TAMPERE UNIVERSITY OF TECHNOLOGY

High-quality Light Field Interpolation in Epipolar Plane Image Domain by DCT-based Shearing

Take down policy
If you believe that this document breaches copyright, please contact cris.tau@tuni.fi, and we will remove access to the work immediately and investigate your claim.

High-quality Light Field Interpolation
in Epipolar Plane Image Domain by DCT-based Shearing

Leonid Bilevich[1], Suren Vagharshakyan[2], and Atanas Gotchev[2, a]

[1]*Department of Physical Electronics, Faculty of Engineering,*
Tel Aviv University, 69978, Tel Aviv, Israel.
[2]*Department of Signal Processing, Faculty of Computing and Electrical Engineering,*
Tampere University of Technology, PO Box 527, FI-33101 Tampere, Finland.

Abstract. Light field (LF) is a 4-D function representing the light radiance in free space in terms of rays. A dense grid of cameras, parameterized by two planes, captures an LF: the sensor plane and the focal plane. Multi-camera images stack in volumes, and when sliced, form epipolar plane images (EPI). In EPI, lines with the same slope represent object points at the same scene depth. Thus, the synthesis of new views can be formulated as a problem of interpolation in EPI domain. In this contribution, we propose a three-pass algorithm for LF interpolation in EPI domain. Each EPI is sheared by a factor determined by the dominant slope, employing fast DCT-based fractional shifts. In sheared EPI, dominant lines appear vertical and can be easily interpolated. 'De-shearing' reconstructs EPI's original geometry. We analyze the performance of the proposed LF interpolation in Fourier domain and illustrate it with examples of virtual view synthesis from dense multi-camera imagery.

INTRODUCTION

A modern approach in 3D scene capture and visualization is to deal with the light emanated or reflected by 3D objects. A modern framework for light sensing and reproduction is based on the notation of the plenoptic function [1] and its 4D LF approximation [2]. As light is sensed by discrete sensors, the problem of LF manipulation becomes a problem of reconstruction of a continuous LF from a (relatively) small number of measurements [3], using no explicit geometry [2] or limited geometry information about the scene [4]. In such a framework, the notation of *Epipolar Plane Image (EPI)* plays a key role [5, 6, 7, 8].

In this article, we adopt the LF and EPI formalism and address the LF reconstruction problem making the following assumptions: (a) the light measurements are taken from a set of parallel-optical-axes cameras uniformly positioned on a fixed line; (b) some geometrical information about the 3D scene depth layers is available (as in [5]). Our aim is to develop a technique for synthesizing intermediate camera views based on LF reconstruction. We propose a reconstruction method based on DCT-based shearing and present and discuss important properties of this method along with experimental results demonstrating its effectiveness.

THE LIGHT FIELD AND EPI REPRESENTATION

The plenoptic function, as defined by Adelson and Bergen in [1], describes the structure of light emanated by 3D scenes in terms of collection of rays. It is a multi-dimensional function of seven variables $P(\theta, \phi, \lambda, t, V_x, V_y, V_z)$, where θ, ϕ are azimuthal and polar angles (spherical coordinates) of the viewing direction, V_x, V_y, V_z are Cartesian coordinates of the viewing position, λ is the wavelength of the observed light and t is the time. For practical use, this function can be degraded to what is called 4D LF, assuming a static scene, bounded by a convex hull, which emanates light rays with constant wavelength [2]. Thus, an LF sample represents the intensity along a geometrical ray. One of the possible parameterizations of the LF is the two-plane parameterization [2], where each ray in space is indexed by four coordinates of its intersection with two parallel planes $L(u, v, s, t)$, where (s, t) are the coordinates of the intersection point on the first ("camera") plane, whereas (u, v) are the coordinates of the intersection point on the second ("image") plane relative to (s, t). The LF formalism allows analyzing data captured by multi-camera setups [9] as well as developing novel plenoptic cameras [10]. It is a native and elegant formulation

especially for the case of light sensing by a set of rectified cameras positioned on a fixed line providing horizontal parallax only (fixed $s = s_0$). This kind of acquisition can be described as a discrete sampling of the continuous function $L(u, v, s, t)$ by a set of parallel-(to-z)-optical-axes pinhole cameras positioned on discrete coordinates t_s on. For each pinhole camera position, a 2D image is captured. These images are stacked together along the t-axis and the captured LF data appears as a set of discrete image samples $L(u_l, v_n, s_0, t_m)$, where (u_l, v_n) are located at the vertices of the image plane uniform sampling grid: $u_l = l\Delta u$; $l = 0,1, \dots, L - 1, v_n = n\Delta v$; $n = 0,1, \dots, N - 1$. Assume uniformly placed cameras over the t-axis at a distance Δt between neighboring cameras: the coordinates of the sensing (given) cameras are given by $t_m = m\Delta t$; $m = 0,1, \dots, M - 1$. The intermediate view generation problem can be formulated as reconstruction of the continuous (over t-axis) function $L(u_l, v_n, s_0, t)$ given the available set of measurements $L(u_l, v_n, s_0, t_m)$. Figure 1 demonstrates the LF parameterization for a 3D scene composed of three dices at different depths (Fig. 1(a)). The stack of images along t is shown in Fig. 1(b), and a slice of the stack (an EPI) is shown in Fig. 1(c).
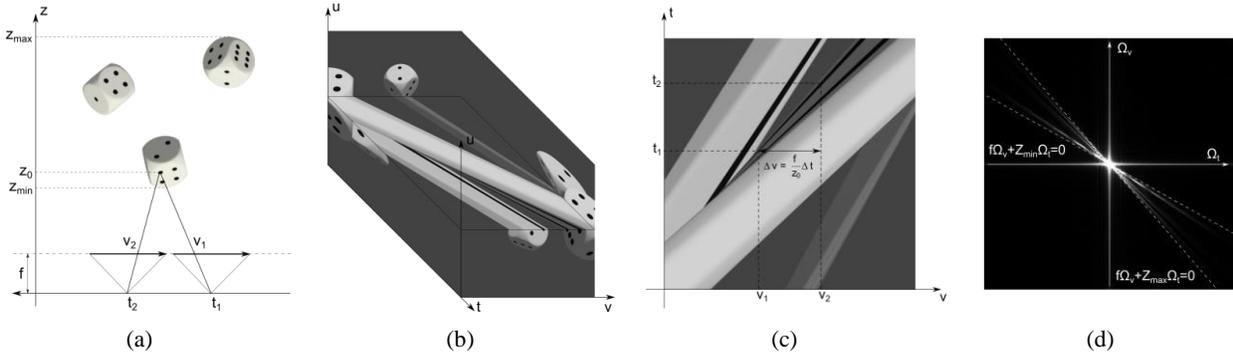

FIGURE 1. (a) Test 3D scene captured by a row of pinhole cameras with focal distance f, where t-axis is the camera plane and v-axis is the image plane. (b) The stack of 2D images captured by cameras positioned along the t-axis. (c) EPI obtained for certain fixed value of u. Each point feature in scene appears in EPI as a line with slope depending on its depth. (d) Magnitude of Fourier transform of an EPI, one can see the dominating constant depths as lines with corresponding slopes in Fourier domain.

The EPI has a distinct structure such that for each point of the 3D scene there is a corresponding line in EPI with slope depending on the depth of this point. For a fixed scene point with depth value z_0 the disparity in the image plane between two cameras positioned at t_1 and t_2 is $\Delta v = v_2 - v_1 = \frac{f}{z_0}(t_2 - t_1)$ (Fig. 1(a, c)). With Lambertian reflectance assumption on the 3D scene, the sample values of the LF function are constant along the line with slope $\frac{f}{z_0}$ in the (v, t)-axis plane (i.e., in the EPI). For a scene limited in the z-axis by $z_{min} \leq z \leq z_{max}$, the spectrum of the LF's *epipolar-plane* slice $L(u_0, v, s_0, t)$ is limited by a characteristic bow-tie shape as shown in Fig. 1(d) [5]. Note, that the slopes of the epipolar lines depend essentially on the focal length of the cameras. By changing the position of the focal plane, one can get the same imagery with different EPI structure, where features in focus would appear as vertical lines in the EPI. The same approach can be applied also computationally by applying numerical re-parameterization of the LF. For fixed depth z_0, the corresponding lines in the EPI can be transformed to vertical lines by resampling from the axes (v, t) to $(v', t') = \left(v - \frac{f}{z_0}t, t\right)$ (i.e. by refocusing to the z_0 plane). The resampling procedure is, actually, a shearing operation that is implemented by horizontal shifting of EPI rows by arbitrary (i.e., non-integer) shifting factors determined by the line slope $\frac{f}{z_0}$. Focused points in desired intermediate views can be generated by a subsequent 1D interpolation along the vertical axis t' where LF samples are supposed to be constant according to the assumption of Lambertian scene reflectance. After interpolation, an inverse shearing would return the imagery to its original view topology. Virtual view synthesis by interpolation using EPI plane shearing requires knowledge of the dominating depths as well as an accurate shearing algorithm. As a useful approximation of a 3D scene depth one can assume truncated or layered and piece-wise constant depth segmentation. Depth layering can be achieved either by an accurate depth estimation and subsequent layering [6, 11, 12] or by a direct piece-wise constant depth layering [7]. Depth estimation can be computed by sequential shears of EPI for different candidate values and estimation of gradient in sheared EPI [11, 13]. A similar technique can be

used for synthetic aperture images generation from LF [14]. In this article, we assume given dominant depth layers, and focus on the shearing computed by high-accuracy interpolation.

DCT-BASED EPI SHEARING

The proposed intermediate view generation approach assumes several depth layers, which are known. In EPIs, they are represented by strips with certain slopes determined by the dominating depth of the layer. Our aim is to make this strip predominantly vertical so that one can apply 1D interpolation method [15] along the vertical direction in order to create new view samples. Inverse shearing (de-shearing) restores the initial image positions. 'Verticalization' is achieved by horizontal shifting line-by-line (shearing). Horizontal shearing in EPI domain corresponds to vertical shearing in spectral domain so that the spectrum of the shared EPI layer gets predominantly horizontal. Since the EPI's spectrum is densely sampled in horizontal ("view") direction and sparsely sampled in vertical ("angular") direction, the resulting (predominantly horizontal) spectrum becomes resilient to aliasing. Shearing is applied iteratively with increasing shift factors. This solves the problem with occlusions automatically, as small shift factors turn to vertical layers corresponding to "far" objects and bigger shift factors turn to vertical layers corresponding to "near" objects.

Consider an EPI $a(m,n)$, where $m = 0,1,\dots,M-1$ and $n = 0,1,\dots,N-1$ are the vertical and horizontal indexes of the EPI, respectively. In (v,t) coordinates, this gives $\begin{pmatrix} v_n \\ t_m \end{pmatrix} = \begin{pmatrix} \Delta v & 0 \\ 0 & \Delta t \end{pmatrix}\begin{pmatrix} n \\ m \end{pmatrix}$. Horizontal shearing relative to the EPI's central row corresponds to EPI resampling at the following points:

$$\begin{pmatrix} v_s' \\ t_s' \end{pmatrix} = \begin{pmatrix} 1 & \Delta \\ 0 & 1 \end{pmatrix}\begin{pmatrix} n \\ m - (M-1)/2 \end{pmatrix} + \begin{pmatrix} 0 \\ (M-1)/2 \end{pmatrix}, \tag{1}$$

where $\Delta = \frac{f}{z_0}\frac{\Delta t}{\Delta v}$ is the shift factor, related to z_0-depth and sampling intervals along v and t. Note that the shearing is performed with respect to the central row with index $m = (M-1)/2$. In case of odd M the central row is not shifted. The EPI's row number $m = 0,1,\dots,M-1$ is shifted by the shift factor

$$\Delta_m = [m - (M-1)/2]\Delta. \tag{2}$$

Denote the EPI row number m by a_n $(n = 0,\dots,N-1)$. The horizontal shearing of the EPI is implemented by 1D translation of EPI row a_n by the shift factor Δ_m corresponding to this row. This translation can be realized by discrete sinc interpolation in transform domain (e.g., DFT or DCT). The DFT-based row translation algorithm, based on the Inverse Shifted DFT (IShDFT), follows directly from the Fourier shift theorem:

$$\tilde{a}_k = \text{IShDFT}[\text{DFT}(a_n)] \triangleq \frac{1}{\sqrt{N}}\sum_{r=0}^{N-1}\text{DFT}(a_n)\exp\left[-i2\pi\frac{(k-\Delta_m)r}{N}\right] = \text{IDFT}\left[A_r\exp\left(i2\pi\frac{\Delta_m r}{N}\right)\right], \tag{3}$$

where A_r is the DFT spectrum of the signal a_n. This algorithm is simple and fast, however it introduces disturbing Gibbs effects due to the periodic nature of DFT and the discontinuity of the signal a_n between successive periods. In order to eliminate the Gibbs effects, it is recommended to mirror-reflect the signal a_n, creating the signal a_n^{MR} that is continuous between successive periods. Applying the DFT-based algorithm (Eq. (3)) to the mirror-reflected signal a_n^{MR}, one can obtain the output signal $\tilde{a}_k$ that is free of Gibbs effects. However, this algorithm requires computation of DFT and IDFT of the signal of double length $2N$. Noting that the DFT of the mirror reflected signal a_n^{MR} is connected to the DCT of the input signal a_n, we can convert this "modified" Inverse Shifted DFT to the Inverse Shifted DCT (IShDCT) operating on the original signal a_n of length N. The DCT-based row translation algorithm, based on the IShDCT, is defined as follows:

- Compute the DCT transform of vector a_n:

$$A_r^C = \text{DCT}(a_n) = \sqrt{\frac{2}{N}}\sum_{n=0}^{N-1}a_n\cos\left[\pi\frac{(n+1/2)r}{N}\right]. \tag{4}$$

- Modify the last DCT coefficient as follows:

$$A_r^{C(1/2)} = \begin{cases} A_r^C & \text{if} \quad r = 0, \dots, N-2 \\ A_r^C/2 & \text{if} \quad r = N-1 \end{cases}.$$

(5)

- Compute the Inverse Shifted DCT (IShDCT) of the spectrum $A_r^{C(1/2)}$:

$$\tilde{a}_k = \text{IShDCT}\left(A_r^{C(1/2)}\right) \triangleq \frac{1}{\sqrt{2N}} \left\{ A_0^{C(1/2)} + 2 \sum_{r=1}^{N-1} A_r^{C(1/2)} \cos\left[\pi \frac{(k+1/2-\Delta_m)r}{N}\right] \right\}$$

$$= \text{IDCT}\left[A_r^{C(1/2)} \cos\left(\pi \frac{\Delta_m r}{N}\right)\right] + \text{IDcST}\left[A_r^{C(1/2)} \sin\left(\pi \frac{\Delta_m r}{N}\right)\right],$$

(6)

where IDCT is the Inverse Discrete Cosine Transform:

$$\text{IDCT}(A_r) = \frac{1}{\sqrt{2N}} \left\{ A_0 + 2 \sum_{r=1}^{N-1} A_r \cos\left[\pi \frac{(k+1/2)r}{N}\right] \right\}$$

(7)

and IDcST is the Inverse Discrete Sine Transform:

$$\text{IDcST}(A_r) = \frac{1}{\sqrt{2N}} \left\{ (-1)^k A_N + 2 \sum_{r=1}^{N-1} A_r \sin\left[\pi \frac{(k+1/2)r}{N}\right] \right\}.$$

(8)

(In our case $A_N = 0$).

This algorithm, originally proposed for integer shifts [16] and later extended for arbitrary shifts [17], provides a perfect discrete sinc interpolation of the input signal a_n. It is free of Gibbs effects and has moderate computational complexity (one DCT, one IDCT and one IDcST, all of length N). We illustrate the border effects for the cases of DFT-based and DCT-based translation by shifting a linear (ramp) signal as shown in Fig. 2. The output of the DFT-based translation has ripples on the rising line and also a strong jump on the left border (due to the Gibbs effect). On the other side, the output of the DCT-based translation has no ripples on the rising line and also has smooth behavior at the left border (due to a mirror-reflection property inherent in DCT).

For completeness, one can represent the shift operation in terms of the interpolation kernel, which is based on the discrete sinc function [17]

$$\text{sincd}(M, N, x) = \frac{\sin\left(\pi \frac{Mx}{N}\right)}{N \sin\left(\pi \frac{x}{N}\right)},$$

(9)

$$\tilde{a}_k = \sum_{n=0}^{N-1} a_n \left\{ \begin{array}{l} \text{sincd}(2N-2, 2N, k+1/2-n-1/2-\Delta_m) \cos\left[\frac{\pi}{2N}(k+1/2-n-1/2-\Delta_m)\right] \\ + \quad \text{sincd}(2N-2, 2N, k+1/2+n+1/2-\Delta_m) \cos\left[\frac{\pi}{2N}(k+1/2+n+1/2-\Delta_m)\right] \end{array} \right\}.$$
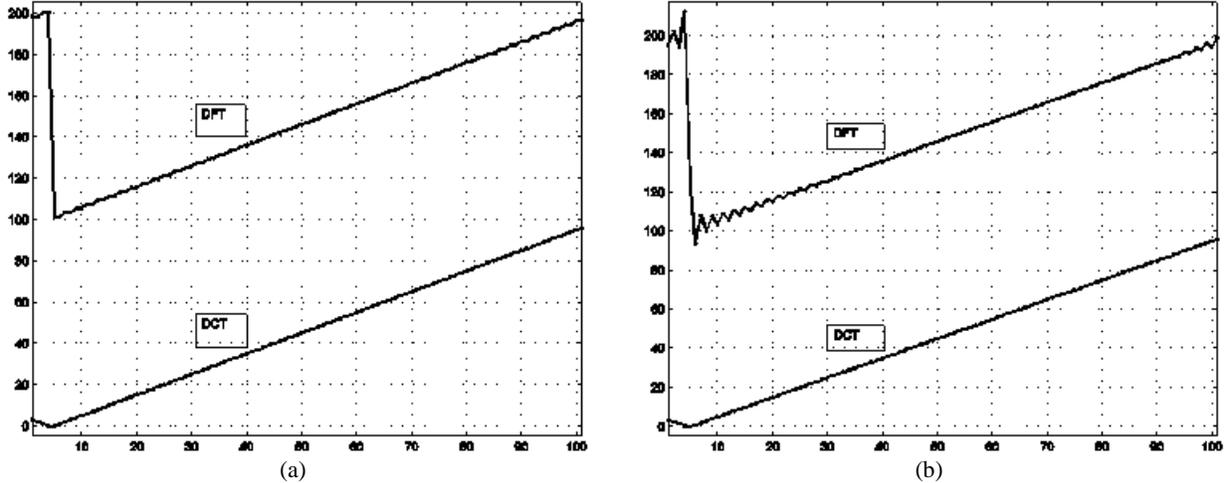
(10)

(a)   (b)

FIGURE 2. Comparison of DFT-based translation and DCT-based translation of linear test signal (ramp) by integer shift factor $\Delta_m = 4$ (a) and by non-integer shift factor $\Delta_m = 3\sqrt{2}$ (b). The output of the DFT-based translation has ripples on the rising line and also a strong jump on the left border. The output of the DCT-based translation has no ripples on the rising line and also has smooth behavior at the left border.

EXPERIMENTAL RESULTS

Our first test scene consists of three dices at different depths, as illustrated in Fig. 1(a). Refocused images of the same scene are shown in Fig. 3. A horizontal slice of the image stack (an EPI) is shown in Fig. 4(a). The disparity map of the same EPI is shown in Fig. 4(b). It consists of three layers determined by three dominant values as visible from the disparity histogram in Fig. 4(c). Correspondingly, these three values determine the shearing values Δ_m for the DCT-based shearing algorithm.

First, we demonstrate the *reversibility* of the DCT-based shearing algorithm by forward and inverse shifting of each EPI row by the same shift factor and by different methods. Figure 5 illustrates the results in the spatial and Fourier domains. PSNRs calculated between the original EPI and the reconstructed EPI give the following results: linear interpolation: 42.69 dB, piecewise cubic spline interpolation: 44.62 dB, fifth order interpolating B-spline [18]: 63.67 dB, and DCT-based interpolation: 83.20 dB. Note that while all PSNRs are relatively high, the reconstructed EPI by the DCT-based interpolation is practically indistinguishable from the original one. The above results demonstrate, both visually and numerically, that the DCT-based interpolation is a very accurate resampling algorithm for the case of EPI shearing. A result of new view synthesis is shown in Fig. 4(d), where the EPI is interpolated along t axis applying the three-pass algorithm in DCT domain.
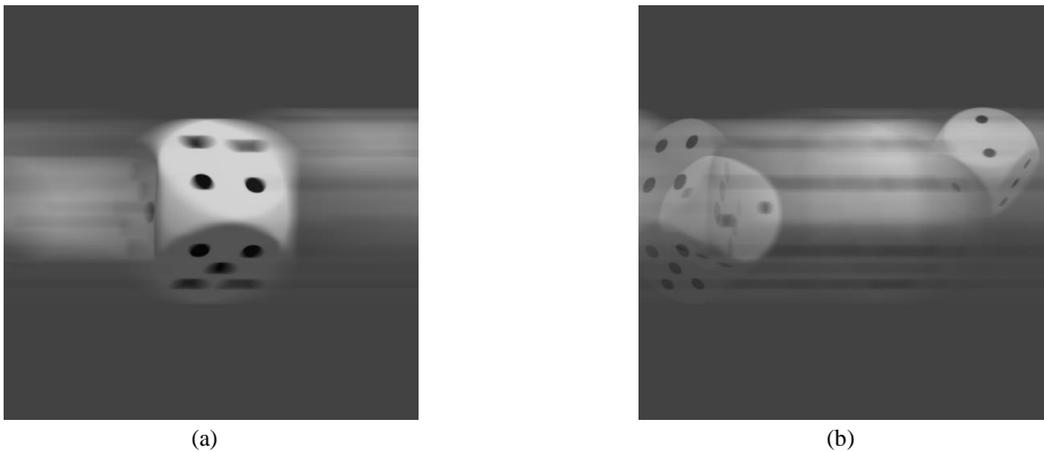

(a)   (b)

FIGURE 3. Demonstration of the refocusing effect. (a) Refocusing on the central dice. (b) Refocusing on the right dice.
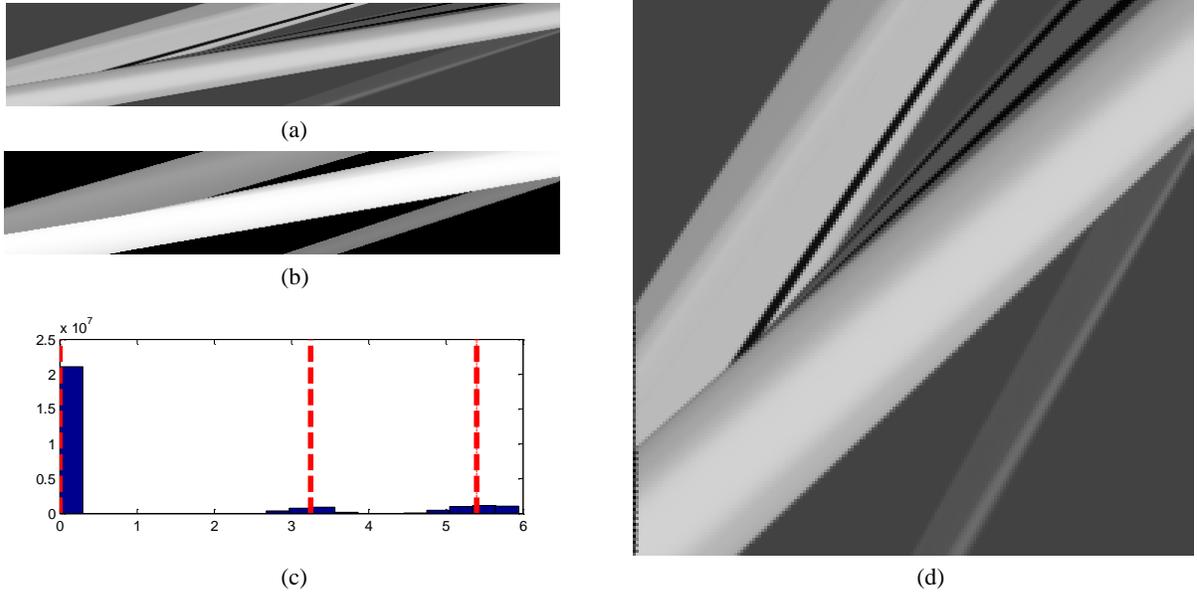
(a)

(b)

(c)

(d)

FIGURE 4. Three pass EPI interpolation algorithm. (a) EPI of the test scene. (b) Disparity map of the EPI of the test scene. (c) Histogram of disparity values in the test scene. The selected shearing values are marked with vertical dashed lines. (d) Three pass interpolated EPI of the test scene ($\sigma = 4$). DCT-based shearing in first and last passes of the algorithm was applied independently on EPI segments corresponding to the selected shearing values.
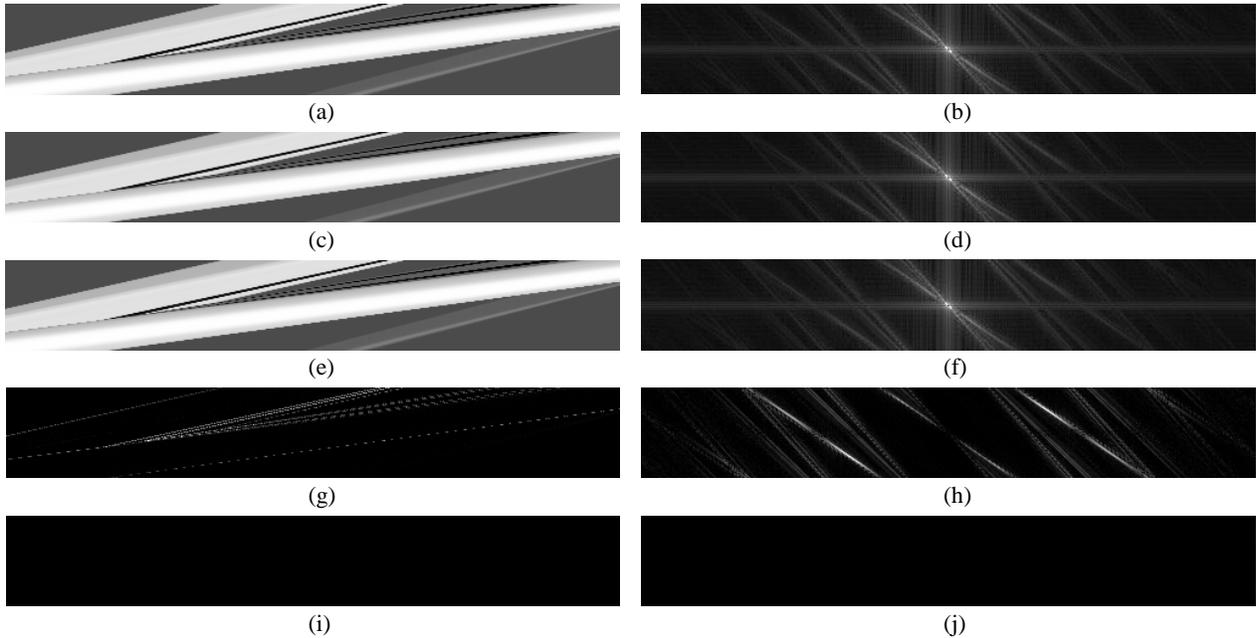

(a)

(b)

(c)

(d)

(e)

(f)

(g)

(h)

(i)

(j)

FIGURE 5. Shearing and de-shearing – comparison of linear and DCT interpolation in spatial domain (left column) and spectral domain (right column). (a), (b) Test EPI. (c), (d) Test EPI after shearing and 'de-shearing' using linear interpolation. (e), (f) Test EPI after shearing and 'de-shearing' using DCT interpolation. (g), (h) Difference in spatial and spectral domains between the test EPI and the result of linear interpolation. (i), (j) Difference in spatial and spectral domains between the test EPI and the result of DCT interpolation.

Second, we demonstrate the robustness of the proposed method against deviations of the true disparities from the applied depth layering. The proposed algorithm assumes knowledge about depth and depth layering. Unfortunately, both depth-from-stereo algorithms and depth sensors are imprecise in depth estimation or measurement. This inaccurate approximation of depth emphasizes the need for reconstruction algorithm that is robust against incorrect

depth estimation, which might lead to incorrect shearing factors. Furthermore, while for typical scenes, there are only few dominating depths (few objects at different depths), for more complex scenes, there might be considerable depth variations within each layer which will cause imperfect 'verticalization' of epipolar lines. Therefore, we study the robustness of the proposed algorithm against different incorrect shearing factors for different spatial frequencies. In our experiment we used sinusoidal textures (Fig. 6(a)) with different frequencies placed on a plane (constant depth), such that its ground true shearing factor is 5 pixels. We synthesize the four intermediate views between adjacent views by the proposed algorithm and by linear interpolation applying different shearing factors, starting from the true factor of 5 and deviating from it as far as 3. Figure 6(b, c) shows the PSNR between the ground truth and the reconstructed intermediate four views for the interpolation methods. Both linear and DCT based algorithms are accurate when the shearing factors are exact or close to the true one. However, with increasing the shearing factor deviation, the DCT-based shift algorithm shows better performance than the linear interpolation shift algorithm (c.f. Fig. 6(d, e)). Also in case of small deviation of the shearing factor from its true value, the DCT-based interpolation method results in reconstruction quality exceeding 30 dB for all scene texture frequencies, while the linear method gives satisfactory results only for textures with low frequencies (Fig. 6(f, g)).
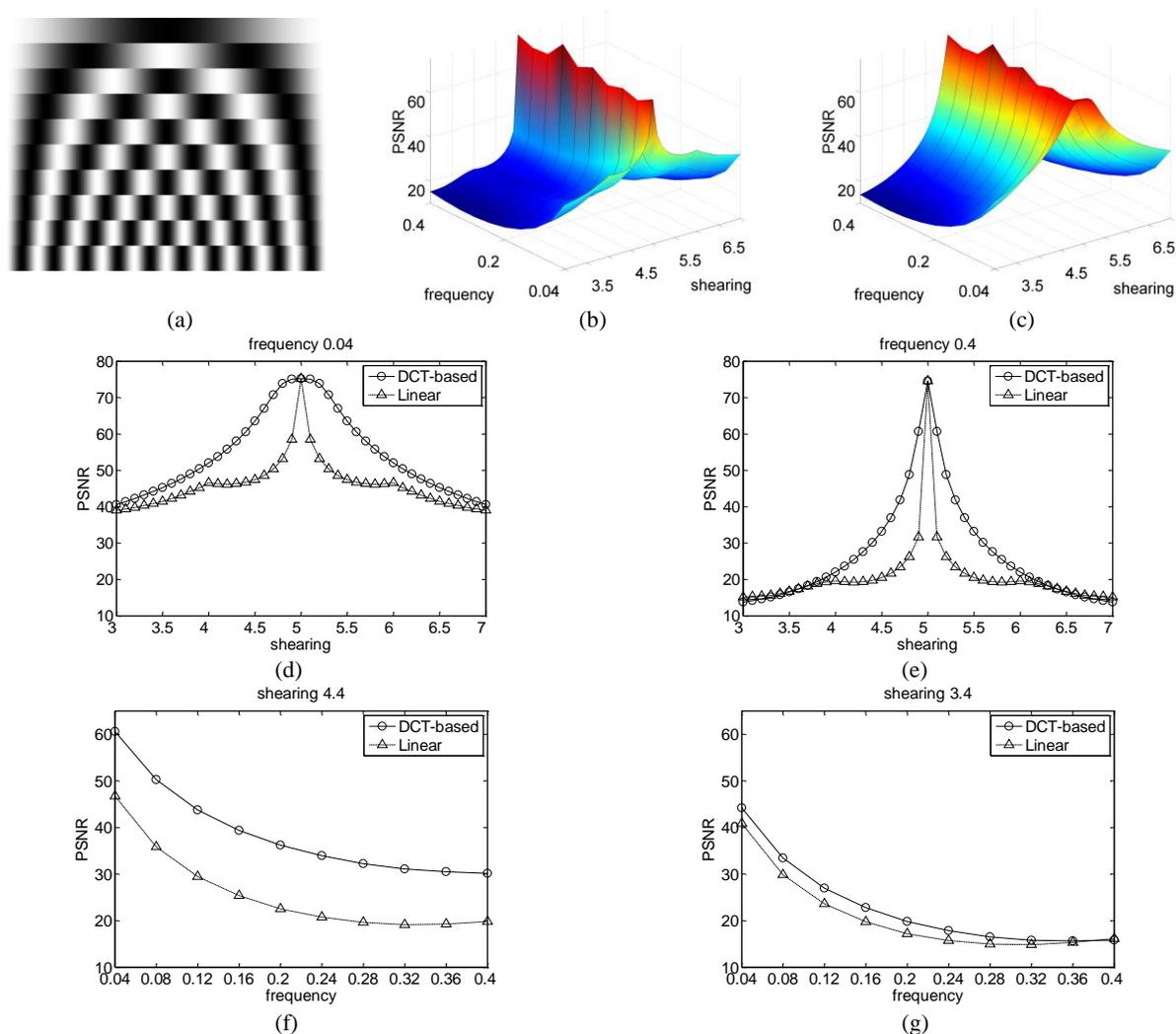

FIGURE 6. The quality of reconstruction depends on shearing parameter and texture frequency. (a) Sinusoidal texture of the test scene placed on the flat plane positioned parallel to camera plane. In this setup plane has constant depth and correct shearing factor is 5. (b), (c) Reconstruction quality in PSNR for different shearing factors and texture frequencies. (d), (e) Comparison of the reconstruction quality of different interpolation methods depending on shearing factor for fixed frequencies 0.4 and 0.04. (f), (g) Comparison of the reconstruction quality depending on frequencies for fixed shearing factors 4.4 and 3.4.

CONCLUSION

In this article we have presented a new method for intermediate view generation using DCT-based EPI-domain shearing. Shearing in EPI domain is equivalent to refocusing of LF imagery, which re-parameterizes its presentation so that the epipolar lines corresponding to the targeted depth appear vertical in the sheared EPIs. Vertical epipolar lines with constant intensity are easy to interpolate with simple 1D interpolation kernels. Inverse shearing reconstructs the original view topology. The method works on successive depth layers from back to front thus dealing naturally with possible occlusions. Few dominant layers are sufficient since the method is insensitive to small depth variations within the layer. The modified DCT transform is an effective arbitrary-factor shift operator, which gives free of boundary distortions results. The experiments have shown that the DCT-based shearing method better handles the direct and inverse shearing operation, in particular by preserving the frequency properties of the processed views. The demonstrated robustness to inaccurate estimated shearing factors for varying frequency components makes the proposed method very practical, as one can safely use information about depth layers obtained from (possibly inaccurate) depth sensing and depth estimation algorithms.

REFERENCES

1. E. H. Adelson and J. R. Bergen, "The plenoptic function and the elements of early vision," in *Computational Models of Visual Processing*, edited by M. Landy and J. A. Movshon, MIT Press, Cambridge, MA, 1991, pp. 3–20.
2. M. Levoy and P. Hanrahan, *Proc. SIGGRAPH*, New Orleans, LA, 1996, pp. 31–42.
3. H.-Y. Shum and S. B. Kang, *Proc. SPIE* **4067**, Perth, Australia, 2000, pp. 2–13.
4. Z. Lin and H.-Y. Shum, *Int. J. Comput. Vision* **58**, 121–138 (2004).
5. Chai, X. Tong, S.-C. Chan, and H.-Y. Shum, *Proc. SIGGRAPH*, New Orleans, LA, 2000, pp. 307–318.
6. S. Wanner and B. Goldluecke, *IEEE T. Pattern Anal.* **36**, 606–619 (2014).
7. J. Pearson, M. Brookes, and P. L. Dragotti, *IEEE T. Image Process.* **22**, 3405–3419 (2013).
8. R. C. Bolles, H. H. Baker, and D. H. Marimont, *Int. J. Comput. Vision* **1**, 7–55 (1987).
9. B. Wilburn, M. Smulsky, H.-H. K. Lee, and M. Horowitz, *Proc. SPIE* **4684**, San Jose, CA, 2001, pp. 29–36.
10. R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera", Tech. Rep. CSTR 2005-02, Stanford University, 2005.
11. A. Criminisi, S. B. Kang, R. Swaminathan, R. Szeliski, and P. Anandan, *Comput. Vis. Image Und.* **97**, 51–85 (2005).
12. T. E. Bishop and P. Favaro, *Proc. ICCV*, Kyoto, Japan, 2009, pp. 1622–1629.
13. M. Diebold and B. Goldluecke, "Epipolar plane image refocusing for improved depth estimation and occlusion handling," in *Vision, Modeling, and Visualization*, edited by M. Bronstein, J. Favre, and K. Hormann, Eurographics, 2013, pp. 145–152.
14. A. Isaksen, L. McMillan, and S. J. Gortler, *Proc. SIGGRAPH*, New Orleans, LA, 2000, pp. 297–306.
15. G. Wolberg, *Digital Image Warping*, IEEE Computer Society Press, 1990.
16. P. Yip and K. R. Rao, *IEEE T. Acoust. Speech* **35**, 404–406 (1987).
17. L. Yaroslavsky, "Fast discrete sinc-interpolation: a gold standard for image resampling," in *Advances in Signal Transforms: Theory and Applications*, edited by J. Astola and L. Yaroslavsky, Hindawi, 2007, pp. 337–405.
18. A. Gotchev, K. Egiazarian, and T. Saramäki, "Image interpolation by optimized spline-based kernels," in *Advances in Signal Transforms: Theory and Applications*, edited by J. Astola and L. Yaroslavsky, Hindawi, 2007, pp. 285–335.