IMAGE BASED RENDERING TECHNIQUE VIA SPARSE REPRESENTATION IN SHEARLET DOMAIN

Suren Vagharshakyan, Robert Bregovic, Atanas Gotchev

Department of Signal Processing, Tampere University of Technology, Tampere, Finland

ABSTRACT

In this paper we propose a method for reconstructing a densely sampled light field from a given sparse set of perspective views from rectified cameras without an explicit estimation of the scene depth. The desired intermediate views are synthesized by inpainting of epipolar-plane images, utilizing their sparsity in the shearlet domain. For the purpose of shearlet-domain representation, compactly supported shearlets have been constructed using different directional filters for different scales in an attempt to provide better directional selectivity at lower scales. The reconstruction procedure with shearlet-domain sparsity condition is implemented through an iterative thresholding algorithm. The performance of the method is quantified by tests on synthetic and real visual data and compared favorably against depth-image based rendering.

Index Terms— Light field, sparse reconstruction, shearlet, image based rendering

1. INTRODUCTION

Modern image based rendering (IBR) methods are based on two, fundamentally different, approaches. First approach is based on estimating the scene geometry, e.g. in the form of depth map(s), from a given set of images (views) [1], [2], [3] and synthesizing the desired views using the estimated depth maps and the given images [4], [5]. Second approach is based on the light field (LF) concept as introduced by Levoy and Hanrahan [6]. This concept considers each pixel of the given views as a sample of a multidimensional LF function, therefore the view synthesis problem transforms to the problem of continuous LF reconstruction and subsequent interpolation at the desired points, performed with no use of explicit depth estimation. In [7], different kernels for interpolation with the usage of available geometrical information are considered. However, this interpolation technique requires a substantial number of samples (images), as discussed in [8] where Lin and Shum derive precise bounds of the LF sampling.

In order to synthesize novel views without ghosting artefacts based only on linear interpolation one needs to sample the LF such that the disparity between nearby views is less than *one* pixel. Hereafter, we refer to this kind of sampled LF as *densely sampled*. Densely sampled LF provides sufficient information about scene's visual content for all practical image-based applications such as refocused image generation [9], depth estimation [10], [11], novel view generation for free viewpoint television [12] and holographic stereogram [13].

In order to capture a densely sampled LF, the required distance between nearby camera positions can be estimated based on the lower bound of the depth of the scene and the camera resolution. Furthermore, camera resolution should provide enough samples to properly capture highest spatial texture frequency in a scene [14].

In [15], it has been shown that seismic data from limited number of measurements can be efficiently reconstructed by using an inpainting technique based on shearlet-domain representation. We employ this idea and present a method for reconstruction of a densely sampled LF from a given sparse set of views, which requires no explicit depth information. The proposed method is based on a sparse representation in shearlet domain of every decimated epipolar-plane image (EPI) slice of the densely sampled LF. Available data (captured views) can be interpreted as known rows in the EPI's. By applying inpainting technique on every EPI, we can reconstruct all unknown samples of the densely sampled LF. The proposed method enables one to capture the scene with a smaller number of cameras and still be able to reconstruct the densely sampled LF.

2. EPIPOLAR-PLANE IMAGES

Epipolar-plane image was first introduced by Bolles *et al*. in [16]. In comparison with regular photo images, an EPI has a specific and distinct structure, see Fig. 1(b). Any captured point of the scene is revealed in one of the EPIs as a line whose slope relates to disparity and directly depends on the distance of the point from the capturing plane (depth). The intensity over the line is related with the intensity of emanated light from that scene point. Within the pinhole camera model assumption, the disparity is defined as $\Delta d = \frac{f}{z}\Delta t$, where f is the focal distance in pixel size, z is the depth of the point, and Δt is the distance between nearby camera positions (see [14] for more details). The corresponding line slope in the EPI is f/z.

The Lambertian reflectance model (any point in the scene emanates light in every direction with the same intensity) drives the distinct structure of EPI formed by lines with constant intensity distribution. Chai *et al*. presented a spectral analysis of the EPI slices of a LF depending on the scene depth and LF sampling rates in different dimensions [14]. It is interesting to point out that the spectrum of the EPI has a bow-tie type shape. Densely sampled LF guaranties that the spectrum of each EPI is always contained in a region similar to the one highlighted in Fig. 1(d). As shown in [14], the visual information of each depth slice is contained in a line passing through DC component in the frequency domain representation of the EPI. In order to obtain space of functions where EPI data will be presented sparsely, we need to provide an analysis tool for identification and separation of the lines in the frequency domain corresponding to different depth slices. While in spatial domain analysis atoms should be similar to lines with different slopes, their spectrum should have bow-tie type shape, as shown with different colors in Fig. 1(d).

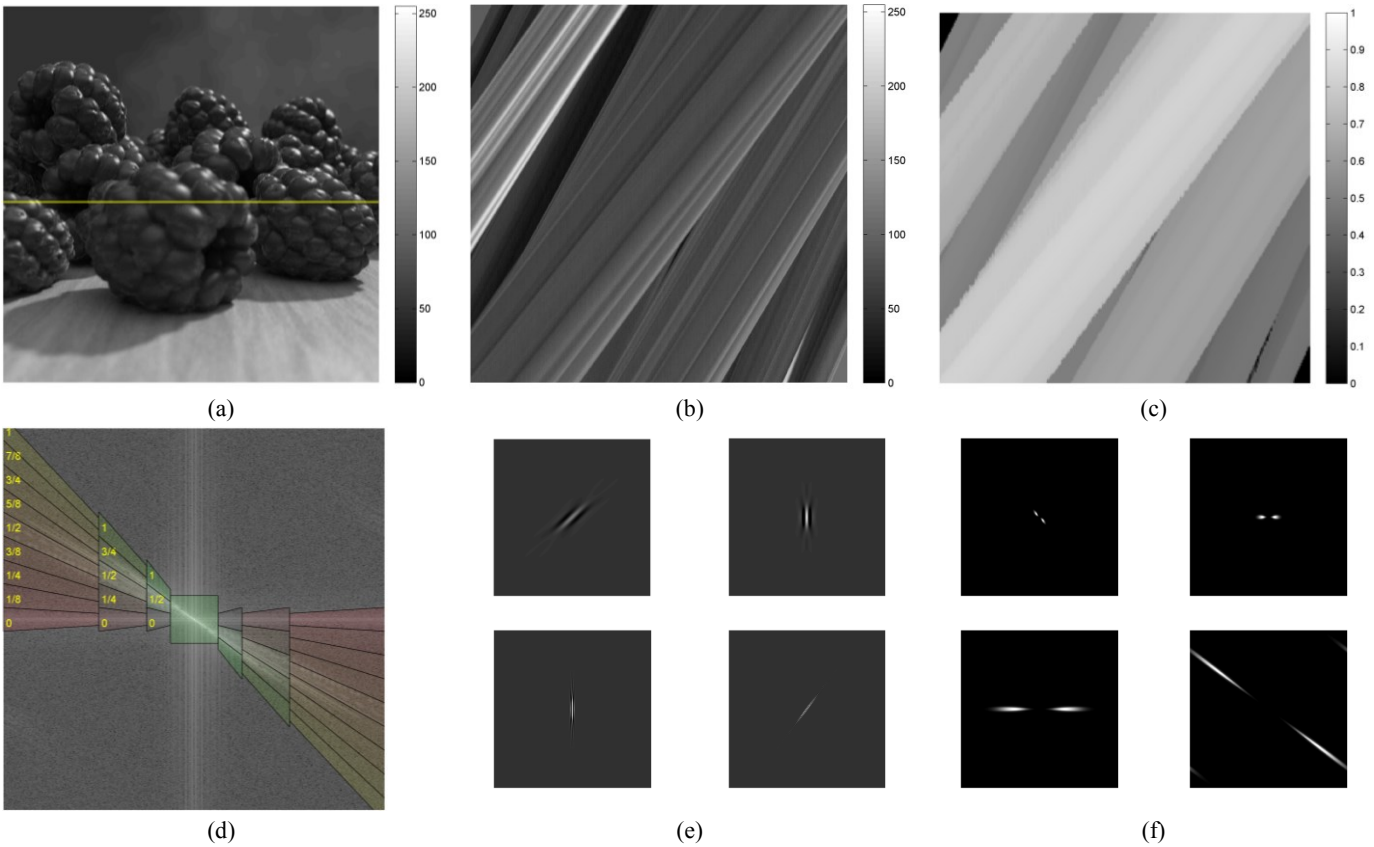


Figure 1. (a) Example of scene image.(b) Example of densly sampled light field EPI corresponding to row highlighted in yellow in (a). (c) EPI of disparity map. (d) Frequnecy domain characteristics of EPI with desirable frequency domain truncation, presented in 3 scales and central low pass filter with disparity values of corresponding shears. (e, f) Example of several constructed sheralet atoms in spatial and frequency domains.

3. SPARSE REPRESENTATION IN SHEARLET DOMAIN

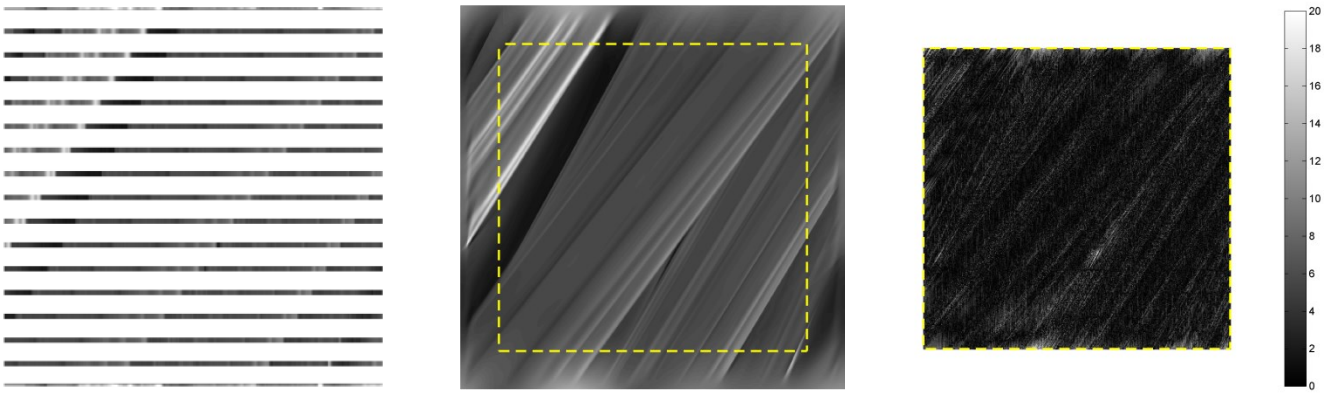
Shearlet frames, as developed in [17], [18], [19], are a perfect tool for the aforementioned sparse representation of the EPI. The elements of shearlet frames are translation-invariant functions whose spectrum covers a region similar to the one presented in Fig 1(f). Shearlet frame is described by number of scales and number of shears (directions) in each scale. An example is the Fast Finite Shearlet Transform (FFST) presented in [17]. FFST consists of a set of atoms that build a tight frame. Those atoms give almost perfect behavior in the frequency domain. However, in the spatial domain non-compact support of the atoms leads to ringing type artifacts. As a result, the approximation quality around the edges, where EPI does not comply with the band limited function condition, is drastically reduced. Another example of basis elements are the so-called compactly supported shearlets, as presented in [18]. Compactly supported shearlets are constructed in spatial domain using scaling and shearing operators. The compact support of the atoms was achieved by slightly changing the behavior in the frequency domain in comparison to atoms of the FFST.

In order to provide good directional properties at lower scales in frequency domain we propose to use different directional filters for different scales in the process of constructing a frame of compactly supported shearlet. Our construction follows the method proposed in [18], [19]. Fig. 1(e, f) presents examples of several constructed frame elements for different scales and shears.

4. RECONSTRUCTION ALGORITHM

We can interpret the set of captured views as given measurements of the unknown densely sampled EPI, as illustrated in Fig. 2(a). The problem tackled in this paper is to find (reconstruct) all missing data in the EPI. In order to simplify the notations, in this paper we assume rectangular size of EPI (in most case the horizontal resolution of the camera is higher than number of cameras, however, the corresponding EPI can be partially processed using overlapping rectangle windows with the size of the number of cameras).

Let $f \in \mathbb{R}^{N\times N}$ be the unknown complete EPI matrix, where each row represents corresponding image row and $g \in \mathbb{R}^{N\times N}$ be incomplete EPI where only rows with available views are presented, while everywhere else is 0. Further, f and g are used in their column-wise reshaped $\mathbb{R}^{N^2}$ vector version with keeping same notations for f and g. Let the mask matrix (measuring matrix) $H \in \mathbb{R}^{N^2\times N^2}$ be $H(i,i) = 1$ if $g(i) \neq 0$ and 0 otherwise. Analysis and synthesis matrix of the shearlet frame will be denoted as $S \in \mathbb{R}^{M\times N^2}$ and $S^* \in \mathbb{R}^{N^2\times M}$, respectively, where $M = \eta N^2$ and η is the number of all shears in all scales of the shearlet.



(a) (b) (c)

Figure 2. (a) Example of the input data for the proposed algorithm, where the original data was decimated by factor 32. (b) Reconstructed EPI, yellow square representing the region which was used for reconstruction quality estimation. (c) Absolute difference between the ground truth and reconstructed EPI.

Reconstruction of missing rows of g can be formulated as an inpainting problem, with prior condition to have sparse solution in the shearlet domain, i.e.

$$\min_{f\in\mathbb{R}^{N^2}} \|Sf\|_0, \text{ subject to } g = Hf \tag{1}$$

It was shown in [20] that the problem (1) can be efficiently solved through the following iterative thresholding algorithm

$$f_{n+1} = S^*\left(H_{\lambda_n}\left(S(f_n + \alpha(g - Hf_n))\right)\right) \tag{2}$$

where $H_\lambda(x) = \begin{cases} x, |x| \geq \lambda \\ 0, |x| < \lambda \end{cases}$, is a hard thresholding operator and α is a chosen relaxation parameter. The thresholding parameter λ_n decreases with the iteration number. Initial value of f_0 can be chosen as 0 everywhere. After sufficient iterations, f_n reaches a satisfying solution for the problem (1). More details can be found in [20], [21], [22], [23].

5. EXPERIMENTAL RESULTS

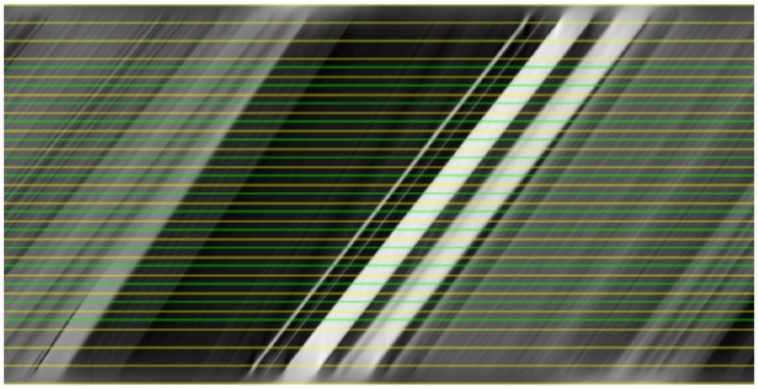
We will illustrate the proposed method on synthetic data as well as on a real-world dataset captured by cameras.

5.1. Synthetic Data

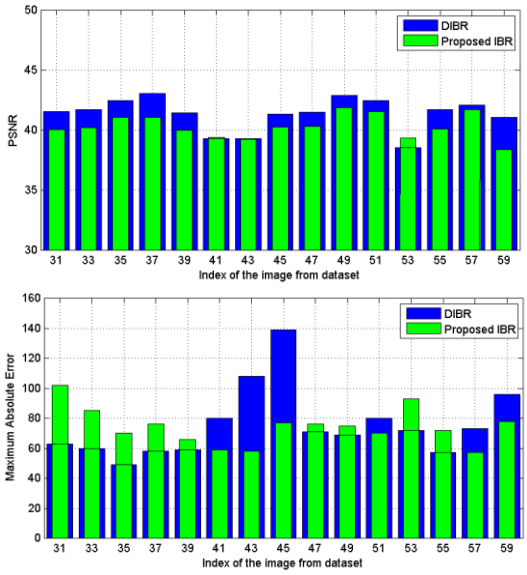
To construct synthetic data we used Blender (open source shareware, www.blender.org). It enables simulating a desired parallel positioned camera capturing system. Our generated synthetic data consists of 511 images with 511×511 resolution. Captured views provide horizontal parallax with disparity values in the range of $[0, 1]$ pixels between views. One of the EPIs generated from the rendered images is shown in Fig. 1(b) with the corresponding frequency domain characteristic in Fig 1(d) and the corresponding ground truth disparity map in Fig 1(c). As an input data for the reconstruction algorithm we use every 32nd view, thus 17 views. An example of the input data for the proposed algorithm is shown in Fig 2(a). In that case the input dataset consist of images with disparity values in the range $[0, 32]$ pixels between two consecutive images. Shearlet frame is constructed using 6 scales and a central low pass filter. In each scale from low to high we have $[2, 3, 5, 9, 17, 33]$ shears respectively. Each set of shears for fixed scale uniformly covers the $[0,1]$ range of disparities. Example of a similar separation (fewer scales) is illustrated in Fig 1(d). Shearlet is a translation-invariant frame thus its synthesis and analysis transforms are easy to implement using convolution operator. Convolution implemented through Fourier transform implicitly assumes circular replication of the signal. This increases the undesirable border effects and decreases the algorithm performance around image borders. In this paper, a Kaiser window is used to reduce these border effects. In Fig. 2(a) example of sparse EPI (input data) is presented, Fig. 2(b) shows the corresponding reconstructed result and Fig. 2(c) shows the residual calculated only over the region within the yellow rectangle. In the presented case, the mean-square-error (MSE) is 8 and the mean-absolute-error (MAE) is 25. Both are calculated with respect to the ground truth data. This example shows that by using the proposed method, a densely sampled LF can be reconstructed by using only a small number of captured views.

5.2. Real Data

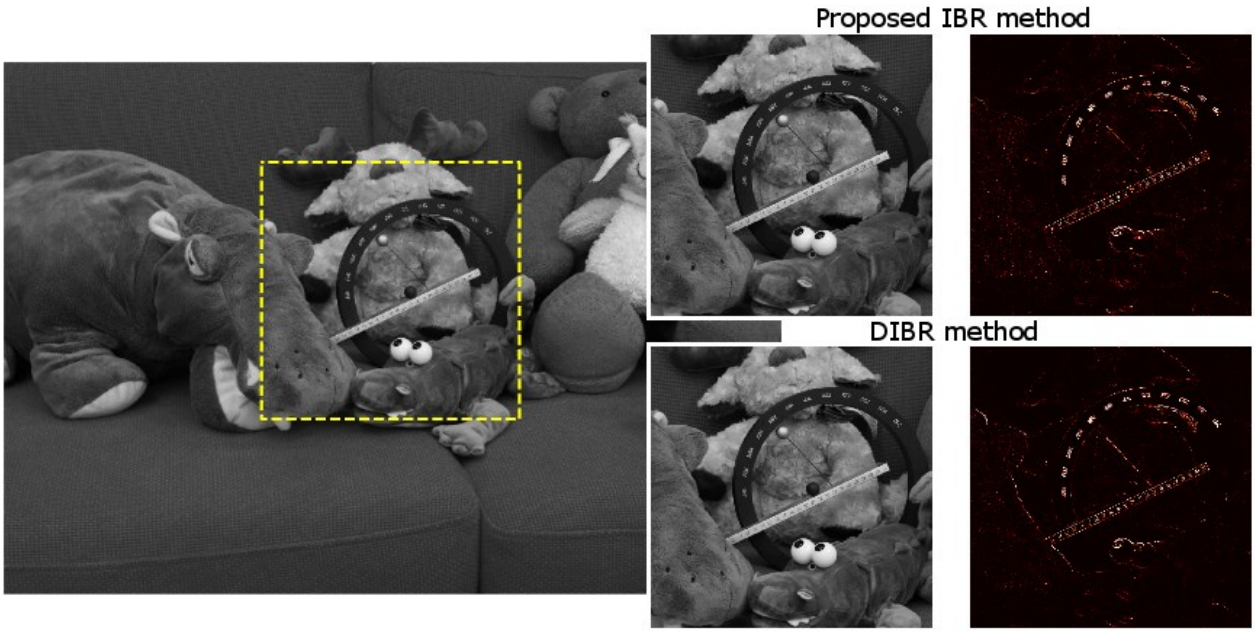
As a real captured dataset we use the "Couch" dataset used in [3]. It consists of 101 images with 2679x4020 resolution as well as 51 estimated disparity maps for the central views obtained by the algorithm proposed in [3] using the whole set of images. Given disparity estimation shows that maximal disparity between consecutive images is about 11px. We applied the presented algorithm to the grayscale images. 15 views were reconstructed using the odd number indexed views from the dataset. An example of a reconstructed EPI is presented in Fig. 3(a), where input (selected) rows for the reconstruction algorithm are highlighted in yellow and rows in green represent views used for assessing the algorithm performance. Same input data is used for depth image based rendering algorithm based on 3D warping and blending implemented as described in [5]. The reconstruction quality of the two algorithms is presented in Fig. 3(b, c). As seen in the figure, both approaches result in reconstructed images with good PSNR with respect to reference captured images whereas the proposed algorithm has in average a lower maximum absolute error.



(a)



(b)



(c)

Figure 3. (a) Reconstructed EPI for the real dataset. Rows which are highlighted with yellow color represent odd indexed views from original data set which were used as an input data for the algorithm and green color represents even indexed views from original dataset which are used for algorithm quality estimation. (b) Evaluation of intermediate views reconstruction in PSNR(top) and MAE (bottom) for regular depth image based rendering (DIBR) algorithm and prosposed algorithm. (c) Ground truth image from dataset(left), reconstruction results for highlighted part of the image and absolute differences between the ground truth and reconstructed images for DIBR (bottom) and proposed algorithm (top).

6. CONCLUSION

In this paper we presented a method for reconstructing densely sampled LF from a given sparse set of views by processing the corresponding EPI images in shearlet domain. We have shown, by using synthetic and real data examples, that the proposed method is very effective in reconstructing densely sampled LFs out of small number of given views. The strength of the proposed method lies in its ability to reconstruct the complete dataset (whole LF) at ones in comparison with classical IBR techniques where each view has to be reconstructed individually. The proposed method establishes a new approach for LF interpolation.

7. REFERENCES

[1] A. Gelman, P. L. Dragotti and V. Velisavljevic, "Multiview image compression using a layer-based representation," in *17th IEEE International Conference on Image Processing (ICIP),* 2010, pp. 493-496.

[2] J. Berent, P. L. Dragotti and M. Brookes, "Adaptive layer extraction for image based rendering," in *Multimedia Signal Processing, 2009. MMSP'09. IEEE International Workshop on,* 2009, pp. 1-6.

[3] C. Kim, H. Zimmer, Y. Pritch, A. Sorkine-Hornung and M. H. Gross, "Scene reconstruction from high spatio-angular resolution light fields." *ACM Trans. Graph.,* vol. 32, pp. 73, 2013.

[4] C. Fehn, "Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV," in *Electronic Imaging 2004,* 2004, pp. 93-104.

[5] M. Li, H. Chen, R. Li and X. Chang, "An improved virtual view rendering method based on depth image," in *Communication Technology (ICCT), 2011 IEEE 13th International Conference on,* 2011, pp. 381-384.

[6] M. Levoy and P. Hanrahan, "Light field rendering," in Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques, 1996, pp. 31-42.

[7] S. J. Gortler, R. Grzeszczuk, R. Szeliski and M. F. Cohen, "The lumigraph," in *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, 1996, pp. 43-54.

[8] Z. Lin and H. Shum, "A geometric analysis of light field rendering," *International Journal of Computer Vision,* vol. 58, pp. 121-138, 2004.

[9] R. Ng, "Fourier slice photography," in *ACM Transactions on Graphics (TOG)*, 2005, pp. 735-744.

[10] I. Tosic and K. Berkner, "Light field scale-depth space transform for dense depth estimation," in *2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2014, pp. 441-448.

[11] S. Wanner and B. Goldluecke, "Variational light field analysis for disparity estimation and super-resolution," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 36, pp. 606-619, 2014.

[12] M. Tanimoto, "Overview of free viewpoint television," *Signal Process Image Commun,* vol. 21, pp. 454-461, 2006.

[13] J. Jurik, T. Burnett, M. Klug and P. Debevec, "Geometry-corrected light field rendering for creating a holographic stereogram," in *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW),* 2012, pp. 9-13.

[14] J. Chai, X. Tong, S. Chan and H. Shum, "Plenoptic sampling," in *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques,* 2000, pp. 307-318.

[15] S. Hauser and J. Ma, *Seismic Data Reconstruction Via Shearlet-Regularized Directional Inpainting*, http://www.mathematik.uni-kl.de/uploads/tx_sibibtex/seismic.pdf, accessed 15 May 2012.

[16] R. C. Bolles, H. H. Baker and D. H. Marimont, "Epipolar-plane image analysis: An approach to determining structure from motion," *International Journal of Computer Vision,* vol. 1, pp. 7-55, 1987.

[17] S. Häuser, "Fast finite shearlet transform," *ArXiv Preprint arXiv:1202.1773,* 2012.

[18] P. Kittipoom, G. Kutyniok and W. Lim, "Construction of compactly supported shearlet frames," *Constructive Approximation,* vol. 35, pp. 21-72, 2012.

[19] G. Kutyniok, W. Lim and R. Reisenhofer, "Shearlab 3D: Faithful digital shearlet transforms based on compactly supported shearlets," *ArXiv Preprint arXiv:1402.5670,* 2014.

[20] M. Elad, J. Starck, P. Querre and D. L. Donoho, "Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA)," *Applied and Computational Harmonic Analysis,* vol. 19, pp. 340-358, 2005.

[21] J. Starck and J. M. Fadili, "An overview of inverse problem regularization using sparsity," in *Ieee Icip,* 2009, pp. 1453-1456.

[22] M. Fadili, J. Starck and F. Murtagh, "Inpainting and zooming using sparse representations," *The Computer Journal,* vol. 52, pp. 64-79, 2009.

[23] M. Kowalski, "Thresholding rules and iterative shrinkage/ thresholding algorithm: A convergence study," in *International Conference on Image Processing (ICIP) 2014,* 2014.