https://doi.org/10.2352/ISSN.2470-1173.2020.2.SDA-098

Dynamic Zero-Parallax-Setting Techniques for Multi-View Autostereoscopic Display

Yuzhong Jiao; United Microelectronics Centre; Hong Kong, China
Mark Ping Chan Mok; United Microelectronics Centre; Hong Kong, China
Kayton Wai Keung Cheung; United Microelectronics Centre; Hong Kong, China
Man Chi Chan; United Microelectronics Centre; Hong Kong, China
Tak Wai Shen; United Microelectronics Centre; Hong Kong, China
Yiu Kei Li; United Microelectronics Centre; Hong Kong, China

Abstract

The objective of this paper is to research a dynamic computation of Zero-Parallax-Setting (ZPS) for multi-view autostereoscopic displays in order to effectively alleviate blurry 3D vision for images with large disparity. Saliency detection techniques can yield saliency map which is a topographic representation of saliency which refers to visually dominant locations. By using saliency map, we can predict what attracts the attention, or region of interest, to viewers. Recently, deep learning techniques have been applied in saliency detection. Deep learning-based salient object detection methods have the advantage of highlighting most of the salient objects. With the help of depth map, the spatial distribution of salient objects can be computed. In this paper, we will compare two dynamic ZPS techniques based on visual attention. They are 1) maximum saliency computation by Graphic-Based Visual Saliency (GBVS) algorithm and 2) spatial distribution of salient objects by a convolutional neural networks (CNN)-based model. Experiments prove that both methods can help improve the 3D effect of autostereoscopic displays. Moreover, the spatial distribution of salient objects-based dynamic ZPS technique can achieve better 3D performance than maximum saliency-based method.

Introduction

With the advancement of computer graphics and display technologies, more multi-view autostereoscopic displays have been manufactured in recent years [1-4]. Depth image-based rendering (DIBR) techniques have been widely employed to generate 3D multi-view images. For instance, a popular input format for 3DTV called 2D plus depth (or 2D+Z) uses a conventional color video and a depth map [5].

In the stereoscopic displays, it is common to adjust the location of the scene relative to the screen plane by manipulating the zero-parallax setting (ZPS) [6-10]. The ZPS defines the position of the convergence plane in the scene. The ZPS adjustment is often used for artistic purposes to control how much of the scene is positioned in front of or behind the screen plane. The most important role of the dynamic ZPS techniques is to reduce the visual discomfort from the vergence-accommodation conflict (VAC). Comparatively, autostereoscopic displays don't have serious VAC problem because of tens of viewpoints, even more.

However, increasing viewpoints provides smooth motion parallax but it may lead to blurry 3D vision for images with large disparity. Fig. 1 shows spatial resolution loss when disparity is not zero. One spatial object is composed of one red point and one green point. When the spatial object is displayed on the zero-parallax-plane, its whole picture can be seen in the front of screen, as shown in Fig. 1a. However, when it's not shown on the zero-parallax-plane, just part of the whole picture can be visible, as shown in Fig.1b and 1c. Because of the limitation of screen resolution, just the red point can be seen in Fig. 1b and just the green point can be seen in Fig. 1c. The bigger the absolute value of disparity is, the larger the spatial resolution loss is. With the increase of viewpoint, the loss rises at the same time. Dynamic ZPS method is still good idea to balance the conflict between viewpoint and spatial resolution in autostereoscopic displays. In [11-13], scholars proposed dynamic disparity control methods for multi-view images.

The objective of this paper is to research a dynamic computation of ZPS for autostereoscopic displays in order to effectively alleviate blurry 3D vision for images with large disparity. Saliency detection techniques can yield saliency map which is a topographic representation of saliency which refers to visually dominant locations [14-18]. By using saliency map, we can predict what attracts the attention, or region of interest, to viewers. It's feasible to generate high-quality 3D images by moving the regions of interest to the zero-parallax plane. Recently, deep learning techniques have been applied in saliency detection [19-24]. Deep learning-based salient object detection methods have the advantage of highlighting most of the salient objects. With the help of depth map, the spatial distribution of salient objects can be computed. It is possible to alleviate blurry 3D vision by moving the spatial center of salient objects to the zero-parallax plane. In this paper, we will compare two dynamic ZPS techniques for multi-view autostereoscopic display. They are 1) maximum saliency computation by graph-based visual saliency (GBVS) algorithm and 2) spatial distribution of salient objects by a convolutional neural networks (CNN)-based model.

The rest of the paper is organized as follows. Section 2 introduces two saliency detection methods: 1) saliency map generation and 2) salient objection detection. Then Section 3 describes two dynamic ZPS techniques for multi-view autostereoscopic displays based on the two saliency detection methods. Experimental results are reported in Section 4. Finally, Section 5 concludes the paper.

Saliency Detection

There are two kinds of saliency detection techniques, namely saliency map generation and salient object detection. The main goal of saliency map generation is to compute a saliency map that topographically represents the level of saliency for visual attention, while salient object detection usually generates bounding boxes, binary foreground and background segmentation. Fig. 2 shows the difference of two kinds of saliency maps from saliency map generation and salient object detection.

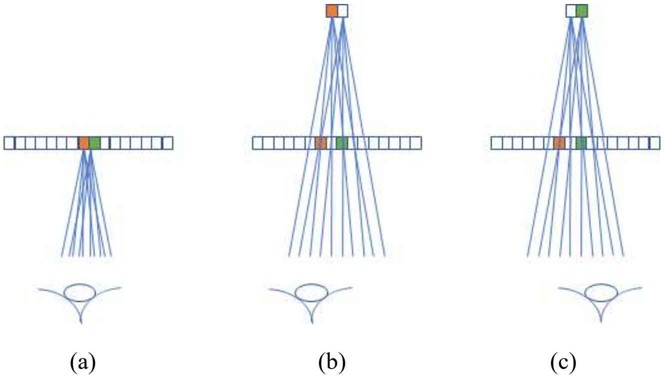

Figure 1. Spatial resolution loss, (a) when disparity is zero, (b) and (c) with two different viewing angles when disparity is not zero

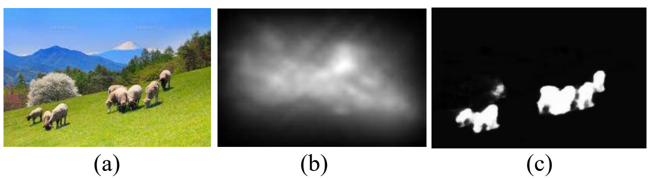

Figure 2. Saliency detection example, (a) original image, (b) the result from saliency map generation and (c) that from salient object detection

Saliency Map Generation

A saliency map is computed for each color channel independently by using the algorithm of GBVS [14]. As a simple bottom-up visual saliency model, GBVS is very powerful for making reliable predictions on human fixation. Referring to the structure of the RGB-Signature algorithm [17], a saliency detection algorithm with the 2D+Z information is designed for 3DTV applications in [25]. The input 2D+Z image is decomposed into four channels, including three color channels and one depth channel. The final saliency map is simply the average of the four-channel saliency maps. From saliency map generation, it's easy to obtain the most salient region in an image. Based on the depth value of the most salient region for 3D-format image, we can move the region to the zero-parallax plane to alleviate blurry 3D images in autostereoscopic displays.

Salient Object Detection

Compared with traditional saliency map generation algorithms, deep learning-based salient object detection methods can generate saliency map with sharp boundaries and highlighting most of the salient objects. From the saliency map, we can easily get the locations and shapes of salient objects. With the help of the depth information from 3D-format image, the spatial distribution of salient objects can be extracted.

A CNN-based model [24] is used in this paper for salient object detection. This method introduces short connections to the skip-layer structures within the holistically-nested edge detection architecture. It takes full advantage of multi-level and multi-scale features extracted from fully convolutional neural networks (FCNs), providing more advanced representations at each layer. This method produces state-of-the-art results with the advantages of high efficiency and simplicity.

Saliency Detection-based Schemes

In the first scheme, we use GBVS algorithm to obtain saliency map, from which the depth corresponding to the maximum saliency can be gotten. The depth value with zero parallax, d_{zps}, is simply set to the depth of the most salient position, the new d_{zps} can be denoted as

$$d_{zps} = d(x_{ms}, y_{ms}) \tag{1}$$

where (x_{ms}, y_{ms}) is the pixel location with the maximum saliency. This means the most salient position is moved to the zero-parallax plane as shown in Fig. 3a. In the figure, Position A1 owns the most saliency. After dynamic ZPS, the position is moved to screen.

Another kind of saliency map which shows salient objects is computed by using a CNN-based model. On this ground, in the second scheme, the average of the depth values corresponding to the pixels of salient objects is employed as d_{zps}. The new d_{zps} can be denoted as

$$d_{zps} = \frac{1}{N}\sum_{i=1}^{N} d_i(x_s, y_s) \tag{2}$$

where (x_s, y_s) is the pixel located at salient objects and N is the total number of pixels in salient objects. The average with the information of the spatial distribution of salient objects can be considered as the center of salient objects in the depth direction perpendicular to screen. Fig. 3b shows the setting of zero-parallax considering spatial distribution. In the scene, there are two positions B1 and B2, corresponding to two salient objects. The center of the two objects is away from screen. After dynamic ZPS, the center position is moved to screen.

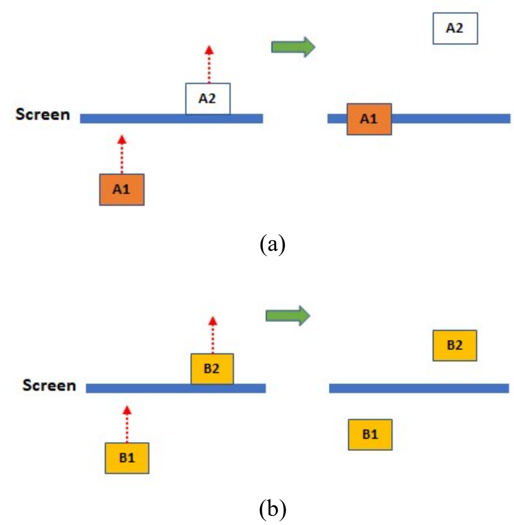

Figure 3. Dynamic ZPS principles by using (a) the most salient region and (2) the spatial distribution of salient objects

Dynamic ZPS Technique

In this paper, we utilize two dynamic ZPS schemes based on saliency detection. They are 1) maximum saliency scheme by GBVS algorithm and 2) spatial distribution scheme by a CNN-based model. In order to illustrate the feasibility of the schemes, a DIBR method is presented for autostereoscopic displays. Fig. 4 shows the system structure of dynamic ZPS-based multi-view autostereoscopic display. The system includes three parts: dynamic ZPS, disparity map generation and multi-view rendering. In the part of dynamic ZPS, two kinds of saliency detection algorithms are used to generate saliency maps and then dynamic schemes are employed to obtain d_{zps}. In disparity map generation, depth map is converted into disparity map by the adjustment of d_{zps}. Multi-view rendering is utilized to convert 2D RGB image and disparity map into 3D multi-view image, which can be shown in autostereoscopic display.

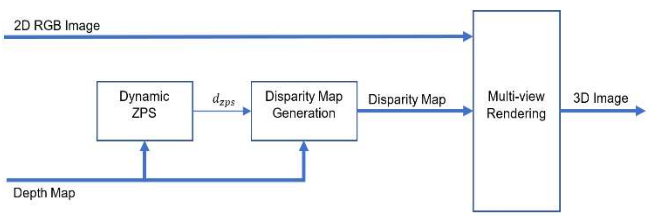

Figure 4. System structure of dynamic ZPS-based multi-view autostereoscopic display

Fig. 5 gives the algorithm flowchart of dynamic ZPS techniques. After one primary d_{zps} is computed by using saliency detection algorithm, we need to check whether the d_{zps} is suitable. Depth map for 2D+Z format usually have a range from 0 to 255. The two numbers denote the nearest and farthest distances from viewer. The distances are not absolute but relative. For common autostereoscopic displays, 128, the middle value of 0 and 255, is usually considered as the position of zero-parallax plane. If the primary d_{zps} estimated by saliency detection is larger than 128, we need move the visual objects to viewer with the distance of d_{zps}-128. If the visual point with the minimum depth is still in the range from 0 to 255 after moving, we must limit the value of d_{zps} by the following equation

$$d_{min} - (d_{zps} - 128) \geq 0 \quad (3)$$

Similarly, if d_{zps} is smaller than 128, there is another limitation as shown below

$$d_{max} + (128 - d_{zps}) \leq 255 \quad (4)$$

Eq. 3 and Eq. 4 can be shortened by

$$d_{max} - 127 \leq d_{zps} \leq d_{min} + 128 \quad (5)$$

Eq. 5 means that the maximum d_{zps}, zps_{max} equals to d_{min}+128 and the minimum d_{zps}, zps_{min} equals to d_{max}-127. The ZPS range is from zps_{min} to zps_{max}.The equations can help prevent 3D scene from moving out of the depth range from 0 to 255. Otherwise, the depth range of 3D scene would be shortened. This maybe leads to the deformation of 3D scene.

Experiment Results

In the test, we compare three ZPS methods: depth=128, GBVS-based maximum saliency, and CNN-based spatial distribution. Totally thirty-six sets of multi-view images are prepared before test. However, nine sets are excluded because of two following situations:

- The d_{zps} values for the methods of both maximum saliency and spatial distribution are same. This means there are no difference in the multi-view images with the two methods.
- The salient regions for the two methods don't overlap. Both GBVS algorithm and CNN-based model can generate wrong saliency map. Non-overlapping of salient regions may be caused by the incorrect results of one or both algorithms.

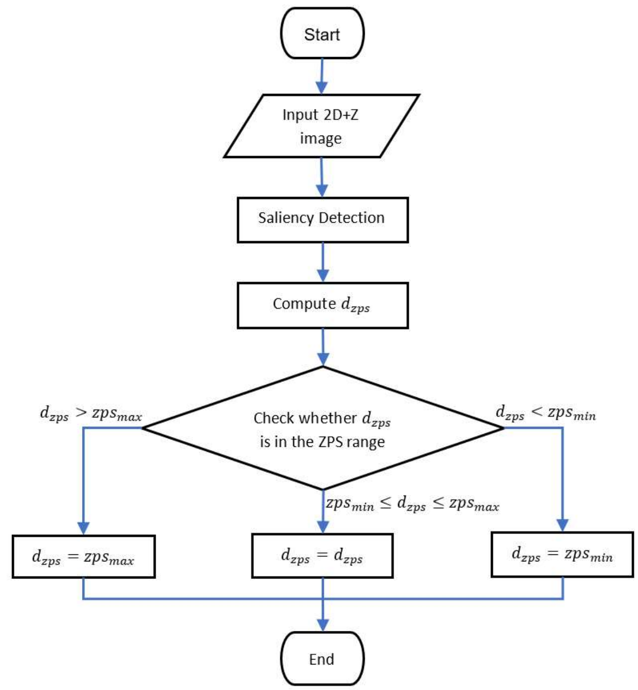

Figure 5. Algorithm flowchart of dynamic ZPS techniques

Thus, there are twenty-seven sets left to compare for each viewer. Two of the three version images are selected randomly for each set. Each participant just needs to compare 27 pairs of 3D images to avoid serios visual fatigue. Double-stimulus continuous quality-scale (DSCQS) method [26] is used to assess the subjective quality of each pair of multi-view 3D images. For the subjective assessment aiming at the comparisons of the three ZPS methods in 3D image quality, the following question is designed

- What the level of the 3D effect of the image is.

There are five levels from 1 to 5, representing bad, poor, fair, good and excellent. After the assessment of thirty persons with the ages from 15 to 52, the mean scores for the images processed by different methods are obtained.

Fig. 6 shows the subjective assessment results for the three ZPS methods. Both GBVS-based maximum saliency and CNN-based spatial distribution can help improve the 3D effect of autostereoscopic displays. Moreover, CNN-based spatial distribution method has the advantage over GBVS-based maximum saliency method.

Furtherly, we analyze the influence of the d_{zps} difference between GBVS-based maximum saliency and CNN-based spatial distribution methods on the subjective assessment results. Fig. 7

gives the analysis results that the score increments are large when the d_{zps} difference is in the middle range from 10 to 20, while the score increments are small in both the little d_{zps} difference range from 0 to 10 and the large difference range larger than 20. It's easy to understand that little d_{zps} difference causes small score increments. Interesting is that the scores for large d_{zps} difference are small. We find that the salient objects for the samples with large d_{zps} difference cover larger depth ranges.

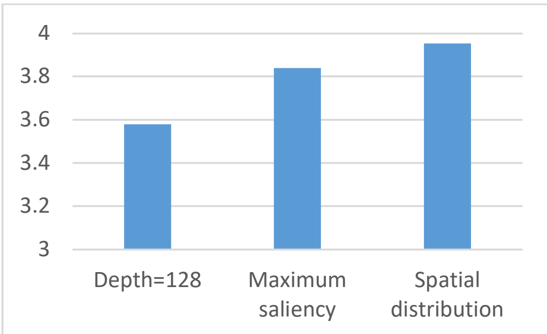

Figure 6. Subjective assessment results of three ZPS methods

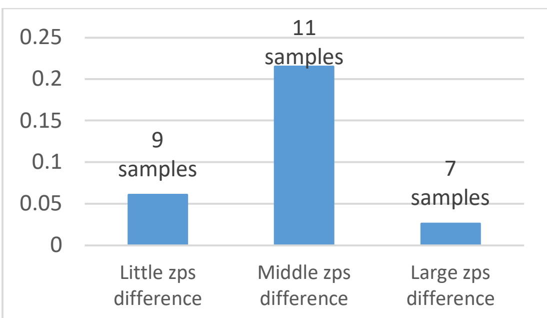

Figure 7. Increment of subjective assessment scores for maximum saliency and spatial distribution based ZPS methods according to the difference of their d_{zps} values.

Conclusions

In common autostereoscopic displays, a fixed d_{zps} value is usually used for different 2D+Z images and videos. This paper presents two dynamic ZPS techniques. They are based on the maximum saliency computed by GBVS algorithm and the spatial distribution of salient objects obtained from a CNN-based model respectively. Experiments prove that the multi-view rendering system with dynamic ZPS techniques have the advantage of generation high-quality 3D images over that with fixed d_{zps} value and the spatial distribution of salient objects-based dynamic ZPS technique can achieve better 3D performance than maximum saliency-based method for autostereoscopic displays.

References

[1] P. Merkle, A. Smolic and etc., "Multi-view Video plus Depth Representation and Coding," in IEEE International Conference on Image Processing, San Antonio, USA, 2007.

[2] A. Smolic, K. Muller and etc., "Intermediate View Interpolation based on Multiview Video plus Depth for Advanced 3D Video Systems," in the 15th IEEE International Conference on Image Processing, San Diego, USA, 2008.

[3] E. K. Lee, Y. S. Ho, "Generation of High-quality Depth Maps Using Hybrid Camera System for 3-D Video," Journal of Visual Communication and Image Representation, vol. 22, iss. 1, pp. 73-84, 2010.

[4] Yuzhong Jiao, Chan Man Chi Chan and Mark, P. C. Mok, "Dynamic Multi-View Autostereoscopy," in IS&T International Symposium on Electronic Imaging, Stereoscopic Displays and Applications XXX, Burlingame, USA, 2019.

[5] A. Redert, M. O. de Beeck and etc., "ATTEST: Advanced Three-dimensional Television System Technologies," in First International Symposium on 3D Data Processing Visualization and Transmission, Padova, Italy, 2002.

[6] Quan Huynh-Thu, Marcus Barkowsky and Patrick Le Callet, "The Importance of Visual Attention in Improving the 3D-TV Viewing Experience: Overview and New Perspectives," IEEE Transactions on Broadcasting, Vol. 57(2), pp.421-431, 2011.

[7] Karim Benzeroual, Robert S. Allison and Laurie M. Wilcox, "3D display size matters: Compensating for the perceptual effects of S3D display scaling," in IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Providence, RI, USA, 2012.

[8] Jae Gon Kim and Jun-Dong Cho, "Optimizing a Virtual Re-Convergence System to Reduce Visual Fatigue in Stereoscopic Camera," IEICE Trans. Inf. & Syst., Vol.E95-D, No.5, pp.1238-1247, 2012.

[9] Petr Kellnhofer, Piotr Didyk and etc. "GazeStereo3D: seamless disparity manipulations." ACM Transactions on Graphics, Vol. 35(4), pp. 1-13, 2016.

[10] Feng Shao, Wenchong Lin and etc., "Optimizing Multiview Video Plus Depth Retargeting Technique for Stereoscopic 3D Displays," Optics Express, Vol. 25, No. 11, pp. 12478-12492, 2017.

[11] Jianjun Lei, Shiqiu Li and etc., "Stereoscopic Visual Attention Guided Disparity Control for Multiview Images," Journal of Display Technology, Vol. 10, no. 5, pp. 373-379, 2014.

[12] Songlin Xie, Peng Wang and etc., "Depth-tunable three-dimensional display with interactive light field control," Optics Communications Volume, Vol. 371, pp. 166-172, 2016.

[13] Dongxue Han, Hui Chen and etc., "View synthesis using foreground object extraction for disparity control and image inpainting," Journal of Visual Communication and Image Representation, Vol. 56, pp. 287-295, 2018.

[14] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," Advances in Neural Information Processing Systems, no. 19, pp. 545–552, 2007.

[15] Y. Fang, Z. Chen and etc., "Saliency Detection in the Compressed Domain for Adaptive Image Retargeting," IEEE Transactions on Image Processing, vol. 21, no. 9, pp. 3888-3901, 2012.

[16] S. Goferman, L. Zelnik-Manor and A. Tal, "Context-Aware Saliency Detection," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, iss. 10, pp. 1915-1926, 2012.

[17] X. Hou, J. Harel and etc., "Image Signature: Highlighting Sparse Salient Regions," IEEE Transactions on Pattern Analysis and Machine

Intelligence, vol. 34, no. 1, pp. 194-201, 2012.

[18] H. Cheng, J. Zhang and etc., “A Novel Saliency Model for Stereoscopic Images,” in 2005 International Conference on Digital Imaging Computing: Techniques and Applications (DICTA), Adelaide, Australia, 2015.

[19] Qiong Yan, Li Xu and etc., “Hierarchical Saliency Detection,” in the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Portland, USA, 2013.

[20] Chuan Yang, Lihe Zhang and etc., “Saliency Detection via Graph-Based Manifold Ranking,” in the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Portland, USA, 2013.

[21] Xiang Wang, Huimin Ma and Xiaozhi Chen, “Salient Object Detection via Fast R-CNN and Low-level Cues,” in 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 2016.

[22] Guanbin Li and Yizhou Yu, “Deep Contrast Learning for Salient Object Detection,” in the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, USA, 2016.

[23] Nian Liu and Junwei Han, “DHSNet: Deep Hierarchical Saliency Network for Salient Object Detection,” in the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, USA, 2016.

[24] Qibin Hou, Ming-Ming Cheng and etc, “Deeply Supervised Salient Object Detection with Short Connections,” in the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Hawaii, USA, 2017.

[25] Yuzhong Jiao, Chan Man Chi Chan and Mark, P. C. Mok, “Saliency Map based Multi-View Rendering for Autostereoscopic Displays,” in IS&T International Symposium on Electronic Imaging, Stereoscopic Displays and Applications XXX, Burlingame, USA, 2019.

[26] Recommendation ITU-R BT.500-7 (Revised), 1996. Methodology for the subjective assessment of the quality of television pictures.

Author Biography

Yuzhong Jiao received his MS in test and measure technology and instrument from Xiamen University (2001) and his Ph. D. in micro-electronics and solid-state electronics from Peking University (2009). From 2001 to 2004, he worked in the State Key Laboratory of Optical Technologies for Micro-fabrication in Chinese Academy of Sciences (CAS). And from 2010 to 2019, he worked in the IC Design Group (ICD) and the Advanced Digital Systems Division (ADS) at the Hong Kong Applied Science and Technology Research Institute (ASTRI). Now he works in United Microelectronics Centre (Hong Kong). His current research interests include auditory/visual immersion, AI chip design and AI applications.

Mark Ping Chan Mok received his Bachelor’s and Master degrees in Electronic Engineering from The Hong Kong University of Science and Technology in 1997 and 2000 respectively. He has over 22 years working experience in various industries, including IC design, EDA and R&D. He was a outstanding technical leader in stereoscopic/autostereoscopic 3D related technologies and has led the team to successfully deliver many glasses-free 3D display products in the market. Currently, Mark is the Director of United Microelectronics Centre (Hong Kong). His work mainly focuses on the research and development of AI chip and diverse applications of edge computing in hardware platform.

Kayton Wai Keung Cheung received his Bachelor’s and Ph.D. degrees in Electronic Engineering from City University of Hong Kong in 2008 and 2012. He has around 8 years R&D experience in image processing, 3D rendering and computer vision. Currently, Kayton is a Senior Engineer of United Microelectronics Centre (Hong Kong) and his research work focuses on the algorithm design and development for AI applications.

Chan Man Chi received his Ph.D. degree in Information Engineering from The Chinese University of Hong Kong in 2000. He graduated in Electronic Engineering from Hong Kong Polytechnic University at 1995 where he received his M.Phil. degree in 1997. He is currently a senior principal engineer at United Microelectronics Centre (Hong Kong) and focuses on edge computing software system design and development. Prior to joining UMEC, he worked in ASTRI responsible for SoC firmware and application system development. Before that, he was a visiting lecturer at The Chinese University of Hong Kong in 2000 and has worked in a local startup company REnex Technology from 2001 to 2005. Dr Chan has over fifteen years of software development experience in various embedded systems including wireless sensor networks, ultra wide band (UWB) MAC chips, multimedia processor SoC and power line communications systems.

Tak-Wai Shen received the B. Eng (Hons.) degree in Electronic Engineering from Oxford Brookes University (England) in 1992. Afterwards, he obtained his M.Phil. and Ph.D. degree in Electronic and Information Engineering from the Hong Kong Polytechnic University in 1998 and 2015, respectively. He has over 20 years working experience in various industries, including Manufacturing, Quality Engineering and R&D. Currently, he is senior engineer of United Microelectronics Centre (Hong Kong). His research interests include image and video processing, fast algorithms and speech enhancement. He is member of the IEEE, USA and IET, UK.

Yiu Kei Li earned his Master’s Degree in Computer and Electrical Engineering at the University of Iowa and his MBA at the University of San Diego in the USA. He also earned an Executive Certificate in Management and Leadership from the Massachusetts Institute of Technology. Mr. Li has more than 30 years of management experience in multinational and regional operations in IC design and semiconductor-related industries. Before joining UmecHK, he held important marketing and technical positions in several IC design and manufacturing companies in Hong Kong, China, and the US. He was responsible for strategic planning, supervising commercialization and product innovation activities. Mr. Li is active in his participation in activities organized in the industry. He holds several external appointments including a member of Hong Kong Information and Technology Joint Council, visiting professor of Peking University, Fellow of The Hong Kong Association for the Advancement of Science and Technology, Advisory committee member on Electronic Engineering in CUHK and Senior Member of IEEE.