our device is not restricted to rectangular grids.

5 Post-Processings

The photometric distortion and aliasing due to undersampling have to be addressed before the captured light field can be applied.

5.1 Photometric Calibration

The vignetting effect causes a scene point to have different intensities in the light field images. While being termed as vignetting collectively, this photometric distortion is attributed to several sources: the cosine-fourth falloff, the blocking of the lens diaphragm or the hood, and the pupil aberrations [Aggarwal et al. 2001; Goldman and Chen 2005]. This distortion must be removed or it can obstruct view interpolation and depth estimation.

The exact physical model of the vignetting effect is difficult to construct. In general, a simplified model that describes the ratio between the distorted light field image $I_u^d(x)$ and the clean image $I_u(x)$ by a $2(D-1)$ -degree polynomial function $f_u(x)$ is adopted:

$$I_u^d(x) = f_u(x)I_u(x) = \left(\sum_{i=0}^{D-1} a_{ui} \|x - c_u\|_2^{2i} \right) I_u(x), \quad (6)$$

where $\{a_{ui}\}$ are the polynomial coefficients, c_u is the vignetting center, and $\|\cdot\|_2$ is the Euclidean distance (the coordinates are normalized to $(0, 1)$). The function $f_u(x)$, called *vignetting field*, is a smooth field across the image. It is large when the distance between x and c_u is small and gradually decreases as the distance increases.

Existing photometric calibration methods generally make two assumptions to make the problem tractable [Goldman and Chen 2005]: the scene points have multiple registered observations, and the vignetting center c_u is known. However, these assumptions are inappropriate for the light field images for two reasons. First, the registration of light field images taken from different viewpoints requires a per-pixel disparity map that is difficult to obtain from the distorted inputs. Second, in each light field image, the parameters, $\{a_{ui}\}$ and c_u , of the vignetting function, are image-dependent and coupled. Therefore, simultaneously estimating the parameters and the clean image is an under-determined nonlinear problem. Another challenge specific to our camera is that the vignetting function changes with the lens and the aperture settings (such as the size of the pre-filter kernel).

Here we propose an algorithm to automatically calibrate the photometric distortion of the light field images. The key idea is that the light field images closer to the center of the optical path have less distortion. Therefore, we can assume $I_0^d \approx I_0$, then approximate other I_u 's by properly transforming I_0 to estimate the vignetting field. This way, the problem is greatly simplified. The approach can also be generalized to handle the distortions of other computational cameras, particularly previous light field cameras.

The flowchart of the algorithm is shown in Figure 6 (a) along with an example. For an input I_u^d , we first use the SIFT method, which is well immune to local photometric distortions [Lowe 2004], to detect the feature points and find their valid matches in I_0 (Figures 6 (b) and (c)). Next, we apply the Delaunay triangulation to the matched points in I_u^d to construct a mesh (Figure 6 (d)). For each triangle A of the mesh, we use the displacement vectors of its three vertices to determine an affine transform. By affinely warping all triangles, we obtain an image I_u^w from I_0 (Figure 6 (e)). I_u^w is close enough to the clean image I_u unless there are triangles including objects of different depths or incorrect feature matchings. Such erroneous cases can be effectively detected and

removed by measuring the variance of the associated displacement vectors (Figure 6 (f)).

After we have a good approximation of I_u for a plenty of pixels, we estimate the parameters of Equation 6. Given an initial estimate, we first fix the vignetting center c_u . This makes Equation 6 linear in $\{a_{ui}\}$, which can be easily solved by least square estimation. Then we fix $\{a_{ui}\}$ and update c_u by gradient descent. These two steps are performed iteratively. Finally, we divide I_u^d by f_u (Figure 6 (g)) to recover the clean image I_u (Figure 6 (h)).

5.2 Multi-View Depth Estimation

Images corresponding to new viewpoints or focus settings can be rendered from the captured light field by re-sampling. However, the quality of the rendered image is dictated by the bandwidth of the light field, which strongly depends on the scene geometry [Chai et al. 2000; Isaksen et al. 2000]. Generally speaking, a scene with higher depth range requires a higher angular resolution for aliasing-free rendering. Although one can adjust the angular resolution of the programmable aperture camera, a high angular sampling rate requires a long capture duration and a large storage, which may not be always affordable.

To solve this problem, we propose a multi-view depth estimation algorithm to generate view-dependent depth maps for view interpolation. By depth-dependent view interpolation, we can greatly reduce the angular sampling rate for the near-Lambertian scene.

The multi-view depth estimation problem is similar to the traditional stereo correspondence problem [Scharstein and Szeliski 2002]. However, the visibility reasoning is extremely important for multi-view depth estimation since the occluded views should be excluded from the depth estimation. Previous methods that determine the visibility by hard constraint [Kolmogorov and Zabih 2002] or greedy progressive masking [Kang and Szeliski 2004] can easily be trapped in local minima because they cannot recover from incorrect occlusion guess. Inspired by the symmetric stereo matching algorithm [Sun et al. 2005], we alleviate this problem by iteratively optimizing 1) a view-dependent *depth map* D_u for each image I_u and 2) an *occlusion map* O_{uv} for each pair of neighboring images I_u and I_v . If a scene point projected onto a point x in I_u is occluded in I_v , it does not have a valid correspondence. When this happens, we can set $O_{vu}(x)=1$ to exclude it from the matching process. On the other hand, if the estimated correspondence x' of x_u in I_v is marked as invisible, that is, $O_{vu}(x')=1$, the estimate is unreliable.

Depth and occlusion estimation is formulated as a discrete labeling problem. For each pixel x_u , we need to determine a discrete depth value $D_u(x) \in \{0, 1, \dots, d_{max}\}$ and a binary occlusion value $O_{uv}(x) \in \{0, 1\}$. More specifically, given a set of light field images $\mathcal{I} = \{I_u\}$, we want to find a set of depth maps $\mathcal{D} = \{D_u\}$ and a set of occlusion maps $\mathcal{O} = \{O_{uv}\}$ to minimize the energy functional defined by

$$\begin{aligned} E(\mathcal{D}, \mathcal{O} | \mathcal{I}) &= \sum_u \{E_{dd}(D_u | \mathcal{O}, \mathcal{I}) + E_{ds}(D_u | \mathcal{O}, \mathcal{I})\} \\ &+ \sum_u \sum_{v \in \mathcal{N}(u)} \{E_{od}(O_{uv} | D_u, \mathcal{I}) + E_{os}(O_{uv})\}, \end{aligned} \quad (7)$$

where E_{dd} and E_{ds} are the data term and the smoothness (or regularization) term, respectively, of the depth map, and E_{od} and E_{os} are the data term and the smoothness term, respectively, of the occlusion map. $\mathcal{N}(u)$ is the set of eight viewpoints that are closest to u . The energy minimization is performed iteratively. In each iteration, we first fix the occlusion maps and minimize $E_{dd} + E_{ds}$ by updating the depth maps and then fix the depths maps and minimize $E_{od} + E_{os}$ by updating the occlusion maps.

Each term of the above equation is defined as follows. First, let $\alpha, \beta, \gamma, \zeta$, and η denote the weighting coefficients and K and T the thresholds. These parameters are empirically determined and fixed

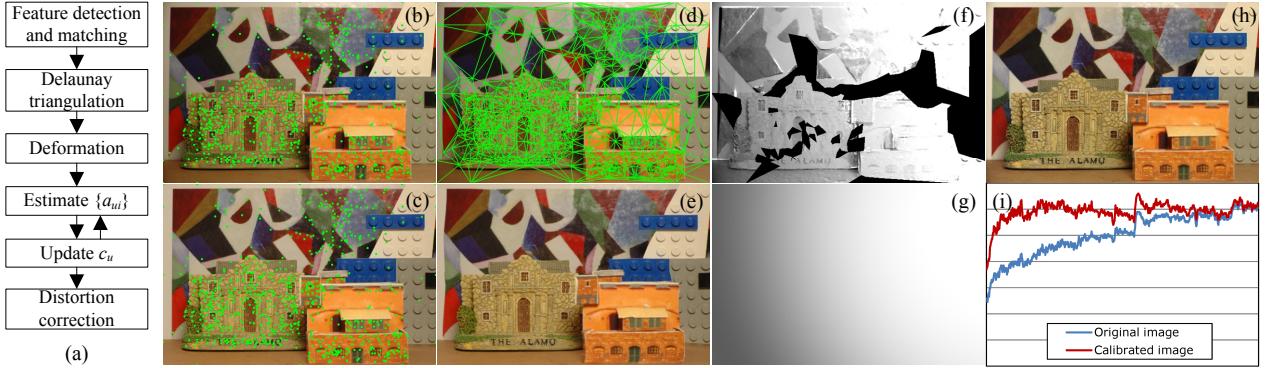


Figure 6: (a) The flowchart of the vignetting correction algorithm. (b) The image to be corrected I_u^d . Note the left side is darker. (c) The reference image I_0 . Matched features in (b) and (c) are marked as green points. (d) Triangulation of the matched features in (b). (e) Image I_u^w warped from (c) based on the triangular mesh. (f) The approximated vignetting field with the suspicious areas (black) removed. (g) The estimated vignetting field. (h) The calibrated image I_u . (i) The intensity profile of the 420th scanline before and after calibration.

in the experiments. The data term E_{dd} is a unary function,

$$E_{dd}(D_u | \mathcal{O}, \mathcal{I}) = \sum_x \left\{ \sum_{v \in \mathcal{N}(u)} (\bar{O}_{uv}(x) C(I_u(x) - I_v(x + D_{uv}(x))) + \alpha O_{vu}(x + D_{uv}(x))) \right\}, \quad (8)$$

where $\bar{O}_{uv}(x) = 1 - O_{uv}(x)$, $D_{uv}(x)$ is the disparity corresponding to the depth value $D_u(x)$, and $C(k) = \min(|k|, K)$ is a truncated linear function. For each pixel x_u , the first term measures the similarity between the pixel and its correspondence in I_v , and the second term adds a penalty to an invalid correspondence.

The pairwise smoothness term E_{ds} is based on a generalized Potts model:

$$E_{ds}(D_u | \mathcal{O}, \mathcal{I}) = \sum_{\substack{(x,y) \in \mathcal{P}, \\ O_u(x) = O_u(y)}} \beta \min(|D_u(x) - D_u(y)|, T), \quad (9)$$

where \mathcal{P} is the set of all pairs of neighboring pixels and $O_u = \bigcap O_{uv}$, which is true only when x_u is occluded in all other images. This term encourages the depth map to be piecewise smooth.

Next, we describe the energy terms involved in the second step of each iteration. Because the depth maps \mathcal{D} are fixed in this step, the prior of an occlusion map can be obtained by warping the depth map. Specifically, let W_{uv} denote a binary map. The value of $W_{uv}(x)$ is 1 when the depth map D_v warped to the viewpoint u is null at x_u and 0 otherwise. If $W_{uv}(x) = 1$, x_u might be occluded in I_v . With this prior, the data term E_{od} is formulated as:

$$E_{od}(O_{uv} | D_u, \mathcal{I}) = \sum_x (\bar{O}_{uv}(x) C(I_u(x) - I_v(x + D_{uv}(x))) + \gamma O_{uv}(x) + \zeta |O_{uv}(x) - W_{uv}(x)|). \quad (10)$$

The first term above biases a pixel to be non-occluded if it is similar to its correspondence. The second term penalizes the occlusion ($O = 1$) to prevent the whole image from being marked as occluded, and the third term favors the occlusion when the prior W_{uv} is true. Finally, the smoothness term E_{os} is based on the Potts model:

$$E_{os}(O_{uv}) = \sum_{(x,y) \in \mathcal{P}} \eta |O_{uv}(x) - O_{uv}(y)|. \quad (11)$$

The solution of the energy minimization problem is a maximum a posteriori estimate of a Markov random field (MRF), for which



Figure 7: Digital refocusing without depth information.

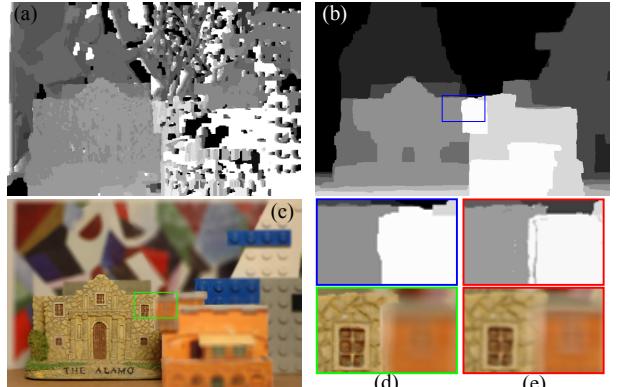


Figure 8: (a) Depth map estimated without photometric calibration and occlusion reasoning. (b) Depth map estimated without occlusion reasoning. (c) Defocusing by the Photoshop Lens Blur tool. (d) Close-up of (b) and (c). (e) Corresponding close-up of Figures 1(b) and 1(c).

high-performance algorithms have been recently developed. We use the MRF optimization library from Middlebury [Szeliski et al. 2006]. Among all the algorithms in the library, both the alpha-expansion graph cut [Boykov et al. 2001] and the tree-reweighted message passing [Kolmogorov 2006] perform well, but the latter gives slightly better results at the cost of execution time. Finally, we apply a modified cross bilateral filtering to the depth maps at the end of each iteration to improve their quality and make the iteration converge faster [Yang et al. 2007].

Discussion: The light field images captured by our programmable aperture camera have several advantages for depth estimation. First, the viewpoints of the light field images are well aligned with the 2D grid on the aperture, and thus the depth estimation can be performed without camera calibration. Second, the disparity corresponding to a depth value can be adjusted by changing the camera parameters without any additional rectification as required in camera array systems. Finally, unlike depth-from-defocus methods [Green et al. 2007; Levin et al. 2007], there is no ambiguity in the scene points behind and in front of the in-focus object.

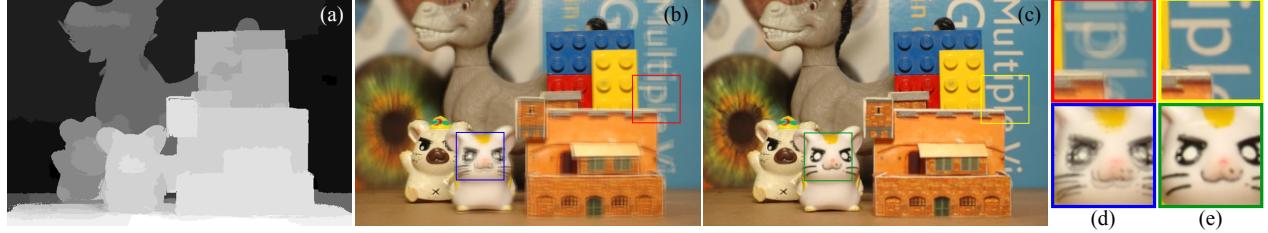


Figure 9: (a) An estimated depth map. (b) Image interpolated without depth information. (c) Image interpolated with depth information. (d-e) Close-up of (b-c). The angular resolution of the light field is 3×3 .

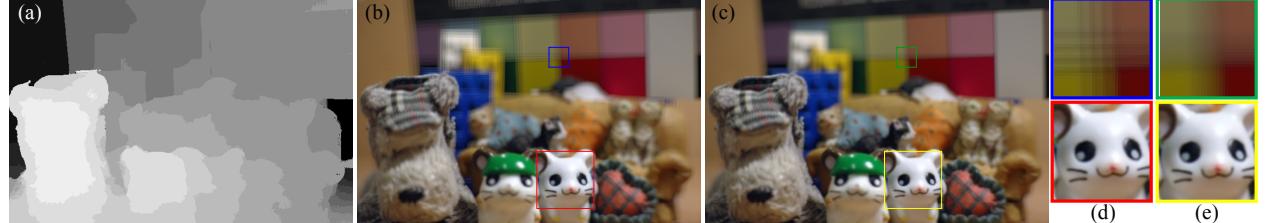


Figure 10: (a) An estimated depth map. (b) Digital refocused image with the original angular resolution 4×4 . (c) Digital refocused image with the angular resolution 25×25 boosted by view interpolation. (d-e) Close-up of (b-c).

6 Results

All data in the experiments are captured indoors. The shutter speed of each exposure is set to 10ms for images shown in Figures 1, 4, and 10, and 20ms for the rest. These settings are chosen for the purpose of fair comparison. For example, it takes 160ms with an aperture setting of f/8 to capture a clean and large depth of field image for the scene in Figure 1, so we choose 10ms for our device. Images shown in Figures 1 and 7 are captured using the first prototype and the rest are captured using the second one.

All the computations are performed on a Pentium 4 3.2GHz computer with 2GB memory. Demultiplexing one light field dataset takes 3 to 5 seconds. To save the computational cost, the light field images are down-sampled to 640×426 after demultiplexing. The photometric calibration takes 30 seconds per image, and the multi-view depth estimation takes around 30 minutes. In the following we demonstrate still images with various effects generated from the captured light field and the associated depth maps. Please watch the supplemental video to see the results in action.

Figure 7 shows a scene containing a transparent object in front of a nearly uniform background. The geometry of this scene is difficult to estimate. However, since our acquisition method does not impose any restriction on the scene, we can capture the light field with 4×4 angular resolution and generate faithful refocused images by dynamic reparameterization [Isaksen et al. 2000].

We use the dataset shown in Figure 1 to evaluate the performance of our post-processing algorithms. Here a well-known graph cut stereo matching algorithm without occlusion reasoning is implemented for comparison [Boykov et al. 2001]. The photo-consistency assumption is violated in the presence of the photometric distortion, and thus poor result is obtained (Figure 8 (a)). With the photometric calibration, the graph cut algorithm generates a good depth map but errors can be observed at the depth discontinuities (Figure 8 (b)). On the contrary, our depth estimation algorithm can successfully identify these discontinuities and generate a more accurate result (Figure 1 (b)).

For quantitative comparison, we also apply our multi-view depth estimation algorithm without individual fine tuning to the four test datasets on the Middlebury site¹. As compared to other top-ranked algorithms, our algorithm does not perform over-segmentation and plane fitting. Yet it achieves an average rank score of 7.0, which



Figure 11: Application of the proposed post-processing algorithms to the dataset in [Veeraraghavan et al. 2007]. (Left) The original image. The inset shows the estimated vignetting field. (Right) The processed image. Image courtesy of Ashok Veeraraghavan.

is the fifth best one in a pool of 41 algorithms at the time of the submission.

Both the light field data and the post-processing algorithms are indispensable for generating plausible photographic effects. To illustrate this, we apply a single light field image and its associated depth map to the Photoshop Lens Blur tool to generate a defocused image. The result shown in Figure 8 (c) contains many errors, particularly at the depth discontinuities, Figure 8 (d). In contrast, our results, Figure 1 (c) and (d), are more natural. The boundaries of the defocused objects are semi-transparent and thus the objects behind can be partially seen.

Figure 9 shows the results of view interpolation. The raw angular resolution is 3×3 . If a simple bilinear interpolation is used, ghosting effect due to aliasing is observed (Figure 9 (b)). While previous methods use filtering to remove the aliasing [Levoy and Hanrahan 1996; Isaksen et al. 2000], we use a modified projective texture mapping [Debevec et al. 1996]. Given a viewpoint, three closest images are warped according to their associated depth maps. The warped images are then blended; the weight of each image is inversely proportional to the distance between its viewpoint and the given viewpoint. This method greatly suppresses the ghosting effect without blurring (Figure 9 (c)).

Figure 10 shows another digital refocusing result. The raw angular resolution is 4×4 . Though the in-focus objects are sharp, the out-of-focus objects are subject to the ghost effect due to aliasing (Figure 10 (b)). With the estimated depth maps, we first increase the angular resolution to 25×25 by view interpolation described above and then perform digital refocusing. As we can see in Figure 10 (c), the out-of-focus objects are blurry while the in-focus objects are unaffected.

¹<http://vision.middlebury.edu/stereo/>

To illustrate the robustness of the proposed algorithms, we apply them to the noisy and photometrically distorted data captured by the heterodyned light field camera [Veeraraghavan et al. 2007]. We pick four clear images from the data, perform photometric calibration and multi-view depth estimation, and synthesize the whole light field by view interpolation. As seen in Figure 11, the interpolated image is much cleaner than the original one.

7 Discussion

In this section we discuss the performance and limitations of the proposed camera and directions for future research.

Performance Comparison: We compare three different devices: a conventional camera, a plenoptic camera [Adelson and Wang 1992; Ng et al. 2005], and a programmable aperture camera. Because no light ray is blocked or attenuated in the plenoptic camera, it is superior to other mask-based light field cameras [Ive 1930; Veeraraghavan et al. 2007]. Without loss of generality, we assume the default number of sensors in these devices is M^2 , and the angular resolution of the two light field cameras is N^2 . The total exposure duration for capturing a single dataset is fixed. Therefore, each exposure in our device is $1/N^2$ of the total exposure.

We make a signal-to-noise ratio (SNR) analysis of these devices by using a simple noise model. There are typically two zero-mean noise sources in the imaging process: one with a constant variance σ_c^2 and another with a variance proportional to the received irradiance of the sensor. Let σ_p^2 be the variance of the second noise when the received irradiance value is P .

The results of the SNR analysis are listed in Table 1. The image captured by a conventional camera with a large aperture has the best quality, but it has a shallow depth of field. A light field image is equivalent to the image captured by a conventional camera with a small aperture and thus its quality is lower. However, this can be improved by digital refocusing. Light rays emitted from an in-focus scene point are recorded by N^2 light field samples. The refocusing averages these samples and thus increases the SNR by N .

The plenoptic camera is slightly better than the programmable aperture camera at the same angular and spatial resolutions. Nevertheless, it requires $N^2 M^2$ sensors. To capture a light field of the same resolution as the dataset shown in Figure 1, the plenoptic camera requires an array of nearly 100 million sensors, which is expensive, if not difficult, to make.

Limitation and Future Direction: The proposed device has great performance and flexibility, but it requires that the scene and the camera be static because the data are captured sequentially. However, as we mentioned in Sec. 4.4, the sharpness of in-focus objects are unaffected by multiplexing. Hence our system can capture a moving in-focus object amid static out-of-focus objects and then recover the light field and scene geometry of the static objects.

On the other hand, other devices capture the light field in one exposure at the expense of spatial resolution. However, it should be pointed out that the proposed method is complementary to the existing ones. We can place a cosine mask or a microlens array near the image plane to capture a coarse angular resolution light field and use the programmable aperture to provide the fine angular resolution needed.

Multiplexing a light field is equivalent to transforming the light field to another representation by basis projection. While our goal is to obtain a reconstruction with minimal error from a fixed number of projected images ($M_u(x)$ in Equation 4), an interesting direction of future research is to reduce the number of images required for reconstruction. The compressive sensing theory states that if a signal of dimension n has a sparse representation, we can use fewer than n projected measurements to recover the full signal [Donoho 2006]. Finding a proper set of bases to perform compressive sensing is worth pursuing in the future.

8 Conclusion

We have described a system for capturing light field using a programmable aperture camera with an optimal multiplexing scheme. Along with the programmable aperture, we have also developed two post-processing algorithms for photometric calibration and multi-view depth estimation. To our best knowledge, this system is the first single-camera system that generates light field at the same spatial resolution as that of the sensor, with adjustable angular resolution, and free of photometric distortion. In addition, the programmable aperture is fully backward compatible with conventional apertures.

While we have focused on the light field acquisition in this work, the programmable aperture camera can be further exploited for other applications. For example, it can be used to realize a computational camera with a fixed mask.

Acknowledgements

We would like to thank Prof. Yung-Yu Chuang for his help on the organization and presentation of the paper, Ching-Chang Liao and Aaron Hu for their assistance on the prototype development, Netanel Ratner for helpful discussion on optimal multiplexing, Li-Yi Wei for encouragements, Chris Liao for narrating the voice-over of the supplemental video, and the anonymous reviewers for their valuable comments. This project was supported in part by grants from the Excellent Research Projects of National Taiwan University under the contract 95R0062-AE00-02 and from the National Science Council of Taiwan under the contract NSC 96-2628-E-002-005-MY2.

References

- ADELSON, E. H., AND WANG, J. Y. A. 1992. Single lens stereo with a plenoptic camera. *IEEE Trans. Pattern Anal. Mach. Intell.* 14, 2, 99–106.
- AGGARWAL, M., HUA, H., AND AHUJA, N. 2001. On cosine-fourth and vignetting effects in real lenses. In *Proc. ICCV '01: Proc. the Eighth IEEE International Conference on Computer Vision*, vol. 1, 472–479.
- BOYKOV, Y., VEKSLER, O., AND ZABIH, R. 2001. Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.* 23, 11, 1222–1239.
- CHAI, J.-X., CHAN, S.-C., AND TONG, H.-Y. S. X. 2000. Plenoptic sampling. In *SIGGRAPH '00: Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, 307–318.
- DEBEVEC, P. E., TAYLOR, C. J., AND MALIK, J. 1996. Modeling and rendering architecture from photographs: a hybrid geometry- and image-based approach. In *SIGGRAPH '96: Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, ACM, New York, NY, USA, 11–20.
- DONOHO, D. 2006. Compressed Sensing. *IEEE Trans. Information Theory* 52, 4, 1289–1306.
- DURAND, F., HOLZSCHUCH, N., SOLER, C., CHAN, E., AND SILLION, F. X. 2005. A frequency analysis of light transport. In *SIGGRAPH '05: ACM SIGGRAPH 2005 Papers*, ACM Press, New York, NY, USA, 1115–1126.
- FARID, H., AND SIMONCELLI, E. P. 1998. Range estimation by optical differentiation. *Journal of the Optical Society of America A* 15, 7, 1777–1786.
- GEORGIEV, T., ZHENG, K. C., CURLESS, B., SALESIN, D., NAYAR, S., AND INTWALA, C. 2006. Spatio-angular resolution tradeoff in integral photography. In *EGRW '06: Proc. the 17th Eurographics workshop on Rendering*.

Device	Aperture size	#shot	Single exposure duration	SNR of the light field samples	SNR of the refocused image	Angular \times spatial resolution
Conventional camera with small aperture	A	1	T	$P/\sqrt{\sigma_p^2 + \sigma_c^2}$	—	$1 \times M^2$
Conventional camera with large aperture	$N^2 A$	1	T	$N^2 P/\sqrt{N^2 \sigma_p^2 + \sigma_c^2}$	—	$1 \times M^2$
Plenoptic camera	$N^2 A$	1	T	$N^2 P/\sqrt{N^2 \sigma_p^2 + \sigma_c^2}$	$N^3 P/\sqrt{N^2 \sigma_p^2 + \sigma_c^2}$	$N^2 \times (M^2/N^2)$
Plenoptic camera with $N^2 M^2$ sensors	$N^2 A$	1	T	$P/\sqrt{\sigma_p^2 + \sigma_c^2}$	$NP/\sqrt{\sigma_p^2 + \sigma_c^2}$	$N^2 \times M^2$
Programmable aperture camera	A	N^2	T/N^2	$N^{-2} P/\sqrt{\sigma_p^2/N^2 + \sigma_c^2} (S_1)$	$N^{-1} P/\sqrt{\sigma_p^2/N^2 + \sigma_c^2}$	$N^2 \times M^2$
PAC with multiplexing	$\approx N^2 A/2$	N^2	T/N^2	$\approx NS_1/2$	$\approx N^2 S_1/2$	$N^2 \times M^2$

Table 1: Performance comparison between the conventional camera, the plenoptic camera, and the programmable aperture camera (PAC).

- GEORGIEV, T., INTWALA, C., AND BABACAN, D. 2007. Light-field capture by multiplexing in the frequency domain. Adobe technical report, Adobe Systems Incorporated.
- GOLDMAN, D. B., AND CHEN, J.-H. 2005. Vignette and exposure calibration and compensation. In *Proc. ICCV '05: Proc. the 10th IEEE International Conference on Computer Vision*, 899–906.
- GORTLER, S. J., GRZESZCZUK, R., SZELISKI, R., AND COHEN, M. F. 1996. The lumigraph. In *SIGGRAPH '96: Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, ACM Press, New York, NY, USA, 43–54.
- GREEN, P., SUN, W., MATUSIK, W., AND DURAND, F. 2007. Multi-aperture photography. *ACM Trans. Graph.* 26, 3, 68.
- HARWIT, M., AND SLOANE, N. J. 1979. *Hadamard Transform Optics*. Academic Press, New York.
- HP COMPONENTS GROUP. 1998. Noise sources in CMOS image sensors. Technical report, Hewlett-Packard Company.
- ISAKSEN, A., McMILLAN, L., AND GORTLER, S. J. 2000. Dynamically reparameterized light fields. In *SIGGRAPH '00: Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, 297–306.
- IVE, H. E. 1930. Parallax panoramagrams made with a large diameter lens. *Journal of the Optical Society of America* 20, 6 (June), 332–342.
- JOSHI, N., MATUSIK, W., AND AVIDAN, S. 2006. Natural video matting using camera arrays. *ACM Trans. Graph.* 25, 3, 779–786.
- KANG, S. B., AND SZELISKI, R. 2004. Extracting view-dependent depth maps from a collection of images. *International Journal of Computer Vision* 58, 2, 139–163.
- KOLMOGOROV, V., AND ZABIH, R. 2002. Multi-camera scene reconstruction via graph cuts. *Proc. European Conference on Computer Vision* 3, 82–96.
- KOLMOGOROV, V. 2006. Convergent tree-reweighted message passing for energy minimization. *IEEE Trans. Pattern Anal. Mach. Intell.* 28, 10, 1568–1583.
- LEVIN, A., FERGUS, R., DURAND, F., AND FREEMAN, W. T. 2007. Image and depth from a conventional camera with a coded aperture. *ACM Trans. Graph.* 26, 3, 70.
- LEVOY, M., AND HANRAHAN, P. 1996. Light field rendering. In *SIGGRAPH '96: Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, ACM Press, New York, NY, USA, 31–42.
- LIANG, C.-K., LIU, G., AND CHEN, H. H. 2007. Light field acquisition using programmable aperture camera. In *Proc. IEEE International Conference on Image Processing*, vol. 5, 233–236.
- LIPPMANN, M. G. 1908. Epreuves reversibles donnant la sensation du relief. *J. Phys* 7, 821–825.
- LOWE, D. G. 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60, 2, 91–110.
- NAYAR, S. K., AND BRANZOI, V. 2003. Adaptive dynamic range imaging: Optical control of pixel exposures over space and time. In *Proc. ICCV '03: Proc. the Ninth IEEE International Conference on Computer Vision*, vol. 2, 1168–1175.
- NG, R., LEVOY, M., BRÉDIF, M., DUVAL, G., HOROWITZ, M., AND HANRAHAN, P. 2005. Light field photography with a hand-held plenoptic camera. CSTR 2005-02, Stanford University, April.
- NG, R. 2005. Fourier slice photography. In *SIGGRAPH '05: ACM SIGGRAPH 2005 Papers*, ACM, New York, NY, USA, 735–744.
- OKOSHI, T. 1976. *Three-dimensional imaging techniques*. Academic Press New York.
- RASKAR, R., TAN, K.-H., FERIS, R., YU, J., AND TURK, M. 2004. Non-photorealistic camera: depth edge detection and stylized rendering using multi-flash imaging. In *SIGGRAPH '04: ACM SIGGRAPH 2004 Papers*, ACM Press, New York, NY, USA, 679–688.
- RASKAR, R., AGRAWAL, A., AND TUMBLIN, J. 2006. Coded exposure photography: motion deblurring using fluttered shutter. *ACM Trans. Graph.* 25, 3, 795–804.
- RATNER, N., AND SCHECHNER, Y. Y. 2007. Illumination multiplexing within fundamental limits. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 1–8.
- SCHARSTEIN, D., AND SZELISKI, R. 2002. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision* 47, 1-3, 7–42.
- SCHECHNER, Y. Y., AND NAYAR, S. K. 2004. Uncontrolled modulation imaging. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, 197–204.

SCHECHNER, Y. Y., NAYAR, S. K., AND BELHUMEUR, P. N. 2003. A theory of multiplexed illumination. In *Proc. ICCV '03: Proc. the Ninth IEEE International Conference on Computer Vision*, vol. 2, 808–815.

SENKICHI, C., TOSHIO, M., TOSHINORI, H., YUICHI, M., AND HIDETOSHI, K., 2003. Device and method for correcting camera-shake and device for detecting camera shake. JP Patent No. 2003-138436.

SUN, J., LI, Y., KANG, S. B., AND SHUM, H.-Y. 2005. Symmetric stereo matching for occlusion handling. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, 399–406.

SZELISKI, R., ZABIH, R., SCHARSTEIN, D., VEKSLER, O., KOLMOGOROV, V., AGARWALA, A., TAPPEN, M., AND ROTHER, C. 2006. A comparative study of energy minimization methods for markov random fields. In *Proc. European Conference on Computer Vision*, vol. 2, 16–29.

TSIN, Y., RAMESH, V., AND KANADE, T. 2001. Statistical calibration of CCD imaging process. In *Proc. ICCV '01: Proc. the Eighth IEEE International Conference on Computer Vision*, 480–487.

VEERARAGHAVAN, A., RASKAR, R., AGRAWAL, A., MOHAN, A., AND TUMBLIN, J. 2007. Dappled photography: mask enhanced cameras for heterodyned light fields and coded aperture refocusing. *ACM Trans. Graph.* 26, 3, 69.

WENGER, A., GARDNER, A., TCHOU, C., UNGER, J., HAWKINS, T., AND DEBEVEC, P. 2005. Performance relighting and reflectance transformation with time-multiplexed illumination. In *ACM SIGGRAPH '05: ACM SIGGRAPH 2005 Papers*, ACM Press, New York, NY, USA, 756–764.

WILBURN, B., JOSHI, N., VAISH, V., TALVALA, E.-V., ANTUNEZ, E., BARTH, A., ADAMS, A., HOROWITZ, M., AND LEVOY, M. 2005. High performance imaging using large camera arrays. In *SIGGRAPH '05: ACM SIGGRAPH 2005 Papers*, ACM, New York, NY, USA, 765–776.

YANG, J. C., EVERETT, M., BUEHLER, C., AND McMILLAN, L. 2002. A real-time distributed light field camera. In *EGRW '02: Proc. the 13th Eurographics workshop on Rendering*, 77–86.

YANG, Q., YANG, R., DAVIS, J., AND NISTER, D. 2007. Spatial-depth super resolution for range images. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 1–8.

ZOMET, A., AND NAYAR, S. K. 2006. Lensless imaging with a controllable aperture. *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 339–346.

ZWICKER, M., MATUSIK, W., DURAND, F., AND PFISTER, H. 2006. Antialiasing for automultiscopic 3d displays. In *EGSR'06: Proc. the 17th Eurographics Symposium on Rendering*.