SHEARLET-BASED LIGHT FIELD RECONSTRUCTION OF SCENES WITH NON-LAMBERTIAN PROPERTIES

Sergio Moreschini, Robert Bregovic, Atanas Gotchev

Faculty of Information Technology and Communication Sciences
Tampere University, Finland

ABSTRACT

Light fields representing Lambertian scenes are characterized by specific epipolar plane images, which support in Fourier domain is limited by the minimum and maximum scene depth. This feature of the light field has been taken into account when designing the shearlet-based light field reconstruction algorithm. However, this feature does not hold for the case of non-Lambertian scenes. Therefore, the original algorithm performs sub-optimally for such scenes.

In this work, we extend the shearlet based reconstruction algorithm to enable light fields reconstruction for scenes composed of surfaces with non-Lambertian properties. Instead of assuming Lambertianity and the corresponding Fourier-domain support, we estimate the real Fourier-domain support of the given light field of a non-Lambertian scene. Based on the estimated support, we adapt the Shearlet frame to properly span that support. The results demonstrate that the new method yields a faithful reconstruction of densely sampled light fields of non-Lambertian scenes from a small number of captured views.

Index Terms— light field, view-reconstruction, optical flow, non-Lambertian surface, shearlet transform

1. INTRODUCTION

In the pursuit of describing faithfully the world around us researchers make use of the Plenoptic function (PF), a 7D continuous function which formalizes the visual world as a set of light rays impinging from different directions [1]. Because of its complexity, it is typically approximated by its 4D representation referred to as Light Field (LF) [2]. Moreover, as sensing devices are predominantly digital, visual sensing of 3D scenes can be formulated as sampling its LF. In this scenario, a complete reconstruction of the underlying PF from its samples would be necessary to calculate and render any possible view of the scene.

Plenoptic sampling theory has been developed in [4], where the authors performed a spectral analysis of the PF and designed tailored anti-aliased filters that guaranteed the minimum sampling rate for which the Plenoptic function of a Lambertian scene[1] does not show aliasing. This work was extended in [5] and [6] to include non-Lambertian scenes. In [5], Zhang and Chen performed a detailed analysis of the spectrum of the PF including examples for non-Lambertian scenes and scenes with occlusions showing that effects such as depth variation, occlusions and non-Lambertianity spread the spectrum, which can be analyzed through an extension of the truncating window. In [6], Do et al. showed that an analysis of scenes with smooth surface is possible through the concept of essential bandwidth. The essential bandwidth is defined as the region in frequency domain where the most of the signal energy resides. This concept can be exploited to reduce the computational burden of frequency-based LF reconstruction algorithms.

One possible approach to reconstruct the PF from its sampled version is by making use of Image-Based Rendering (IBR) systems which have as a main goal the rendering of novel views from a set of sparsely sampled images composing the LF [7]. A very widely used simplification in IBR is the use of the Lambertian reflectance model whose adoption has enabled the development of multiple LF reconstruction algorithms [8], [9]. However, in real life scenarios it is very unlikely to deal with pure Lambertian scenes, therefore it is vital to extend the reconstruction process to non-Lambertian scenes.

There are several LF reconstruction approaches capable of dealing with non-Lambertian surfaces. In [10], Wang et al, performed a *"learnable interpolation"* followed by a restoration process. The latter process, allowed the authors to recover high frequency details that were missing after the first step. The process was implemented by using a 3D Convolutional Neural Network (CNN). In [11], a *"blur-restoration-deblur"* framework is implemented. For every Epipolar Plane Image (EPI[2]), after blurring the input, a bicubic interpolation to upsample the starting EPI to the desired resolution is performed. Once at the correct resolution, the details in the angular dimension of the EPI are restored by use of a CNN.

[1] An object is defined as Lambertian if the intensity of the light ray leaving every point of its surface is the same regardless the viewing angle between the object itself and the observer.

[2] An EPI is a representation which gathers all the information of an LF epipolar plane. It is discussed in Section 2.1.

As a last step, the spatial details of the EPI are recovered by using a non-blind deblur operation. In [12] two sequential neural networks have been employed to model the disparity and color estimation component. In [13], two CNNs have been used for view synthesis and refinement. In [9], after detecting non-Lambertian regions, the author makes use of this information within an optimization framework to retrieve the missing views as an energy minimization problem by exploiting the concept of structure tensor. A different approach is the one proposed in [14] where use is made of the sparsity in the Fourier domain to reconstruct the missing views of the LF. In [15], Gao *et al.* proposed Parallax-Interpolation Adaptive Separable Convolution *(PIASC)* as a method for LF reconstruction. PIASC implements a fine-tuning process by adjusting coefficient values for two 2D convolution kernels that are computed for each pixel in the interpolated intermediate image. As the method is based on video frame interpolation it can reconstruct also non-Lambertian frames. Each convolution kernel is generated by a deep neural network. PIASC is nowadays the method which provides the best results for LF reconstruction.

In contrast to aforementioned methods, we seek a more robust approach which does not require any training or prior knowledge of the scene and provides a good quality for each view in the whole LF. Based on works [16] [17], in which a shearlet-based reconstruction algorithm has been introduced for reconstruction of LFs, in this paper we show that is possible to extend that approach to non-Lambertian scenes. We achieve such result by performing an analysis of the spectrum of a decimated LF in order to reconstruct the essential bandwidth of the LF before decimation.

The rest of this paper is organized as follows: In Section II, the concepts of LF, EPI, shearlet transform and the optical flow are introduced. In Section III, the proposed method to deal with non-Lambertian scenes is presented. In Section IV, two examples are shown and commented. Conclusions are given in Section V.

2. PRELIMINARIES

2.1. Light field formalization

Following the two-plane parameterization [2], each ray composing the LF $L(u, v, s, t)$ can be described by the intersection with two parallel planes as shown in Figure 1 (a). The planes are typically defined as camera plane (s, t) and image plane (u, v), and are distant from each other by a value f equal to the focal length. In order to simplify our notations we are going to take into account a camera which moves only on a single trajectory t (Figure 1 (b)). Such simplification leads to a single horizontal slice of the previously defined LF $H(u, v, t) = L(u, v, s_0, t)$. In this scenario, by stacking all captured images for a single row $u = u_0$ we achieve a representation $E(v, t) = H(u_0, v, t)$ which is referred as EPI

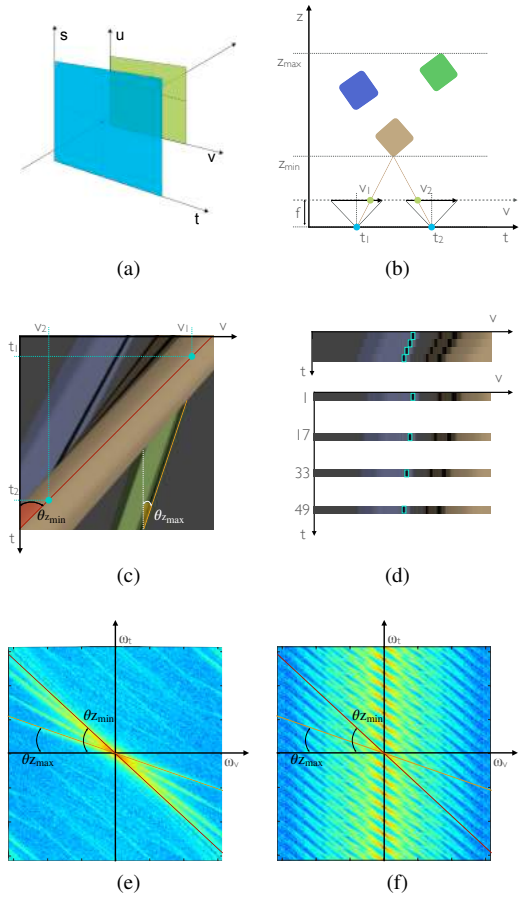


Fig. 1: LF and EPI formalization. (a) Two-plane LF parameterization. (b) Capturing setup. (c) DSEPI. (d) Sparsely sampled EPI. (e) Frequency support of an EPI from a DSLF. (f) Frequency support of a sparsely sampled EPI.

(Figure 1 (c)) [3].

For an ideal horizontal camera motion, like the one depicted in Figure 1 (b), a specific point at distance z from the camera plane generates in the image plane a disparity equal to:

$$\Delta v = v_1 - v_2 = \frac{f}{z}(t_1 - t_2) = \frac{f}{z}\Delta t. \tag{1}$$

To better understand the concept of disparity we can make use of the EPI where every point in space is mapped into a line with slope $-\Delta v/\Delta t$. In this EPI representation, the disparity for a single point between two consecutive images is represented as the horizontal displacement between a point and its occurrence in the adjacent row. When the Nyquist sampling criteria is satisfied, a point at distance z_{min} results in a line with angle $\theta_{z_{min}} \leq 45°$ (red line in Figure 1 (c)). An LF generated respecting this criteria is usually referred to as a densely sampled LF (DSLF) and in particular, an EPI related to a DSLF is defined as densely sampled EPI (DSEPI). Following the same calculation, an object placed at the maximum

distance shows a slope with angle $\theta_{z_{max}} \leq \theta_{z_{min}}$ (yellow line in Figure 1 (c)). Furthermore, when the Lambertian reflectance model is satisfied such lines have a constant value of intensity.

2.2. Frequency domain analysis

For Lambertian scenes, the spectrum of the DSEPI from Figure 1 (c) is shown in Figure 1 (e). As demonstrated in [4], the spectral support of a DSEPI is limited by the minimum and maximum depth of the scene. On the contrary, if we decimate the rows from the DSEPI by a factor 16, as illustrated in Figure 1 (d), we would generate a condition for which the Nyquist criteria is not satisfied. When the Nyquist criteria is not satisfied, the spectrum of the EPI consist of multiple replicas of the real spectrum that are generated because of the discrete sampling in t. An example is illustrated in Figure 1 (f).

In the case of Non-Lambertian scenes, the spectrum of the DSEPI is more widespread than the one of a Lambertian scene, since the limits of the spectral support are no more the minimum and maximum physical depth of the scene, but those generated by the non-Lambertianities. This condition will be further illustrated and discussed in Section 3.

2.3. Shearlet Reconstruction

Our goal is to reconstruct the original spectrum (Figure 1 (e)) out of the available aliased spectrum (Figure 1 (f)). This is equivalent to reconstruct the missing rows for a DSEPI and, consequently, the missing views of a DSLF. One way to do so is through a shearlet-based approach [16]. Starting from an aliased representation, we can reconstruct the DSEPI E^* by solving a sparsification problem. We assume that E^* is a squared image such as $E^* \in \mathbb{R}^{N^2}$, where $N = (R-1)d_{max}+1$. R is the number of available rows in the EPI and

$$d_{max} = \frac{f}{z_{min}} \Delta t_d \tag{2}$$

is the maximum disparity, in pixels, between nearby available views at distance Δt_d. In the EPI domain such available views corresponds to rows $E \in \mathbb{R}^{N^2}$, which are indexed as: $E(i,l) = H(i,l)E^*(i,l)$, where $H \in \mathbb{R}^{N^2} : H(rd_{max}, \cdot) = 1, r = 1, ..., R$ and 0 elsewhere. The main tool employed for retrieving the missing rows is a variation of a specific kind of shearlet system known as: cone-adapted shearlet system [18]. The reconstruction problem is solved as an inpainting problem with solution which needs to be sparse in the shearlet domain, i.e.:

$$x^* = \underset{x \in \mathbb{R}^{N^2}}{\arg\min} ||S(x)||_1, \text{subject to } y = Hx. \tag{3}$$

The direct and inverse shearlet transforms are defined as $S : \mathbb{R}^{N^2} \rightarrow S : \mathbb{R}^{N^2 \times \eta}$ and $S^* : \mathbb{R}^{N^2 \times \eta} \rightarrow S : \mathbb{R}^{N^2}$, where η is the number of the generated shears for all possible scales. Such shears are divided in different scales which are selected as: $J = \lceil \log_2 \Delta d \rceil$, where $\Delta d = d_{max} - d_{min}$. This makes Δd the most important parameter to compute in order to correctly describe the properties of the scene. Therefore, we are going to focus on retrieving d_{min} and d_{max}.

2.4. Optical Flow

Since disparity is essential for our approach, we need to estimate it. More specifically, our target is to compute the minimum and maximum disparity at which every pixel, including those related to the non-Lambertianities, varies between two adjacent images in the LF. This can be achieved by exploiting the Optical Flow (OF) between LF views.

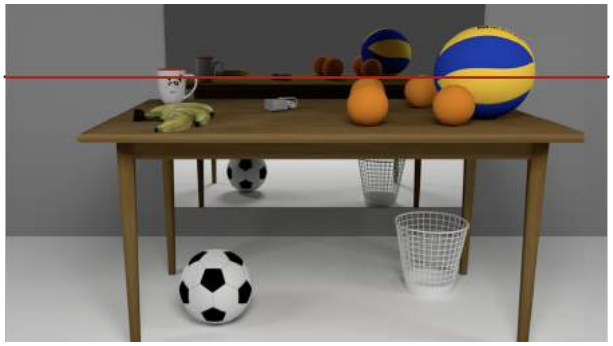
The OF represents an approximation of the 2D motion field generated from the projections of the 3D velocities of the surfaces points. We interpret the values obtained from the OF as disparities. Most of the OF algorithms have usually three stages of processing in common. First, a prefiltering is performed, second, basic measurements are extracted, third, measurements are integrated [19]. In this work, after computing the minimum and maximum disparities we arrange them in a histogram generated by taking the disparity values for every pixel in every successive couple of images. A vital part of this approach is the outliers removal, performed by exploiting the 3-σ rule [20]. In our case this consists in taking the 99.7% of disparity occurrences between the minimum and the 0 disparity value, and the 0 and the maximum. The peculiarity of this approach is the ability of the algorithm to retrieve both positive and negative values of disparity in a single iteration.

Modern disparity algorithms [9] prove that it is possible to compute accurately d_{min} and d_{max} in the scene reducing the probability of obtaining outliers to less than 1%. The achieved parameters are those that we define as $d_{min}^{(nL)}$ and $d_{max}^{(nL)}$ for non-Lambertian scenes and we use these as the input for the reconstruction process. By doing so, we tailor our tool to the frequency response of the scene.

3. PROPOSED RECONSTRUCTION METHOD FOR NON-LAMBERTIAN SCENES

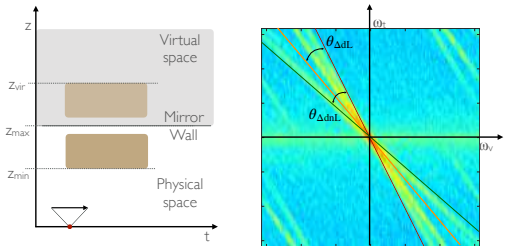
Lambertian objects, are very rare in real world scenarios. Usually the light rays reflected on real surfaces tend to show small variations based on the viewer's position. A fundamental task of the reconstruction algorithm is to take into consideration these variations. In order to illustrate the motivation behind our proposed method we make use of an example that includes a highly reflective object: a mirror.

Following the notation introduced in the previous section, we assume a capturing device moving over the t axis, at a distance z_{min} from the first object and at a distance z_{max} from the farthest object, a mirror mounted on a wall. The scene under consideration is shown in Figure 2 (a) and its geometry



(a)

(b)



(c) (d)

Fig. 2: Non-Lambertian scene composed of a highly reflective surface. (a) Central view. (b) EPI corresponding to the row represented in red in Fig 3 (a). (c) Geometry of the scene. (d) Spectrum of the targeted DSEPI.

in Figure 2 (c). As the scene is part of a DSLF, we illustrate in Figure 2 (d) the spectrum from one of the EPIs, shown in Figure 2 (b) where lines of both positive and negative slopes θ are present. Starting from the given distance values z_{min} and z_{max}, by making use of (2) it is possible to calculate $d_{min}^{(L)}$ and $d_{max}^{(L)}$ in the scene. Those estimates are used in the original shearlet reconstruction algorithm [16] to determine the reconstruction area. As illustrated in Figure 2 (d), in Fourier domain this area is a cone delineated by a red line proportional to $d_{max}^{(L)}$ and an orange line related to $d_{min}^{(L)}$. The aperture of this cone is an angle proportional to the Lambertian disparity of the scene: $\theta_{\Delta dL} \propto d_{max}^{(L)} - d_{min}^{(L)}$. Obviously, from this figure it is clear that by performing such reconstruction (inside such cone) we would not cover the whole spectrum. Therefore, we propose a modified approach which goes beyond the geometric interpretation of the scene, which was defined in [16].

The proposed algorithm performs the reconstruction by taking into account the Fourier domain support of the LF. This support relies on the estimation of the correct disparity range. The algorithm can be divided in 7 steps.

1. Estimate disparities between consecutive views by computing the OF between the views. Arrange the estimated disparities in a histogram.

2. Remove outliers in the disparity histogram by taking the 3-σ confidence interval.

3. Generate the shearlet frame to perform the reconstruction composed of $J = \lceil \log_2 \Delta d_{nL} \rceil$ levels, where $\Delta d_{nL} = d_{max}^{(nL)} - d_{min}^{(nL)}$. Here Δd_{nL} corresponds to the disparity range for non-Lambertian scenes. These atoms, introduced in Section 2.3, are those necessary for the reconstruction algorithm in point 7.

4. Shear the EPI is sheared such that the area showing the minimum slope is aligned with angle $\theta = 0°$. This is the step that enables the non-Lambertian reconstruction: it performs a rotation of the signal in frequency such that the signal area gets covered by the frequency support of the cone-adapted shearlet system.

5. The rows composing the EPI are spaced apart by Δd rows and 0-padded in between such that the final number of rows in t is equal to $N_t = (R-1)\Delta d_{max}^{(nL)} + 1$.

6. Perform the accelerated shearlet-based reconstruction algorithm [17] using the parameters as described in the steps above.

7. Shear the DSEPI to its original position to compensate for the shearing performed in step 4.

Considering once again Figure 2 (d), the new reconstruction area is equal to $\theta_{\Delta dnL} \propto \Delta d_{nL}$.

4. EXAMPLES

The scene introduced in the previous section and illustrated in Figure 2 (a) is a Horizontal Parallax Only (HPO) scene composed of 97 images generated with Blender [21], 12 of those images (decimation by 8) have been used as input for reconstruction. The reconstructed images have been compared with the state-of-the-art algorithm for LF reconstruction [15]. The means of comparison were minimum and average Peak Signal-to-Noise Ratio (PSNR). We define the minimum PSNR as the minimum value achieved when comparing the RGB channels of each reconstructed view with the ground truth (GT). Similarly, the average PSNR is the average value of all the PSNR achieved when comparing the reconstructed views with the GT-available views. The results of the reconstruction process are shown in Table 1.

Results show that the proposed algorithm performs better than state-of-the-art. Figure 3 presents a visual comparison of the results presented in Table 1. In the figure multiple cutouts are presented; it is important to specify that the results presented in the figure are related to different views as

	ST [17]	$PIASC$ [15]	Proposed ST
Min PSNR.	32.66	40.94	41.40
Av. PSNR.	35.96	43.54	43.99

Table 1: LF reconstruction quality presented by minimum and average PSNR in dB for the dataset: *Mirror*.

	ST [17]	$PIASC$ [15]	Proposed ST
Min PSNR.	20.37	20.11	20.78
Av. PSNR.	25.09	25.31	25.82

Table 2: LF reconstruction quality presented by minimum and average PSNR in dB for the dataset: *Tarot Cards and Crystal Ball*.

the minimum PSNR is not achieved for the same view. Figure 3 (a), (b) and (c) depict part of the reconstructed views (the reflection of the basket), while Figure 3 (d), (e) and (f) represent the difference in the Y-channel between the reconstructed view and the GT. The figures show that the main difference in the reconstruction result is in areas composed of repetitive patterns which are occluded in parts of the LF. In this case the shearlet based algorithm overcomes the PIASC algorithm as the latter does not rely on all the available images. As a second mean of comparison we used datasets from [22]. In particular, we used the dataset known as: *Tarot Cards and Crystal Ball (large angular extent)*. The reason behind our choice are the multiple reflections and the see-through effect given by the crystal ball. The original dataset is composed of 289 views on a 17x17 grid with image resolution 1024x1024. We perform our test by following the same procedure in [16] where the dataset has been resized to half of its original resolution. The decimation factor for both the horizontal and the vertical direction has been chosen as 4. The reconstructions have been performed by recreating missing views from those available along rows at first, then views from columns are reconstructed and as a last step the views from the rows which were not available at the beginning are reconstructed. The results of the reconstruction process are shown in Table 2 and depicted in Figure 4. It should be noted, similar to the previous case the minimum PSNR is achieved for different views therefore, different views are depicted in Figure 4.

5. CONCLUSIONS

In this paper we extended the shearlet based reconstruction algorithm to non-Lambertian scenes. By analyzing the spectrum of the EPI we demonstrated that it is possible to perform LF reconstruction using the shearlet-based algorithm in non-Lambertian scenes when the properties of the objects in the scene are properly taken into account.

Instead of computing the disparities based on the physical distances of the objects, we estimated the real Fourier-domain support of the LF of a non-Lambertian scene by making use of the OF.

During the experiments we compared the result achieved by the new algorithm against the state-of-the-art algorithm for LF reconstruction. To this goal we made use of two datasets: in the first case the synthetic scene has been generated to exemplify the concepts behind the introduced method, while in the second case we chose a well known problematic dataset for reconstruction from [22]. In our experiments the proposed method performs better in both cases.

6. ACKNOWLEDGMENT

The work in this paper was funded from the European Union's Horizon 2020 research and innovative program under the Marie Sklodowska-Curie grant agreement No. 676401, European Training Network on Full Parallax Imaging.

7. REFERENCES

[1] E. Adelson and J. Bergen, "The plenoptic function and the elements of early vision." *Computational Models of Visual Processing,* vol. 1, MIT Press, 1991.

[2] M. Levoy and P. Hanrahan. "Light field rendering." *Proc. 23rd Annu. Conf. Compu. Graphics Interactive Techn.,* 1996, pp. 31-42.

[3] R. Bolles, H. Baker and D. Marimont. "Epipolar-plane image analysis: An approach to determining structure from motion," *Int. J. Comput. Vis.,* vol. 1, no. 1, pp. 7-55, Mar. 1987.

[4] J.-X.Chai, X. Tong, and H.-Y. Shum. "Plenoptic sampling." *Proc. 27rd Annu. Conf. Compu. Graphics Interactive Techn.,* 2000, pp. 307-318.

[5] C. Zhang and T. Chen. "Spectral analysis for sampling image-based rendering data." *IEEE Trans. Circuits Syst. Video Techn. 13,* 2003, pp. 1038-1050.

[6] M. N. Do, D. Marchand-Maillet and M. Vetterli. "On the bandwidth of the plenoptic function." *IEEE Trans. Image Process. 21.2,* 2012, pp. 708-717.

[7] H. Shum, S. Chan and S. Kang, *Image-Based Rendering*. New York: Springer, 2007.

[8] C. Kim, H. Zimmer, Y. Pritch, A. Sorkine-Hornung, and M. Gross, Scene Reconstruction from High Spatio-Angular Resolution Light Fields, *ACM Trans. on Graphics,* vol. 32, no. 4, pp. 112, Jul. 2013.

[9] 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.

[10] Y. Wang, F. Liu, Z. Wang, G. Hou, Z. Sun, and T. Tan, "End-to-End View Synthesis for Light Field Imaging with Pseudo 4DCNN." *Eur. Conf.on Comput. Vision,* Sept. 2018, pp. 340-355. Springer.

[11] G. Wu, Y. Liu, L. Fang, Q. Dai, and T. Chai. "Light Field Reconstruction Using Convolutional Network on EPI and Extended Applications." *IEEE Trans. Pattern Anal. Mach. Intell.,* Jun. 2018,

[12] N. K. Kalantari, T.-C. Wang, and R. Ramamoorthi. "Learning-based view synthesis for light field cameras." *ACM Trans. Graphics,* vol. 35 no. 6, 2016.

[13] H. W. F. Yeung, J. Hou, J. Chen, Y. Y. Chung, and X. Chen. "Fast light field reconstruction with deep coarse-to-fine modelling of spatial-angular clues." *Eur. Conf.on Comput. Vision,* Sept. 2018, pp. 138-154. Springer.

[14] L. Shi, H. Hassanieh, A. Davis, D. Katabi, and F. Durand, "Light Field Reconstruction Using Sparsity in the Continuous Fourier Domain," *ACM Trans. on Graphics,* vol. 34, no.1, 2014.

[15] Y. Gao and R. Koch, "Parallax view generation for static scenes using parallax-interpolation adaptive separable convolution," *IEEE Int. Conf. on Multimedia & Expo Workshop,* pp. 1-4, Jul. 2018.

[16] S. Vagharshakyan, R. Bregovic and A. Gotchev. "Light Field reconstruction using Shearlet transform," *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. 40, no. 1, pp. 133-147, Jan. 2018.

[17] S. Vagharshakyan, R. Bregovic and A. Gotchev. "Accelerated Shearlet-domain light field reconstruction." *IEEE J. Sel. Topics Signal Process.,* vol. 11, no. 7, pp. 1082-1091, Oct. 2017.

[18] G. Kutyniok et al., "Shearlets: Multiscale Analysis for Multivariate Data," *Birkhuser Basel*, 2012.

[19] J. L. Barron, D. J. Fleet, and S. S. Beauchemin. "Performance of optical flow techniques." *Int. J. Comput. Vis.,* vol. 12. no. 1, pp. 43-77, Feb. 1994.

[20] F. Pukelsheim. "The three sigma rule." *The American Statistician,* vol. 48, no. 2, pp. 88-91, 1994.

[21] Blender Foundation, "Free and open source 3D animation suite," 3D rendering software in www.blender.org.

[22] V. Vaish and A. Adams, "The (New) Stanford Light Field Archive," *http://lightfield.stanford.edu*, 2008.

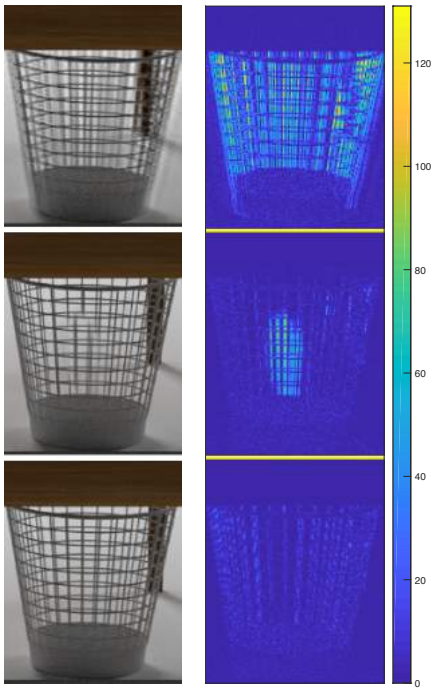


Fig. 3: Comparison of Y-channel for Mirror scene reconstruction. First column: reconstructed cutouts for Shearlet Transform (ST), PIASC, Proposed Shearlet Transform. Second column: Difference between ground truth and first column for the Y-channel.

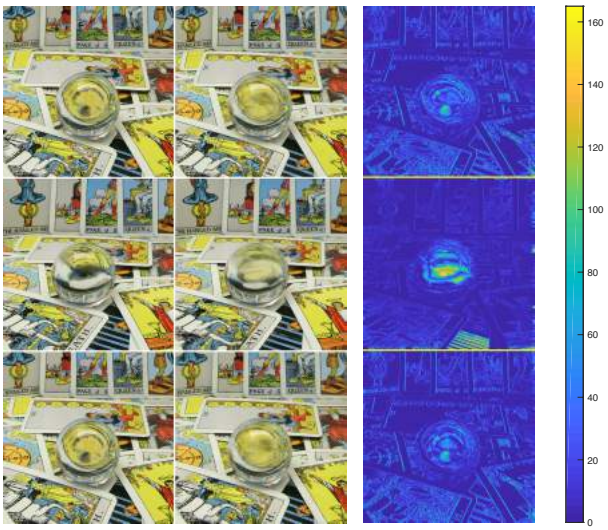


Fig. 4: Comparison for the dataset *Tarot Cards and Crystal Ball*. First column: Ground Truth. Second column: reconstructed views for Shearlet Transform (ST), PIASC, Proposed Shearlet Transform. Third column: Difference between first and second column for the Y-channel.