LIGHT FIELD ACQUISITION USING PROGRAMMABLE APERTURE CAMERA

Chia-Kai Liang, Gene Liu, and Homer H. Chen∗

Multimedia Processing and Communication Laboratory
Graduate Institute of Communication Engineering
National Taiwan University

ABSTRACT

We propose a new device, Programmable Aperture Camera (PAC), to capture 4D light field in a camera. PAC can adjust the shape of the aperture in each exposure. This allows us to capture the angular information of the light field, which is lost in regular photography. Although multiple exposures are needed to obtain a light field, the total exposure time remains the same as that of taking a single regular photograph at the same image quality level. As opposed to previous techniques that seriously reduce the spatial resolution, PAC captures the image at full spatial resolution and allows adjustable angular resolution. Also its manufacturing cost is much lower than previous techniques. We describe the PAC prototype and demonstrate how digital refocusing is made possible by using the captured light field.

Index Terms— light field, digital refocusing, computational photography

1. INTRODUCTION

Research on the capture and visualization of three-dimensional information began more than a century ago. The ultimate goal is to digitize the real-world scene and display it. However, the illumination, geometry, and surface reflectance are hard to obtained and stored in most cases. It has been found that light rays can be represented as a 5D *plenoptic function* without any knowledge of the scene properties. Furthermore, if light rays travel in free-space without occlusion, 5D plenoptic function can be reduced to a 4D light field [1, 2]. This elegant formulation inspires many new interpretations of light-object interaction and enables novel applications [3].

Various techniques have been proposed for light field acquisition [3], but the tradeoff between efficiency, device size, and resolution remains an open issue. To address it, this paper proposes a new device called *Programmable Aperture Camera* (PAC) that can be created by inserting a programmable aperture into a regular camera. PAC can be as compact as a normal camera and can capture high resolution light field by changing the shape (pattern) of the aperture without moving the camera. For efficiency, the aperture pattern is designed based on the *Hadamard transform optics*. The total exposure time remains the same as that of taking a single regular photograph at the same image quality level.

The rest of the paper is organized as follows. Section 2 gives a review of some related work. Section 3 describes the concept of the programmable aperture camera, and Section 4 describes the prototype implementation in detail. Section 5 demonstrates how the captured light field can be used for digital refocusing. The conclusion and future work are drawn in Section 6.

∗We thank Ching-Chang Liao for developing a prototype of the programmable aperture camera. Chia-Kai Liang was supported in part by the MiTac Technology Scholarship. This project was supported in part by Ministry of Education, R.O.C., under the contract 95R0062-AE00-02.

2. RELATED WORK

Among the light field acquisition methods that have been proposed, the most straightforward one takes many images at different view positions and then rectifies the resulting data using structure from motion [2]. However, this method is slow and prone to error. Another method builds a highly calibrated camera array [4]. However, camera array is expensive to make and cumbersome.

Many portable light field cameras [5, 6] that have been developed based on old techniques [7, 8] have the following drawbacks. First, they reduce the spatial resolution by a factor of tens to hundreds. Second, they can not function as traditional cameras to capture normal photographs. Third, they require many costly optical components. PAC has many advantages over these cameras: it maintains the spatial resolution, it can operate as a regular camera, it is much easier to manufacture and calibrate, and it enables many new imaging techniques.

Our idea is to program the shape of the aperture. This is inspired by some previous work in computational cameras. The most related ones are uncontrolled modulation imaging (UMI) [9] and lensless imaging with a controllable aperture (LICA) [10]. Both devices place a light modulator (or attenuator) in front of the sensor array, which is somewhat similar to PAC. However, PAC is different from them in many perspectives. UMI uses a light modulator to control the global camera response. It can capture high-dynamic-range or hyper-spectral images. By contrast, PAC adjusts the shape of the aperture to capture the light field. LICA uses some attenuating layers to create some novel functions, such as view point switch and multiple field-of-view imaging. However, LICA is essentially a pinhole camera that requires long exposure. On the contrary, PAC takes the lens module of the camera as is, which results in superior image quality. A closely related work has been developed in [11]. The 4D light field is recovered from a single photo captured by a camera with a specially designed mask. However, this method still seriously degrades the spatial resolution.

3. PROGRAMMABLE APERTURE CAMERA

We first describe the definition of the light field in a camera, and show how previous methods acquire it. Then we show how a programmable aperture camera can be used to capture the light field sequentially. Finally we introduce Hadamard transform optics to increase the efficiency.

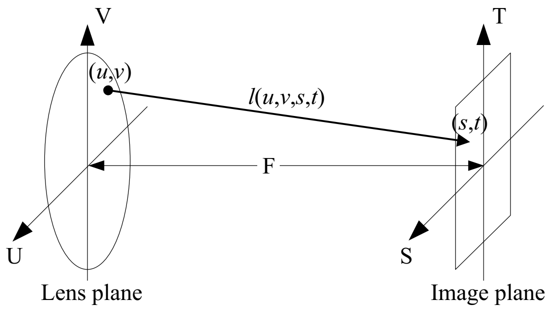

Fig. 1. The light field in a camera.

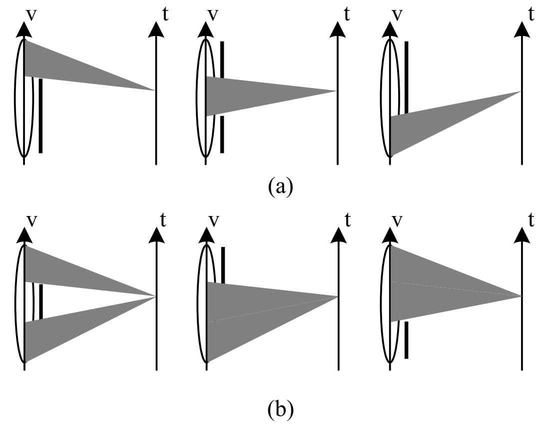

Fig. 2. The programmable aperture can control the shape of the aperture. (a) Only a single region on the lens plane is open in each exposure. (b) Multiple regions on the lens plane are open in each exposure using S-matrix.

3.1. Light field in a camera

A light ray within a camera must enter the camera from the exit pupil (lens plane) of the lens module and fall somewhere on the image plane, as shown in Fig. 1. Because each light ray only intersects each plane once, it can be represented by the intersections points (u,v) and (s,t). Therefore, each light ray in a camera is a sample of the 4D light field $l(u,v,s,t)$ [1]. Two assumptions are made in the following analyses. First the scene is static. Second, both the lens and the image planes are discreteized; (u,v) and (s,t) are sample indices.

In a traditional camera, a sensor element at (s,t) collects all light rays coming from the lens plane within the exposure time T, to form a pixel $p(s,t)$:

$$p(s,t) = \frac{T}{F^2}\sum_{u=0}^{N_u-1}\sum_{v=0}^{N_v-1} l(u,v,s,t), \tag{1}$$

where F denotes the distance between the lens and image planes, and N_u and N_v denote the numbers of horizontal and vertical samples respectively and determine the *angular resolution*. The angular attenuation term is omitted here for simplicity. The information of the light field on the lens plane is lost due to the summation. Therefore, the light field l can not solely be recovered from the image $p(s,t)$.

To preserve the angular information, the plenoptic camera (or light field camera) places a microlens array on the original image plane and shifts the sensor array a little further from the lens [5]. Consequently, the rays convergent to the image plane become divergent; each sensor collects only one direction of rays from the lens plane. Many equivalent optical designs were proposed in [6]. However, such designs seriously degrade the spatial resolution because the product of angular and spatial resolution is limited by the number of sensor elements.

3.2. Programmable aperture

If in each exposure only a portion of the aperture around (u_0,v_0) is open, the image is a subset of the light field samples $l(u_0,v_0,s,t)$, where $s = 0,1,...,N_s-1$ and $t = 0,1,...,N_t-1$. The complete light field is obtained by taking many images with different open apertures. This operation is possible if a programmable light blocker is inserted at the lens plane, as shown in Fig. 2(a). We call the light blocker *programmable aperture* (PA) and a camera equipped with the PA programmable aperture camera. The shape of the aperture is called *blocking pattern*. Note that PAC can adjust the shape of the aperture, but the conventional cameras can only adjust its size.

However, if only a single region is open in each exposure, the resulting samples would be much noisier than the photo captured at full aperture. This is because the obtained light intensity becomes smaller and the noise level remains unchanged. As a result, the signal-to-noise ratio (SNR) is reduced. We solve this problem by using the Hadamard transform optics [12]. In each exposure, many regions are open instead of one. The light field is then estimated from the captured images. The Hadamard transform optics has been widely applied in many areas, such as multi-spectral imaging, single sensor imaging, and image-based relighting [13].

The 1D example shown in Fig. 2 is used to explain how the SNR is improved using Hadamard transform optics. Because $N_v = 3$, we need to shoot 3 times to capture the light field. If the total exposure is still T, the exposure of each photo is $T/3$. Suppose each sensor measures the light intensity with an independent additive noise having variance σ^2, the SNR of the light field sample is:

$$SNR_0(v,t) = T \cdot l(v,t)/3\sigma. \tag{2}$$

Note that the spatial index t will be omitted for simplicity.

Now if in each exposure many regions are open, each pixel p in the i-th photo is a linear combination of the light field:

$$p(v) = \sum_{j=0}^{2} w_{vj} l(j) + \sigma^2, v,j = 0,1,2, \tag{3}$$

where $0 \leq w_{vj} \leq 1$ because physically we can only attenuate the incoming light. All $p(i)$'s can be arranged into a vector:

$$\mathbf{p} = \mathbf{W}\mathbf{l} + \mathbf{n}, \tag{4}$$

where $\mathbf{W}(v,j) = w_{vj}$ and $n_v = \sigma^2$.

Because the noise vector $\mathbf{n}$ is independent of $\mathbf{l}$, it is simple to show that the best estimation $\widetilde{\mathbf{l}}$ in the least-square-error sense is

$$\widetilde{\mathbf{l}} = \mathbf{W}^{-1}\mathbf{p}, \tag{5}$$

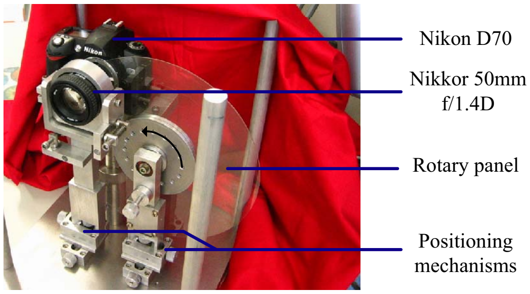

Fig. 3. The PAC prototype.

and the mean squared error (MSE) of each $\widetilde{l}_v$ is

$$\begin{aligned} MSE &= \tfrac{1}{3}\text{Trace}(E(\widetilde{\mathbf{l}} - E(\mathbf{l}))E(\widetilde{\mathbf{l}} - E(\mathbf{l}))^T) \\ &= \tfrac{1}{3}\text{Trace}(\sigma^2(\mathbf{W}^T\mathbf{W})^{-1}) = \tfrac{\sigma^2}{3}\text{Trace}((\mathbf{W}^T\mathbf{W})^{-1}), \end{aligned} \quad (6)$$

where $E()$ denotes the expectation.

Many matrices can minimize the MSE (6) under the constraint $0 \leq \mathbf{W}(i,j) \leq 1$, but not all of them can be formulated analytically. A well-formulated matrix is the *S-matrix* [12]. For an $n \times n$ S-matrix, its row is based on an $n+1$ *Hadamard code*. In this example the 3×3 S-matrix is $[(101)^T, (011)^T, (110)^T]$. As shown in Fig. 2(b), in each exposure the blocking pattern corresponds to a row of the S-matrix.

The optimal MSE using a $n \times n$ S-matrix is:

$$MSE_s = 4n\sigma^2/(n+1)^2, \quad (7)$$

which is $3\sigma^2/4$ for $n = 3$. Therefore the SNR using S-matrix, SNR_s, is increased from SNR_0 by a factor of

$$\frac{SNR_s}{SNR_0} = \frac{\sqrt{n} + (1/\sqrt{n})}{2}, \quad (8)$$

which is 1.155 for $n = 3$. In light field acquisition, $n = N_u \times N_v$, which usually ranges from tens to hundreds. Hence the SNR improvement is significant. In PAC, n is an adjustable parameter. User can specify it according to the storage limitation and applications while the spatial resolution is not affected.

Comparing the image quality of the PCA with that of the light field camera directly is not meaningful. In a light field camera, each sensor collects the light within a much larger area on the image plane and thus SNR is increased. However, the spatial resolution is seriously degraded. If the total exposure time and the sampling rates of the PCA and the light field camera are the same, the SNR are comparable. However, in that case the light field camera would require much more sensors which may exceed the physical limitation.

4. PROTOTYPE

Our PAC prototype is shown in Fig. 3. A Nikon DSLR D70 is adopted as the main camera body. The sensor array is 3040×2014 with Bayer color filter. The lens module is Nikkor 50mm f/1.4D. In the experiments the original stop is set at the maximal value.

We use a rotary panel to implement a programmable aperture. The rotary panel is a circular transparent plastic plane. The designed blocking patterns are pasted on the panel. 16 blocking patterns can be placed on the panel. All the pieces are aligned using the positioning mechanisms, which can translate and rotate independently.

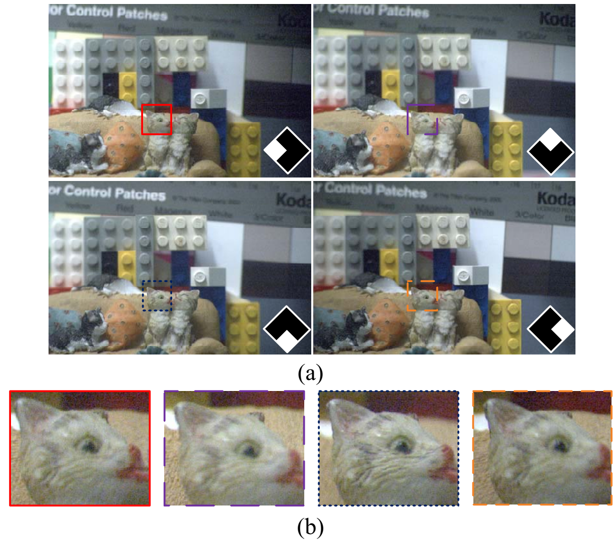

Fig. 4. (a) Light field captured by the PAC prototype. The insets illustrate the corresponding blocking patterns. (b) The close-up images of (a). Note the disparity of the cat ears.

The PAC prototype can capture the light field with multiple exposures. It can also capture regular pictures by setting the programmable aperture fully open.

In the original D70, there are many image processing steps which corrupt the linearity of the light intensity, and thus disables the estimation process described in Section 3.2. To avoid this problem, we dump the 12-bit linear raw data. Only demosaicing, white balance, and gamma correction are performed.

A $2 \times 2 \times 3040 \times 2019$ light field captured by the prototype is shown in Fig. 4. The total exposure time is only $40ms$. In Fig. 4(b), the disparity of the cat ears is around 20 pixels. The disparity magnitude can be adjusted by changing F and the lens parameters.

5. DIGITAL REFOCUSING

The captured light field can be used in many applications. Due to the page limit, only digital refocusing is demonstrated.

Light rays emitted from objects at different depths converge to different locations in the camera. Only the objects with rays converging on the image plane result in sharp images, and the other objects appear blurred. The in-focus range (depth-of-field) is determined by the camera parameters, such as the aperture size.

Focusing is essentially a filtering of the 4D light field [14]. If we have a complete 4D light field, we can specify the filter kernel; thus the focus depth and aperture size can be changed after shooting. This technique is called synthetic aperture imaging, or digital refocusing.

In previous work, the new focal plane can only be parallel or tilted to the original image plane [15], unless we assign the depth of each pixel and perform ray tracing, which is impossible in most cases. However, even specifying a focal plane is not an intuitive thing, users need to understand the scene structure and perspective

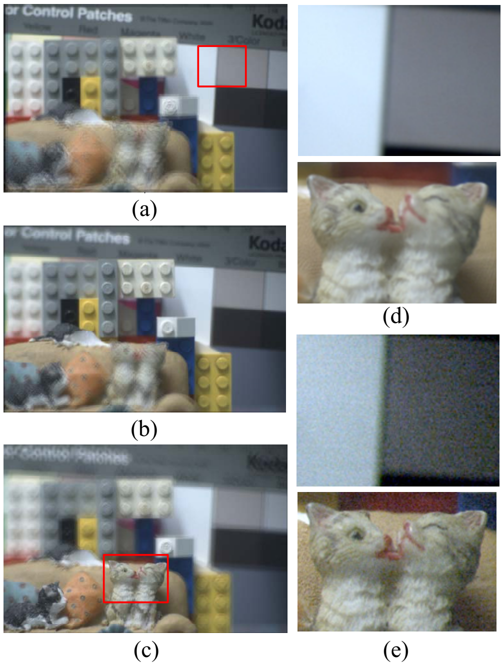

Fig. 5. Digital refocusing on (a) the color checker, (b) the Lego, and (c) the cats. (d) The close-up images of (a) and (c). (e) The close-up of a regular photo at a smaller aperture. With the same exposure and depth-of-field settings, refocus images have superior SNR.

transformation. Here we propose a novel method to perform refocusing based on feature matching [16].

Our algorithm has two phases: preprocess and interactive refocusing. In preprocessing, the light field is organized as a 2D array of light field images, as shown in Fig. 4(a). For each light field image, the SIFT features are extracted, and the kd-tree of detected features is established for the fast search in the next phase.

In interactive refocusing, a user first marks a region-of-interest (ROI) to be focused on in one image. Next, the features in the ROI are identified and their matches in other images are found. Because matches are not always correct, RANSAC is performed to remove the outliers and then obtain accurate translational motions (disparities) among images. Finally the images are shifted and added together to generate the refocused image. The in-focus regions can be specified arbitrarily. That is, they need not be parallel to the original image plane. However, this requires additional warping that may degrade the image quality.

Examples of the refocused images are shown in Fig. 5 (a)-(c). Compared with the photos taken at a smaller aperture (for the same depth-of-field) and with the same exposure time, the noise is greatly suppressed in the refocused ones, as shown in Fig. 5 (d) and (e). The little blur of the refocused images is due to the limited precision in shifting images. Except the preprocessing, the interactive refocusing operates in real-time on a Pentium 4 3.2GHz computer.

6. CONCLUSION AND FUTURE WORK

We have proposed a programmable aperture camera to capture the 4D light field. Using Hadamard transform optics, PAC can efficiently capture quality light field data. We have built a prototype using a conventional camera and a rotary panel. We have presented a new method to perform digital refocusing using feature matching, and demonstrated it using the captured light field.

A more compact way to make the programmable aperture is using a liquid crystal array (LCA) which can be embedded in the camera body and controlled electronically. However, previous work has showed that current LCAs have some drawbacks that degrade the image quality [10]. As a result, a mechanical approach is chosen in our prototype. With the improvement of liquid crystal techniques, it is possible to build a compact PAC using LCA in the near future.

Our work can be extended in several ways. First, the captured light field needs to be compressed. For a single light field with 3×3 angular resolution, the raw data occupy 83MB. Although many lossy light field compression methods have been proposed, lossless compression is required because high precision and linearity of the data is required for refocusing and other post-processing.

Second, many other advanced imaging techniques can be applied to PAC. For example, we observe that by slightly changing the blocking patterns and capture a number of pictures, the depth discontinuities of the scene can be extracted. The programmable aperture can be used as a shutter if the operation speed is fast enough, and thus coded exposure photography can be applied to benefit the deblurring process. Finally, all computational cameras that attempt to shape the aperture can be implemented by PAC.

7. REFERENCES

[1] M. Levoy and P. Hanrahan, "Light field rendering," in *ACM SIGGRAPH*, 1996, pp. 31–42.

[2] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen, "The lumigraph," in *ACM SIGGRAPH*, 1996, pp. 43–54.

[3] M. Levoy, C.-K. Liang, and anonymities, "Light field," 2006, http://en.wikipedia.org/wiki/Light_field.

[4] B. Wilburn et al., "High performance imaging using large camera arrays," *ACM Trans. Graph.*, vol. 24, no. 3, pp. 765–776, 2005.

[5] R. Ng et al., "Light field photography with a hand-held plenoptic camera," CSTR 2005-02, Stanford University, 2005.

[6] T. Georgeiv et al., "Spatio-angular resolution tradeoff in integral photography," in *EGSR*, 2006.

[7] M. G. Lippmann, "Epreuves reversible donnant la sensation du relief," *J. Phys*, vol. 7, pp. 821–825, 1908.

[8] H. E. Ive, "Parallax panoramagrams made with a large diameter concave mirror," *JOSA*, vol. 18, pp. 332–342, June 1930.

[9] Y. Y. Schechner and S. K. Nayar, "Uncontrolled modulation imaging," in *CVPR*, Jun 2004, vol. 2, pp. 197–204.

[10] A. Zomet and S. K. Nayar, "Lensless imaging with a controllable aperture," in *CVPR*, Jun 2006, vol. 1, pp. 339–346.

[11] A. Veeraraghavan, R. Raskar, A. Agrawal, A. Mohan, and J. Tumblin, "Mask based modulation for light field capture by conventional cameras," in *ACM SIGGRAPH*, 2007.

[12] M. Harwit and N. J.A. Sloane, *Hadamard Transform Optics*, Academic Press, New York, 1979.

[13] Y. Y. Schechner, S. K. Nayar, and P. N. Belhumeur, "A theory of multiplexed illumination," in *ICCV*, 2003, pp. 808–815.

[14] A. Isaksen, L. McMillan, and S. J. Gortler, "Dynamically reparameterized light fields," in *ACM SIGGRAPH*, 2000, pp. 297–306.

[15] V. Vaish et al., "Synthetic aperture focusing using a shear-warp factorization of the viewing transform," in *A3DISS*, 2005, vol. 3, pp. 129–136.

[16] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *IJCV*, vol. 60, no. 2, pp. 91–110, 2004.