IEICE TRANS. INF. & SYST., VOL.E101–D, NO.9 SEPTEMBER 2018

PAPER Special Section on Picture Coding and Image Media Processing

Designing Coded Aperture Camera Based on PCA and NMF for Light Field Acquisition

Yusuke YAGI[†], ***Nonmember***, **Keita TAKAHASHI**[†a)], **Toshiaki FUJII**[†], ***Members***, **Toshiki SONODA**[††], ***Nonmember***, ***and*** **Hajime NAGAHARA**[†††], ***Member***

SUMMARY 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), where sparse representations on a pre-trained dictionary or basis are explored to reconstruct the light field. In contrast, we formulated this problem from the perspective of principal component analysis (PCA) and non-negative matrix factorization (NMF), where only a small number of basis vectors are selected in advance based on the analysis of the training dataset. From this formulation, we derived optimal non-negative aperture patterns and a straight-forward reconstruction algorithm. Even though our method is based on conventional techniques, it has proven to be more accurate and much faster than a state-of-the-art CS-based method.

key words: *light field, coded aperture, principal component analysis, non-negative matrix factorization*

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 [3], [6], and 3D displays [7]–[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 [5], [13]–[15] and coded aperture/mask cameras [16]–[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]–[24], which provides a sophisticated framework for signal reconstruction from a limited number of samples, where sparse representations on a pre-trained dictionary or basis are explored to reconstruct the target signal. In contrast, we formulate this problem from the perspective of principal component analysis (PCA) and non-negative matrix factorization (NMF), where only a small number of basis vectors are selected in advance based on the analysis of the training dataset. From this formulation, we derived 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].

This paper is an extension of our conference presentation [25], where the aperture pattern and reconstruction algorithm were derived via PCA. In the present paper, we established a general framework where both PCA and NMF can be used to derive optimal non-negative aperture patterns and the corresponding reconstruction algorithms. Correspondingly, new experimental results that were not included in [25] have been added to the present paper.

The remainder of this paper is organized as follows. Section 2 describes the background of this paper including light fields, coded aperture cameras and compressed sensing. In Sect. 3, we present first a general framework in which

Table 1 List of variables

Name	Variable	Size
Number of viewpoints	M	scalar
Number of acquisitions ($N < M$)	N	scalar
Number of pixels in a pixel block	P	scalar
Angular light field for each pixel	$\boldsymbol{x}$	$1 \times M$
Spatial-angular light field block	$\dot{\boldsymbol{x}}$	$1 \times MP$
Observation vector w.r.t. $\boldsymbol{x}$	$\boldsymbol{y}$	$1 \times N$
Observation vector w.r.t. $\dot{\boldsymbol{x}}$	$\dot{\boldsymbol{y}}$	$1 \times NP$
Observation matrix w.r.t. $\boldsymbol{x}$	$\boldsymbol{A}$	$N \times M$
Observation matrix w.r.t. $\dot{\boldsymbol{x}}$	$\dot{\boldsymbol{A}}$	$NP \times MP$
Number of atoms	Q	scalar
Number of training data	K	scalar
Dictionary for representing $\dot{\boldsymbol{x}}$	$\boldsymbol{D}$	$MP \times Q$
Coefficient vector w.r.t. $\dot{\boldsymbol{x}}$	$\dot{\boldsymbol{\theta}}$	$1 \times Q$
Basis for representing $\boldsymbol{x}$	$\boldsymbol{B}$	$M \times M$
Reduced basis for representing $\boldsymbol{x}$	$\tilde{\boldsymbol{B}}$	$M \times N$
Coefficient vector w.r.t. $\boldsymbol{x}$	$\boldsymbol{\theta}$	$1 \times M$
Reduced coefficient vector w.r.t. $\boldsymbol{x}$	$\tilde{\boldsymbol{\theta}}$	$1 \times N$

Manuscript received December 15, 2017.
Manuscript revised March 7, 2018.
Manuscript publicized June 20, 2018.

[†]The authors are with the Graduate School of Engineering, Nagoya University, Nagoya-shi, 464–8603 Japan.
[††]The author is with the Graduate School of Information Science and Electrical Engineering, Kyushu University, Fukuoka-shi, 819–0395 Japan.
[†††]The author is with the Institute for Datability Science, Osaka University, Osaka-shi, 565–0871 Japan.
a) E-mail: keita.takahashi@nagoya-u.jp

DOI: 10.1587/transinf.2017PCP0007

Copyright © 2018 The Institute of Electronics, Information and Communication Engineers

only a limited number of basis vectors are selected for designing the aperture patterns and reconstruction algorithm, followed by specific methods derived via PCA and NMF. In Sect. 4, experimental results are presented to validate our method. Section 5 concludes the paper. For convenience, we present a table of variables used in this paper in Table 1.

2. Background

2.1 Light Field and Coded Aperture Camera

A mathematical formulation for light field acquisition using a coded-aperture camera is presented using the illustration in Fig. 1, which is similar to the cameras in previous work [16]–[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 $(n = 1, \ldots, N)$. 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 parameterized with (s, t) is quantized into a finite number of blocks, they are numbered by an integer m $(m = 1, \ldots, M)$. In this case, Eq. (1) is rewritten as

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

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 $l(s, t, u, v)$ with a fixed (s, t)) 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 $y_n(u, v)$.

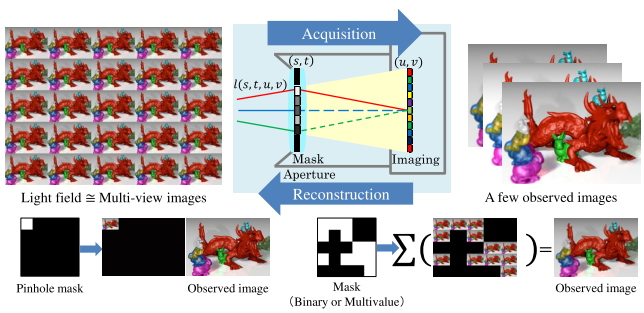


Fig. 1 Coded aperture camera model. Each light ray passing through (s, t) on the aperture plane is attenuated by the transmittance at (s, t) and reaches a pixel (u, v) on the imaging plane.

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 $y_n(u, v)$ and $\boldsymbol{A} \in \mathbb{R}^{N \times M}$ is called an observation matrix whose (n, m) element is given by $a_{n,m}$. Symbol $\boldsymbol{x}_{u,v} \in \mathbb{R}^M$ is a column vector representing the *angular image* of a pixel (u, v), whose m-th element is $x_m(u, v)$.

To simplify the notation, we drop the subscript (u, v) in the remainder of this paper and rewrite the observation model as

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}. \tag{4}$$

The goal here is to obtain $\boldsymbol{x} \in \mathbb{R}^M$ from the given $\boldsymbol{A} \in \mathbb{R}^{N \times M}$ and $\boldsymbol{y} \in \mathbb{R}^N$. 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. (4) an under-determined system.

2.2 Compressed Sensing Based Approach

The aforementioned problem is often considered in the context of compressive sensing (CS) [22]–[24]. A general framework utilized in several studies [19]–[21], [26], [27] 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 $\dot{\boldsymbol{x}} \in \mathbb{R}^{MP}$ is defined to contain M viewpoints on the aperture plane and a pixel block with P pixels on the imaging plane, where the elements are aligned in pixel-major order. When $P = 1$, each of the angular images is used as a processing unit.

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

$$\dot{\boldsymbol{y}} = \dot{\boldsymbol{A}}\dot{\boldsymbol{x}} \tag{5}$$

where $\dot{\boldsymbol{y}} \in \mathbb{R}^{NP}$ contains the observations on the P pixels inside the block with N times acquisitions in the pixel-major order. The observation matrix $\dot{\boldsymbol{A}} \in \mathbb{R}^{NP \times MP}$ is given as $\boldsymbol{A} \otimes \boldsymbol{I}$, where $\boldsymbol{I} \in \mathbb{R}^{P \times P}$ 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

$$\dot{\boldsymbol{x}} \simeq \boldsymbol{D}\dot{\boldsymbol{\theta}}, \tag{6}$$

where $\boldsymbol{D} = [\boldsymbol{d}_1 \dots \boldsymbol{d}_Q] \in \mathbb{R}^{MP \times Q}$ is a basis or a dictionary having Q elements, and $\dot{\boldsymbol{\theta}} \in \mathbb{R}^Q$ is a sparse coefficient vector where only a few elements can take non-zero values. The dictionary is often made over-complete; Q is set greater than MP. The estimation of $\dot{\boldsymbol{\theta}}$ is formulated as a minimization problem with a non-negative weight λ:

$$\dot{\boldsymbol{\theta}}_{opt} = \arg\min_{\dot{\boldsymbol{\theta}}} \frac{1}{2}\|\dot{\boldsymbol{y}} - \dot{\boldsymbol{A}}\boldsymbol{D}\dot{\boldsymbol{\theta}}\|_2^2 + \lambda\|\dot{\boldsymbol{\theta}}\|_1^1, \tag{7}$$

from which the reconstructed signal is obtained as

$$\dot{\boldsymbol{x}}_{rec} = \boldsymbol{D}\dot{\boldsymbol{\theta}}_{opt}. \tag{8}$$

Solving this minimization problem often takes significant time due to the complexity.

An important question here is how the aperture pattern $\boldsymbol{A}$ can be determined to achieve the best reconstruction quality. It is thought that using a random pattern is not the best but sufficient. Another possible answer can be found in the work done by Marwah et al. [20]. Matrix $\boldsymbol{G}_{\boldsymbol{A}} \in \mathbb{R}^{NP \times Q}$ is defined as $\boldsymbol{G}_{\boldsymbol{A}} = \dot{\boldsymbol{A}}\boldsymbol{D}$, and each column of $\boldsymbol{G}_{\boldsymbol{A}}$ is normalized to the unit length to yield $\bar{\boldsymbol{G}}_{\boldsymbol{A}}$. The optimal $\boldsymbol{A}$ is obtained as

$$\boldsymbol{A} = \arg\min_{\boldsymbol{A}} \|\boldsymbol{I} - \bar{\boldsymbol{G}}_{\boldsymbol{A}}^T \bar{\boldsymbol{G}}_{\boldsymbol{A}}\|_F. \tag{9}$$

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

3. Proposed Method

3.1 General Framework

We first mention the signal model used in our proposal. Using M linearly independent vectors, $\boldsymbol{b}_1, \dots, \boldsymbol{b}_M$, as the basis, any instance of $\boldsymbol{x}$ can be represented as

$$\boldsymbol{x} = \boldsymbol{B}\boldsymbol{\theta} \tag{10}$$

where $\boldsymbol{B} = [\boldsymbol{b}_1 \dots \boldsymbol{b}_M] \in \mathbb{R}^{M \times M}$. The key idea of our method is to reduce the basis vectors to the most important ones. More specifically, we use only N vectors when the number of observations is N. In this case, $\boldsymbol{x}$ can be approximated as

$$\boldsymbol{x} = \tilde{\boldsymbol{B}}\tilde{\boldsymbol{\theta}} + \boldsymbol{\varepsilon} \tag{11}$$

where $\tilde{\boldsymbol{B}} = [\boldsymbol{b}_1 \dots \boldsymbol{b}_N] \in \mathbb{R}^{M \times N}$ is called a *reduced basis* that is composed of the most important N basis vectors, and $\tilde{\boldsymbol{\theta}} \in \mathbb{R}^N$ is the corresponding coefficient vector. It is expected that with a sufficiently small N we can achieve $\boldsymbol{\varepsilon} \simeq \boldsymbol{0}$ because $\boldsymbol{x}$ is not a random signal, but an angular image that has some degree of redundancy inside. To chose N important vectors, we use principal component analysis (PCA) and non-negative matrix factorization (NMF), which will be detailed in Sect. 3.2 and 3.3.

It should be noted that Eq. (11) seems to be similar to Eq. (6). Our method and CS-based approach shares the same assumption that a target signal would be represented by combining only a small number of *basic components* (atoms, basis vectors). However, they are absolutely different in how to *select* the basic components for the target signal. In the CS-based approach, the number of basic components Q is much larger than the number of observations N, and determining which of $[\boldsymbol{d}_1 \dots \boldsymbol{d}_Q]$ are used is a part of the reconstruction problem. Meanwhile, in our method, the number of basic components is reduced to N in advance; used/unused elements are determined at the point where $\boldsymbol{B}$ is reduced to $\tilde{\boldsymbol{B}}$. 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.

We then describe how to apply our signal model to the configuration of a coded aperture camera with an aperture pattern $\boldsymbol{A} \in \mathbb{R}^{M \times N}$. By substituting Eq. (11) into Eq. (4), we obtain

$$\boldsymbol{y} = \boldsymbol{A}\tilde{\boldsymbol{B}}\tilde{\boldsymbol{\theta}} + \boldsymbol{A}\boldsymbol{\varepsilon}. \tag{12}$$

Here, we consider two conditions for the right hand side:

1. in the first term, $\boldsymbol{A}\tilde{\boldsymbol{B}}$ ($\in \mathbb{R}^{N \times N}$) is invertible
2. the second term $\boldsymbol{A}\boldsymbol{\varepsilon}$ ($\in \mathbb{R}^N$) is zero.

If the aperture pattern $\boldsymbol{A}$ is designed to meet the two conditions, we can straightforwardly obtain the coefficient vector $\tilde{\boldsymbol{\theta}}$ for the target signal from the observation $\boldsymbol{y}$ and reconstruct the signal in accordance with Eq. (11).

Our proposal is to set the aperture pattern $\boldsymbol{A}$ to

$$\boldsymbol{A} = \boldsymbol{C}\tilde{\boldsymbol{B}}^T \tag{13}$$

where $\boldsymbol{C} \in \mathbb{R}^{N \times N}$ is an invertible matrix, which is used to make the aperture patterns physically feasible as will be detailed in Sect. 3.2 and 3.3. Importantly, this aperture pattern satisfies the two conditions as mentioned below. Therefore, the observation model of Eq. (12) is rewritten as

$$\boldsymbol{y} = \boldsymbol{C}\tilde{\boldsymbol{B}}^T\tilde{\boldsymbol{B}}\tilde{\boldsymbol{\theta}} \tag{14}$$

from which we can immediately obtain

$$\tilde{\boldsymbol{\theta}} = (\tilde{\boldsymbol{B}}^T\tilde{\boldsymbol{B}})^{-1}\boldsymbol{C}^{-1}\boldsymbol{y} \tag{15}$$

and reconstruct the signal in accordance with Eq. (11) as

$$\boldsymbol{x}_{rec} = \tilde{\boldsymbol{B}}(\tilde{\boldsymbol{B}}^T\tilde{\boldsymbol{B}})^{-1}\boldsymbol{C}^{-1}\boldsymbol{y} \tag{16}$$

Once $\tilde{\boldsymbol{B}}$ and $\boldsymbol{C}$ have been determined in the off-line process, this reconstruction is extremely simple and computationally easy compared with that of the CS-based approach (compare Eqs. (15) and (7)).

Finally, we check that the two conditions mentioned above are satisfied with the aperture pattern given by Eq. (13).

1. The first condition is satisfied if $\tilde{\boldsymbol{B}}^T\tilde{\boldsymbol{B}} \in \mathbb{R}^{N \times N}$ is invertible. This can be confirmed as follows. It was assumed that all the columns in $\tilde{\boldsymbol{B}}$ are linearly independent with each other. Equivalently, $\tilde{\boldsymbol{B}}\boldsymbol{v} \neq \boldsymbol{0}$ with $\boldsymbol{v} \notin \boldsymbol{0}$. Accordingly, we obtain $|\tilde{\boldsymbol{B}}\boldsymbol{v}|^2 = \boldsymbol{v}^T\tilde{\boldsymbol{B}}^T\tilde{\boldsymbol{B}}\boldsymbol{v} \neq \boldsymbol{0}$. Therefore, we derive $\tilde{\boldsymbol{B}}^T\tilde{\boldsymbol{B}}\boldsymbol{v} \neq \boldsymbol{0}$ with $\boldsymbol{v} \notin \boldsymbol{0}$, which is equivalent to the proposition that $\tilde{\boldsymbol{B}}^T\tilde{\boldsymbol{B}}$ is invertible.

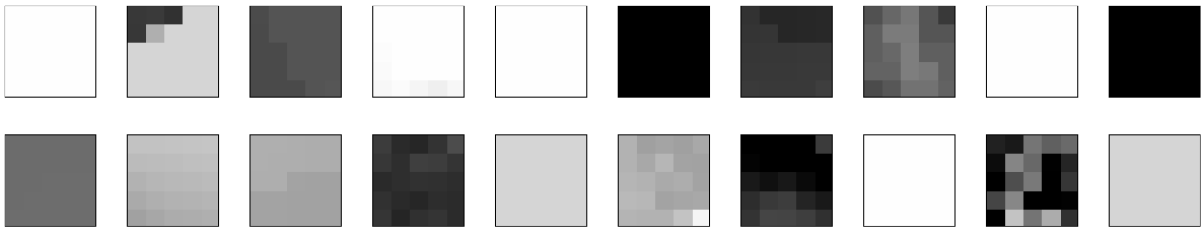
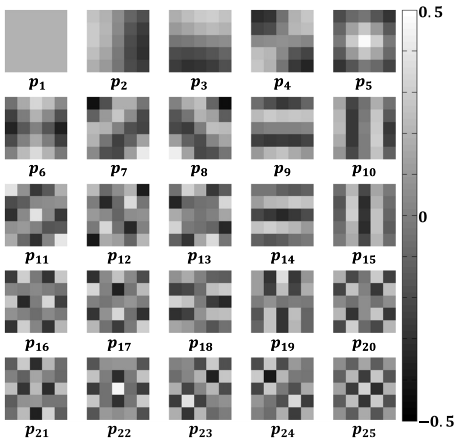


Fig. 2 Samples of angular images.



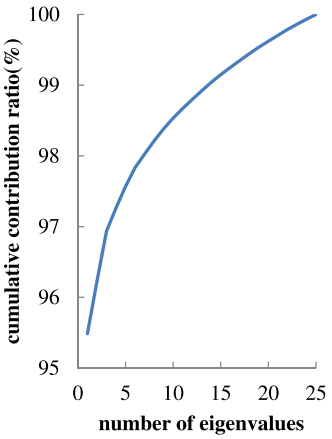


Fig. 3 Basis vectors derived via PCA and their cumulative contribution ratio.

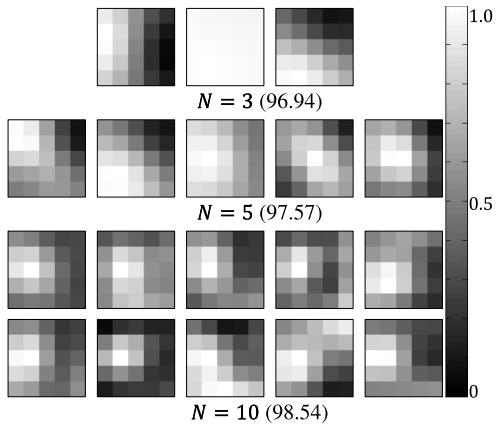


Fig. 4 Aperture patterns derived via PCA.

2. The second condition is satisfied if $\tilde{\boldsymbol{B}}^T\boldsymbol{\varepsilon}$ is zero. This can be confirmed as follows. In Eq. (11), the coefficient vector $\tilde{\boldsymbol{\theta}}$ is determined to meet

$$\tilde{\boldsymbol{\theta}} = \arg\min_{\tilde{\boldsymbol{\theta}}} E(\tilde{\boldsymbol{\theta}}),\ E(\tilde{\boldsymbol{\theta}}) = \|\boldsymbol{x} - \tilde{\boldsymbol{B}}\tilde{\boldsymbol{\theta}}\|_2^2. \tag{17}$$

The optimal $\tilde{\boldsymbol{\theta}}$ should satisfy the extrema condition:

$$\frac{\partial E(\tilde{\boldsymbol{\theta}})}{\partial \tilde{\boldsymbol{\theta}}} = -\tilde{\boldsymbol{B}}^T(\boldsymbol{x} - \tilde{\boldsymbol{B}}\boldsymbol{\theta}) = -\tilde{\boldsymbol{B}}^T\boldsymbol{\varepsilon} = 0 \tag{18}$$

3.2 Basis Derived from PCA

A basis with an ordered significance is derived from principal component analysis (PCA). We first prepare a training dataset that is composed of K angular images: $\boldsymbol{X} = [\boldsymbol{x}_1 \dots \boldsymbol{x}_K] \in \mathbb{R}^{M\times K}$[†]. As will be detailed in Sect. 4, the dataset consists of 5,977,440 angular images taken from 4 light fields with 25 (5×5) viewpoints, which means $M = 25$ and $K = 5,977,440$. Several samples of $\boldsymbol{x_k}$ are shown in Fig. 2. We then obtain a covariance matrix as

$$\boldsymbol{S} = E[\boldsymbol{x}\boldsymbol{x}^T] = \frac{1}{K}\boldsymbol{X}\boldsymbol{X}^T, \tag{19}$$

which has M non-negative eigen values ρ_m and corresponding eigen vectors $\boldsymbol{b}_m$ satisfying

$$\boldsymbol{S}\boldsymbol{b}_m = \rho_m\boldsymbol{b}_m. \tag{20}$$

Without loss of generality, we assume that the eigenvectors are normalized and sorted in descending order of the eigenvalues. The eigenvectors and their cumulative contribution ratio obtained from a dataset are visualized in Fig. 3.

We take the first N eigenvectors and use them as the reduced basis.

$$\tilde{\boldsymbol{B}}_{PCA} = [\boldsymbol{b}_1 \dots \boldsymbol{b}_N] \in \mathbb{R}^{M\times N}. \tag{21}$$

This basis gives the optimal approximation of the target signal $\boldsymbol{x}$ in the sense of the least squared error as far as only N components can be used. The observation model of Eq. (12) is simplified into

$$\boldsymbol{y} = \boldsymbol{C}\tilde{\boldsymbol{\theta}} \tag{22}$$

because each of the basis vectors $\boldsymbol{b}_m$ are orthogonal with each other.

The matrix $\boldsymbol{C} \in \mathbb{R}^{N\times N}$ is designed to make physically feasible aperture patterns. In accordance with Eq. (13), the aperture pattern is given as $\boldsymbol{A} = \boldsymbol{C}\tilde{\boldsymbol{B}}_{PCA}^T \in \mathbb{R}^{M\times N}$. All the elements of $\boldsymbol{A}$ should be limited within $[0, 1]$, because they correspond to the transmittance of the aperture. Meanwhile, the basis vectors constituting $\tilde{\boldsymbol{B}}_{PCA}$ may take negative elements. As can be seen from Fig. 3, all the basis vectors except for $\boldsymbol{b}_1$ (which corresponds to the DC component) have negative elements. By mixing the basis vectors using a carefully chosen matrix $\boldsymbol{C}$, we aim to make all the elements of $\boldsymbol{A}$

[†]Generally, PCA is applied to a zero-mean dataset; otherwise, the mean vector is subtracted from the dataset to make it zero-mean. In our case, the light field signal is inherently non-negative and never zero-mean. However, mean subtraction is not compatible with the physical imaging process of a coded aperture camera. Therefore, we used PCA without mean subtraction. This results in a large contribution ratio of the first PCA component because it includes the mean vector over the entire training dataset.

non-negative. Moreover, we should consider the stability of the solution $\tilde{\boldsymbol{\theta}}$ obtained from Eq. (22); matrix $\boldsymbol{C}$ should not merely be invertible, but the solution $\tilde{\boldsymbol{\theta}}$ should be robust to noise on the measurement vector $\boldsymbol{y}$.

Taking the above issues into account, we determine the mixing matrix $\boldsymbol{C}$ (and in return, the observation matrix $\boldsymbol{A}$) as follows.

1. We randomly generate the elements of $\boldsymbol{C}^{init}$ and retain the matrix as a candidate if all the elements of $\boldsymbol{A}^{init} = \boldsymbol{C}^{init}\tilde{\boldsymbol{B}}_{PCA}^{T}$ are non-negative. We generate a sufficient number of candidates.
2. We normalize each of the candidates of $\boldsymbol{C}^{init}$. In the resulting observation matrix, all the elements should be limited within $[0, 1]$. Therefore, we take the maximum of each row in $\boldsymbol{A}^{init}$, and denote it as $a_n^{\max}$ $(n = 1, \ldots, N)$. The modified matrix $\boldsymbol{C}$ is given as $\boldsymbol{C} = \mathrm{diag}[1/a_1^{\max}, \ldots, 1/a_N^{\max}]\boldsymbol{C}^{init}$.
3. Choose one from the candidates of $\boldsymbol{C}$ that minimizes the condition number, which is an index of stability of matrix inversion.

The aperture patterns derived via PCA for $N = 3, 5, 10$ are shown in Fig. 4, each of which was generated with the best $\boldsymbol{C}$ out of 10,000 random samples. We experimentally confirmed in Appendix A that unless a significantly bad $\boldsymbol{C}$ is chosen, the resulting reconstruction quality is fine.

To handle the negative elements of the basis vectors, Ashok et al. [28] took another approach called a dual rail measurement [29], where each basis vector $\boldsymbol{b}_m$ is divided into positive and negative parts, $\boldsymbol{b}_m^+ > \boldsymbol{0}$ and $\boldsymbol{b}_m^- > \boldsymbol{0}$, satisfying $\boldsymbol{b}_m = \boldsymbol{b}_m^+ - \boldsymbol{b}_m^-$. Two measurements using $\boldsymbol{b}_m^+$ and $\boldsymbol{b}_m^-$ were conducted, and the results were combined to obtain a measurement that would be virtually conducted with $\boldsymbol{b}_m$. An obvious disadvantage of this method is the increase in the number of acquisitions; if we use N basis vectors, the number of acquisitions is $2N-1$ because all basis vectors except for $\boldsymbol{b}_1$ require two measurements for each. In Appendix A, we demonstrate that our method can achieve equivalent reconstruction quality to the dual rail measurement while keeping the number of acquisitions N.

3.3 Basis Derived from NMF

Another approach to select a set of basis vectors are derived from non-negative matrix factorization (NMF). Similar to the case of PCA, we first prepare a training dataset $\boldsymbol{X} \in \mathbb{R}^{M\times K}$. We then try to find a representation given as

$$\boldsymbol{X} = \tilde{\boldsymbol{B}}_{NMF}\boldsymbol{\Theta} + \boldsymbol{\varepsilon} \tag{23}$$

where $\tilde{\boldsymbol{B}}_{NMF} \in \mathbb{R}^{M\times N}$ and $\boldsymbol{\Theta} \in \mathbb{R}^{N\times K}$ are non-negative matrices, where all the elements are greater than or equal to zero. This representation can be obtained using an iterative method based on multiplicative update rule [30].

Our proposal here is to take $\tilde{\boldsymbol{B}}_{NMF}$ as the reduced basis for the angular images, which is written as

$$\tilde{\boldsymbol{B}}_{NMF} = [\boldsymbol{b}_1 \ldots \boldsymbol{b}_N]. \tag{24}$$

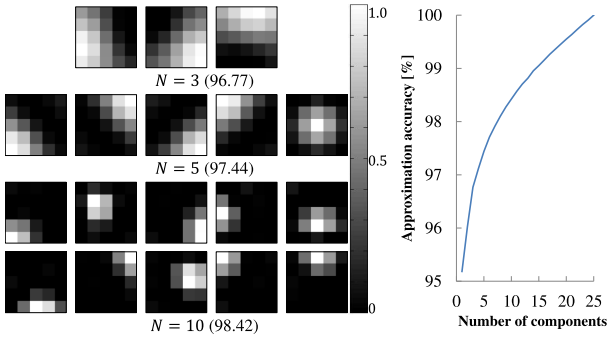


Fig. 5 Aperture patterns derived via NMF.

Equation (23) is considered as a rank N approximation of $\boldsymbol{X}$, where each column of $\tilde{\boldsymbol{B}}_{NMF}$ is a basis vector, and each column of $\boldsymbol{\Theta}$ corresponds to the coefficients for each of the angular images in the training dataset $\boldsymbol{X}$. By making the residual term $\boldsymbol{\varepsilon}$ close to zero, it is expected that best basis vectors for representing the training dataset are selected under the constraint that the number of basis vectors is limited to N. It is also highly expected that with a sufficient small N all the selected vectors $\boldsymbol{b}_n$ are linearly independent. However, being different from PCA, the basis vectors $\boldsymbol{b}_n$ $(n = 1, \ldots, N)$ selected via NMF have no order in their significance.

Similar to the case with PCA, matrix $\boldsymbol{C} \in \mathbb{R}^{N\times N}$ is designed so as to make physically feasible aperture patterns. In accordance with Eq. (13), the aperture pattern is given as $\boldsymbol{A} = \boldsymbol{C}\tilde{\boldsymbol{B}}_{NMF}^{T} \in \mathbb{R}^{N\times M}$. This time, it is ensured that all the components of $\tilde{\boldsymbol{B}}_{NMF}^{T}$ are already non-negative. Only the remaining constraint is that all the elements of $\boldsymbol{A}$ should be limited within $[0, 1]$, because they correspond to the transmittance of the aperture. Therefore, we take the maximum of each row in $\tilde{\boldsymbol{B}}_{NMF}^{T} \in \mathbb{R}^{N\times M}$, and denote it as $b_n^{\max}$ $(n = 1, \ldots, N)$. The matrix $\boldsymbol{C}$ is given as

$$\boldsymbol{C} = \mathrm{diag}(1/b_1^{\max}, \ldots, 1/b_N^{\max}). \tag{25}$$

The aperture patterns derived via NMF for $N = 3$, 5, 10 are shown in Fig. 5. Here, we use the same dataset as Section 3.2 that has 25 (5×5) viewpoints. We also present the approximation accuracy (corresponding to the cumulative contribution ratio for PCA) for each N. We can see that even with $N = 3$, we can achieve fine approximation quality for the given dataset.

3.4 Discussions

The basis vectors obtained by PCA (Fig. 3) seem to be similar to those of the 2-D discrete cosine transform (2-D DCT) basis. This is reasonable because DCT corresponds to PCA for the first order autoregressive model, which can describe smooth signals including angular images where neighboring symbols are closely correlated.

One might think that the 2-D DCT basis can be used instead of PCA. This intuition is correct, as verified by the cumulative contribution curve shown in Fig. 6. Here, we

sorted the 2-D DCT basis vectors in descending order with respect to the contribution ratio on the training dataset $\boldsymbol{X}$. It is clearly seen that the cumulative contribution curve of the 2-D DCT basis is very close to that of the PCA basis, which indicates that the 2-D DCT basis can achieve equivalent approximation accuracy to the PCA basis. This point was further confirmed in the experiments in Sect. 4. It should be noted that the 2-D DCT basis vectors (except for the first vector) also take negative values, for which we adopted the same strategy as the one for PCA; we chose matrix $\boldsymbol{C}$ so as to avoid negative transmittance values.

The potential of our method can also be analyzed in the frequency domain. It is expected that with a limited number of basis vectors, high frequency components in the angular images cannot be reconstructed accurately. This is because both PCA and NMF try to minimize the squared error over the entire signal, but high frequency components are insignificant in terms of the spectral power. Figure 7 shows power distributions over the 5×5 individual frequency components, which were obtained from the original dataset $\boldsymbol{X}$ and reconstructed ones using either PCA or NMF. The number of basis vectors was set to 3 and 5. It is clearly seen that some high frequency components of the original signal are lost by approximating the signal only with a small number of selected basis vectors. A deeper analysis on this effect is presented in Appendix B.

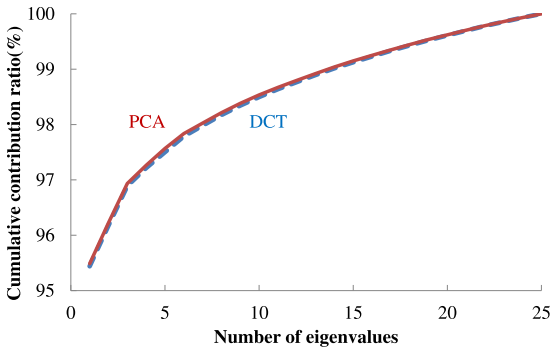


Fig. 6 Cumulative contribution ratio via DCT and PCA.

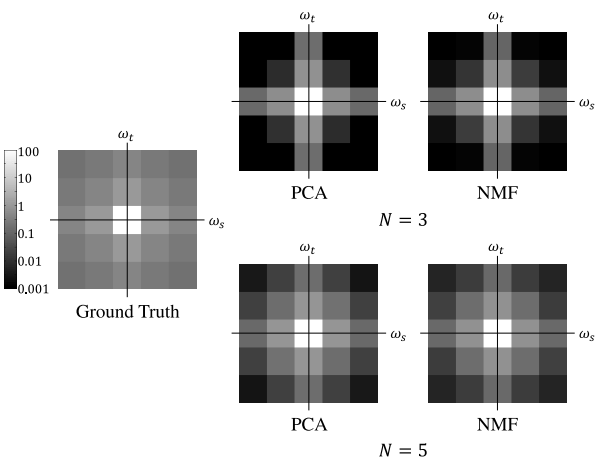


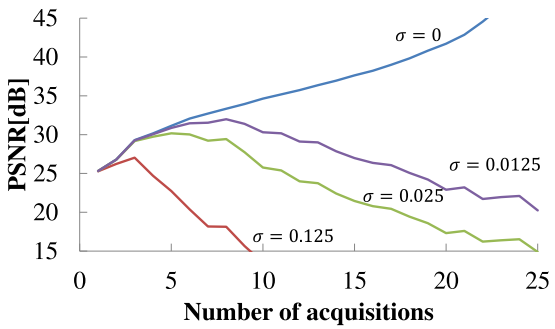
Fig. 7 Power distributions individual frequency components.

4. Experiment

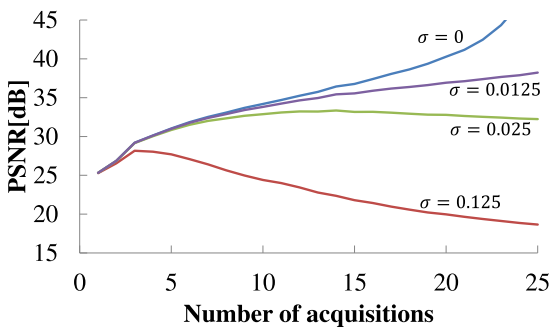
4.1 Simulative Experiment

We conducted simulative experiments using several light field datasets [31], [32]. Throughout the experiments, the pixel values of the original datasets and the transmittances for the aperture patterns were normalized within the range of $[0, 1]$. Therefore, the elements of $\boldsymbol{x}$ and $\boldsymbol{A}$ in Eq. (4) can take values within 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 training dataset $\boldsymbol{X}$ used in our method (both with PCA and NMF) was calculated from four datasets (Dice, Fish, Messerschmitt, and Shrubbery) [31] 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 8 shows the performance of our methods obtained with a different number of acquisitions and noise levels. The dataset used was *Dragon and Bunnies*. When $\sigma = 0$, the reconstruction quality increased as the number of acquisitions increased, which is an expected behavior in accordance with the principle of PCA and NMF. However, when $\sigma > 0$ with PCA, the reconstruction quality initially increased but then decreased beyond a certain point. This

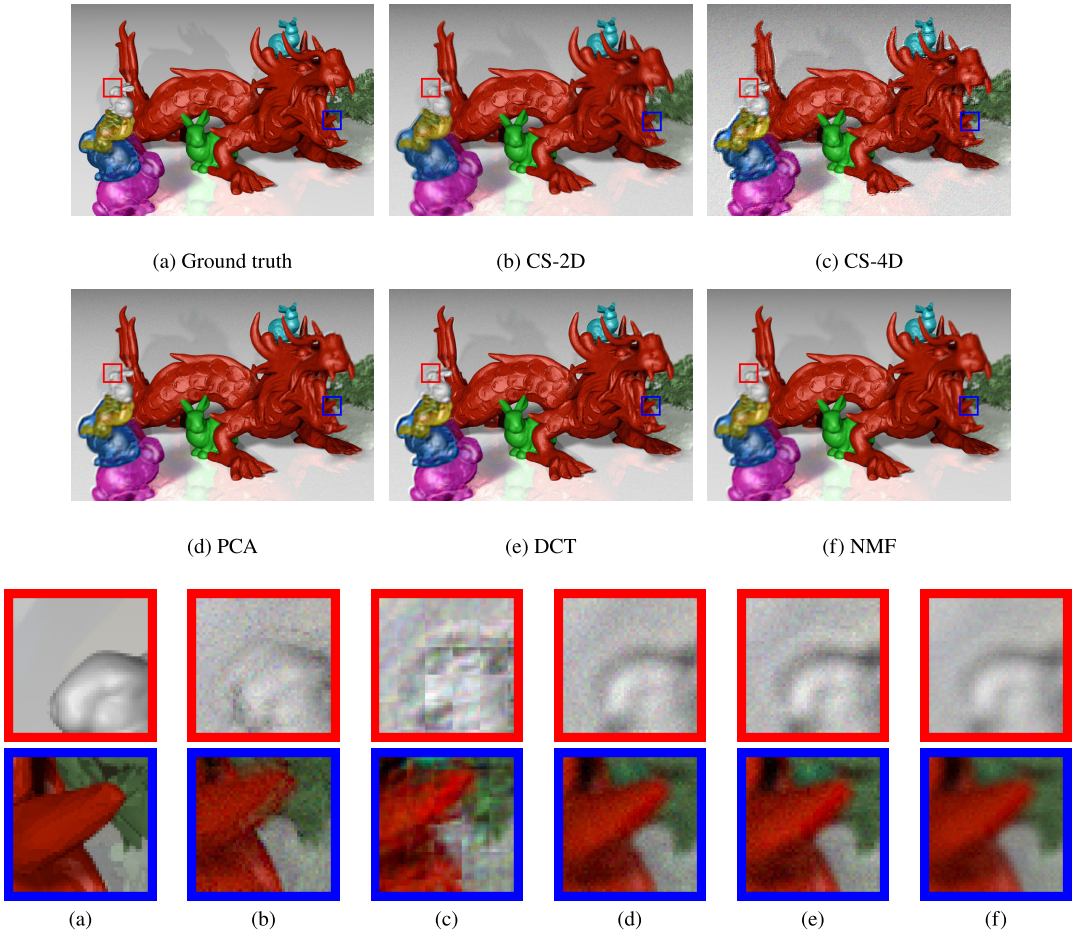


(a) PCA basis



(b) NMF basis

Fig. 8 Reconstruction quality vs. number of acquisitions while changing the noise level σ.



(a) Ground truth (b) CS-2D (c) CS-4D

(d) PCA (e) DCT (f) NMF

(a) (b) (c) (d) (e) (f)

(g) Close-ups.

Fig. 9 Reconstructed and ground truth images at a top-left view ($N = 3$, $\sigma = 0.125$).

Table 2 Light field datasets from [31] and [32], and comparison of the reconstruction quality with $N = 3$. The noise level was set to (left) $\sigma = 0.025$ and (right) $\sigma = 0.125$.

Dataset	Dir. Res. (s,t)	Pos. Res. (u,v)	PSNR [dB] ($\sigma = 0.025$)					PSNR [dB] ($\sigma = 0.125$)				
			CS-2D	CS-4D	PCA	DCT	NMF	CS-2D	CS-4D	PCA	DCT	NMF
DragonsAndBunnies			25.86	27.15	29.19	29.21	29.15	24.46	25.57	27.24	27.46	28.16
dice			24.40	26.17	27.49	27.51	27.45	23.29	24.85	26.02	26.19	26.73
fishi	5×5	840×593	22.36	24.50	25.95	26.00	25.84	21.95	23.93	25.35	25.47	25.53
messerschmitt			25.22	25.72	27.96	27.98	27.93	23.95	24.55	26.37	26.55	27.14
shrubbery			13.36	13.40	14.42	14.39	14.41	13.26	13.26	14.33	14.32	14.37
Amethyst		384×512	31.57	31.78	35.85	35.94	35.95	28.39	28.86	30.87	31.27	32.77
Bracelet		512×320	22.38	22.18	24.58	24.55	24.57	21.65	21.42	23.76	23.83	24.18
Bulldozer		768×576	23.05	24.53	26.34	26.35	26.27	22.33	23.61	25.27	25.40	25.73
Bunny		512×512	31.42	33.79	36.81	37.06	36.87	28.03	29.63	30.76	31.23	32.92
Chess	5×5	700×400	27.74	28.62	31.91	32.02	31.83	25.70	26.55	28.71	29.06	30.10
Chest	out of	768×640	23.50	23.87	25.77	25.78	25.74	22.84	23.15	24.98	25.08	25.34
Flowers	17×17	640×768	29.64	29.39	32.02	32.03	32.08	27.41	27.55	29.27	29.52	30.48
Knights		512×512	23.69	25.27	27.60	27.65	27.51	22.73	24.03	26.11	26.31	26.78
Tarot(large)		512×512	16.66	17.04	18.28	18.26	18.27	16.43	16.74	18.08	18.08	18.17
Tarot(small)		512×512	21.25	21.33	24.09	24.07	24.03	20.66	20.63	23.36	23.43	23.69
Truck		640×480	30.86	30.06	34.94	35.05	35.02	27.52	27.62	29.99	30.41	32.01
Average			24.80	25.62	27.97	28.01	27.95	23.16	23.87	25.65	25.85	26.51

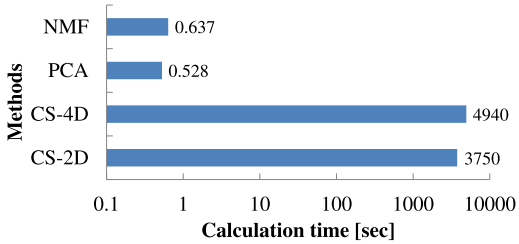
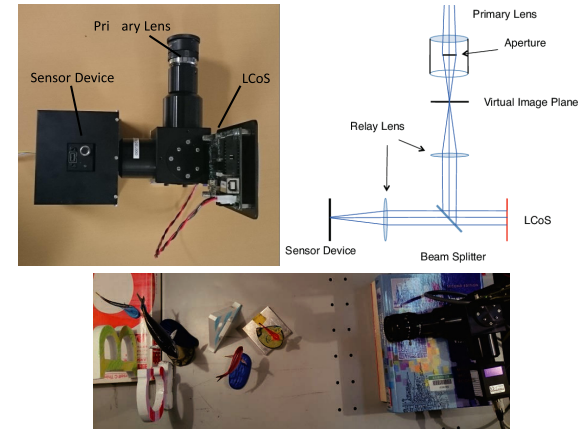


Fig. 10 Calculation time for each method.

phenomenon reflects the fact that less significant basis 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. Meanwhile, in case of NMF, the reconstruction quality was less affected by the noise compared with PCA. It indicates that the NMF-based method is more robust against noise during observations.

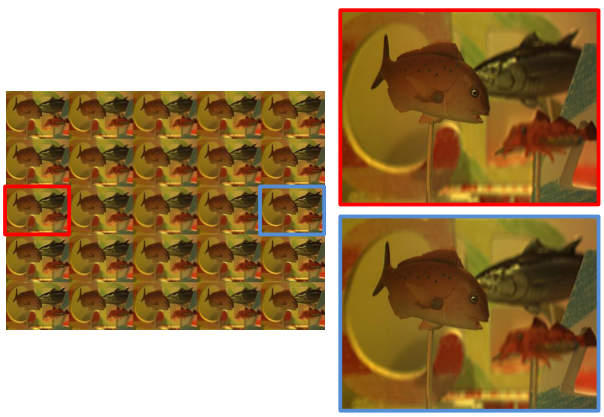
Next, we compared our method with a state-of-the-art CS-based method that uses a learned dictionary [20]. The number of acquisitions N was set to 3. The details of the CS-based method are described as follows. We used the software available from [31]. We used two kinds of dictionaries: one is a 2-dimensional dictionary whose sizes were set to $5\times5\times1\times1$ (5×5 for viewpoints and 1×1 for pixels) which was trained by ourselves, the other is a 4-dimensional dictionary whose sizes were set to $5\times5\times9\times9$ which was originally distributed with the software. In the reconstruction process, each light field dataset was divided into non-overlapping processing units. After several tests, we found that the default value 0.01 for λ in Eq. (7) 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 Eq. (9). Table 2 summarizes the specifications of 16 datasets and the reconstruction quality obtained using the CS-based and our methods (PCA and NMF). In addition, the 2-D DCT basis was also incorporated into our method; the DCT basis was multiplied by the chosen matrix $\boldsymbol{C}$ to make a feasible set of aperture patterns similarly to the case with PCA. The average PSNR values are also presented at the bottom. Clearly, our methods outperformed the CS-based methods with considerable margins for all datasets. Moreover, when the noise is large, our method with NMF exhibited better performance than those with PCA and DCT. Interestingly, the DCT basis performed slightly better than the PCA basis, which indicates that the PCA basis might overfit the training data. Figures 9 shows the reconstructed and ground truth images for the top-left viewpoint of *Dragon and Bunnies*, where we can see that better visual quality was achieved using our methods.

Moreover, in terms of the computational complexity, our methods are much simpler than the CS-based methods. Figure 10 shows calculation time of each methods with $N = 3$ (Dragon and Bunnies). Once the aperture patterns were determined, our methods were able to reconstruct an entire light field in only approximately 0.5 – 0.7 seconds with our unoptimized MATLAB implementation. Meanwhile, the CS-based methods took approximately 3700 – 5000 seconds for the same reconstruction task on the same computer.



(a) Coded aperture camera and experimental setup.

(b) Observed images and corresponding aperture patterns



(c) Reconstructed images

Fig. 11 Experiments with a real coded-aperture camera

4.2 Experiment with Real Camera

We applied our methods to a programmable aperture camera [18], where the aperture was implemented using a LCOS display device. The appearance of the camera and experimental setup are shown in Fig. 11(a). We set the number of acquisition N to 3, and used the aperture pattern derived via

NMF, which is more robust to noise than those derived via PCA and DCT. The three captured images and corresponding aperture patterns are shown in Fig. 11(b). The reconstructed light field (5×5 multi-view images) is shown in Fig. 11(c), where we can observe both horizontal and vertical disparities. To see natural transitions among the viewpoints, refer to the supplemental video on our website [33].

5. Conclusion

We investigated efficient light field acquisition using a coded aperture camera. We formulated the problems of aperture pattern design and signal reconstruction from the perspective of signal representation using a only small number of basis vectors that are selected in advance. Specifically, we selected basis vectors using PCA and NMF, with which we derived optimal non-negative aperture patterns and straightforward reconstruction algorithm. Experimental results validated the effectiveness of our proposal; our method is superior to the state-of-the-art CS-based method [25] in speed and accuracy of light field reconstruction.

Acknowledgments

This work was supported by JSPS kakenhi under Grant 15H05314.

References

[1] M. Levoy and P. Hanrahan, "Light field rendering," Proc. 23rd annual conference on Computer graphics and interactive techniques, pp.31–42, ACM, 1996.
[2] L. Shi, H. Hassanieh, A. Davis, D. Katabi, and F. Durand, "Light field reconstruction using sparsity in the continuous fourier domain," ACM Trans. Graphics (TOG), vol.34, no.1, p.12, 2014.
[3] S. Wanner and B. Goldluecke, "Variational light field analysis for disparity estimation and super-resolution," IEEE Trans. Pattern Anal. Mach. Intell., vol.36, no.3, pp.606–619, 2014.
[4] T.C. Wang, A.A. Efros, and R. Ramamoorthi, "Depth estimation with occlusion modeling using light-field cameras," IEEE Trans. Pattern Anal. Mach. Intell., vol.38, no.11, pp.2170–2181, 2016.
[5] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," Computer Science Technical Report CSTR, vol.2, no.11, pp.1–11, 2005.
[6] T.E. Bishop, S. Zanetti, and P. Favaro, "Light field superresolution," Computational Photography (ICCP), 2009 IEEE International Conference on, pp.1–9, IEEE, 2009.
[7] G. Wetzstein, D. Lanman, M. Hirsch, and R. Raskar, "Tensor Displays: Compressive Light Field Synthesis using Multilayer Displays with Directional Backlighting," ACM Trans. Graph. (Proc. SIGGRAPH), vol.31, no.4, pp.1–11, 2012.
[8] F.C. Huang, K. Chen, and G. Wetzstein, "The light field stereoscope: immersive computer graphics via factored near-eye light field displays with focus cues," ACM Trans. Graphics (TOG), vol.34, no.4, p.60, 2015.
[9] S. Lee, C. Jang, S. Moon, J. Cho, and B. Lee, "Additive light field displays: realization of augmented reality with holographic optical elements," ACM Trans. Graphics (TOG), vol.35, no.4, Article No. 60, 2016.
[10] T. Saito, Y. Kobayashi, K. Takahashi, and T. Fujii, "Displaying real-world light fields with stacked multiplicative layers: Requirement and data conversion for input multiview images," J. Display Technol., vol.12, no.11, pp.1290–1300, Nov. 2016.
[11] B. Wilburn, N. Joshi, V. Vaish, E.V. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy, "High performance imaging using large camera arrays," vol.24, no.3, pp.765–776, 2005.
[12] T. Fujii, K. Mori, K. Takeda, K. Mase, M. Tanimoto, and Y. Suenaga, "Multipoint measuring system for video and sound-100-camera and microphone system," 2006 IEEE International Conference on Multimedia and Expo, pp.437–440, IEEE, 2006.
[13] E.H. Adelson and J.Y.A. Wang, "Single lens stereo with a plenoptic camera," IEEE Trans. Pattern Anal. Mach. Intell., vol.14, no.2, pp.99–106, Feb. 1992.
[14] J. Arai, F. Okano, H. Hoshino, and I. 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] R. Ng, Digital light field photography, Ph.D. thesis, Stanford University, 2006.
[16] A. Veeraraghavan, R. Raskar, A. Agrawal, A. Mohan, and J. Tumblin, "Dappled photography: Mask enhanced cameras for heterodyned light fields and coded aperture refocusing," ACM Trans. Graph., vol.26, no.3, p.69, 2007.
[17] C.K. Liang, T.H. Lin, B.Y. Wong, C. Liu, and H.H. Chen, "Programmable aperture photography: multiplexed light field acquisition," vol.27, no.3, Article No. 55, 2008.
[18] H. Nagahara, C. Zhou, T. Watanabe, H. Ishiguro, and S.K. Nayar, "Programmable aperture camera using LCoS," European Conference on Computer Vision, pp.337–350, Springer, 2010.
[19] S.D. Babacan, R. Ansorge, M. Luessi, P.R. Mataran, R. Molina, and A.K. Katsaggelos, "Compressive light field sensing," IEEE Trans. Image Process., vol.21, no.12, pp.4746–4757, Dec. 2012.
[20] K. Marwah, G. Wetzstein, Y. Bando, and R. Raskar, "Compressive light field photography using overcomplete dictionaries and optimized projections," ACM Trans. Graphics (TOG), vol.32, no.4, Article No. 46, 2013.
[21] S. Tambe, A. Veeraraghavan, and A. Agrawal, "Towards motion aware light field video for dynamic scenes," Proc. IEEE International Conference on Computer Vision, pp.1009–1016, 2013.
[22] D.L. Donoho, "Compressed sensing," IEEE Trans. Inf. Theory, vol.52, no.4, pp.1289–1306, April 2006.
[23] E.J. Candès and M.B. Wakin, "An introduction to compressive sampling," IEEE Signal Process. Mag., vol.25, no.2, pp.21–30, March 2008.
[24] E.J. Candès, Y.C. Eldar, D. Needell, and P. Randall, "Compressed sensing with coherent and redundant dictionaries," Applied and Computational Harmonic Analysis, vol.31, no.1, pp.59–73, 2011.
[25] Y. Yagi, K. Takahashi, T. Fujii, T. Sonoda, and H. Nagahara, "PCA-coded aperture for light field photography," IEEE International Conference on Image Process. (ICIP), 2017.
[26] M.H. Kamal, M. Golbabaee, and P. Vandergheynst, "Light field compressive sensing in camera arrays," 2012 IEEE International Conference on Acoustics, Speech Signal Process. (ICASSP), pp.5413–5416, Ieee, 2012.
[27] Y. Miyagi, K. Takahashi, and T. Fujii, "Weighted 4D-DCT basis for compressively sampled light fields," IEICE Trans. Fundamentals of Electronics, Communications and Computer Sciences, vol.E99-A, no.9, pp.1655–1664, Sept. 2016.
[28] A. Ashok and M.A. Neifeld, "Compressive light field imaging," SPIE Defense, Security, and Sensing, International Society for Optics and Photonics, 2010.
[29] M.A. Neifeld and P. Shankar, "Feature-specific imaging," Applied optics, vol.42, no.17, pp.3379–3389, 2003.
[30] D.D. Lee and H.S. Seung, "Algorithms for non-negative matrix factorization," Advances in neural information processing systems, pp.556–562, 2001.
[31] MIT Media Lab's Camera Culture Group, "Compressive light field camera," 2013. http://cameraculture.media.mit.edu/projects/

compressive-light-field-camera/.
[32] Stanford Graphics Laboratory, "The (new) stanford light field archive," 2008. http://lightfield.stanford.edu.
[33] Fujii Laboratory, Nagoya University, "Computational camera project," 2017. http://www.fujii.nuee.nagoya-u.ac.jp/~takahasi/Research/CompCam/index.html.

Appendix A: Analysis on matrix C for PCA basis

We evaluated the significance of the condition number of the mixing matrix $\boldsymbol{C}$ that is used with the PCA based basis. We randomly generated 100 matrices for three acquisitions ($N = 3$). The standard deviation σ for the observation noise was fixed to 0.025. Other experimental conditions were the same as Section 4. The reconstruction quality for the *Dragon and Bunnies* dataset are plotted against the condition number of $\boldsymbol{C}$ in Fig. A· 1. 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 basis vectors) and $N = 5$ (three basis vectors), respectively. The line with $N = 5$ gives the upper-bound quality that can be achieved with three basis 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.

Appendix B: Angular Frequency in Light Field

As discussed in Sect. 3.4, our method loses some high frequency components in the angular domain. These components correspond to angular dependent factors such as non-Lambertian reflections and occlusion/dis-occlusion boundaries. It should also be noted that even Lambertian textures (whose appearance is direction independent) are affected by the angular frequency components, as explained below.

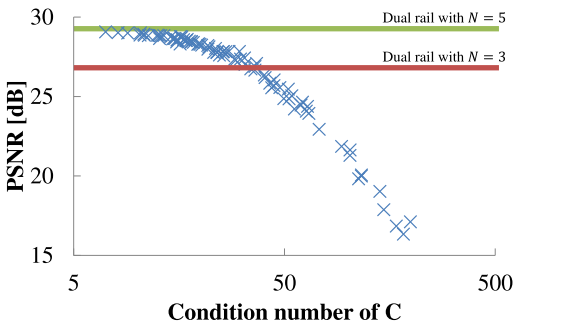


Fig. A· 1 Reconstruction quality vs. condition number of $\boldsymbol{C}$.

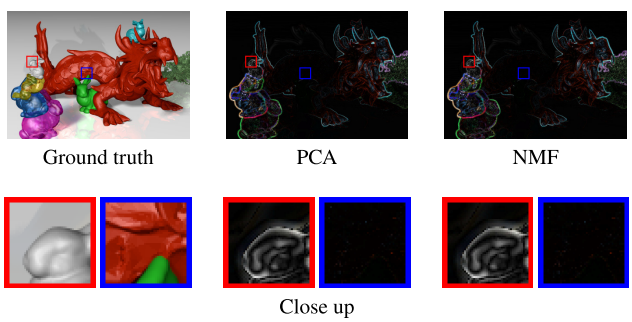


Fig. A· 2 The reconstruction errors (magnified by 3 for visualization) of the central viewpoint.

Let $I(u, v)$ be the image from the viewpoint of $(s, t) = (0, 0)$, which is associated with the 4-D light field $l(s, t, u, v)$ as follows:

$$I(u, v) = l(0, 0, u, v). \tag{A· 1}$$

We assume that an object is located at a depth whose disparity is d among the neighboring viewpoints, and the surface of the object is Lambertian. The light field generated by this object satisfies:

$$l(s, t, u, v) = I(u - sd, v - td). \tag{A· 2}$$

The image signal $I(u, v)$ is generally decomposed into several frequency components. Here, we only consider a unit wave with a spatial frequency (ω_u, ω_v):

$$I(u, v) = \exp(j\omega_u u) \exp(j\omega_v v). \tag{A· 3}$$

Substituting Eq. (A· 3) into Eq. (A· 2), we obtain

$$\begin{aligned} l(s, t, u, v) \;=\; & \exp(j\omega_u u) \cdot \exp(j\omega_v v) \cdot \\ & \exp(-j(d\omega_u)s) \cdot \exp(-j(d\omega_v)t). \end{aligned} \tag{A· 4}$$

This equation shows that the frequency along the angular domains is given as $(-d\omega_u, -d\omega_v)$; i.e., the angular frequency is represented as the product of the texture's frequency (ω_u, ω_v) and the disparity d. This indicates that a textured object with a non-zero disparity value induces some frequency components in the angular domain, even if the object's surface is Lambertian. Inversely, the loss of high frequency components in the angular domain affects the appearance of the texture if the object has a non-zero disparity value.

The point mentioned above can be observed from the experimental results. Figure A· 2 visualizes the reconstruction errors of the central viewpoint obtained with PCA and NMF ($N = 3, \sigma = 0$). In this dataset, the belly of the dragon is located at the zero-disparity depth, and nearer and farther areas from there have positive and negative disparity values, respectively. We can see that significant errors are caused around highly textured objects with non-near-zero disparity values.

Yusuke Yagi received his B.E. and M.E. degrees in Graduate School of Engineering from Nagoya University in 2016 and 2018. He has been engaged in computational photography.

Keita Takahashi received his B.E., M.S., and Ph.D. degrees in information and communication engineering from the University of Tokyo, Japan, in 2001, 2003, and 2006. He was a project assistant professor at the University of Tokyo from 2006–2011 and was an assistant professor at the University of Electro-Communications, Japan from 2011–2013. He is currently an associate professor at Graduate School of Engineering, Nagoya University, Japan. His research interests include light field processing, image-based rendering, computational photography, and 3D display. He is a member of IEEE SPS & CS, ITE, IPSJ, and IEICE.

Toshiaki Fujii received B.E., M.E., and Dr.E. degrees in electrical engineering from the University of Tokyo in 1990, 1992, and 1995. From 1995, he has been with the Graduate School of Engineering, Nagoya University apart from 2008 to 2010, when he was with the Graduate School of Science and Engineering, Tokyo Institute of Technology. He is currently a professor in the Graduate School of Engineering, Nagoya University. His current research interests include multi-dimensional signal processing, multi-camera systems, multi-view video coding and transmission, free-viewpoint television, and their applications.

Toshiki Sonoda received his B.E. and M.E. degrees from Kyushu University in 2012 and 2014. He has been engaged in computational photography.

Hajime Nagahara is a professor at Institute for Datability Science, Osaka university, since 2017. He received Ph.D. degree in system engineering from Osaka University in 2001. He was a research associate of the Japan Society for the Promotion of Science from 2001 to 2003. He was an assistant professor at the Graduate School of Engineering Science, Osaka University, Japan from 2003 to 2010. He was an associate professor in faculty of information science and electrical engineering at Kyushu University from 2010 to 2017. He was a visiting associate professor at CREA University of Picardie Jules Verns, France, in 2005. He was a visiting researcher at Columbia University in 2007–2008 and 2016–2017. Computational photography, computer vision are his research areas. He received an ACM VRST2003 Honorable Mention Award in 2003, IPSJ Nagao Special Researcher Award in 2012, ICCP2016 Best Paper Runners-up, and SSII Takagi Award in 2016.