https://doi.org/10.2352/ISSN.2470-1173.2018.04.SDA-144

Conversion of sparsely-captured light field into alias-free full-parallax multiview content

Erdem Sahin[1]***, Suren Vagharshakyan***[1]***, Robert Bregovic***[1]***, Gwangsoon Lee***[2]***, Atanas Gotchev***[1]***;***
[1] ***Tampere University of Technology; Tampere, Finland,*** [2] ***Electronics and Telecommunications Research Institute (ETRI); Daejeon, Republic of Korea***

Abstract

We propose shearlet decomposition based light field (LF) reconstruction and filtering techniques for mitigating artifacts in the visualized contents of 3D multiview displays. Using the LF reconstruction capability, we first obtain the densely sampled light field (DSLF) of the scene from a sparse set of view images. We design the filter via tiling the Fourier domain of epipolar image by shearlet atoms that are directionally and spatially localized versions of the desired display passband. In this way, it becomes possible to process the DSLF in a depth-dependent manner. That is, the problematic areas in the 3D scene that are outside of the display depth of field (DoF) can be selectively filtered without sacrificing high details in the areas near the display, i.e. inside the DoF. The proposed approach is tested on a synthetic scene and the improvements achieved by means of the quality of the visualized content are verified, where the visualization process is simulated using a geometrical optics model of the human eye.

Introduction

Autostereoscopic multiview displays provide comfortable 3D viewing experience without requiring wearing glasses. Furthermore, they are cost effective due to their simple design that consists of a lens array or parallax barrier placed on top of a conventional 2D display. The content preparation for multiview displays includes two critical problems. The first one is the difficulty in the capture of scene information which usually requires a capture setup consisting of a very dense set of cameras. The problem of intermediate view synthesis from a given set of captured views is critical to relieve the capture process for several reasons such as to reduce the data rate. The second problem, on the other hand, is related to capability of the multiview display to reproduce a given 3D scene, which is usually defined by the concepts of display bandwidth or display depth of field (DoF). Thus, the multiplexed image to be written on the 2D display should be properly calculated from the captured content such that it is suitable for the given multiview display, i.e. the captured content should be properly filtered to prevent aliasing artifacts.

The problem of sparse view capture and then synthesis of intermediate views can be classified into two categories. In the first one, the approaches rely on depth estimation of the scene where the estimated depth is utilized together with captured sparse set of images to synthesize novel views [1, 2, 3]. On the other hand, in the second approach the problem is formulated as sampling and reconstruction of the underlying light field (LF) of the scene, where the scene depth information is not used as an auxiliary information [4, 5, 6]. These two categories of approaches have their own pros and cons which is a topic that deserves an extensive discussion, but it is out of scope of this paper. Here we rather address our previously proposed advanced LF reconstruction algorithm that can produce the densely sampled light field (DSLF) of a scene based on the shearlet decomposition [7]. It has been demonstrated to be superior compared to most of the existing techniques especially owing to the fact that it does require depth estimation and thus does not suffer from problems related to inaccurate depth estimation. Therefore, in this paper we utilize the same method to relieve the content capture process of multiview displays.

Antialiased LF sampling (rendering) is another extensively discussed problem in the literature [8, 9, 10]. The work presented in [11] provides derivations of display bandwidth and DoF for a given (conventional) multiview display that are then used as a guidance to design antialiasing filter in the ray-space tailored for the given display. There are several other works that deal with filter design and filtering of the multiview content for various purposes such as for reducing aliasing artifacts due to non-rectangular sampling grid [12], cross-talk mitigation [13], or better compression of the content [14]. Here we approach the antialiasing problem again by utilizing the shearlet decomposition in the epipolar image (EPI) domain of the DSLF, since the distribution of shearlet atoms is particularly suited for designing directional (depth-dependent) filter in this domain that is tailored for the given display. Thus, altogether, we provide a structured and unified framework for multiview displays addressing both the view synthesis and content filtering problems.

DSLF Reconstruction

The sparse LF capture setup that we consider is illustrated in Fig. 1 in relation to the multiview display, where a horizontal cross-section of the 3D space is shown for simplicity. (x,y) and (p,q) planes represent the display panel and lens array, respectively, which together form the multiview display. Δx and Δp denote the pixel size of the display panel and the lens pitch, respectively. On the other hand, (s,t) and (u,v) planes are used to parametrize the 4D LF $L(s,t,u,v)$, where they represent the camera view position and sensor image coordinates, respectively. The distance between adjacent camera positions, Δs, and the pixel size of capture camera images, Δu, define the sampling rates of the continuous LF on these two planes. l_x denotes the gap between the display panel and lens array, and l_s denotes distance between the center of projections (CoP) of cameras and their sensor planes.

Let us assume that the sensor images of the cameras are re-centered at the furthest scene point at $z_s + z^+$, i.e. the disparity between two adjacent camera images is 0 for scene points on this plane. We refer the case of LF sampling where the total depth

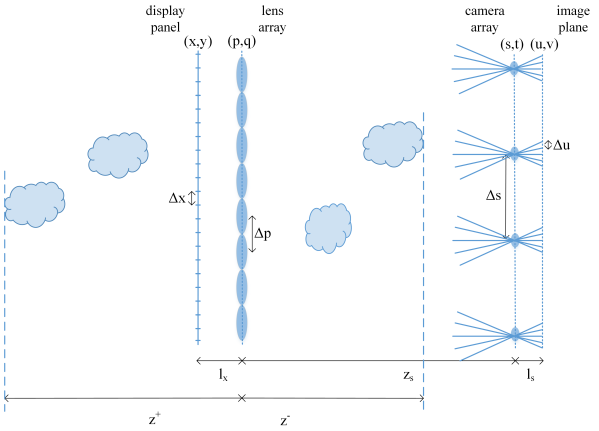

Figure 1. *Sparse LF capture setup in relation to the display parametrization.*

range of scene is at most one pixel as the DSLF. In other words, for a DSLF capture setup, we assume that the (recentered) disparity corresponding to the nearest scene point at $z_s - z^-$ is smaller than 1px. The DSLF is considered as an equivalent representation to the continuous LF, since any desired ray can be accurately obtained by resampling the DSLF via linear interpolation [9]. Please note that in the sparse LF capture scenario that we consider in Fig. 1, the abovementioned DSLF conditions are not satisfied. Therefore, our approach consists of first reconstruction of DSLF from captured sparse LF and then application of directional filtering to the DSLF to obtain display specific LF. For the given camera parameters Δu, l_s and scene setup in Fig. 1, we specify $\tilde{\Delta} s$ as the view sampling step that satisfy the DSFL requirement by equality, i.e. the disparity for the depth $z_s - z^-$ is 1px. We will use this parameter in the following discussion.

Let us consider the epipolar-plane image (EPI) $E(s,u)$, or $E(t,v)$ depending on the direction of analysis, that is formed by taking slices from 4D LF $L(s,t,u,v)$. The problem of DSLF reconstruction can be formulated as reconstruction of each densely sampled EPI slice from a sparse (decimated) set of samples as illustrated in Fig. 2(a). As previously presented in [7], this problem can be efficiently solved utilizing sparsity of LFs in the shearlet domain. The tiling of the frequency domain of EPI by the shearlet atoms is illustrated in Fig. 2(b), where the shearlet atoms are distributed directionally and each direction corresponds to a different depth layer of the scene.

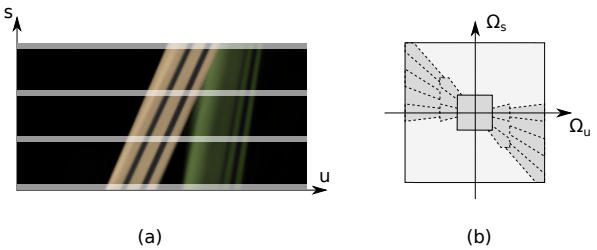

Figure 2. *Reconstruction of the densely sampled EPI. (a) Sparsely sampled EPI of the DSLF. (b) Tiling of the frequency domain by the shearlet atoms.*

Let us denote the vectorized versions of the densely sampled and decimated EPIs as $\boldsymbol{a}$ and $\boldsymbol{b}$, respectively, and the masking matrix representing the known samples positions as $\boldsymbol{H}$. The reconstruction of unknown EPI samples can then be modeled as estimation of $\boldsymbol{a}$ given that $\boldsymbol{b} = \boldsymbol{H}\boldsymbol{a}$. This inverse problem can be solved by iterative hard thresholding procedure with decreasing threshold as [7]

$$\boldsymbol{a}_{n+1} = \boldsymbol{S}^* \left\{ \boldsymbol{T}_{\lambda_n} \left\{ \boldsymbol{S} [\boldsymbol{a}_n + \alpha_n (\boldsymbol{b} - \boldsymbol{H}\boldsymbol{a}_n)] \right\} \right\}, \tag{1}$$

where $\boldsymbol{S}$ and $\boldsymbol{S}^*$ are the shearlet analysis and synthesis transform matrices, respectively, $\boldsymbol{T}_{\lambda_n}$ is the hard thresholding operator with threshold λ_n, and α_n is the acceleration coefficient. A sufficient number of iterations produce the final result $\boldsymbol{a}_n$ as the solution. The corresponding sparse representation is given as $\boldsymbol{S}\boldsymbol{a}_n$. The same procedure is repeated first for each horizontal parallax set and then for each vertical parallax set to obtain full parallax DSLF reconstruction. For a more detailed discussion of the shearlet transform based LF reconstruction algorithm, the reader is referred to [7].

Antialiasing Filter Design Using Shearlet Decomposition

We aim at designing the antialiasing filter tailored to the given multiview display. The filter should be designed such that it takes into account the depth-dependent LF reproduction capability of the display and, thus, it is able to be applied in a depth-dependent manner. The distribution of shearlet atoms has desirable features for designing directional (i.e. depth-dependent) filter. Thus, we construct the display specific depth-dependent filter also by utilizing the shearlet decomposition. The design of the filter starts with estimation of the display DoF. Here, we consider again the simple 2D scenario illustrated in Fig. 3.

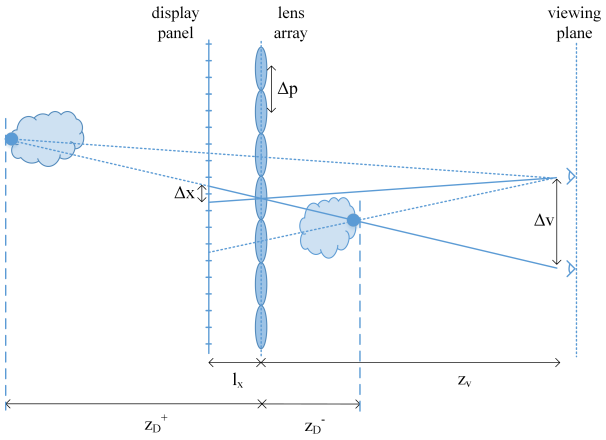

Figure 3. *Display depth of field.*

As shown in Fig. 3, the depth limits $z_D{}^+$ and $z_D{}^-$ behind and in front of the display, respectively, correspond to -1px and $+1$px display pixel disparities perceived by the viewer at z_v, when moved between adjacent views that are Δv apart [11]. Noting that $\Delta v = z_v \Delta x / l_x$, these depth limits of display DoF are found as

$$\begin{aligned} z_D^- &= \frac{z_v \Delta p}{\Delta v + \Delta p}, \\ z_D^+ &= \frac{z_v \Delta p}{\Delta v - \Delta p}. \end{aligned} \tag{2}$$

Please note that the pixel and lens pitches used in Fig. 3 represent equivalent pitches for this illustration, which may not be equal to real physical parameters. The relation between these equivalent and physical parameters depend on how is the view multiplexing done while producing the multiplexed 3D content on the 2D display.

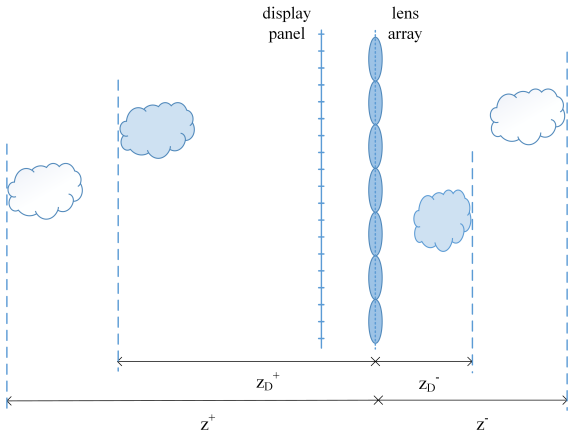

Figure 4. *Representation of display DoF in relation to the scene depth range.*

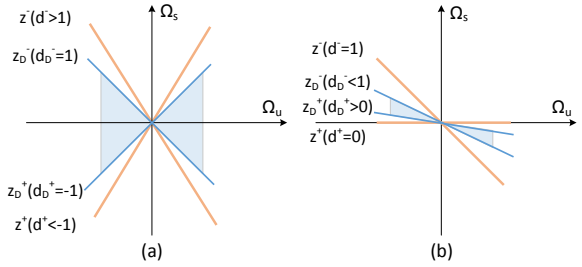

Figure 5. *(a) Frequency domain support of captured LF, where $[d^+, d^-]$px is the disparity range of the scene that falls outside the range $[0,1]$px corresponding to display DoF (b) DoF in the frequency domain of DSLF.*

Since the display is incapable of accurately reproducing the scene content outside the display DoF, the visual content in the LF data corresponding to the objects outside of the defined DoF of the display should be properly filtered, i.e. antialiased, beforehand. Such a scenario is illustrated in Fig. 4, where the scene extends beyond the display depth budget, i.e. DoF. Let us consider the previously reconstructed DSLF of the scene, where the distance between the adjacent reconstructed views is $\tilde{\Delta}s$ and the corresponding recentered disparity range of the scene is $[0,1]$px. Thus, the nearest and furthest depth boundaries of the scene, z^- and z^+, respectively, are revealed in the EPI representation of the DSLF as tilted planes corresponding to 1px and 0px disparities, respectively. On the other hand, the corresponding (recentered) disparity values for the depth limits of display DoF, $z_D{}^-$ and $z_D{}^+$, are found respectively as

$$d_D{}^- = \left(\frac{1}{z_s - z_D{}^-} - \frac{1}{z_s + z^+}\right)\frac{\tilde{\Delta}s l_s}{\Delta u},$$
$$d_D{}^+ = \left(\frac{1}{z_s + z_D{}^+} - \frac{1}{z_s + z^+}\right)\frac{\tilde{\Delta}s l_s}{\Delta u}. \quad (3)$$

In the recentered DSLF frequency domain, the support of the filter tailored for such display is illustrated Fig. 5. Thus, the directional filter that we design is to approximate the frequency plane support using shearlet transform atoms such that it keeps frequencies corresponding to disparities in the range of $[d_D{}^+, d_D{}^-]$px (cone-shaped colored area) and suppress all other high spatial frequencies (corresponding to disparities outside of the cone-shaped area). Although, due to spatial localization, construction of a filter with such ideal frequency domain support is not possible, we can sufficiently approximate the required filter bandwidth using shearlet decomposition atoms.

Experimental Results

The proposed shearlet decomposition based LF reconstruction and filtering algorithms are together tested on a simulated 3D display with hexagonal microlens array for which the parameters are given in Fig. 6.

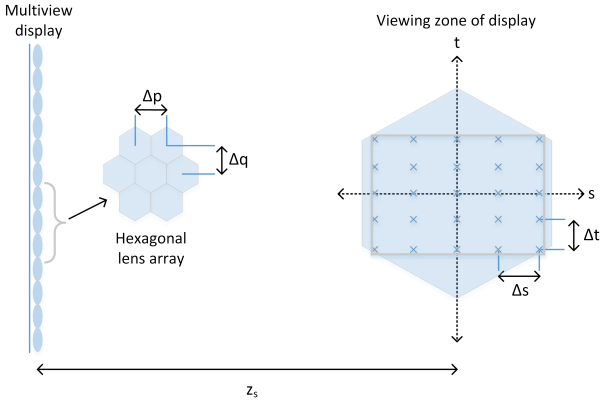

Figure 6. *Simulation setup.*

The 3D display constitutes a 7680×4320 resolution RGB stripe panel with subpixels of size $9.7\mu m \times 29.1\mu m$ and corresponding pixel sampling steps of $\Delta x = \Delta y = 29.1\mu m$, and an hexagonal microlens array with sampling steps of $\Delta p = 250\mu m$, $\Delta q = 217\mu m$ as shown in Fig. 6. The gap between the display panel and lens array is set to $l_x = 0.636mm$ and the multiview subpixel mapping is done such that there are 15×9 views within the hexagonal viewing zone at $z_v = z_s = 50cm$. The camera parameters used in the LF capture are as following: the resolution is 1280×720, the camera pixel pitches are $\Delta u = \Delta v = 7\mu m$ and the distance between the CoPs and sensors of the cameras is $l_s = 20mm$. Different capture scenarios are considered depending on the camera view sampling distance Δs. The synthetic test scene designed in Blender is confined within $z^+ = 58.10cm$ and $z^- = 17.48cm$. For this depth range, the view sampling distance $\tilde{\Delta}s = 0.163mm$ satisfies the DSLF constraint with equality, i.e. the corresponding (recentered) disparity range of the scene is [0,1]px.

Several different reconstructions corresponding to different sets of (captured or synthesized) view images are compared. In each case the captured/synthesized images are confined within the same central rectangular sub-region of size $21cm \times 10.5cm$ shown in Fig. 6. The reconstruction process is implemented by simulating the human eye viewing process employing geometric optics principles, where the eye is modeled as a camera with thin and aberration-free lens of $8mm$ aperture size and $2.9\mu m$ sensor pixel pitch. The perceived image is found by integrating rays on the retina that are back-projected from multiple viewpoints over the aperture.

Figure 7. *Eye view from multiplexed image which is obtained by quadri-linear interpolation from original 15x9 LF.*

In the first scenario, we render 15x9 view images that is close to what is required by the display. In the naive way of multiplexed image calculation, one could simply map these view images to sub-pixels of display via nearest-neighbor interpolation. However, to be more accurate, here we prefer to find the ray intensity corresponding to each sub-pixel of the display via quadri-linear interpolation applied on both (s,t) and (u,v) planes. The result of reconstruction (i.e. the perceived image by the viewer) at the novel viewpoint $(s,t) = (13mm, 22mm)$ is shown in Fig. 7. As can be observed in the zoomed-in image, there exist visual artifacts. This is due to insufficient view sampling rate provided by 15x9 views, i.e. it is too coarse for the given depth range.

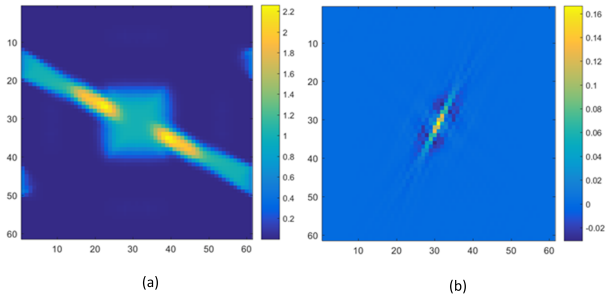

Figure 8. *(a) Spectral and (b) spatial representations of the designed directional filter using shearlet decomposition.*

In the second scenario, we consider the sparse LF capture setup which consists of 5×5 camera array. From rendered sparse view images, we first reconstruct the DSLF of the scene using our shearlet decomposition based LF reconstruction algorithm. The DSLF consists of 129×129 view images within the same central sub-region of viewing zone confining the original sparse view images. Then, the proposed shearlet decomposition based filter is designed based on the display DoF that is estimated to be limited by the depths $z_D{}^- = 1.07cm$ and $z_D{}^+ = 1.12cm$ according to Eq. 2. The spectral and spatial representations of the designed filter are shown in Fig. 8.

We generate two sets of multiplexed images this time by using nearest-neighbor interpolation from the reconstructed DSLF and its filtered version. The corresponding reconstructed eye view images at the novel viewpoint $(13mm, 26mm)$ are shown in Fig. 9 and Fig. 10, respectively. As seen in Fig. 9, there are still some visible artifacts in the reconstructed view image. Those artifacts are due to fact that the scene spreads beyond the display DoF. Indeed, the region shown in the zoomed-in image is actually outside of the display DoF. Therefore, the display is not capable of reproducing such depth ranges. However, as shown in Fig. 10, when the DSLF of the 3D scene is prefiltered with the tailored-designed filter given in Fig. 8, the details of the scene content near the display plane (inside the DoF) are kept and the content further away from the display plane (outside of DoF) is smoothened. Thus, we are able to suppress the artifacts at deep regions without over-smoothing the shallow regions being close to the display plane.

Figure 9. *Eye view from multiplexed image which is obtained by nearest-neighbor interpolation from original reconstructed DSLF.*

Figure 10. *Eye view from multiplexed image which is obtained by nearest-neighbor interpolation from filtered DSLF.*

Conclusions

We have demonstrated a DSLF reconstruction and LF filtering framework which enables appropriately filtering a 3D multiview content that is captured by a sparse set of cameras. In particular, one can sample (render) the scene employing a smaller number of cameras (e.g. 5×5) than the total number of views provided by the multiview display (e.g. 15×9), and then use our shearlet decomposition based LF reconstruction algorithm to obtain the DSLF of the scene. Afterward, with the provided filter designed also utilizing shearlet decomposition, one can prefilter the scene (i.e. DSLF) in a depth-dependent manner before calcu-

lating the multiplexed image. By this way, most of the artifacts that appears in the areas outside the display DoF are eliminated, and at the same time the high details in the areas of the scene that fall inside the DoF are kept.

Acknowledgments

This work was supported by "The Cross-Ministry Giga KOREA Project" grant funded by the Korea government(MSIT) GK17C0200, Development of Full-3D Mobile Display Terminal and its Contents)

References

[1] C. Kim, H. Zimmer, Y. Pritch, A. Sorkine-Hornung , and M. Gross, Scene reconstruction from high spatio-angular resolution light fields, ACM Trans. Graph., vol. 32, no. 4, pp. 1-12, 2013.

[2] J. Pearson, M. Brookes, and P. Dragotti, Plenoptic layer-based modeling for image based rendering, IEEE Trans. Image Process., vol. 22, no. 9, pp. 3405-3419, 2013.

[3] 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.

[4] D. C. Schedl, C. Birklbauer, and O. Bimber, Directional super-resolution by means of coded sampling and guided upsampling, Proc. IEEE Conf. Comput. Photography, 2015, pp. 1-10.

[5] S. Heber and T. Pock, Convolutional networks for shape from light field, Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2016, pp. 3746-3754.

[6] N. K. Kalantari, T.-C. Wang, and R. Ramamoorthi, Learning based view synthesis for light field cameras, ACM Trans. Graph., vol. 35, no. 6, pp. 1-10, 2016.

[7] S. Vagharshakyan, R. Bregovic and A. Gotchev, Light Field Reconstruction Using Shearlet Transform, in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 40, no. 1, pp. 133-147, 2018.

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

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

[10] J. Stewart, J. Yu, S. J. Gortler, L. McMillan, A new reconstruction filter for undersampled light fields, EGRW '03: Proceedings of the 14th Eurographics workshop on Rendering, pp. 150-156, 2003.

[11] M. Zwicker, W. Matusik, F. Durand, and H. Pfister, Antialiasing for automultiscopic 3d displays, in Rendering Techniques 2006: 17th Eurographics Workshop on Rendering, 2006, pp. 73-82.

[12] A. Boev, R. Bregovic, D. Damyanov and A. Gotchev, Anti-aliasing filtering of 2D images for multi-view auto-stereoscopic displays, 2009 International Workshop on Local and Non-Local Approximation in Image Processing, Tuusula, 2009, pp. 87-97.

[13] V. Ramachandra, K. Hirakawa, M. Zwicker and T. Nguyen, Spatioangular Prefiltering for Multiview 3D Displays, in IEEE Transactions on Visualization and Computer Graphics, vol. 17, no. 5, pp. 642-654, 2011.

[14] M. Zwicker, S. Yea, A. Vetro, C. Forlines, W. Matusik, and H. Pfister, Display pre-filtering for multi-view video compression, in Proceedings of the 15th International Conference on Multimedia, pp 1046-1053, 2007.

Author Biography

Erdem Sahin received the Ph.D. degree from the Electrical and Electronics Engineering Department of Bilkent University, Turkey (2013). He has been post-doctoral researcher at Tampere University Technology, Finland, since 2014. His current research interests are computational holography, plenoptic cameras, and multiview displays.

Suren Vagharshakyan received the M.Sc. in mathematics from Yerevan State University (2008). He is a Ph.D. student at the Department of Signal Processing at Tampere University of Technology since 2013. His research interests are in the area of light field capture and reconstruction.

Robert Bregović received the M.Sc. in electrical engineering from University of Zagreb (1998) and the Dr.Sc.(Tech) in information technology from Tampere University of Technology (2003). He has been working at Tampere University of Technology since 1998. His research interests include the design and implementation of digital filters and filterbanks, multirate signal processing, and topics related to acquisition, processing/modeling and visualization of 3D content.

Gwangsoon Lee received the M.Sc. and Ph.D. degrees, all in electronics engineering from Kyungpook National University, Daegu, Korea, in 1995 and 2004. He joined Electronics and Telecommunications Research Institute (ETRI) in 2001, and he is currently the Principal Researcher of Tera-media Research Group in ETRI. His current research interests include light-field-based signal process and autostereoscopic display.

Atanas Gotchev received the M.Sc. degrees in radio and television engineering (1990) and applied mathematics (1992) and the Ph.D. degree in telecommunications (1996) from the Technical University of Sofia, and the D.Sc.(Tech.) degree in information technologies from the Tampere University of Technology (2003). He is a Professor at Tampere University of Technology. His recent work concentrates on algorithms for multisensor 3-D scene capture, transform-domain light-field reconstruction, and Fourier analysis of 3-D displays.