Multi-Shot Single Sensor Light Field Camera Using a Color Coded Mask

Ehsan Miandji
dept. of Science and Technology
Linköping University, Sweden

Jonas Unger
dept. of Science and Technology
Linköping University, Sweden

Christine Guillemot
INRIA, Campus de Beaulieu
Rennes, France

Abstract—We present a compressed sensing framework for reconstructing the full light field of a scene captured using a single-sensor consumer camera. To achieve this, we use a color coded mask in front of the camera sensor. To further enhance the reconstruction quality, we propose to utilize multiple shots by moving the mask or the sensor randomly. The compressed sensing framework relies on a training based dictionary over a light field data set. Numerical simulations show significant improvements in reconstruction quality over a similar coded aperture system for light field capture.

Index Terms—light field camera, coded aperture

I. INTRODUCTION

Light field imaging has emerged as a promising technology for a variety of applications including photorealistic rendering, computational photography and computer vision. Many acquisition devices have been designed to capture light fields, ranging from camera arrays, [1], to single cameras mounted on moving gantries, [2], and plenoptic cameras [3], [4]. Plenoptic cameras rely on multiplexing and use an array of microlenses placed in front of the sensor to capture multiple low resolution views in one 2D sensor image [4], [5]. This is an efficient and easy to implement way of capturing multiple view-points, but reduces the spatial resolution by orders of magnitude compared to the a raw sensor image.

Recent research efforts have been directed towards overcoming the spatio-angular trade-off inherent to plenoptic cameras. In [6], Liang et al. used a programmable non-refractive mask placed at the aperture to sequentially capture the light field as a linear combination subsets of light rays. This series of 2D measurements, however, usually requires a large number of captured images and long exposure times. Two architectures are instead described in [7] for compressive acquisition of light fields, one exploiting correlations along the spatial dimensions and the other one along the angular dimensions, allowing for capture using short exposure times.

In [8], an architecture is proposed to compute optically coded projections on a fixed 2D sensor using two attenuation masks separately placed at the aperture plane and in front of the sensor. Given the measurements recorded on the sensor, the light field is then reconstructed using a least square minimization with a total variation regularization constraint. Similarly, the authors in [9] place a randomly coded mask in front of the aperture to obtain incoherent measurements of the light field. Multiple shots are captured as random linear combinations of angular images by separately opening one region of the aperture and blocking light in the others.

A camera architecture is also proposed in [10] based on optically coded projections on a single image sensor. The light field is reconstructed using a compressive sensing framework, assuming that the light field is sparse in a domain defined by an overcomplete dictionary. In [11], incoherent spatio-angular measurements are extracted using a random binary mask and the light field is reconstructed using an ensemble of 2D separable dictionaries. While the results significantly outperform [10] for general purpose light field sensing, the special structure of the sensing matrix required for a single sensor light field camera is not considered.

In this paper, we propose a camera architecture which captures incoherent measurements of the light field via a controllable color mask placed in front of the sensor. In order to increase the incoherence, we take multiple shots, where for each shot the mask configuration is changed to create a new random pattern. To reduce computations and increase the incoherence, we optionally also perform a random sampling of the spatial domain. The reconstruction algorithm is computationally efficient and existing ℓ_1 or ℓ_0 solvers [12]–[16] can be used. In the analysis and evaluation we use RGB color filters, but the theory also applies to multi-spectral filters.

Throughout the paper we use the following notation. Vectors and matrices are denoted by boldface lower-case ($\mathbf{a}$) and boldface upper-case ($\mathbf{A}$) letters, respectively. Elements of a vector or matrix are denoted using subscripts, e.g. $\mathbf{a}_i$ or $\mathbf{A}_{i,j}$. The jth column of a matrix is denoted $\mathbf{A}_j$. Finally, we denote a set of algebraic objects by superscripts; e.g. $\mathbf{\Phi}^{i,j}$, $i \in \{1, \ldots, m\}$, $j \in \{1, \ldots, n\}$, denotes a set of mn matrices.

II. LIGHT FIELD COMPRESSIVE ACQUISITION

A popular approach for light field photography is coded aperture [6], [17]. By employing a geometric optics perspective, this method constructs several views of a scene using different regions of the aperture [9]. Let the function $l(\mathbf{r}_i, \mathbf{t}_j, \mathbf{u}_\alpha, \mathbf{v}_\beta)$ describe the two plane light field parametrization [18] defined by a pair of (discrete) locations on the sensor plane $(\mathbf{r}_i, \mathbf{t}_j)$ and the aperture plane $(\mathbf{u}_\alpha, \mathbf{v}_\beta)$. The resolution of the sensor and the aperture plane are $|\mathbf{r}| \times |\mathbf{t}|$ and $|\mathbf{u}| \times |\mathbf{v}|$, respectively. The coding of the aperture is achieved by using a mask of resolution $|\mathbf{u}| \times |\mathbf{v}|$ on the aperture plane. Each region of this mask produces an *angular* image on the sensor, which

$$
\begin{bmatrix}
\mathbf{y}^{R,1} \\
\mathbf{y}^{G,1} \\
\mathbf{y}^{B,1} \\
\mathbf{y}^{R,2} \\
\mathbf{y}^{G,2} \\
\mathbf{y}^{B,2} \\
\vdots \\
\mathbf{y}^{R,s} \\
\mathbf{y}^{G,s} \\
\mathbf{y}^{B,s}
\end{bmatrix}
=
\begin{bmatrix}
\mathbf{\Phi}^{1,R,1} & \dots & \mathbf{\Phi}^{\nu,R,1} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\
\mathbf{0} & \dots & \mathbf{0} & \mathbf{\Phi}^{1,G,1} & \dots & \mathbf{\Phi}^{\nu,G,1} & \mathbf{0} & \dots & \mathbf{0} \\
\mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{\Phi}^{1,B,1} & \dots & \mathbf{\Phi}^{\nu,B,1} \\
\mathbf{\Phi}^{1,R,2} & \dots & \mathbf{\Phi}^{\nu,R,2} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\
\mathbf{0} & \dots & \mathbf{0} & \mathbf{\Phi}^{1,G,2} & \dots & \mathbf{\Phi}^{\nu,G,2} & \mathbf{0} & \dots & \mathbf{0} \\
\mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{\Phi}^{1,B,2} & \dots & \mathbf{\Phi}^{\nu,B,2} \\
\vdots & \vdots & & \ddots & & \vdots & & & \\
\mathbf{\Phi}^{1,R,s} & \dots & \mathbf{\Phi}^{\nu,R,s} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\
\mathbf{0} & \dots & \mathbf{0} & \mathbf{\Phi}^{1,G,s} & \dots & \mathbf{\Phi}^{\nu,G,s} & \mathbf{0} & \dots & \mathbf{0} \\
\mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{\Phi}^{1,B,s} & \dots & \mathbf{\Phi}^{\nu,B,s}
\end{bmatrix}
\underbrace{
\begin{bmatrix}
\mathbf{x}^{1,R} \\
\vdots \\
\mathbf{x}^{\nu,R} \\
\mathbf{x}^{1,G} \\
\vdots \\
\mathbf{x}^{\nu,G} \\
\mathbf{x}^{1,B} \\
\vdots \\
\mathbf{x}^{\nu,B}
\end{bmatrix}}_{},
\tag{1}
$$

$$\underbrace{\qquad}_{\triangleq \mathbf{y} \in \mathbb{R}^{\omega \lambda s}} \qquad \underbrace{\qquad}_{\triangleq \Phi \in \mathbb{R}^{\omega \lambda s \times \omega \nu \lambda}} \qquad \underbrace{\qquad}_{\triangleq \mathbf{x} \in \mathbb{R}^{\omega \nu \lambda}}$$

we denote by $\mathbf{y}^i$, where $i \in \{1, \dots, \nu\}$, and $\nu = |\mathbf{u}||\mathbf{v}|$. If the mask is fully transparent, then the image formed on the sensor is the integrated radiance from all aperture regions, i.e. $\mathbf{y} = \sum_{i=1}^{\nu} \mathbf{y}^i$. Assuming the mask has values $\mathbf{a}_i$, the sensing model for a coded aperture can be formulated as follows

$$
\mathbf{y} = \begin{bmatrix} \mathbf{a}_1 \mathbf{I} & \mathbf{a}_2 \mathbf{I} & \dots & \mathbf{a}_\nu \mathbf{I} \end{bmatrix}
\begin{bmatrix} \mathbf{x}^1 \\ \mathbf{x}^2 \\ \vdots \\ \mathbf{x}^\nu \end{bmatrix},
\tag{2}
$$

where $\mathbf{x}^i \in \mathbb{R}^\omega$, with $\omega = |\mathbf{r}||\mathbf{t}|$, is the incoming radiance corresponding to the aperture region i, and $\mathbf{I} \in \mathbb{R}^{\omega \times \omega}$ is the identity matrix. To maximize incoherence of the sensing matrix and a deterministic basis function, we would like the elements of Φ to be independent and identically distributed random variables [19]–[21]. Unfortunately, the sensing matrix $\begin{bmatrix} \mathbf{a}_1 \mathbf{I} & \dots & \mathbf{a}_\nu \mathbf{I} \end{bmatrix}$ in (2) is highly structured, and hence leads to coherent measurements. To alleviate this problem, Babacan et al. [9] propose to use multiple shots, where for each shot a different set of mask values is used. Assuming s shots are taken, equation (2) becomes

$$
\begin{bmatrix} \mathbf{y}^1 \\ \mathbf{y}^2 \\ \vdots \\ \mathbf{y}^s \end{bmatrix}
=
\begin{bmatrix}
\mathbf{A}_{1,1}\mathbf{I} & \mathbf{A}_{1,2}\mathbf{I} & \dots & \mathbf{A}_{1,\nu}\mathbf{I} \\
\mathbf{A}_{2,1}\mathbf{I} & \mathbf{A}_{2,2}\mathbf{I} & \dots & \mathbf{A}_{2,\nu}\mathbf{I} \\
\vdots & \vdots & \ddots & \vdots \\
\mathbf{A}_{s,1}\mathbf{I} & \mathbf{A}_{s,2}\mathbf{I} & \dots & \mathbf{A}_{s,\nu}\mathbf{I}
\end{bmatrix}
\begin{bmatrix} \mathbf{x}^1 \\ \mathbf{x}^2 \\ \vdots \\ \mathbf{x}^\nu \end{bmatrix}.
\tag{3}
$$

To further improve the incoherence, Marwah et al. [10] propose to place the mask at a small distance d_m from the sensor, such that $d_m < d_a$, where d_a is the distance of the aperture plane from the sensor. In this setup, the pixel value at location $(\mathbf{r}_i, \mathbf{t}_j)$ over the sensor is calculated by integrating the incoming radiance from different directions while modulating them by the mask:

$$
y(\mathbf{r}_i, \mathbf{t}_j) = \sum_{\alpha=1}^{|\mathbf{u}|} \sum_{\beta=1}^{|\mathbf{v}|} f\left(\mathbf{r}_i + \gamma(\mathbf{u}_\alpha - \mathbf{r}_i), \mathbf{t}_j + \gamma(\mathbf{v}_\beta - \mathbf{t}_j)\right)
$$
$$
l(\mathbf{r}_i, \mathbf{t}_j, \mathbf{u}_\alpha, \mathbf{v}_\beta), \tag{4}
$$

where $\gamma = d_m / d_a$. The function $f(.)$ describes the mask value given a spatial location. Indeed there is a one to one
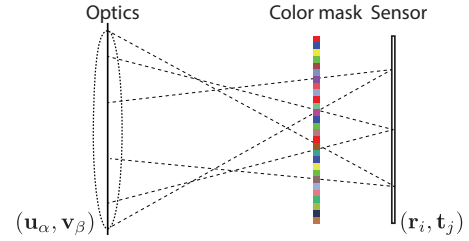

Fig. 1: Single sensor light field camera using a color coded mask

correspondence between the pixels on the sensor and on the mask. The sensing model according to (4) is

$$
\mathbf{y} = \begin{bmatrix} \mathbf{\Phi}^1 & \mathbf{\Phi}^2 & \dots & \mathbf{\Phi}^\nu \end{bmatrix}
\begin{bmatrix} \mathbf{x}^1 \\ \mathbf{x}^2 \\ \vdots \\ \mathbf{x}^\nu \end{bmatrix},
\tag{5}
$$

where $\mathbf{\Phi}^i \in \mathbb{R}^{\omega \times \omega}$ are diagonal matrices, each containing sheared mask values obtained by evaluating the function $f(.)$ in (4) for the corresponding view direction $(\mathbf{u}_\alpha, \mathbf{v}_\beta)$ and all the spatial locations over the sensor. The nonzero elements of the mask are typically drawn from an appropriate distribution (e.g. a nonzero Gaussian) or optimized for maximal incoherence with the sparsifying dictionary [10].

In all the aforementioned acquisition frameworks, a monochrome mask is used. As a result, the same sensing matrix Φ is applied to all the color channels. This indeed leads to coherent measurements. In the following section, a new acquisition framework will be described that increases the measurement incoherence by randomizing the sensing of different spectral bands of the light field.

III. MULTI-SHOT COLOR CODED ACQUISITION

Unlike [10] that uses a monochrome mask, we utilize a random color mask (see Fig. 1). Recent advances in micro-lithography techniques [22]–[24] has enabled the utilization of colored micro-filters in many imaging applications such as multi-spectral imaging [25], depth extraction [26], and image enhancement [27]. To further increase the measurement matrix incoherence, we use multiple shots, each with a different random colored mask. To achieve this, similar to [28], we propose to use a piezo system for rapid mask movement. Depending

on the spatial patch size (see Section IV) a movement of few pixels is enough for a new random mask pattern in each shot. Such systems are capable of capturing more than 60 frames per second, hence moving objects can be faithfully reconstructed. The pixel pitch of the mask is the same as the sensor; however, the mask has a higher resolution to facilitate movements.

The new light field sensing model incorporating per-shot colored mask patterns is formulated in (1), shown on the previous page. Without loss of generality, and to simplify our notation, we consider three color channels (RGB). The entries $\mathbf{0}$ denote zero-valued matrices of appropriate size. Additionally, the sensing matrix $\mathbf{\Phi}$ takes into account all the shots. Hence, the reconstruction method, see (7), can recover the light field $\mathbf{x}$ in a single pass. Alternatively, one can recover each shot individually and compute the final result as the average of all the recovered shots (see section V for a comparison of these approaches).

As it can be seen in (1), the total number of samples, i.e. the number of rows in $\mathbf{\Phi}$, is $\omega\lambda s$, where λ is the number of color channels. Since ω and λ are determined by the hardware implementation, the number of samples is mainly determined by s, which in turn affects the computational complexity of the reconstruction algorithm (see section IV). To provide flexibility over the trade off between computational complexity and reconstruction quality, we use spatial subsampling in the measurement model (1). This can be done by a sampling matrix $\mathbf{P} \in \mathbb{R}^{r\omega\lambda s \times \omega\lambda s}$ as follows

$$\mathbf{y} = \mathbf{P}\mathbf{\Phi}\mathbf{x}, \tag{6}$$

where $r \in (0,1]$ is the sampling ratio. The matrix $\mathbf{P}$ can be constructed by sampling $r\omega\lambda s$ rows from $\mathbf{I} \in \mathbb{R}^{\omega\lambda\nu \times \omega\lambda\nu}$ uniformly at random without replacement. Using (6), the total number of samples is proportional to sr. When s is fixed, e.g. due to hardware limitations, one can use r to reduce the reconstruction time (see Table II). On the other hand, spatial sub-sampling can be used to increase the reconstruction quality when the total number of samples, or the reconstruction time, is constant. To illustrate this, we consider two cases: 1. We take one shot without sub-sampling 2. We take two shots and set $r = 0.5$ so that the total number of samples is equal to the previous case. Results of this experiment are summarized in Table I. Using this simple approach we gain 0.24dB in PSNR without an increase in reconstruction time.

TABLE I: Effect of sub-sampling on reconstruction quality. Angular resolution is 5×5 and spatial patch size is 9×9 (non-overlapping). The data set used here is described in section V. Timing results are in seconds.

ν	ω	λ	s	r	total # samples	PSNR (dB)	Time
25	81	3	1	1	243	27.52	6
25	81	3	2	0.5	243	27.76	6

The nonzero values of the matrix $\mathbf{\Phi}$ in (1) can be drawn from various distributions. A common approach is to use a nonzero centered Gaussian distribution $|\mathcal{N}(0,\sigma)|$. Subgaussian distributions, e.g. Bernoulli distribution with probability 0.5,

can also be used. We will compare the reconstruction quality for different distributions in section V.

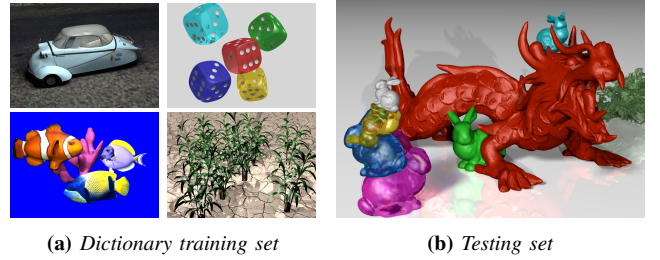**(a)** *Dictionary training set*   **(b)** *Testing set*

Fig. 2: Data sets used in our experiments: (a) Four light fields used for training an overcomplete dictionary, and (b) the light field we used in our simulations.

IV. RECONSTRUCTION ALGORITHM

The reconstruction algorithm takes as input the measured light field ($\mathbf{y}$) and the sensing matrix ($\mathbf{P}\mathbf{\Phi}$) in order to obtain the original light field $\mathbf{x}$, see (6). Since in practice $\nu \gg s$, the sensing matrix is wide and there exists infinitely many solutions to (1). Using compressed sensing [29], the measurement model in (1) becomes $\mathbf{x} = \mathbf{P}\mathbf{\Phi}\mathbf{D}\alpha$, where $\mathbf{D} \in \mathbb{R}^{\omega\nu\lambda \times \rho\omega\nu\lambda}$ is an overcomplete dictionary with ρ as the overcompleteness factor, and α defines the sparse coefficients. By solving the following optimization problem

$$\hat{\mathbf{c}} = \min_{\mathbf{b}} \|\mathbf{b}\|_1 \quad \text{s.t.} \quad \hat{\mathbf{p}} = \mathbf{\Phi}\mathbf{D}\mathbf{b}, \tag{7}$$

we can recover or estimate the original light field as $\hat{\mathbf{x}} = D\hat{\mathbf{c}}$. Equation (7) is known as Basis Pursuit Denoising (BPDN) and there exist several approaches for solving this problem [12], [15], [16]. In this paper we use the SL0 algorithm [16] since we observed superior reconstruction quality. Moreover, similar to [10], we use a training based method [30] to obtain an overcomplete dictionary.

Due to the large size of the vectorized light field, $\mathbf{x}$, it is a common practice to divide the light field over the spatial domain [10], [11], while including all the angular and spectral information. As a result, the dimensionality of each vectorized light field patch will be $\hat{\omega}\nu\lambda$, where $\hat{\omega}$ is the spatial patch size over the sensor. In addition, one can create overlapping patches [31] from the light field. After the reconstruction, the final light field elements are calculated by averaging the corresponding values in overlapped patches.

V. EXPERIMENTAL RESULTS

In this section we will report simulation results for our light field capture method. For the purpose of comparison, the synthetic light field data set of [10] is used. This data set consists of five light fields, which we have divided into training and testing sets as shown in Fig. 2. The angular resolution of each light field is 5×5, covering a field of view of 20 degrees; i.e. the angle separation is 4 degrees. The training set is used to learn a 1.5 times overcomplete, 10-sparse dictionary. The spatial patch size is set to 9×9. Hence the signal dimensionality is $9 \times 9 \times 5 \times 5 \times 3 = 6075$, where

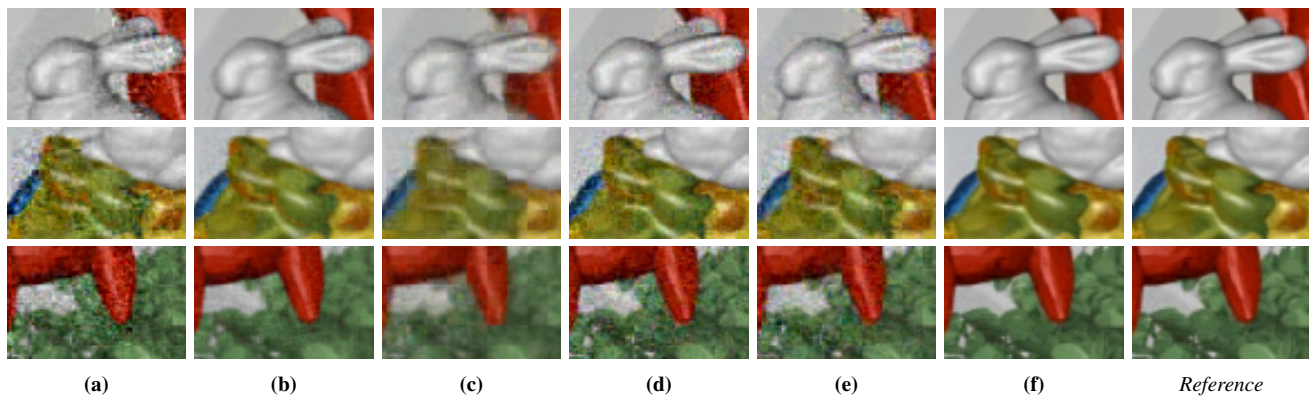| | **(a)** | **(b)** | **(c)** | **(d)** | **(e)** | **(f)** | *Reference* |

Fig. 3: Visual quality comparison: (a) and (b) are obtained by the method of Marwah et al. [10] using non-overlapping and overlapping patches, respectively. And (c)-(f) show the results of our method using the parameters listed in Table II. For all these cases we used 5 shots.

TABLE II: Reconstruction quality for various compressive coded aperture methods. For these results we used uniform distribution for nonzero values of (1).

	Mask	Method	s	r	overlap	PSNR (dB)	Time (min.)	Figure
Marwah et al. [10]	Grayscale	Simultaneous	5	1	no	27.91	43	3a
	Grayscale	Simultaneous	5	1	yes	33.59	380	3b
Our method	RGB	Independent	5	1	no	29.23	30	3c
	RGB	Simultaneous	5	1	no	31.50	43	3d
	RGB	Simultaneous	5	0.5	no	30.14	24	3e
	RGB	Simultaneous	5	1	yes	37.63	381	3f

the last component is for color. To simulate mask movements, we generate a new random mask pattern for each shot.

Figure 3 compares our method with that of Marwah et al. [10]. Parameter values, along with PSNR and timing results, are summarized in Table II. We trained one dictionary for both methods. It should be noted that in [10], a dictionary of size $\omega\nu \times \rho\omega\nu$ is trained and used for all the color channels independently. In our implementation of their method, we trained a dictionary of size $\omega\nu\lambda \times \rho\omega\nu\lambda$. This dictionary improves the results of [10] by about 3dB. We compared our results with the improved implementation of [10]. Our method achieves about 3.6dB to 4dB improvement over [10], using non-overlapping and overlapping patches, respectively. Comparing figures 3b and 3f, we see that the grayscale mask leads to noise-like artifacts in the reconstruction. In contrast, our color coded mask produces results that are almost indistinguishable from the original light field.

When spatial sub-sampling is used, while keeping the number of shots fixed, we observe a drop of 1.5dB in PSNR. However, the computation time is reduced almost in half. Given the small reduction in visual quality (compare 3d and 3e), sub-sampling can be used to accelerate the reconstruction process. On the other hand, increasing the number of shots and decreasing the sub-sampling ratio while keeping the total number of samples constant will lead to a slight increase in reconstruction quality (see Table I). Moreover, we observed that independent reconstruction of shots followed by linear interpolation (Fig. 3c), performs inferior to simultaneous reconstruction (Fig. 3d) as described in section III.

The effect of the number of shots (s) on reconstruction quality is plotted in Fig. 5, along with a visual quality comparison in Fig. 4. It can be seen that when the number

of shots is more than three, the PSNR grows linearly; note that a linear growth in PSNR implies exponential reduction in error. As a result, it is beneficial to increase s as far as the camera's ability to handle fast moving objects is not hindered. Results in figures 4 and 5 were obtained using non-overlapping patches and a Gaussian color mask.

TABLE III: Reconstruction quality (PSNR) for various distributions used for non-zero values of (1). The zero-mean Gaussian distribution with variance σ is denoted $\mathcal{N}(0,\sigma)$.

Uniform	$\mathcal{N}(0,1)$	$\|\mathcal{N}(0,1)\|$	Bernoulli	Rademacher	Halton
30.72	31.50	30.80	30.87	31.50	27.70

Finally, we compare the reconstruction quality for different mask distributions in Table III. The following distributions are considered: uniform in the interval $[0,1]$, Gaussian and nonnegative Gaussian, Bernoulli distribution with probability 0.5, Rademacher distribution, and a quasi random sequence, specifically the Halton sequence [32]. Note that the distributions that take negative values like the Gaussian cannot be implemented as a mask. However, we include these results for comparison since they are commonly used in compressed sensing. Indeed the Halton sequence is inferior to random masks. Moreover, nonnegative distributions are performing very similarly. Here we used non-overlapping patches.

VI. Conclusions

We introduced a compressive light field camera using a color coded mask. Moreover, a multi-shot capture by micro movements of the mask shows significant improvements in reconstruction quality. For future work, we are interested in determining the conditions for exact recovery that take into account the sparsity of the light field and the number of shots.
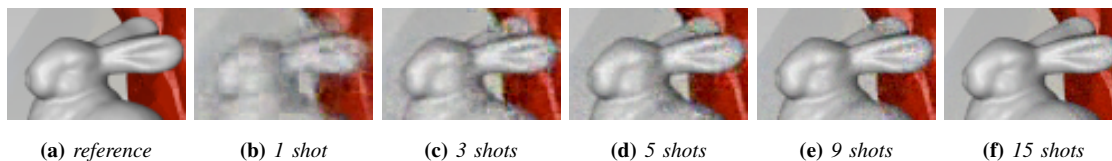
(a) *reference* **(b)** *1 shot* **(c)** *3 shots* **(d)** *5 shots* **(e)** *9 shots* **(f)** *15 shots*

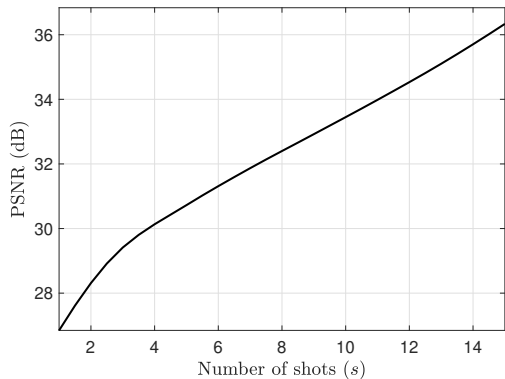Fig. 4: Visual quality comparison for the number of shots (s). See Fig. 5 for the corresponding PSNR values.

Fig. 5: The Effect of number of shots (s) on PSNR. See Fig. 4 for a visual comparison. Non-overlapping patches are used.

ACKNOWLEDGMENT

REFERENCES

[1] 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," *ACM Trans. Graph.*, vol. 24, no. 3, pp. 765–776, jul 2005.

[2] J. Unger, S. Gustavson, P. Larsson, and A. Ynnerman, "Free Form Incident Light Fields," *Computer Graphics Forum*, 2008.

[3] T. Adelson and J. Wang, "Single lens stereo with a plenoptic camera," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 99–106, 1992.

[4] R. Ng, M. Levoy, M. Bredif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a handheld plenoptic camera," *Technical Report CTSR 2005-02, Stanford University*, 2005.

[5] A. Lumsdaine and T. Georgiev, "The focused plenoptic camera," in *Proc. of IEEE International Conf. on Computational Photography, ICCP*, 2009, p. 18.

[6] C.-K. Liang, T.-H. Lin, B.-Y. Wong, C. Liu, and H. H. Chen, "Programmable aperture photography: multiplexed light field acquisition," in *Proc. of ACM SIGGRAPH*, vol. 27, 2008, pp. 1–10.

[7] A. Ashok and M. Neifeld, "Compressive light field imaging," in *Proc. of SPIE 7690A, Three-Dimensional Imaging, Visualization, and Display*, May 2010.

[8] Z. Xu and E. Lam, "A high-resolution lightfield camera with dual-mask design," in *Proc. Image Reconstruction from Incomplete Data VII, SPIE Optical engineering + applications*, vol. 8500, Oct. 2012.

[9] S. Babacan, R. Ansorge, M. Luessi, P. R. Mataran, R. Molina, and A. K. Katsaggelos, "Compressive light field sensing," *IEEE Trans. on Image Processing*, vol. 21, no. 12, pp. 4746–4757, dec. 2012.

[10] K. Marwah, G. Wetzstein, Y. Bando, and R. Raskar, "Compressive light field photography using overcomplete dictionaries and optimized projections," *ACM Trans. on Graphics (Proc. of SIGGRAPH)*, vol. 32, no. 4, pp. 46:1–46:12, 2013.

[11] E. Miandji, J. Kronander, and J. Unger, "Compressive image reconstruction in reduced union of subspaces," *Comput. Graph. Forum*, vol. 34, no. 2, pp. 33–44, may 2015.

[12] S. Wright, R. Nowak, and M. Figueiredo, "Sparse reconstruction by separable approximation," *Signal Processing, IEEE Transactions on*, vol. 57, no. 7, pp. 2479–2493, July 2009.

[13] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *Annals of Statistics*, vol. 32, no. 2, pp. 407–499, 2004.

[14] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 586–597, Dec 2007.

[15] E. van den Berg and M. P. Friedlander, "Probing the pareto frontier for basis pursuit solutions," *SIAM Journal on Scientific Computing*, vol. 31, no. 2, pp. 890–912, 2009.

[16] H. Mohimani, M. Babaie-Zadeh, and C. Jutten, "A fast approach for overcomplete sparse decomposition based on smoothed l0 norm," *Signal Processing, IEEE Transactions on*, vol. 57, no. 1, pp. 289–301, Jan 2009.

[17] C. K. Liang, Y. C. Shih, and H. H. Chen, "Light field analysis for modeling image formation," *IEEE Transactions on Image Processing*, vol. 20, no. 2, pp. 446–460, Feb 2011.

[18] S. Gortler, R. Grzeszczuk, R. Szeliski, and M. Cohen, "The lumigraph," in *Proc. SIGGRAPH*, 1996, pp. 43–54.

[19] D. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *Information Theory, IEEE Transactions on*, vol. 47, no. 7, pp. 2845–2862, Nov 2001.

[20] Z. Ben-Haim, Y. Eldar, and M. Elad, "Coherence-based performance guarantees for estimating a sparse vector under random noise," *Signal Processing, IEEE Transactions on*, vol. 58, no. 10, pp. 5030–5043, Oct 2010.

[21] E. J. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.

[22] P.-J. Lapray, X. Wang, J.-B. Thomas, and P. Gouton, "Multispectral filter arrays: Recent advances and practical implementation," *Sensors*, vol. 14, no. 11, pp. 21 626–21 659, 2014.

[23] J. M. Eichenholz and J. Dougherty, "Ultracompact fully integrated megapixel multispectral imager," 2009.

[24] D. Yi, L. Kong, J. Wang, and F. Zhao, "Fabrication of multispectral imaging technology driven MEMS-based micro-arrayed multichannel optical filter mosaic," in *Advanced Fabrication Technologies for Micro/Nano Optics and Photonics IV*, ser. Proceedings of the SPIE, vol. 7927, feb 2011.

[25] H. Arguello and G. R. Arce, "Colored coded aperture design by concentration of measure in compressive spectral imaging," *IEEE Transactions on Image Processing*, vol. 23, no. 4, pp. 1896–1908, April 2014.

[26] Y. Bando, B.-Y. Chen, and T. Nishita, "Extracting depth and matte using a color-filtered aperture," *ACM Trans. Graph.*, vol. 27, no. 5, pp. 134:1–134:9, dec 2008.

[27] E. Lee, W. Kang, S. Kim, and J. Paik, "Color shift model-based image enhancement for digital multifocusing based on a multiple color-filter aperture camera," *IEEE Transactions on Consumer Electronics*, vol. 56, no. 2, pp. 317–323, May 2010.

[28] D. Kittle, K. Choi, A. Wagadarikar, and D. J. Brady, "Multiframe image estimation for coded aperture snapshot spectral imagers," *Appl. Opt.*, vol. 49, no. 36, pp. 6824–6833, Dec 2010.

[29] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, March 2008.

[30] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding," *J. Mach. Learn. Res.*, vol. 11, pp. 19–60, mar 2010.

[31] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Transactions on Image Processing*, vol. 15, no. 12, pp. 3736–3745, Dec 2006.

[32] J. H. Halton, "Algorithm 247: Radical-inverse quasi-random point sequence," *Commun. ACM*, vol. 7, no. 12, pp. 701–702, dec 1964.