PCA-CODED APERTURE FOR LIGHT FIELD PHOTOGRAPHY

Yusuke Yagi †, *Keita Takahashi* †, *Toshiaki Fujii* †, *Toshiki Sonoda* ††, *Hajime Nagahara* †††

† Department of Electrical Engineering and Computer Science, Nagoya University, Japan
†† Department of Advanced Information Technology, Kyushu University, Japan
††† Institute for Datability Science, Osaka University, Japan

ABSTRACT

A light field, which is often understood as a set of dense multi-view images, has been utilized in various 2D/3D applications. Efficient light field acquisition using a coded aperture camera is the target problem considered in this paper. Specifically, the entire light field, which consists of many images, should be reconstructed from only a few images that are captured through different aperture patterns. In previous work, this problem has often been discussed from the context of compressed sensing (CS). In contrast, we formulated this problem from the perspective of principal component analysis (PCA) to derive optimal non-negative aperture patterns and a straight-forward reconstruction algorithm. Even though it is based on a conventional technique, our method has proven to be more accurate and much faster than a state-of-the-art CS-based method.

***Index Terms*—** PCA, light field, coded aperture

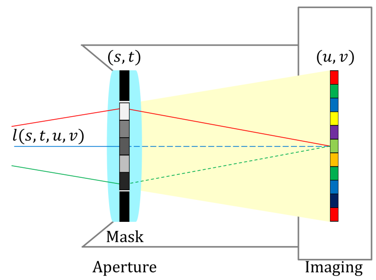

Fig. 1. Coded aperture camera model. Each light ray passing through (s,t) on the aperture plane is attenuated by the transmittance at (s,t) and reaches a pixel (u,v) on the imaging plane.

1. INTRODUCTION

A light field, which is often understood as a set of dense multi-view images, has been utilized in various applications, such as free-viewpoint image synthesis [1, 2], depth estimation [3, 4], synthetic refocusing [5], super resolution [6, 3], and 3D displays [7, 8, 9, 10]. Acquisition of a sufficiently dense light field is a challenging task due to the huge amount of data. Several researchers have used direct approaches such as a moving camera gantry [1] or multiple cameras [11, 12] to capture the target from different viewpoints. Meanwhile, lens-array based cameras [13, 14, 5, 15] and coded aperture/mask cameras [16, 17, 18, 19, 20, 21] have also been utilized to achieve more efficient acquisition of the light field.

This paper focuses on a problem of efficient light field acquisition using a coded aperture camera; the entire light field, which consists of many images, should be reconstructed from only a few images that are captured with different aperture patterns. This problem has often been discussed from the context of compressed sensing (CS) [22, 23, 24], which provides a sophisticated framework for signal reconstruction from a limited number of samples. In contrast, we formulate this problem from the perspective of principal component analysis (PCA) to derive optimal non-negative aperture patterns and a straight-forward reconstruction algorithm. Even though it is based on a rather old-fashioned technique, our method has proven to be more accurate and much faster than a state-of-the-art CS-based method [20].

2. THEORY AND METHODS

In this section, we first describe the problem statement of light field acquisition using a coded aperture camera. Next, a common approach to this problem based on compressed sensing (CS) is mentioned, followed by another approach based on principal component analysis (PCA). Finally, On the basis of the latter approach, we designed a method of deriving optimal non-negative aperture patterns and a straight-forward reconstruction algorithm are proposed.

2.1. Problem Statement

A mathematical formulation for light field acquisition using this camera model is presented supposing a camera model illustrated in Fig. 1, which is similar to the cameras in previous work [16, 17, 18, 19, 20, 21].

All incoming light rays that will be recorded by this camera are parameterized with a tuple of four variables (s,t,u,v), where (s,t) and (u,v) denote the intersections with the aperture and imaging planes, respectively. Therefore, the light field is defined over the 4-D space (s,t,u,v), with which the light intensity is described as $l(s,t,u,v)$. We consider a coded aperture design where the transmittance of the aperture can be controlled for each position and for each acquisition. Let $a_n(s,t)$ be the transmittance at position (s,t) for the n-th acquisition ($n = 1,\dots,N$). The observed image $y_n(u,v)$ is formed as

$$y_n(u,v) = \iint a_n(s,t) l(s,t,u,v) ds dt. \tag{1}$$

When the aperture plane is quantized into a finite number of blocks, they are numbered by an integer m ($m = 1,\dots,M$). In this case, Eq. (1) is rewritten as

$$y_n(u,v) = \sum_{m=1}^{M} a_{n,m} x_m(u,v), \tag{2}$$

978-1-5090-2175-8/17/$31.00 ©2017 IEEE

where M is the total number of blocks, and $a_{n,m}$ (equivalent to $a_n(s,t)$) is the transmittance of the aperture at position m for the n-th acquisition. Symbol $x_m(u,v)$ (equivalent to $l(s,t,u,v)$) is called a sub-aperture image because it is formed only by the light rays that pass through a sub-region (the position denoted by m) on the aperture plane. Sub-aperture images can also be called multi-view images because their viewpoints (denoted by m) are slightly different from each other. Reconstruction of the light field is equivalent to estimating M sub-aperture images $x_m(u,v)$ from the N given observations $y_n(u,v)$.

Putting all the N observations together, the image formation model (observation model) for each pixel (u,v) is expressed in a vector-matrix format as

$$\boldsymbol{y}_{u,v} = \boldsymbol{A}\boldsymbol{x}_{u,v}, \tag{3}$$

where $\boldsymbol{y}_{u,v} \in \mathbb{R}^N$ is a column vector called an observation or measurement whose n-th element is $y_n(u,v)$, $\boldsymbol{A} \in \mathbb{R}^{N\times M}$ is called an observation matrix whose (n,m) element is given by $a_{n,m}$, and $\boldsymbol{x} \in \mathbb{R}^M$ is a column vector representing the target signal whose m-th element is $x_m(u,v)$. The goal here is to obtain $\boldsymbol{x}_{u,v}$ from the given $\boldsymbol{A}$ and $\boldsymbol{y}_{u,v}$. To pursue the efficiency, we should do it with $N < M$: the number of acquisitions should be smaller than the number of viewpoints of the sub-aperture images, which makes Eq. (3) an under-determined system.

2.2. CS-based approach

The aforementioned problem is often considered in the context of compressive sensing (CS) [22, 23, 24]. A general framework utilized in several studies [25, 19, 20, 21, 26] is presented here.

To fully exploit the inherent structure of the light field signal, a processing unit of the light field should cover both the viewpoint (aperture) and pixel (imager) domains. Specifically, a processing unit $\boldsymbol{X} \in \mathbb{R}^{MB}$ is defined to contain M viewpoints on the aperture plane and a pixel block with B pixels on the imaging plane, where the elements are aligned in pixel-major order.

In accordance with the definition of a processing unit, the observation model of Eq. (3) is rewritten as

$$\boldsymbol{Y} = \boldsymbol{\Phi_A}\boldsymbol{X}, \tag{4}$$

where $\boldsymbol{Y} \in \mathbb{R}^{NB}$ contains the observations on the B pixels inside the block with N times acquisitions in the pixel-major order. The observation matrix $\boldsymbol{\Phi_A} \in \mathbb{R}^{NB\times MB}$ is given as $\boldsymbol{A} \otimes \boldsymbol{I}$, where $\boldsymbol{I} \in \mathbb{R}^{B\times B}$ and $\otimes$ denote the identity matrix and Kronecker product operator.

The underlying assumption behind compressive sensing is that the signal can be represented in a sparse form using an appropriate basis or dictionary. Specifically, the signal can be represented as $\boldsymbol{X} = \boldsymbol{D}\boldsymbol{\theta}$, where $\boldsymbol{D} \in \mathbb{R}^{MB\times K}$ is a basis or a dictionary having K elements, and $\boldsymbol{\theta} \in \mathbb{R}^K$ is a sparse coefficient vector where only a few elements can take non-zero values. The estimation of $\boldsymbol{\theta}$ is formulated as a minimization problem with a non-negative weight λ:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \frac{1}{2}\|\boldsymbol{Y} - \boldsymbol{\Phi_A}\boldsymbol{D}\boldsymbol{\theta}\|_2^2 + \lambda\|\boldsymbol{\theta}\|_1^1, \tag{5}$$

from which the reconstructed signal is obtained as $\hat{\boldsymbol{X}} = \boldsymbol{D}\hat{\boldsymbol{\theta}}$. Solving this minimization problem often takes significant time.

An important question here is how the aperture pattern $\boldsymbol{A}$ can be determined to achieve the best reconstruction quality. One possible answer can be found in the work done by Marwah et al. [20]. Matrix $\boldsymbol{G_A} \in \mathbb{R}^{NB\times K}$ is defined as $\boldsymbol{G_A} = \boldsymbol{\Phi_A}\boldsymbol{D}$, and each column of $\boldsymbol{G_A}$ is normalized to the unit length to yield $\tilde{\boldsymbol{G}}_{\boldsymbol{A}}$. The optimal $\boldsymbol{A}$ is obtained as

$$\boldsymbol{A} = \arg\min_{\boldsymbol{A}} \|\boldsymbol{I} - \tilde{\boldsymbol{G}}_{\boldsymbol{A}}^T\tilde{\boldsymbol{G}}_{\boldsymbol{A}}\|_F. \tag{6}$$

This scheme is closely related to the incoherence among the observations, which is a desirable property for compressive sensing [23].

2.3. PCA-based approach

Another approach to the problem mentioned in Sec. 2.1 can be found from the context of principal component analysis (PCA) [27]. Here, we consider the observation model of Eq. (3) again, where the vector $\boldsymbol{x}_{u,v}$ contains M viewpoints for a fixed pixel (u,v). To simplify the notation, we drop the subscript (u,v).

In accordance with the idea of PCA[1], $\boldsymbol{x}$ is described as

$$\boldsymbol{x} = \boldsymbol{P}\boldsymbol{\theta}, \tag{7}$$

where matrix $\boldsymbol{P} = [\boldsymbol{p}_1 \dots \boldsymbol{p}_M] \in \mathbb{R}^{M\times M}$ contains the principal component (PC) vectors $\boldsymbol{p}_m \in \mathbb{R}^M$ in descending order, and vector $\boldsymbol{\theta} = [\theta_1,\dots,\theta_M]^T \in \mathbb{R}^M$ contains the corresponding PC coefficients. As an example, the PC vectors and their cumulative contribution ratio for 25 (5×5) viewpoints, which will be detailed in Sec. 3, are visualized in Fig. 2. When we use only the first N ($N < M$) PC vectors, we have an approximation of $\boldsymbol{x}$ as

$$\boldsymbol{x} \simeq \tilde{\boldsymbol{P}}\tilde{\boldsymbol{\theta}}, \tag{8}$$

where $\tilde{\boldsymbol{P}} = [\boldsymbol{p}_1 \dots \boldsymbol{p}_N] \in \mathbb{R}^{M\times N}$ and $\tilde{\boldsymbol{\theta}} = [\theta_1,\dots,\theta_N]^T \in \mathbb{R}^N$. Equation (8) is the optimal approximation in the sense of the least squared error as far as only N orthogonal components can be used.

Substituting Eq. (7) into the observation model, we derive

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}\boldsymbol{P}\boldsymbol{\theta}. \tag{9}$$

It is easy to see that if $\tilde{\boldsymbol{P}}^T \in \mathbb{R}^{N\times M}$ were set to $\boldsymbol{A}$ (equivalently, each measurement would be conducted with each of the first N PC vectors), the first N PC coefficients could be directly obtained as $\boldsymbol{y} = \tilde{\boldsymbol{P}}^T\boldsymbol{P}\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}$, from which we could reconstruct $\boldsymbol{x}$ using Eq. (8). However, in practice, we cannot use $\tilde{\boldsymbol{P}}^T$ as the actual aperture pattern because all of the PC vectors except for $\boldsymbol{p}_1$ (which corresponds to the DC component) have negative elements. To make feasible aperture patterns, all the elements of $\boldsymbol{A}$ should be non-negative.

To satisfy this non-negativity condition, Ashok et al. [28] used a dual rail measurement [29], where each PC vector $\boldsymbol{p}_m$ is divided into positive and negative parts, $\boldsymbol{p}_m^+$ and $\boldsymbol{p}_m^-$, satisfying $\boldsymbol{p}_m = \boldsymbol{p}_m^+ - \boldsymbol{p}_m^-$. Two measurements using $\boldsymbol{p}_m^+$ and $\boldsymbol{p}_m^-$ were conducted, and the results were combined to obtain a measurement that would be virtually conducted with $\boldsymbol{p}_m$. An obvious disadvantage of this method is the increase in the number of acquisitions; if we need N PC coefficients, the number of acquisitions is $2N-1$ because all PC vectors except for $\boldsymbol{p}_1$ require two measurements for each.

[1]PCA is formulated as an eigenvalue problem: $\boldsymbol{S}\boldsymbol{p} = \rho\boldsymbol{p}$, where $\boldsymbol{S} = E[\boldsymbol{x}\boldsymbol{x}^T]$ is the covariance matrix, and ρ is an eigenvalue. We obtain M non-negative eigenvalues ρ_m and corresponding eigenvectors $\boldsymbol{p}_m$, which are sorted in descending order of the eigenvalues. The eigenvectors satisfy ${\boldsymbol{p}_m}^T\boldsymbol{p}_{m'} = 0$ for $m \neq m'$ and ${\boldsymbol{p}_m}^T\boldsymbol{p}_m = 1$.

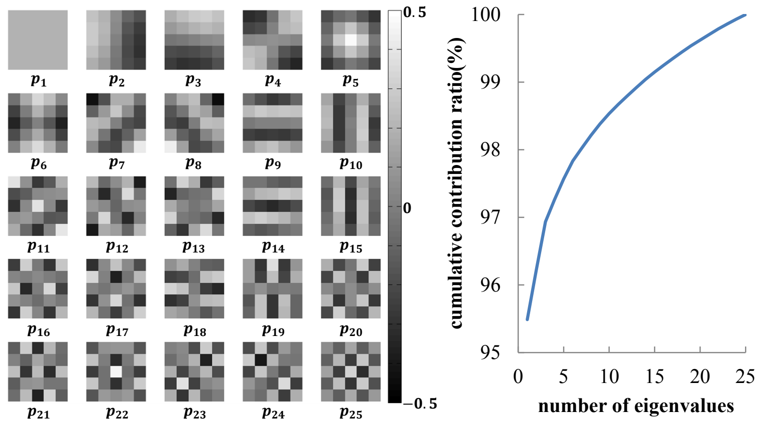

Fig. 2. PC vectors and their cumulative contribution ratio.

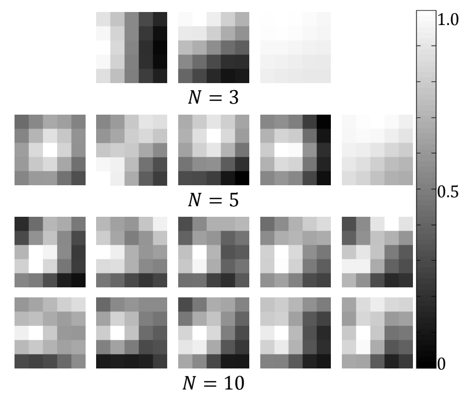

Fig. 3. Aperture patterns derived using our method.

2.4. Proposed Method

On the basis of the PCA-based approach, we propose a method of determining the aperture patterns that can satisfy the non-negativity condition without increasing the number of acquisitions.

We introduce a new matrix $\boldsymbol{C} \in \mathbb{R}^{N \times N}$ called a mixing matrix and define the observation matrix as $\boldsymbol{A} = \boldsymbol{C}\tilde{\boldsymbol{P}}^T$. By mixing PC vectors using a carefully chosen matrix $\boldsymbol{C}$, we aim to make all the elements of $\boldsymbol{A}$ non-negative, enabling us to obtain N PC coefficients while keeping the number of acquisitions N. Moreover, we consider the stability of the solution $\tilde{\boldsymbol{\theta}}$ obtained with $\boldsymbol{C}$. By substituting $\boldsymbol{A} = \boldsymbol{C}\tilde{\boldsymbol{P}}^T$ into the observation model, we have $\boldsymbol{y} = \boldsymbol{C}\tilde{\boldsymbol{P}}^T\boldsymbol{P}\boldsymbol{\theta} = \boldsymbol{C}\tilde{\boldsymbol{\theta}}$. To make this linear system stably invertible, we should make the condition number of $\boldsymbol{C}$ small. Therefore, we determine the mixing matrix $\boldsymbol{C}$ (and in return, the observation matrix $\boldsymbol{A}$) as follows.

1. We randomly generate the elements of $\boldsymbol{C}$ and retain the matrix as a candidate if all the elements of $\boldsymbol{A} = \boldsymbol{C}\tilde{\boldsymbol{P}}^T$ are non-negative. We generate a sufficient number of candidates.

2. Choose one of the candidates that minimizes the condition number.

The final observation matrix, which is generated with the selected mixing matrix, is normalized in each row so that the maximum element of each row equals 1.0. The aperture patterns derived using our method for $N = 3, 5, 10$ are shown in Fig. 3, each of which was selected out of 10,000 random samples.

In accordance with Eq. (8), the light field signal $\boldsymbol{x}$ is reconstructed as $\hat{\boldsymbol{x}} = \tilde{\boldsymbol{P}}\tilde{\boldsymbol{\theta}} = \tilde{\boldsymbol{P}}\boldsymbol{C}^{-1}\boldsymbol{y}$. Once $\tilde{\boldsymbol{P}}$ and $\boldsymbol{C}$ have been obtained in the off-line process, this reconstruction is extremely simple and computationally easy compared with that of the CS-based approach.

Our method is essentially equivalent to the CS-based approach in the underlying assumption that a target signal $\boldsymbol{x}$ would be represented by combining only a small number of *basic components* (atoms, basis vectors, or PC vectors). However, they are absolutely different in how to *select* the basic components for the target signal. In the CS-based approach, determining whether each element of $\boldsymbol{\theta}$ is used or not is a part of the problem. Meanwhile, in our method, used/unused elements are necessarily determined at the point where $\boldsymbol{\theta}$ is reduced to $\tilde{\boldsymbol{\theta}}$ in accordance with the concept of PCA. In the latter case, the problem becomes much simpler, and the quality of the reconstructed signal improves–as will be demonstrated in the next section.

3. EXPERIMENTS

We conducted simulative experiments using several light field datasets [30, 31]. Throughout the experiments, the pixel values of the original datasets and the transmittances for the aperture patterns were normalized within the range of $[0, 1]$. Moreover, to simulate noisy acquisition processes, a zero-means Gaussian noise with a standard deviation σ was added to the observed images. The covariance matrix $\boldsymbol{S}$ used in our method was calculated from four datasets (Dice, Fishi, Messerschmitt, and Shrubbery) [30] because the same datasets were used to learn the light field dictionary of [20]. The reconstruction quality was measured by PSNR against the ground truth.

Figure 4 shows the performance of our method obtained with a different number of acquisitions and noise levels. The dataset used was *DragonsAndBunnies*. When $\sigma = 0$, the reconstruction quality increased as the number of acquisitions increased, which is an expected behavior in accordance with the principal of PCA. However, when $\sigma > 0$, the reconstruction quality initially increased but then decreased beyond a certain point. This phenomenon reflects the fact that less significant PC coefficients are generally small and more likely to be below the noise level; if we try to reconstruct these tiny components, the noise is amplified in return. For the experiments that will be mentioned hereafter, we fixed σ to 0.005.

We also evaluated the significance of the condition number of the mixing matrix $\boldsymbol{C}$ in our method. We randomly generated 100 matrices with three acquisitions ($N = 3$). The reconstruction quality for the DragonsAndBunnies dataset are plotted against the condition number of $\boldsymbol{C}$ in Figure 5. We can see that the reconstruction quality decreased sharply when the condition number exceeded 50. In the same graph, the red and green line indicate the reconstruction quality of the dual rail measurement [28] with $N = 3$ (two PC vectors) and $N = 5$ (three PC vectors), respectively. The line with $N = 5$ gives the upper-bound quality that can be achieved with three PC vectors. Clearly, our method can reach the upper-bound quality if a good $\boldsymbol{C}$ is used. Meanwhile, our method is much better than the dual rail measurement for the same number of acquisitions $N = 3$ unless a significantly bad $\boldsymbol{C}$ is chosen.

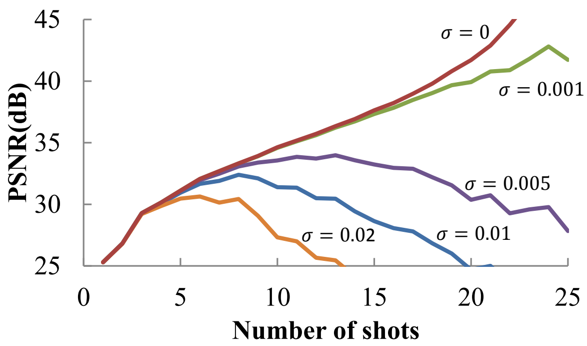

Fig. 4. Reconstruction quality vs. the number of acquisitions while changing the noise level σ.

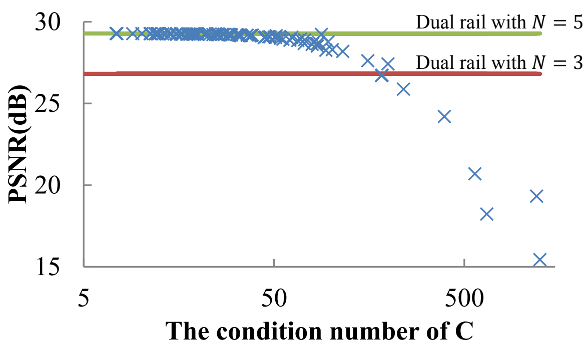

Fig. 5. Reconstruction quality vs. condition number of $\boldsymbol{C}$. The red and green lines indicate the reconstruction quality of dual rail measurements with $N = 3$ and $N = 5$, respectively.

Finally, we compared our method with a state-of-the-art CS-based method that uses a learned dictionary [20]. The number of acquisitions N was set to 3. The details of the CS-based method are described as follows. We used the software and the learned dictionary available from [30]. In the reconstruction process, each light field dataset was divided into non-overlapping processing units, whose sizes were set to $5 \times 5 \times 9 \times 9$ (5×5 for viewpoints and 9×9 for pixels). After several tests, we found that the default value 0.01 for λ in Eq. (5) works well for all the datasets, so we fixed λ to 0.01. As for the aperture pattern, we randomly generated 10,000 samples for $\boldsymbol{A}$ and retained the one that optimized the Eq. (6). Table 1 summarizes the specifications of 17 datasets and the reconstruction quality obtained using the CS-based and our methods. The average PSNR values are also presented at the bottom. Clearly, our method outperformed the CS-based method with considerable margins for all datasets. Figure 6 shows the reconstructed and ground truth images for the central viewpoint of DragonsAndBunnies, where we can see that better visual quality was achieved using our method.

Moreover, in terms of the computational complexity, our method is much simpler than the CS-based method. Once the aperture patterns were determined, our method could reconstruct an entire light field dataset in only 0.1 – 0.5 seconds with our unoptimized MATLAB implementation. Meanwhile, the CS-based method took 500 – 3000 seconds for the same reconstruction task on the same computer.

Table 1. Light field datasets from [30] and [31], and comparison of the reconstruction quality with CS-based approach and our method.

Dataset	Dir. Res. (s,t)	Pos. Res. (u,v)	PSNR [dB] CS	PSNR [dB] Prop.
DragonsAndBunnies			27.22	29.29
Dice			26.23	27.56
Fishi	5×5	840×593	24.52	25.98
Messerschmitt			25.75	28.04
Shrubbery			13.41	14.42
Amethyst		384×512	31.84	36.33
Bracelet		512×320	22.13	24.62
Lego Bulldozer		768×576	24.55	26.39
The Stanford Bunny		512×512	33.95	37.45
Chess		700×400	28.70	32.11
Treasure Chest	5×5	768×640	23.89	25.82
Eucalyptus Flowers	out of	640×768	29.42	32.22
Jelly Beans	17×17	512×256	32.35	34.47
Lego Knights		512×512	25.12	27.68
Tarot (large)		512×512	17.01	18.29
Tarot (small)		512×512	21.21	24.12
Lego Truck		640×480	30.12	35.34
Average			25.65	28.13

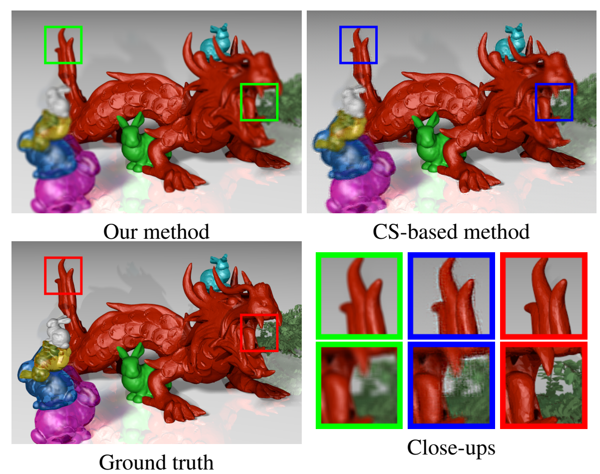

Fig. 6. Reconstructed and ground truth images.

4. CONCLUSION

We investigated efficient light field acquisition using a coded aperture camera from the perspective of PCA and proposed how to determine the optimal non-negative aperture patterns and how to reconstruct the light field from the acquired images. We demonstrated that, with N acquisitions, the proposed method can reach the upper-bound quality that is achievable with the first N PC component. Furthermore, we demonstrated that our method can achieve better reconstruction quality in a much smaller computational time compared to the state-of-the-art CS-based method [20]. In future work, we will implement our method on an actual coded-aperture camera to further validate the effectiveness.

5. REFERENCES

[1] Marc Levoy and Pat Hanrahan, "Light field rendering," in *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*. ACM, 1996, pp. 31–42.

[2] Lixin Shi, Haitham Hassanieh, Abe Davis, Dina Katabi, and Fredo Durand, "Light field reconstruction using sparsity in the continuous fourier domain," *ACM Transactions on Graphics (TOG)*, vol. 34, no. 1, pp. 12, 2014.

[3] Sven Wanner and Bastian Goldluecke, "Variational light field analysis for disparity estimation and super-resolution," *IEEE transactions on pattern analysis and machine intelligence*, vol. 36, no. 3, pp. 606–619, 2014.

[4] Ting-Chun Wang, Alexei A Efros, and Ravi Ramamoorthi, "Depth estimation with occlusion modeling using light-field cameras," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 11, pp. 2170–2181, 2016.

[5] Ren Ng, Marc Levoy, Mathieu Brédif, Gene Duval, Mark Horowitz, and Pat Hanrahan, "Light field photography with a hand-held plenoptic camera," *Computer Science Technical Report CSTR*, vol. 2, no. 11, pp. 1–11, 2005.

[6] Tom E Bishop, Sara Zanetti, and Paolo Favaro, "Light field superresolution," in *Computational Photography (ICCP), 2009 IEEE International Conference on*. IEEE, 2009, pp. 1–9.

[7] Gordon Wetzstein, Douglas R Lanman, Matthew Waggener Hirsch, and Ramesh Raskar, "Tensor displays: compressive light field synthesis using multilayer displays with directional backlighting," 2012.

[8] Fu-Chung Huang, Kevin Chen, and Gordon Wetzstein, "The light field stereoscope: immersive computer graphics via factored near-eye light field displays with focus cues," *ACM Transactions on Graphics (TOG)*, vol. 34, no. 4, pp. 60, 2015.

[9] Seungjae Lee, Changwon Jang, Seokil Moon, Jaebum Cho, and Byoungho Lee, "Additive light field displays: realization of augmented reality with holographic optical elements," *ACM Transactions on Graphics (TOG)*, vol. 35, no. 4, pp. 60, 2016.

[10] Toyohiro Saito, Yuto Kobayashi, Keita Takahashi, and Toshiaki Fujii, "Displaying real-world light fields with stacked multiplicative layers: Requirement and data conversion for input multiview images," *Journal of Display Technology*, vol. 12, no. 11, pp. 1290–1300, 2016.

[11] Bennett Wilburn, Neel Joshi, Vaibhav Vaish, Eino-Ville Talvala, Emilio Antunez, Adam Barth, Andrew Adams, Mark Horowitz, and Marc Levoy, "High performance imaging using large camera arrays," in *ACM Transactions on Graphics (TOG)*. ACM, 2005, vol. 24, pp. 765–776.

[12] Toshiaki Fujii, Kensaku Mori, Kazuya Takeda, Kenji Mase, Masayuki Tanimoto, and Yasuhito Suenaga, "Multipoint measuring system for video and sound-100-camera and microphone system," in *2006 IEEE International Conference on Multimedia and Expo*. IEEE, 2006, pp. 437–440.

[13] Edward H Adelson and John YA Wang, "Single lens stereo with a plenoptic camera," *IEEE transactions on pattern analysis and machine intelligence*, vol. 14, no. 2, pp. 99–106, 1992.

[14] Jun Arai, Fumio Okano, Haruo Hoshino, and Ichiro Yuyama, "Gradient-index lens-array method based on real-time integral photography for three-dimensional images," *Applied optics*, vol. 37, no. 11, pp. 2034–2045, 1998.

[15] Ren Ng, *Digital light field photography*, Ph.D. thesis, stanford university, 2006.

[16] Ashok Veeraraghavan, Ramesh Raskar, Amit Agrawal, Ankit Mohan, and Jack Tumblin, "Dappled photography: Mask enhanced cameras for heterodyned light fields and coded aperture refocusing," *ACM Trans. Graph.*, vol. 26, no. 3, pp. 69, 2007.

[17] Chia-Kai Liang, Tai-Hsu Lin, Bing-Yi Wong, Chi Liu, and Homer H Chen, "Programmable aperture photography: multiplexed light field acquisition," in *ACM Transactions on Graphics (TOG)*. ACM, 2008, vol. 27, p. 55.

[18] Hajime Nagahara, Changyin Zhou, Takuya Watanabe, Hiroshi Ishiguro, and Shree K Nayar, "Programmable aperture camera using lcos," in *European Conference on Computer Vision*. Springer, 2010, pp. 337–350.

[19] S Derin Babacan, Reto Ansorge, Martin Luessi, Pablo Ruiz Mataran, Rafael Molina, and Aggelos K Katsaggelos, "Compressive light field sensing," *IEEE Transactions on image processing*, vol. 21, no. 12, pp. 4746–4757, 2012.

[20] Kshitij Marwah, Gordon Wetzstein, Yosuke Bando, and Ramesh Raskar, "Compressive light field photography using overcomplete dictionaries and optimized projections," *ACM Transactions on Graphics (TOG)*, vol. 32, no. 4, pp. 46, 2013.

[21] Salil Tambe, Ashok Veeraraghavan, and Amit Agrawal, "Towards motion aware light field video for dynamic scenes," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 1009–1016.

[22] David L Donoho, "Compressed sensing," *IEEE Transactions on information theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[23] Emmanuel J Candès and Michael B Wakin, "An introduction to compressive sampling," *IEEE signal processing magazine*, vol. 25, no. 2, pp. 21–30, 2008.

[24] Emmanuel J Candes, Yonina C Eldar, Deanna Needell, and Paige Randall, "Compressed sensing with coherent and redundant dictionaries," *Applied and Computational Harmonic Analysis*, vol. 31, no. 1, pp. 59–73, 2011.

[25] Mahdad Hosseini Kamal, Mohammad Golbabaee, and Pierre Vandergheynst, "Light field compressive sensing in camera arrays," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Ieee, 2012, pp. 5413–5416.

[26] Yusuke Miyagi, Keita Takahashi, and Toshiaki Fujii, "Weighted 4d-dct basis for compressively sampled light fields," *IEICE TRANSACTIONS on Fundamentals of Electronics, Communications and Computer Sciences*, vol. 99, no. 9, pp. 1655–1664, 2016.

[27] Ian Jolliffe, *Principal component analysis*, Wiley Online Library, 2002.

[28] Amit Ashok and Mark A Neifeld, "Compressive light field imaging," in *SPIE Defense, Security, and Sensing*. International Society for Optics and Photonics, 2010.

[29] Mark A Neifeld and Premchandra Shankar, "Feature-specific imaging," *Applied optics*, vol. 42, no. 17, pp. 3379–3389, 2003.

[30] "Compressive light field camera," http://cameraculture.media.mit.edu/projects/compressive-light-field-camera/.

[31] "The (new) stanford light field archive," http://lightfield.stanford.edu.