Light Field Virtual View Rendering Based on EPI-Representations

SUN Yule and YU Lu

(Department of Information Science & Electronic Engineering, Zhejiang University, Hangzhou 310007, China)

◀ Abstract

Image-Based Rendering (IBR) is one powerful approach for generating virtual views. It can provide convincing animations without an explicit geometric representation. In this paper, several implementations of light field rendering are summarized from prior arts. Several characteristics, such as the regular pattern in Epipolar Plane Images (EPIs), of light field are explored with detail under 1D parallel camera arrangement. It is proved that it is quite efficient to synthesize virtual views for Super Multi-View (SMV) application, which is in the third phase of Free-Viewpoint Television (FTV). In comparison with convolutional stereo matching method, in which the intermediate view is synthesized by the two adjacent views, light field rendering makes use of more views supplied to get the high-quality views.

◀ Keywords

depth estimate; Epipolar Plane Image (EPI); light field; view synthesis

1 Introduction

F ree-Viewpoint Television (FTV) enables users to view a 3D scene by freely changing viewpoints as they do naturally in the real world. FTV is ranked as the top of visual media for its infinite number of views to display [1]. It provides a very realistic glasses-free 3D viewing without eye fatigue. Since the first phase of FTV, i.e. Multi-View Video Coding (MVC), was proposed in 2001 [2], the applications of 3DTV has developed from stereo viewing to auto-stereoscopic viewing and Super Multi-View viewing.

With the blooming of Super-Multi-View (SMV) displays, hundreds of linearly or angularly arranged, horizontal parallax ultra-dense views are required. SMV, one specific scenario in the third phase of FTV, aims at displaying hundreds of very

dense and very wide baseline views [3]. Due to the limited bandwidths, it is not realistic to efficiently encode hundreds of views for viewers. In addition, different from 3D video (3DV), depth-image-based-rendering (DIBR) could not be applied for lack of depth map in SMV application.

As a simple and robust IBR method for generating new views, the 4D light field rendering simply combines and resamples the available images to generate the virtual views [4] without explicit depth map and corresponding information. Light field is a promising representation to describe the visual appearance of a scene and the special linear structure of Epipolar Plane Image (EPI) can be exploited to estimate accurate depth map which is essential for rendering quality.

The remainder of the paper is organized as follows. In section 2, we introduce the four-dimensional light field and analyze EPI in light field. Section 3 summarizes the rendering method based on light field, including depth estimation and view synthesis. Finally, conclusions are drawn in section 4.

2 Four-Dimensional Light Field and EPI

As we can describe each image by the location and orientation of the camera, light field is defined as the radiance to represent the related information [4]. The light field can be represented by a 5D function $f(x,y,z,\theta,\phi)$. The five dimensional quantity function describes the flow of light at every 3D space position (x,y,z) for every 2D direction (θ,ϕ) [5], and the intensity of the specified ray is f. When there is no light dispersion in the scene, all the coincident rays along a portion of free space (between solid or refractive objects) have the same color [6]. Under these conditions, the 5D representation can be reduced to 4D function as $L(u,v,s,t)$, which is called 4D light field. 4D light field is composed of two plane (u,v) and (s,t), as shown in **Fig. 1**. The (s,t) plane contains the focal points of the views, and the (u,v) plane means image plane. $L(u,v,s,t)$ can be viewed as an assignment of an intensity value to the ray passing through (u,v) and (s,t) [7].

It is difficult to think about 4D light field, so we fix the row value t to drop our visualization by one dimension. Along the s axis, an array of camera views is stacked. The 2D subspace $L(u,s)$ is called the EPI, and an example is shown in **Fig. 2**. EPI consists of Epipolar lines which are the intersection of epipolar plane and the camera plane in computer vision. In the EPI, adjacent line comes from adjacent view captured by camera. Assuming the scene is Lambertian, the rays contained in such planes have same intensity and color. If cameras are linearly arranged with the same interval, the EPI seems to be composed of many straight lines with different slopes (Fig. 2). With such constraints, let us consider the light field geometry (**Fig. 3**). $p=(x,y,z)$ is an arbitrary point in space, while A and B are the corresponding projection points on the image plane from adjacent camera. If the adjacent camera's distance is $\triangle s$, it is clear that the offset $\triangle x=x_2-x_1$ in the corresponding

Light Field Virtual View Rendering Based on EPI-Representations
SUN Yule and YU Lu
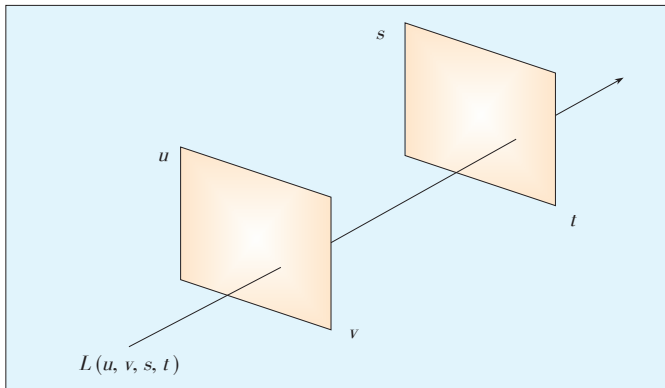
▲Figure 1. 4D light field.

image plane can be computed using (1)

$$\triangle x = -\frac{f}{Z}\triangle s \qquad (1)$$

where f is the focal length which is the distance between two parallel planes and Z is the depth of the point.

Since $d = \triangle x/\triangle s = -f/Z$ is a constant called disparity for one given 3D space point, the slope of the line which is composed of projection points from different views in EPI is related to its depth if the point is visible.

3 Light Field Rendering Analysis

Rendering methods based on light field analysis can be decomposed into 2 parts: depth estimation and view synthesis. Disparity is represented by the slope of lines in EPIs (Fig. 2). Disparity estimation is to detect the lines and calculate the corresponding slope (Sec. 3.1). Given the disparity information, virtual views are easy to synthesize by DIBR (Sec. 3.2).

3.1 Depth Estimation

The quality of the virtual views, and hence the quality of the 3DTV experience, relies heavily on the depth map accuracy [8]. Compared with traditional stereo matching methods, depth estimation through EPI characteristic can make use of more views' information. It is apparent that estimating the depth from more views can be more reliable and accurate. The depth estimation methods in EPI can be classified into two kinds: global line detection and local disparity estimation. Both the two methods estimate the depth map by detecting and computing the slope in EPIs. The main difference of the two methods is whether making full use of rows in EPI.

3.1.1 Global Line Detection Method

Global line detection is the most common and most established depth estimation method in light field analysis. Similar to the stereo matching method which looks for corresponding image features in two or more camera views, the global line detection method is to determine the slope for every pixel in the
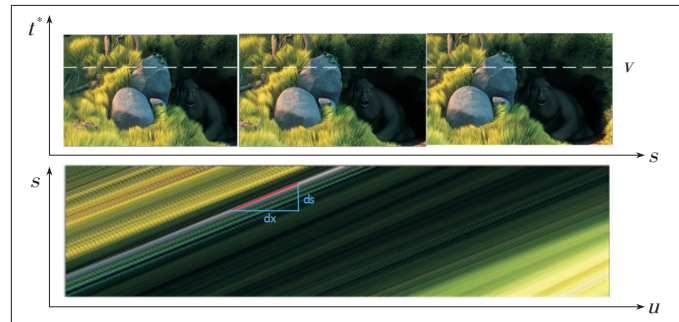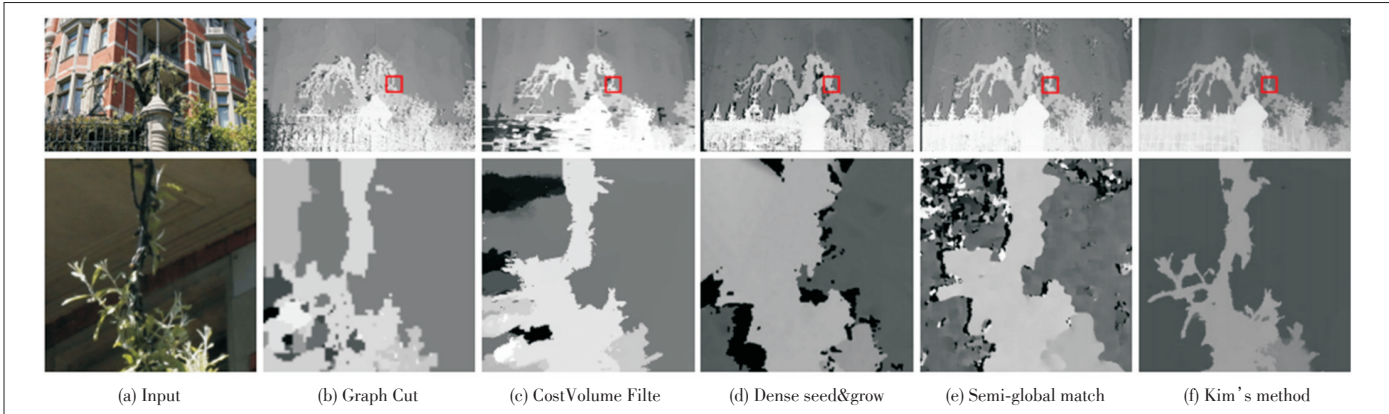

▲Figure 2. Camera views and EPI.
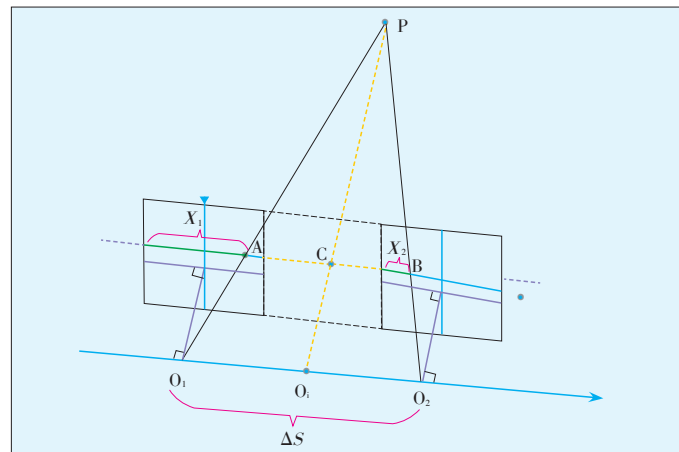
▲Figure 3. Geometry mapping.

EPI, i.e.to find the right corresponding projection points in different rows. The slope of the line is related to disparity described in detail in Section 2.

Although each pixel in EPI has its directionality associated with disparity, the correspondence may not be reliable if pixel is in smooth area because its surroundings are pixels with similar intensity. Based on this analysis, we give the general steps of global line detection method: 1) Extracting the characteristic pixels with high edges confidence in each EPI; 2) defining a cost function according to the coherence of intensity and gradient to determine directionality of each characteristic pixel, and 3) assigning disparity values to the pixels of the homogeneous interior region in EPI, for example, using interpolation [9] or fine to coarse procedure [10].

Compared with stereo matching, the global line detection method has higher reliability for the characteristic points. Many algorithms have been proposed to estimates disparity through locating the optimal slope of each line in EPIs [11]–[19]. Here we introduce two typical algorithms. Kim [10] proposed an algorithm which first estimates the reliable depth around object boundaries and then processes the homogeneous interior regions in a fine-to-coarse procedure, which can yield precise object contours and ensure smoothness in less detailed areas. According to the results in **Fig. 4**, Kim's method focuses on the closeups and has better performance in homogeneous

(a) Input (b) Graph Cut (c) CostVolume Filte (d) Dense seed&grow (e) Semi-global match (f) Kim's method

▲Figure 4. Kim's method compared with two-view stereo methods on the Mansion data set [10].

regions and areas around object contours. The latest method utilizes intensity pixel value, gradient pixel value, spatial consistency as well as reliability measure to select the best slope from a predefined set for every pixel in EPI. In this method, the spatial smoothness constraint can handle the pixels both in edge and in homogeneous regions [20].

In general, the disparity estimation is reliable by computing the slope of the line from plenty of views in the absence of occlusion. The occluded pixels at occlusion boundaries should be detected to alleviate error estimation, for example, re-projecting the EPIs and filling in occluded pixels using depth propagation [20]. However, the global line detection method is highly complex and may be challenging to run in real time with limited processing capabilities since every characteristic pixel needs to search all the rows in EPIs for choosing the optimal direction.

3.1.2 Local Disparity Estimation Method

Instead of detecting straight line with expensive matching, the local disparity estimation method by directly computing the slope in EPIs can obtain fast disparity estimation. Dansereau first proposed a directly computing method to extract depths applying gradient operators on the EPIs obtained from the light field [21]. To achieve higher quality, Wanner and Goldlucke [7] applied a structure tensor to measure pixels' directions on a 3*3 stencil in EPI. The structure tensor is also used in 4D field to estimate depth [22]. The structure tensor J applied in EPI plane (x, s) is presented in (2), where G_σ represents a Gaussion smoothing operator at an outer scale σ and S_x, S_s denote the horizontal and vertical gradients calculated on an inner scale ρ in EPI. Optimal inner and outer parameters are found after testing a number of different parameter combinations, and the disparity can be computed using (3).

$$J = \begin{bmatrix} G_\sigma*(S_xS_x) & G_\sigma*(S_xS_s) \\ G_\sigma*(S_xS_s) & G_\sigma*(S_sS_s) \end{bmatrix} = \begin{bmatrix} J_{xx} & J_{xs} \\ J_{xs} & J_{ss} \end{bmatrix} \qquad (2)$$

$$d_{x,s} = \frac{\Delta x}{\Delta s} = \frac{2J_{xs}}{J_{ss} - J_{xx}} \qquad (3)$$

To yield a higher quality depth map from the 4D light field, we need to add some optimization approaches. Since the local disparity estimation only takes into account the local structure in EPI, if there is no dominant orientation in the neighborhood, i.e. homogeneity or multiple orientations, the structure tensor cannot give a reliable estimation. Li [23] proposed a method to refine the certainty maps produced by the structure tensor. And the disparity map would be more convincing if applying consistent disparity labeling globally on the EPI [24]. However, this is a computationally very expensive procedure, and it is far too slow. In order to obtain quality disparity map, we employ denoising or globally optimization on individual views. **Table 1** shows the speed and accuracy of depth estimation

▼Table 1. Average disparity reconstruction accuracy and speed for this local disparity estimation method compared to a multi-view stereo method on light field benchmark database

Algorithm		Average run time		Accuracy		Disparity error			Depth error		
		All views	Single views	SSIM	MSE	≥1.0	≥0.5	≥0.1	≥1.0%	≥0.5%	≥0.1%
Proposed EPI algorithm	Structure tensor only	6 s	\	0.87	0.029	0.70	1.95	16.60	0.32	3.66	55.69
	TV-L¹ smoothing	1.6 s	0.02 s	0.94	0.018	0.43	1.28	9.85	0.18	2.45	47.83
	Global optimization	5.5 h	240 s	0.93	0.019	0.54	1.4	8.82	0.23	2.44	46.71
Multiview stereo (all 9*9 views)	Dataterm only	101 s	1.25 s	0.77	0.037	0.85	2.89	14.61	0.45	6.93	46.36
	TV-L¹ smoothing	1.6 s	0.02 s	0.91	0.023	0.54	1.71	9.13	0.19	3.68	41.65
	Global optimization	5.5 h	240 s	0.92	0.027	0.71	1.9	7.51	0.08	3.48	37.34
Multiview stereo (center views crosshair)	Dataterm only	44 s	0.55 s	0.74	0.043	0.93	3.48	18.73	0.58	8.77	51.44
	TV-L¹ smoothing	1.6 s	0.02 s	0.91	0.021	0.46	1.64	10.91	0.18	4.14	45.21
	Global optimization	5.5 h	240 s	0.92	0.023	0.57	1.71	8.34	0.09	3.17	39.64

Parameters for all methods were turned to yield an optimal SSIM measure, but for completeness MSE and the percentage of pixels with disparity or depth error larger than a given quantity are shown [7], and TV-L¹ means Total Variation regularization with L¹ norm.

EPI: Epipolar Plane Image   MSE: Mean Squared Error   SSIM: Structural Similarity

with optimization or not for this method compared to a multi-view stereo. It is clear that this method is much faster than the multi-view stereo because it does not need to match pixels at different locations. In addition, the accuracy of the depth map is a little better than the results using the multi-view stereo method since multi-view method is more likely to find the wrong matching patch, leading to lower accuracy.

3.2 View Synthesis Utilizing DIBR

After the step of disparity estimation, we synthesize arbitrary virtual views we need. View synthesis can be done by utilizing DIBR to generate new views with depth map available. Compared with the MPEG depth estimation which uses stereo matching based on 2 or 3 input views and view synthesis reference software, the results of the synthesized views show that light field method is better objectively for most of the scenes and always visually better [25]. The DIBR procedure is briefly introduced below.

For the camera is 1D parallel arrangement, once the depth value z is calculated, we can use DIBR easily by mapping every pixel from the reference view to the virtual view using (4) [26].

$$x_v = x_r - \frac{f \cdot (t_{x,v} - t_{x,r})}{Z} = x_r - d \qquad (4)$$

where x_v and x_r are the positions of x in camera plane in virtual and real view respectively, and d is disparity which is related to the depth z and the baseline $(t_{x,v} - t_{x,r})$.

Applying the DIBR method, every pixel in reference view can be projected to the virtual view with disparity map according to (4). The geometry matching procedure is illustrated in Fig. 3. For a point C in virtual view, its unique corresponding point in reference view is A. If C is a point in integer pixel position, the intensity of C is equal to A. However, during the procedure of mapping, maybe more than one point in the reference view map to one point in the virtual view. To solve this problem, we can choose the corresponding pixel which has smallest depth in reference view. Another problem is that some positions in the virtual view cannot find their corresponding points in the reference views because of occlusion. Therefore, hole filling needs to be done, which has great impact on the quality of synthesized views. We can preprocess the depth data or inpaint the synthesizing view [27] to deal with this problem.

4 Conclusion

In this paper, we summarize several major methods of light field virtual view rendering based on EPI-representations in SMV application. Compared with stereo matching, the method of light field analysis can obtain better depth estimation. With a more accurate depth map, it is apparent that the virtual views' quality is higher. In addition, rendering complexity is also in-

cluded in this paper. Methods based on global line detection to estimate depth is more time-consuming than the methods based on local disparity estimation. Apparently, the local disparity estimation method saves more time at the cost of degrading the quality of virtual views. For current use cases, it is more suitable to use local disparity estimation to compute the depth map for SMV application because of the requirement of real-time.

References
[1] "Draft Call for Evidence on FTV," ISO/IEC JTC1/SC29/WG11 MPEG2015/N15095, Feb. 2015.
[2] M. Tanimoto and T. Fujii, "FTV-Free Viewpoint Television," ISO/IEC JTC1/SC29/WG11, M8595, Jul. 2002.
[3] "Use Cases and Requirements on Free-viewpoint Television (FTV)," ISO/IEC JTC1/SC29/WG11, Oct. 2015.
[4] M. Levoy and P. Hanrahan, "Light field rendering," in *Proc. 23rd ACM Annual Conference on Computer Graphics and Interactive Techniques*, New Orleans, USA, 1996, pp. 31−42. doi: 10.1145/237170.237199.
[5] S. J. Gortler, R. Grzeszczuk, R. Szeliski, et al., "The lumigraph," in *Proc. 23rd ACM Annual Conference on Computer Graphics and Interactive Techniques*, New Orleans, USA, 1996, pp. 43-54. doi: 10.1145/237170.237200.
[6] R. Szeliski, *Computer Vision: Algorithms and Applications*. Berlin Heidelberg, Germany: Springer Science & Business Media, 2010, pp. 628−628.
[7] 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, no. 3, pp. 606−619, 2014.
[8] S. Schwarz, R. Olsson, and M. Sjostrom, "Depth sensing for 3DTV: a survey," *MultiMedia*, vol. 20, no. 4, pp. 10−17, 2013.
[9] G. Jiang, M. Yu, X. Ye, et al., "New method of ray-space interpolation for free viewpoint video," in *IEEE International Conference on Image Processing*, Genova, Italy, 2005. doi: 10.1109/ICIP.2005.1530261.
[10] C. Kim, H. Zimmer, Y. Pritch Y, et al., "Scene reconstruction from high spatio-angular resolution light fields," *ACM Transactions on Graphics*, vol. 32, no. 4, article no. 73. doi: 10.1145/2461912.2461926.
[11] T. Ishibashi, M. P. Tehrani, T. Fujii, et al., "FTV format using global view and depth map," in *IEEE Picture Coding Symposium (PCS)*, Krakow, Poland, 2012, pp. 29−32.
[12] L. Fan, X. Yu, Z. Shu, et al., "Multi-view object segmentation based on epipolar plane image analysis," in *IEEE International Symposium on Computer Science and Computational Technology*, Shanghai, China, Dec. 2008, pp. 602−605.
[13] L. Jorissen, P. Goorts, S. Rogmans, et al., "Multi-camera epipolar plane image feature detection for robust view synthesis," in *IEEE 3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video*, Lisbon, Portugal, 2015, pp. 1−4.
[14] T. Ishibashi, M. P. Tehrani, T. Fujii, et al., "FTV format using global view and depth map," in *IEEE Picture Coding Symposium (PCS)*, Krakow, Poland, 2012, pp. 29−32.
[15] M. Matoušek, T. Werner, and V. Hlaváč, "Accurate correspondences from epipolar plane images," in *Proc. Computer Vision Winter Workshop*, Bled, Slovenia, 2001, pp. 181−189.
[16] A. Criminisi, S. B. Kang, R. Swaminathan, et al., "Extracting layers and analyzing their specular properties using epipolar-plane-image analysis," *Computer Vision and Image Understanding*, vol. 97, no. 1, pp. 51−85, 2005.
[17] 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, no. 1, pp. 7−55, 1987.
[18] F. C. Calderon, C. Parra, and C. L. Niño, "Depth map estimation in light fields using an stereo-like taxonomy," in *IEEE XIX International Symposium on Image, Signal Processing and Artificial Vision*, Armenia, Colombia, 2014, pp. 1−5.
[19] M. Uliyar, G. Putraya, S. V. Basavaraja, "Fast EPI based depth for plenoptic cameras," in *IEEE International Conference on Image Processing (ICIP)*, Melbourne, Australia, 2013, pp. 1−4.
[20] H. Lv, K. Gu, Y. Zhang, et al., "Light field depth estimation exploiting linear structure in EPI," in *IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, Torino, Italy, 2015, 1−6.
[21] D. Dansereau and L. Bruton, "Gradient-based depth estimation from 4d light

fields," in *Proc. IEEE International Symposium on Circuits and Systems*, Vancouver, Canada, 2004, pp. 549−552. doi: 10.1109/ISCAS.2004.1328805.

[22] J. Luke, F. Rosa, J. Marichal, et al., "Depth from light fields analyzing 4D local structure," *Journal of Display Technology*, vol. 11, no. 11, Nov. 2015. doi: 10.1109/JDT.2014.2360992.

[23] J. Li and Z. N. Li, "Continuous depth map reconstruction from light fields," in *IEEE International Conference on Multimedia and Expo*, San Jose, USA, 2013, pp. 1−6.

[24] S. Wanner and B. Goldluecke, "Globally consistent depth labeling of 4D light fields," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Providence, USA, 2012, pp. 41−48.

[25] L. Jorissen, P. Goorts, B. Bex, et al., "A qualitative comparison of MPEG view synthesis and light field rendering," in *IEEE 3DTV-Conference: The True Vision -Capture, Transmission and Display of 3D Video*, Budapest, Hungary, 2014, pp. 1−4.

[26] D. Tian, P. L. Lai, P. Lopez, et al., "View synthesis techniques for 3D video," in *SPIE Optical Engineering Applications, International Society for Optics and Photonics*, San Diego, USA, 2009, 74430T-74430T-11. doi:10.1117/12.829372.

[27] I. Daribo, H. Saito, R. Furukawa R, et al., "Hole filling for view synthesis," in *3D-TV System with Depth-Image-Based Rendering*. New York, USA: Springer, 2013, pp. 169−189. doi: 10.1007/978-1-4419-9964-1_6.

Biographies

SUN Yule (sunyule@zju.edu.cn) received his BEng degree in information and communication engineering from Zhejiang University, China in 2015. He is currently completing his PhD degree at the Institute of Information and Communication Engineering, Zhejiang University. His research interests include 3D video processing and video coding.

YU Lu (yul@zju.edu.cn) received her BEng degree in radio engineering and PhD degree in communication and electronic systems from Zhejiang University, China in 1991 and 1996. She is currently a professor at the Institute of Information and Communication Engineering, Zhejiang University. In 2002, she was a senior visiting scholar at the University of Hannover, German, supported by the China Scholarship Council and German Research Foundation. In 2004, she was a senior visiting scholar at the Chinese University of Hong Kong, supported by the United College Resident Fellow Scheme. She has published more than 100 technical papers and contributed more than 200 proposals to national and international standards. Her research interests include video coding, multimedia communication, and relative application-specific integrated circuit design. She is the chair of the Video Subgroup of the Audio Video Coding Standard (AVS) of China and was previously co-chair of the Implementation Subgroup of AVS.

Call for Papers

ZTE Communications Special Issue on

Channel Measurement and Modeling for Heterogeneous 5G

While cellular networks have continuously evolved in recent years, the industry has clearly seen unprecedented challenges to meet the exponentially growing expectations in the near future. The 5G system is facing grand challenges such as the ever−increasing traffic volumes and remarkably diversified services connecting humans and machines alike. As a result, the future network has to deliver massively increased capacity, greater flexibility, incorporated computing capability, support of significantly extended battery lifetime, and accommodation of varying payloads with fast setup and low latency, etc. In particular, as 5G requires more spectrum resource, higher frequency bands are desirable. Nowadays, millimeter wave has been widely accepted as one of the main communication bands for 5G. As a result, envisioned 5G research and development are inclined to be heterogeneous, with possibly ultra dense network layouts due to their capability to support high speed connections, flexibility of resource management, and integration of distinct access technologies.

Towards the heterogeneous 5G, the first and foremost hurdle lies in the channel measurement and modeling in the broad and diversified 5G scenarios. This special issue is dedicated to providing a platform to share and present the latest views and developments on 5G channel measurement and modeling issues.

Schedule

Submission Deadline: November 1, 2016

Final Decision Due: December 1, 2016

Final Manuscript Due: December 15, 2016

Publication Date: February 25, 2017

Guest Editors

Prof. CUI Shuguang, University of California, Davis, USA. Email: sgcui@ucdavis.edu

Prof. CHENG Xiang, Peking University, China. Email: xiangcheng@pku.edu.cn

Paper Submission

Please directly send to sgcui@ucdavis.edu and xiangcheng@pku.edu.cn, using the email subject "ZTE−CMMH5G−Paper−Submission".