GOOD GROUP SPARSITY PRIOR FOR LIGHT FIELD INTERPOLATION

Keita Takahashi, Shu Fujita, Toshiaki Fujii

Graduate School of Engineering, Nagoya University, Japan

ABSTRACT

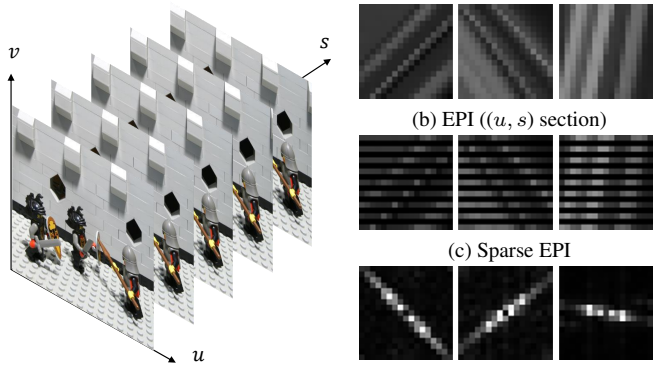
A light field, which is equivalent to a dense set of multi-view images, has various applications such as depth estimation and 3D display. One of the essential problems is light field interpolation, which is obtaining sufficiently dense views from sparser views. The accuracy of interpolation will be enhance by exploiting an inherent property of a light field. Specifically, an epipolar plane image (EPI), which is a 2D subset of the 4D light field, consists of many lines. This structure induces a sparse representation in the frequency domain, where most of the energy resides on a line passing through the origin. On the basis of this observation, we propose a group sparsity prior suitable for light fields to fully exploit their line structure for interpolation. Our experimental results show that the proposed method can achieve better quality than a state-of-the-art shearlet-based method.

Index Terms— Light field interpolation, group sparsity, discrete Fourier transform, EPI, line structure

1. INTRODUCTION

A light field (LF) [1, 2], which is equivalent to a set of multi-view images, is an useful data representation for both computer vision and graphics researchers. 3D visual information is represented as a 4D light field signal with spatial (u, v) and angular (s, t) coordinates at which light rays pass a reference plane. The LF representation is used for a wide range of applications such as digital refocusing [3, 4], free viewpoint videos [5], and 3D displays [6, 7, 8].

One of the essential problems is LF interpolation [9, 10, 11, 12], which is obtaining sufficiently dense views from sparser views. As an example, Fig. 1 (a) shows several views with horizontal viewpoint shifts that constitute a 3D LF on (u, v, s). A section of the original LF with a fixed v is called an epipolar plane image (EPI). As shown in Fig. 1 (b), EPIs consist of many tilted lines. Interpolation of an LF, i.e., view interpolation, is the problem of reconstructing the latent EPIs (Fig. 1 (b)) from sparsely sampled EPIs as shown in Fig. 1 (c). For this purpose, exploiting an inherent property of a light field would help improve the accuracy. The line structure of the latent EPI induces a sparse representation in the frequency domain, where most of the energy concentrates on a line passing through the origin [13], as shown in Fig. 1 (d). On the basis of this observation, we propose a group sparsity prior suitable for light fields to fully exploit their line structure for interpolation.



(a) Light field w.r.t (u, v, s)

(b) EPI ((u, s) section)

(c) Sparse EPI

(d) Frequency domain of (b)

Fig. 1: Light field and its properties.

In the previous literature, most successful approaches to this problem adopted a sparse coding framework based on shearlet transform [11, 12] because the tilted lines in an EPI can be sparsely represented in the shearlet transform domain [14]. In contrast, we revisit the sparsity and group sparsity in the frequency domain representation of an EPI to derive a better prior for LF interpolation, which is described using discrete Fourier transform (DFT) coefficients. Our experimental results show that our method outperforms the shearlet-based method [11, 12] in accuracy.

The remainder of this paper is organized as follows. Section 2 describes a signal reconstruction framework using sparsity and group sparsity. In Sec. 3, we apply this framework to the problem of LF interpolation and derive a group sparsity prior that is suitable for LFs. The proposed prior is experimentally validated in Sec. 4, followed by the conclusion in Sec. 5.

2. SPARSITY AND GROUP SPARSITY

Letting $\boldsymbol{x} \in \mathbb{R}^N$ be a target signal, $\boldsymbol{y} \in \mathbb{R}^N$ is the observation from which $\boldsymbol{x}$ should be reconstructed. The observation model is given by:

$$\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x}. \tag{1}$$

where $\boldsymbol{\Phi} \in \mathbb{R}^{N \times N}$ is an observation matrix. We assume that $\boldsymbol{x}$ is represented using a linear combination of M column vectors $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_M \in \mathbb{R}^N$ and the coefficients $z_1, \ldots, z_M$ are

represented as:

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{z}, \tag{2}$$

where $\boldsymbol{z} = (z_1, \ldots, z_M)^T$ is the coefficient vector, and $\boldsymbol{A} = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_M)$ is generally called a basis or frame.

A key assumption in sparse coding is that vector $\boldsymbol{z}$ should be sparse; only a few elements take non-zero values. With this sparsity prior, the optimal vector $\hat{\boldsymbol{z}}$ satisfying Eq. 2 can be derived by solving a lasso problem [15] as follows:

$$\hat{\boldsymbol{z}} = \arg\min_{\boldsymbol{z}} ||\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{A}\boldsymbol{z}||_2^2 + \lambda||\boldsymbol{z}||_1^1, \tag{3}$$

where λ is a non-negative parameter, and $\|\cdot\|_1$ and $\|\cdot\|_2$ are l_1 and l_2 norms, respectively. To further introduce a group structure to the sparsity, Eq. 3 is extended using a group norm term $\|\cdot\|_{\mathcal{G}}$ to:

$$\hat{\boldsymbol{z}} = \arg\min_{\boldsymbol{z}} ||\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{A}\boldsymbol{z}||_2^2 + \lambda||\boldsymbol{z}||_1^1 + \eta||\boldsymbol{z}||_{\mathcal{G}} \tag{4}$$

$$||\boldsymbol{z}||_{\mathcal{G}} = \sum_{\boldsymbol{g}_i \in \mathcal{G}} w_i ||\boldsymbol{z}_{\boldsymbol{g}_i}||_2^1, \tag{5}$$

where η is a non-negative parameter. Vector $\boldsymbol{z}_{\boldsymbol{g}_i}$ includes all the elements of $\boldsymbol{z}$ that belong to the i-th group. The set of $\boldsymbol{g}_i$ is described as $\mathcal{G} = \{\boldsymbol{g}_1, \ldots, \boldsymbol{g}_{|\mathcal{G}|}\}$. Symbol w_i is the weight for the i-th group. This problem is called sparse group lasso [16] or overlapping group lasso [17, 18] depending on the existence of overlapped coefficients between the groups.

3. PROPOSED METHOD

In this section, we propose a group sparsity prior suitable for light fields to fully exploit their line structure for the problem of LF interpolation.

3.1. Basic Formulation

We apply the aforementioned sparsity and group sparsity to the LF interpolation problem by interpreting the notations used in Sec. 2 as follows. Vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ are the latent and sub-sampled light field signals. In this paper, the processing unit for the light field is a 2D block extracted from a 2D EPI; therefore, it is reshaped from 2D to 1D to yield a vector representation used as $\boldsymbol{x}$ or $\boldsymbol{y}$. The observation matrix $\boldsymbol{\Phi}$ is a sub-sampling operator. We will discuss the detailed design of the groups $\mathcal{G}$ in Sec. 3.2.

The design of $\boldsymbol{A}$ is important since it determines the domain where the sparsity and group sparsity are considered. A learned dictionary [19, 20], weighted discrete cosine transform basis [21], and shearlet frame [11, 12] have been used for sparse representations of LFs previously. Meanwhile, we use the discrete Fourier transform (DFT) basis because it enables us to sparsely represent the line structure of an LF signal in the frequency domain as shown in Fig. 1. Note that the DFT coefficients are complex. In this paper, we simply replace $\|\boldsymbol{z}\|_1^1$ by $\|\mathrm{Re}(\boldsymbol{z})\|_1^1 + \|\mathrm{Im}(\boldsymbol{z})\|_1^1$, where $\mathrm{Re}(\boldsymbol{z})$ and $\mathrm{Im}(\boldsymbol{z})$ are the real and imaginary parts.

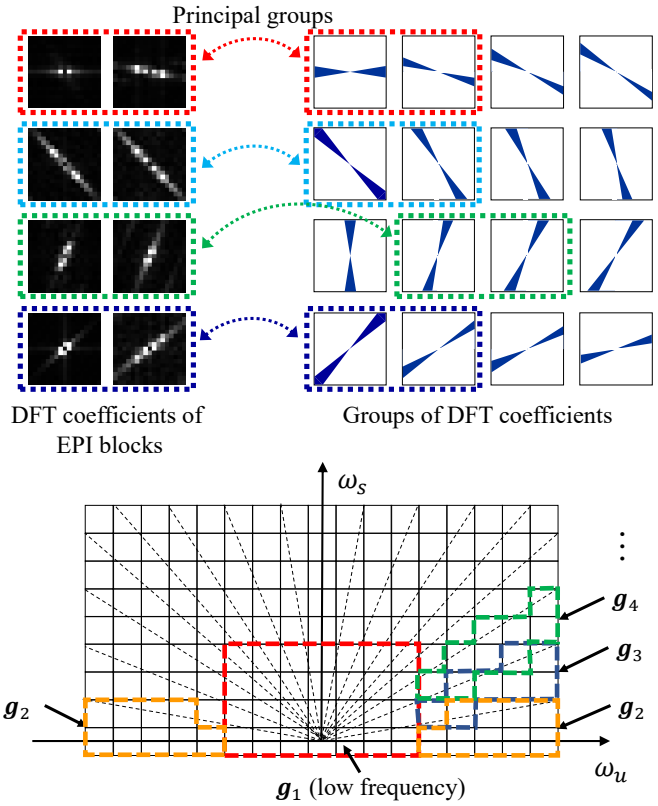


Fig. 2: Group design; (top) concept and (bottom) example.

In this study, the size of processing units is relatively small, e.g., 17×17 pixels, which causes annoying window effects. We found that these window effects disturb the sparseness in the frequency domain. Therefore, to suppress these negative effects, we apply a window function to the observed LF signal $\boldsymbol{y}$ after the mean μ of the non-zero elements of $\boldsymbol{y}$ is subtracted:

$$\boldsymbol{y}' = \boldsymbol{W}(\boldsymbol{y} - \mu), \tag{6}$$

where $\boldsymbol{W} \in \mathbb{R}^{N \times N}$ is a diagonal matrix encoding the Kaiser window function, whose tails are set to non-zero values to make $\boldsymbol{W}$ invertible. Instead of $\boldsymbol{y}$ itself, $\boldsymbol{y}'$ is used as the observation signal for reconstruction, from which the optimal $\hat{\boldsymbol{z}}$ is derived. Consequently, the interpolated LF is obtained as:

$$\hat{\boldsymbol{x}} = \boldsymbol{W}^{-1}\boldsymbol{A}\hat{\boldsymbol{z}} + \mu. \tag{7}$$

3.2. Designing Group and Group Weights

As shown in Fig. 1, the line structure in the spatial domain leads to sparse coefficients in the DFT domain. Moreover, these coefficients mostly concentrate on a line passing through the origin whose direction is perpendicular to that of the lines in the spatial domain. On the basis of this observation, we design the group and group weights that are suitable for representing EPIs.

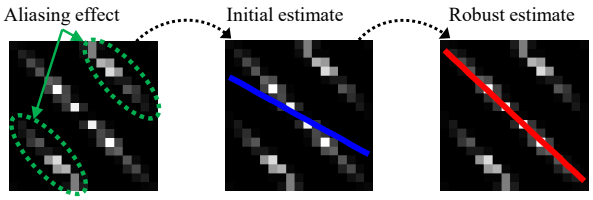


Fig. 3: Estimating direction in frequency domain.

Figure 2 (top) illustrates the basic concept of our group design, where the DFT coefficients are divided into a set of groups in accordance with the directions. It is expected from the aforementioned observation that the non-zero DFT coefficients will exist in only a few groups. In other words, the DFT coefficients will be group sparse over the set of groups mentioned above. This concept is practically implemented as follows. First, we allow the coefficients to overlap between the adjacent groups because the coefficients are defined over only the discrete set of frequencies while the direction is continuous in nature. Second, the lower frequency components are gathered in a group regardless of the directions because these components are likely to exist in any EPI. Finally, we only consider the upper hemisphere in the frequency domain because the DFT coefficients that are symmetric with respect to the origin are complex conjugate. An example of the groups is shown in the bottom of Fig. 2.

The group weights w_i in Eq. 4 are also designed to promote the concentration of the DFT coefficients on a specific direction. We compute w_i using the difference between the direction of each group θ_i and a pre-estimated direction $\hat{\theta}$:

$$w_i = \rho \sin(\,|\hat{\theta} - \theta_i|\,) \quad i = 2, \ldots, |\mathcal{G}| \tag{8}$$

$$\theta_i = \frac{(i-2)\pi}{|\mathcal{G}| - 1}, \tag{9}$$

where ρ is a non-negative amplitude factor. We experimentally determined $\rho = 10$ and $w_1 = 1$. The pre-estimated direction $\hat{\theta}$ is computed from the sub-sampled EPI signal $\boldsymbol{y}'$ given as the input. The derivation details of $\hat{\theta}$ will be mentioned in Sec. 3.3.

3.3. Pre-estimation of Direction

An EPI block is originally defined over the 2D space as $f_{k,l}$ with the discrete spatial coordinate (k, l) $(0 \leq k, l < N)$. The DFT coefficient of $f_{k,l}$ is denoted as $F_{m,n}$, where $-\lfloor \frac{N}{2} \rfloor \leq m, n \leq \lfloor \frac{N}{2} \rfloor$. We want to find a line equation $n = m \tan\theta$ in the DFT domain that fits the distribution of $F_{m,n}$. Specifically, we solve the weighted least square problem as:

$$\hat{\theta} = \arg\min_{\theta} \sum_{m,n} \Psi_{m,n} ||n - m\tan\theta||_2^2. \tag{10}$$

The weight term $\Psi_{m,n}$ is designed as:

$$\Psi_{m,n} = |F_{m,n}|^2 \cdot \left(\frac{1 - \cos(2\pi\sqrt{m^2+n^2})}{2} \right)^2 b_{m,n}. \tag{11}$$

Algorithm 1 LF interpolation using group sparsity prior

Inputs: sub-sampled LF signal $\boldsymbol{Y}$

Parameters: $\boldsymbol{\Phi}$, $\boldsymbol{A}$, $\mathcal{G}$, $\boldsymbol{W}$, λ, η

Extract EPI block $\boldsymbol{y} \in \mathbb{R}^N$ from the input $\boldsymbol{Y}$

for all EPI blocks $\boldsymbol{y}$ **do**

Pre-processes:

- Compute means compute μ from $\boldsymbol{y}$
- Compute input EPI blocks: $\boldsymbol{y}' = \boldsymbol{W}(\boldsymbol{y} - \mu)$

Main processes:

- Determine group weight $\boldsymbol{w}$ from $\boldsymbol{y}'$ (Eq. 8)
- Compute optimal coefficient by solving Eq. 4: $\hat{\boldsymbol{z}} \leftarrow \mathrm{overlappingGroupLasso}(\boldsymbol{y}', \boldsymbol{\Phi}\boldsymbol{A}, \lambda, \eta, \mathcal{G}, \boldsymbol{w})$

Post-process:

- Reconstruct the EPI block $\hat{\boldsymbol{x}} \leftarrow \boldsymbol{W}^{-1}\boldsymbol{A}\hat{\boldsymbol{z}} + \mu$

end for

Reconstruct LF signal $\boldsymbol{X}$ from EPI blocks $\hat{\boldsymbol{x}}$

Output: interpolated LF signal $\boldsymbol{X}$

The first term is the energy of the coefficient $F_{m,n}$ obtained from the sub-sampled EPI $\boldsymbol{y}'$. The second term reduces the weight for higher frequency components to suppress the effect of aliasing, which is caused by the sub-sampling. Exceptionally, the second term is set to 0 when $m^2 + n^2 > \lfloor \frac{N}{2} \rfloor$. The third term is the weight for robust estimation [22], which is iteratively updated using the previous estimation of $\hat{\theta}$ as:

$$b_{m,n} = \begin{cases} \left(1 - (d_{m,n}/\kappa)^2\right)^2 & |d_{m,n}| \leq \kappa \\ 0 & \text{otherwise} \end{cases} \tag{12}$$

$$d_{m,n} = \frac{|n - m\tan\hat{\theta}|}{\sqrt{m^2+n^2}}, \tag{13}$$

where κ is a positive constant. The overview of direction estimation is shown in Fig. 3.

3.4. Overview of Light Field Interpolation

To conclude, the entire procedure of our method is described in Alg. 1. Our method consists of block-wise operations because it is designed for relatively small processing units (2D blocks). Therefore, an EPI block is extracted from the entire LF $\boldsymbol{Y}$, and then it is interpolated using our method. Finally, it is written back to the corresponding position of the target LF $\boldsymbol{X}$. These blocks are made to overlap each other, and on the target LF, the interpolation results are merged using weighted averages, where the pixel-wise weights given by the window function are considered.

4. EXPERIMENTAL RESULTS

We verified the performance of our method using Stanford light field datasets [23] as the input LFs. Each dataset has 17

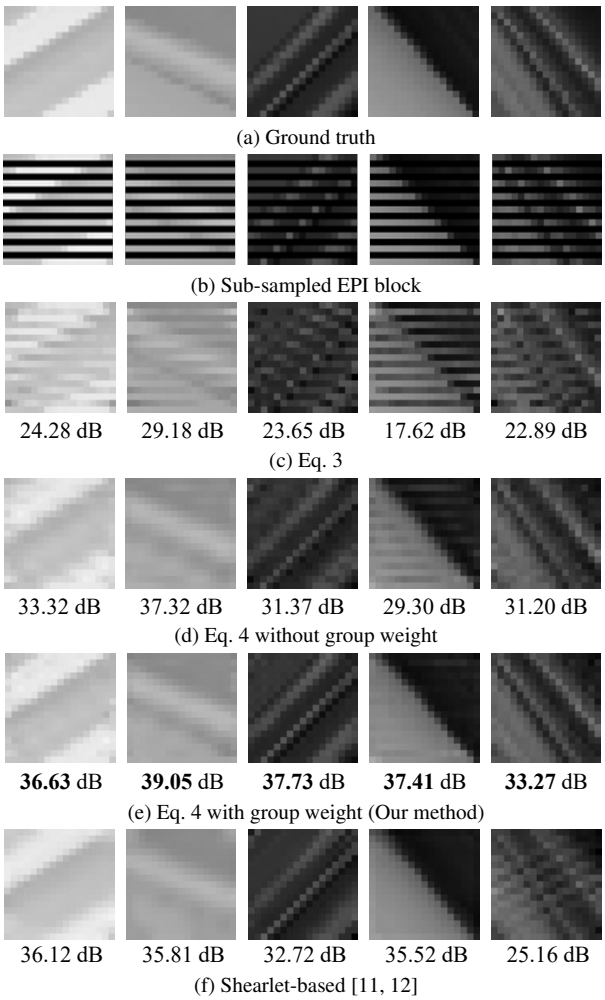


Fig. 4: Interpolation in the EPI domain. The PSNRs are evaluated in 13×13 of the center.

horizontal views (we used the central horizontal line in each dataset), from which views are alternatively sub-sampled to produce the input LF with 9 views. We interpolated the sub-sampled LF and compared it against the original data using the peak signal-to-noise ratio (PSNR).

The processing units were 17×17 EPI 2D blocks. To implement our method, we divided the DFT coefficients into 33 groups, where $\boldsymbol{g}_1$ is allocated to the 7×7 low frequency component around the DC component, and the remaining 32 groups are assigned to directional high frequency components. We compared our method with the shearlet-based method [11, 12], which was implemented using the MATLAB code of ShearLab [24, 25]. The formulation of this method is equivalent to Eq. 3, where $\boldsymbol{A}$ is the shearlet frame. We found that using the pre/post-processings of Eqs 6 and 7 significantly improves the accuracy of both methods. Therefore, we applied these processings to both of them. We sought the best parameters for both of the methods and used them. We used SLEP library [26] to solve Eq. 4.

Figure 4 shows the results of interpolation for several EPI blocks. The ground truth and sub-sampled EPI blocks are

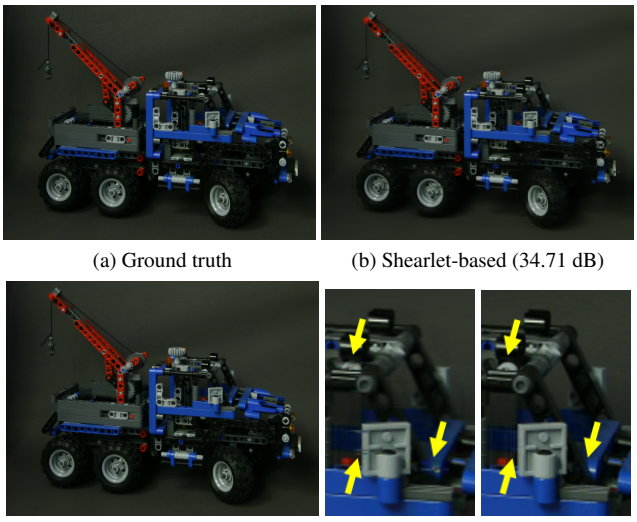


Fig. 5: Interpolated views.

Table 1: Average PSNR for interpolated views.

Dataset	Shearlet-based	Our method
Bunny [23]	31.14 dB	**39.89** dB
Chess [23]	27.61 dB	**38.15** dB
Flower [23]	34.14 dB	**37.31** dB
Knights [23]	26.27 dB	**30.02** dB
Tarot [23]	21.62 dB	**32.04** dB
Truck [23]	34.39 dB	**39.72** dB

shown in Figs. 4 (a) and (b). Figures 4 (c)–(e) are presented to show the contributions of the group sparsity and group weights to the interpolation accuracy of our method. Figure 4 (f) shows the results of the shear-based method, which are inferior to those of our method. Figure 5 shows the examples of the resulting views that were missing in the input LFs but interpolated by our method and the shearlet-based method. The shearlet-based method suffers from artifacts, while they are suppressed in our method. Table 1 summarizes the performance comparison over six datasets, where in each dataset, the PSNR values are averaged over 8 interpolated views. Obviously, our method is superior to the shearlet-based method.

5. CONCLUSION

Aiming to achieve accurate light field interpolation, we proposed a group sparsity prior in the discrete Fourier transform (DFT) domain that can fully exploit the specific line structure in 2D epipolar plane images. Our experimental results show that the proposed method achieved better quality than a state-of-the-art shearlet-based method. In future work, our group sparsity prior will be extended to the full 4D DFT domain of the light field to better utilize their structure. Moreover, the concept behind our method will be applied to problems other than interpolation, such as compressive sensing, denoising, and super-resolution of light fields.

6. REFERENCES

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

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

[3] R. Ng, M. Levoy, M. Bredif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," in *Stanford University Computer Science Tech Report CSTR*, 2005.

[4] R. Ng, "Fourier slice photography," *ACM Transactions on Graphics*, vol. 24, no. 3, pp. 735–744, 2005.

[5] M. Tanimoto, "Overview of free viewpoint television," *Signal Processing: Image Communication*, vol. 21, no. 6, pp. 454–461, 2006.

[6] D. Lanman, G. Wetzstein, M. Hirsch, W. Heidrich, and R. Raskar, "Polarization fields: Dynamic light field display using multi-layer lcds," *ACM Transactions on Graphics*, vol. 30, no. 6, pp. 186:1–186:10, 2011.

[7] G. Wetzstein, D. Lanman, M. Hirsch, and R. Raskar, "Tensor displays: Compressive light field synthesis using multilayer displays with directional backlighting," *ACM Transactions on Graphics*, vol. 31, no. 4, pp. 1–11, 2012.

[8] K. Takahashi, Y. Kobayashi, and T. Fujii, "Displaying real world light fields using stacked lcds," in *International Display Workshops in conjunction with Asia Display*, 2016, pp. 1300–1303.

[9] A. Levin and F. Durand, "Linear view synthesis using a dimensionality gap light field prior," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010, pp. 1831–1838.

[10] 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*, vol. 34, no. 1, pp. 12:1–12:13, 2014.

[11] S. Vagharshakyan, R. Bregovic, and A. Gotchev, "Image based rendering technique via sparse representation in shearlet domain," in *IEEE International Conference on Image Processing (ICIP)*, 2015, pp. 1379–1383.

[12] E. Sahin, S. Vagharshakyan, J. Mäkinen, R. Bregovic, and A. Gotchev, "Shearlet-domain light field reconstruction for holographic stereogram generation," in *IEEE International Conference on Image Processing (ICIP)*, 2016, pp. 1479–1483.

[13] J.-X. Chai, X. Tong, S.-C. Chan, and H.-Y. Shum, "Plenoptic sampling," in *ACM SIGGRAPH*, 2000, pp. 307–318.

[14] G. Kutyniok and D. Labate, *Shearlets: Multiscale Analysis for Multivariate Data*, Birkhäuser Basel, 2012.

[15] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM Journal on Computing*, vol. 24, no. 2, pp. 227–234, 1995.

[16] N. Simon, J. Friedman, T. Hastie, and R. Tibshirani, "A sparse-group lasso," *Journal of Computational and Graphical Statistics*, vol. 22, no. 2, pp. 231–245, 2013.

[17] L. Jacob, G. Obozinski, and J.-P. Vert, "Group lasso with overlap and graph lasso," in *International Conference on Machine Learning (ICML)*, 2009, pp. 433–440.

[18] L. Yuan, J. Liu, and J. Ye, "Efficient methods for overlapping group lasso," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 9, pp. 2104–2116, 2013.

[19] K. Marwah, G. Wetzstein, Y. Bando, and R. Raskar, "Compressive light field photography using overcomplete dictionaries and optimized projections," *ACM Transactions on Graphics*, vol. 32, no. 4, pp. 1–12, 2013.

[20] O. Johannsen, A. Sulc, and B. Goldluecke, "What sparse light field coding reveals about scene structure," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 3262–3270.

[21] Y. Miyagi, K. Takahashi, M. P. Tehrani, and T. Fujii, "Reconstruction of compressively sampled light fields using a weighted 4d-dct basis," in *IEEE International Conference on Image Processing (ICIP)*, 2015, pp. 502–506.

[22] F. Mosteller and J. Tukey, *Exploratory Data Analysis*, Addison Wesley, 1977.

[23] V. Vaish and A. Adams, "The (new) stanford light field archive," http://lightfield.stanford.edu, 2008.

[24] G. Kutyniok, D. Labate, W.-Q Lim, M. Leitheiser, R. Reisenhofer, and X. Zhuang, "Shearlab," http://www.shearlab.org/.

[25] G. Kutyniok, W.-Q Lim, and R. Reisenhofer, "Shearlab 3d: Faithful digital shearlet transforms based on compactly supported shearlets," *ACM Transactions on Mathematical Software*, vol. 42, no. 1, pp. 5:1–5:42, 2016.

[26] J. Liu, S. Ji, and J. Ye, "SLEP: Sparse learning with efficient projections," http://www.public.asu.edu/j̃ye02/Software/SLEP, 2009.