https://doi.org/10.2352/ISSN.2470-1173.2019.3.SDA-632

Dynamic Multi-View Autostereoscopy

Yuzhong Jiao; Hong Kong Applied Science and Technology Research Institute (ASTRI); Hong Kong, China
Chan Man Chi; Hong Kong Applied Science and Technology Research Institute (ASTRI); Hong Kong, China
Mark, P. C. Mok; Hong Kong Applied Science and Technology Research Institute (ASTRI); Hong Kong, China

Abstract

With advantages of motion parallax and viewing convenience, multi-view autostereoscopic displays have attracted increasing attention in recent years. It is obvious that increasing the number of views improves the quality of 3D images/videos and leads to better motion parallax. However, it requires huge amount of computing resources to generate large numbers of view images in real time. In principle, objects appearing near the screen plane have very small absolute disparity. It can use fewer views to present these objects for achieving the same level of motion parallax. The concept of dynamic multi-view autostereoscopy is to dynamically control the number of views to generate for the points in 3D space based on their disparity. Points with larger absolute disparity use more views, while points with smaller absolute disparity use fewer views. As a result, fewer computing resources are required for real-time generation of view images. Subjective assessments show that only slight degradation in 3D experience is resulted on its realization over 2D plus depth based multi-view autostereoscopic display. However, the amount of computation for generating view images can be reduced by about 44.3% when 3D scenes are divided into three spaces.

Introduction

Multi-view autostereoscopic displays show immersive three-dimension (3D) images without the need for the viewer to wear glasses [1-2]. With more views, they can make full use of various depth cues, including stereo parallax, motion parallax, vergence and accommodation [3]. Traditional autostereoscopic displays usually have fewer than ten views. With advances in computer graphics and display technologies, more autostereoscopic displays with tens of views have been manufactured in recent years. In principle, increasing of the number of views improves the quality of 3D images/videos, provides very smooth movement parallax and achieves better viewing comfort.

However, it requires huge amount of computing resources to generate large numbers of view images in real time, which makes the commercialization of autostereoscopic display very challenging. There are two major classes of approaches to multi-view generation [4]. One is computer graphics rendering, also called geometry-based modeling. Rendering of 3D scenes from 3D models is complex and time consuming. The compression and transmission of the created views is also very difficult to achieve in real time. In order to reduce the complexity and difficulty for rendering and transmission of view images, several methods have been proposed in the literature [5-7]. The other approach does not use explicit 3D models but depth or disparity maps with the formats of multi-view video plus depth instead, where multiple depth maps are assigned to the multiple color images [8-10]. The depth information itself can be extracted from a stereo pair or obtained through special range cameras. This approach generates virtual views through depth-based image rendering techniques and has been widely employed in the 3D industry. For instance, a popular input format for 3DTV called 2D plus depth (or 2D+Z) uses a conventional color video and an associated per sample depth map [11].

This paper proposes a novel concept of dynamic multi-view autostereoscopy to reduce the computational complexity of multi-view displays. It considers the basic principle of binocular disparity, where an object to appear at a longer distance in front of or behind the screen plane has larger absolute disparity. For getting smooth movement parallax, objects with larger absolute disparity require more views to present on autostereoscopic display. On the other hand, objects with smaller absolute disparity can be presented by fewer views. Dynamic multi-view autostereoscopy exploits this principle to reduce the computing resources in need for generating the multiple images corresponding to the views. Under the dynamic multi-view autostereoscopy, points in the 3D scene are divided into various spaces based on their disparity. These spaces are rendered with different number of views and synthesized together to form a full-frame multi-view 3D image. Owing to the use of fewer views for the objects with smaller absolute disparity, this concept can greatly reduce computational complexity for multi-view autostereoscopic displays. To illustrate its implementation feasibility, a 2D+Z format based dynamic multi-view rendering system is presented for autostereoscopic displays.

The rest of the paper is organized as follows. Section 2 presents the concept of dynamic multi-view autostereoscopy. Section 3 describes the dynamic multi-view rendering system for 2D+Z format based autostereoscopic displays. Section 4 presents subjective tests for assessing the 3D performance of the dynamic multi-view scheme and analyzes its computational complexity. Section 5 concludes the paper.

Working Principle

On a multi-view lenticular autostereoscopic display, an array of cylindrical lenses attached to a flat-panel display direct the light from adjacent pixel columns to different viewing zones at a pre-defined distance. Massive viewing zones are generated continuously, providing smooth motion parallax. As each of the viewer's eyes can see more than two viewing zones with their width smaller than the pupil diameter, the depth cue of accommodation can be achieved and larger 3D comfort is resulted.

Fig. 1 shows the relationship between disparity and 3D point location. The locations where the points are appeared in front of the screen, on the screen or behind the screen, corresponds to negative, zero and positive disparity spaces, respectively. As shown in the figure, Point b1 and b2 are appeared in the positive disparity space, Point c1 and c2 are in the negative disparity space, and Point a is in the zero disparity space. As the distance between the point and the screen plane in the 3D scene increases, its absolute disparity gets larger. In this example, d_{b2}, the disparity of Point b2, is larger than d_{b1}, the disparity of Point b1. And d_{c2}, the absolute disparity of Point c2, is larger than d_{c1}, the absolute disparity of Point c1.

The concept of disparity density is described as the following equation:

$$s = {|d|}/{N} \quad (1)$$

where |·| is the symbol of absolute value, d is ideal disparity which indicates the disparity resulted from the leftmost view zone to the rightmost zone (which is different from that between two eyes), N is view number denoting the total number of views supported by a multi-view stereoscopic display. Disparity density s illustrates the disparity difference between two adjacent view zones. For the points shown in Fig. 1, the disparity densities of Point b2 and c2 are larger than that of Point b1 and c1. Point a has the lowest disparity density because zero disparity means the viewer will see the same 2D image from different viewing zones. For the multi-view autostereoscopic display to achieve smooth motion parallax, it is necessary to use more views to reduce disparity density. The number of views supported by a multi-view autostereoscopic display is determined by the maximum disparity.

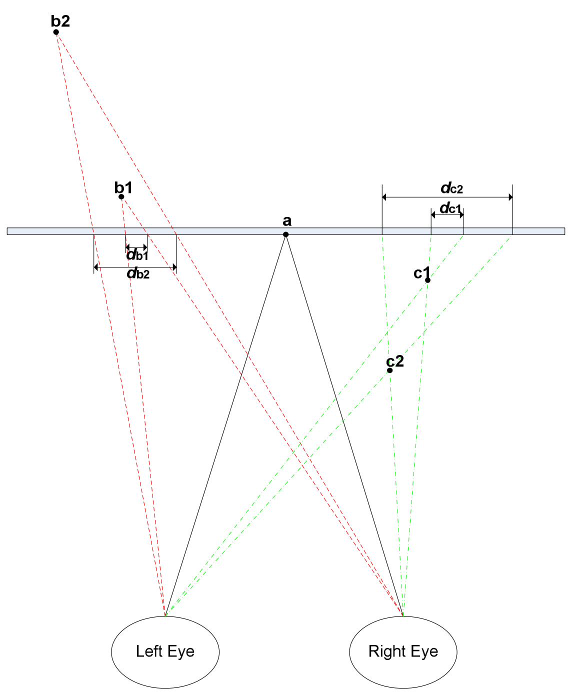


Figure 1. Relation between point location and disparity.

To fully exploit the strengths of multi-view autostereoscopic displays, it is necessary to generate N images in real time. This requires very large amount of computing resources. Conventional solutions are to capture or render multiple images corresponding to the N views. Each full-frame image generated is with the same resolution as the flat-panel display. In this paper, the concept of dynamic multi-view autostereoscopy is proposed to reduce the computational complexity for generating multiple view images. For dynamic multi-view autostereoscopy, the number of views is changeable for the points of a 3D scene. The 3D scene shown in a whole screen is partitioned into several spaces according to the absolute disparity of the 3D points and the defined disparity ranges. More views are generated for the 3D points with higher absolute disparity and fewer views for those belonging to lower disparity range. For the spaces with lower absolute disparity, it is not necessary to use too many views, because the disparity difference between adjacent views is small. The novel concept is useful for handling a uniformly distribution of disparity density for the points in the 3D scene. Of course, fewer views would degrade the 3D performances of a multi-view autostereoscopic display, and reduce the effect of the depth cues, for instance, motion parallax. This paper is to determine whether the extent of the 3D performance degraded is worth with considering the amount of computational complexity saved by the dynamic multi-view scheme. The answer to this question will be addressed in the following sections.

2D+Z Format based Dynamic Multi-View Rendering

This section presents the work flow to realize dynamic multi-view scheme on a 2D+Z format based autostereoscopic display. For the sake of clear presentation, the relationship between disparity and depth is simplified as follows,

$$d = K(z_0 - z), \quad (2)$$

where K is a linear coefficient, z is the depth used in actual image engineering with the range from 0 to 255. $z = 0$ means the farthest point and $z = 255$ means the closest point from viewer. For the points appearing on the screen plane, their depth z is equal to z_0, usually given that $z_0 = 128$. Need to note that z is not the distance from viewer to 3D point but the distance from screen plane to 3D point. By using Eq. 2, we obtain the relationship between points in 3D scene and pixels in depth map. In this case, dividing 3D scene into several spaces can be handled by dividing depth map into several segments.

Fig. 2 shows the work flow of a dynamic multi-view autostereoscopic display for 2D+Z format input. The first step is to partition depth map into several parts shown in partition map. Then the input 2D RGB image and its depth image are divided into several segments according to the partition map. The third step is to generate $\check{N}$ images for each segment. Here, we use $\check{N}$ as the view number for a certain segment. Different number of views are generated for different segments according to the disparity ranges they belong. The next step is to put the image segments together to form N full frame view images. The last step is masking. It is to select sub-pixels in one proper view image to generate a final multi-view 3D image with the same number of viewing zones as the multi-view autostereoscopic display. The main blocks in the process will be described further in the followings.

Depth Partition

Depth partition is to divide the depth map into several parts and reset the value of each pixel to form partition map. Table. 1 shows the ranges of z defined for the three disparity spaces in use. Each depth range includes two disparity regions, called positive and negative disparity regions. Although the pixels in the two disparity regions have different depth values, they belong to the same absolute disparity range or disparity space. Corresponding to the three depth ranges in depth map, the values of pixels in partition map are set to be 0, 128 and 255 respectively. Therefore, partition map can show segments clearly.

2D/Z Segmentation

Based on partition map, 2D RGB image and depth map are partitioned into three segments. Fig. 3a-c show the three segments of depth map corresponding to the three ranges in Table. 1. Fig. 3d-f show the three segments of 2D RGB image following its depth partitions shown in Fig. 3a-c.

View Generation

View generation is to render view images for different view zones. In the example shown in Fig 3, the full frame of depth map and 2D RGB image are separated into three segments according to the defined depth ranges. For each segment, different numbers of views are generated. In this case, they are 7, 14 and 28 for the Segment 1, 2 and 3, respectively. Fig. 4 shows the view distribution for these three parts. Given that View 1, 2, …, 28 refer to the views generated for a true 28-view system. View 2, 6, 10, 14, 18, 22 and 26 are generated for Segment 1. View 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25 and 27 are generated for Segment 2 and View 1, 2, …, 28 are generated for Segment 3. In this paper, the system uses conventional common depth image-based rendering technique for generating the $\check{N}$ images for each pair of 2D RGB and depth segments. Depth image-based rendering has been widely employed for multi-view rendering. It's not the key part in this paper, so we don't present its details here. At the end, the 28-view images are put together to form a full-frame multi-view 3D image by means of masking.

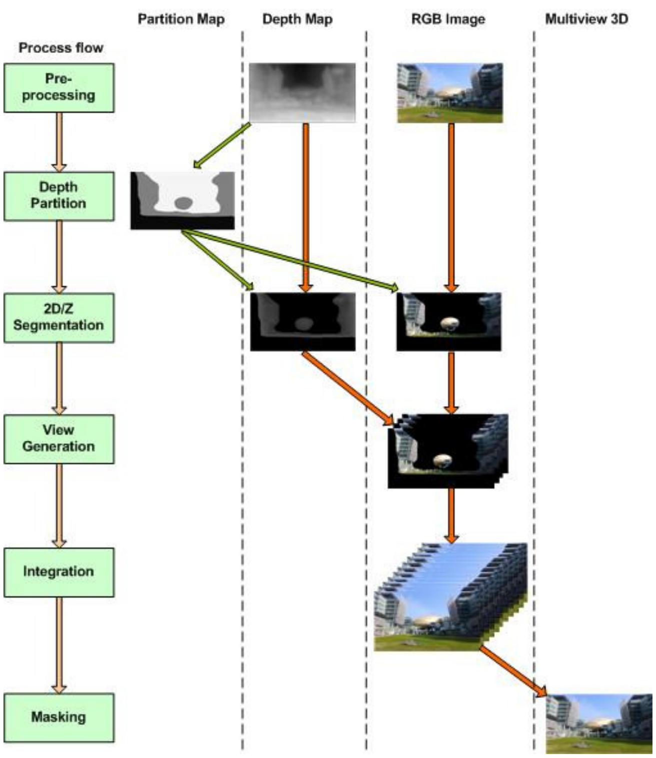


Figure 2. Dynamic multi-view autostereoscopic display with 2D+Z format input

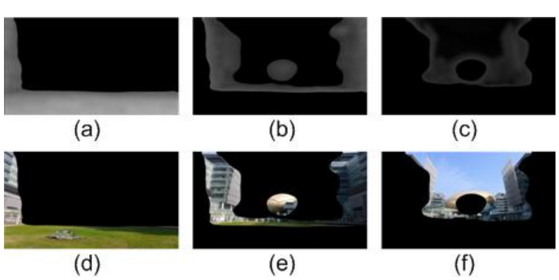


Figure 3. The depth map and 2D RGB image after partition

2	2	2	2	6	6	6	6	10	10	10	10	14	14	14	14	18	18	18	18	22	22	22	22	26	26	26	26

(a)

1	1	3	3	5	5	7	7	9	9	11	11	13	13	15	15	17	17	19	19	21	21	23	23	25	25	27	27

(b)

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28

(c)

Figure 4. Distribution of views for 28-view zones corresponding to (a) Segment 1, (b) Segment 2 and (c) Segment 3

Table 1. Ranges of *z* in use for partition step

Range	Depth in depth map (*z*)		Partition map
	Positive disparity	Negative disparity	
1	86~127	128~169	0
2	43~85	170~212	128
3	0~42	213~255	255

Performance Results & Complexity Analysis

This section presents a subjective assessment method for evaluating the 3D performance of the proposed dynamic multi-view scheme compared with a true 28-view scheme and discusses the results. The dynamic multi-view system under test uses three absolute disparity spaces to partition 2D+Z images. 7, 14 and 28 views are generated for the three segments from small disparity to large disparity. All the 2D+Z images analyzed in this section come from ASTRI and our partners. The basic requirement for selecting the images is that good-quality 3D effect can be achieved with convenient rendering methods by using them. Double-stimulus continuous quality-scale (DSCQS) method [12] is used to assess the subjective quality of multi-view 3D images. In addition, the relative computational complexity of the dynamic multi-view scheme is derived and the extents of the computation complexity saved is evaluated.

Subjective Assessment

Fifteen persons of ages ranged from twenty-three to forty-five years old participated in the tests. Amongst the test participants, five have experiences about 3D projects. There are totally eighteen sets of multi-view 3D images under test. Each set of has two versions of images: one is processed by a true 28-view system, while the other is processed by the dynamic multi-view system with view number up to 28. The order of showing the two version images is not fixed in each test set, so that the participants would not know which one is processed by the proposed multi-view scheme. In each test set, the two versions of test images are shown one after the other and repeats once. Between two successive test sets, there will be a grey image displayed to let the participants' eyes to rest before the new test set to begin. All the test images are shown on a 50-inch 28-view lenticular autostereoscopic display.

The objective of the subjective assessment is to compare the two multi-view schemes in terms of 3D image quality, motion parallax and visual discomfort. Based on this objective, three questions are designed and listed as follows:

Q1. What is the level of the 3D effect of the image without head motion?

Q2. What is the level of the 3D effect of the image with slightly head motion?

Q3. What is the level of the visual comfort of the 3D image?

To quantify the answers for the above questions, five levels from 1 to 5 are defined, representing bad, poor, fair, good and excellent, respectively. The three questions mainly assess the quality of binocular disparity, the quality of motion parallax, and the level of comfort respectively. It is believed that the influence of view number can be assessed by these three factors.

The mean scores of the test participants rated for the true 28-view system and the proposed dynamic multi-view system on the three subjective tests are shown in Fig. 5. The true 28-view system has higher scores in all the three subjective tests. The scores of the dynamic multi-view system are slightly lower than that of the true 28-view system by 0.17%, 1.51% and 1.03%, for the test factor 3D effect with/without head motion and visual comfort, respectively. It's reasonable that dynamic multi-view scheme affects motion parallax and comfort more than 3D effect without head moving. Fewer views can cause larger disparity density and further result in the feeling of jumping between two adjacent views. And non-smooth also has bad influence on comfort. In summary, the assessment results show that the proposed dynamic multi-view scheme just slightly affects the overall 3D quality. It can provide high quality 3D images in general.

Computational Complexity

The computational complexity of the proposed dynamic multi-view scheme is based on the depth map of the input 2D+Z images. We calculate the computation amounts for the three disparity spaces by using the same 2D+Z images tested in the last section. The overall computation amount is defined as below

$$C_P = \frac{1}{S}\sum_{sample=1}^{S}\sum_{segment=1}^{P} pn \times vn \quad (3)$$

where S denotes the number of 2D+Z images used to average the computation amount (equaling to 18 here), P denotes the number of segments used for partitioning the input 2D+Z images (equaling to 3 here), pn and vn mean the pixel number in each segment and the view number generated for each segment. $P = 1$ means that the current method belongs to the convenient true multi-view scheme and there is no image partition performed. C_1, the computation amount for the multi-view scheme when $P = 1$ is easy to calculate, which is equal to $1080 \times 2160 \times 28$ for a 2k 28-view system.

Tab. 2 lists the average percentages of the pixels and computation amounts in the three segments. The pixels in each segment account for about 1/3. Due to fewer views used in low-disparity segments, the computation amounts for corresponding segments decrease to some extent.

Furthermore, we can derive the relative computational complexity as below

$$RC = C_P / C_1. \quad (4)$$

In this paper, the relative computational complexity is used to show how many computing resources can be saved by the new dynamic multi-view scheme compared with the conventional multi-view scheme. Based on the 2D+Z images and the processes described in the previous sections, the relative computational complexity is up to 55.7%.

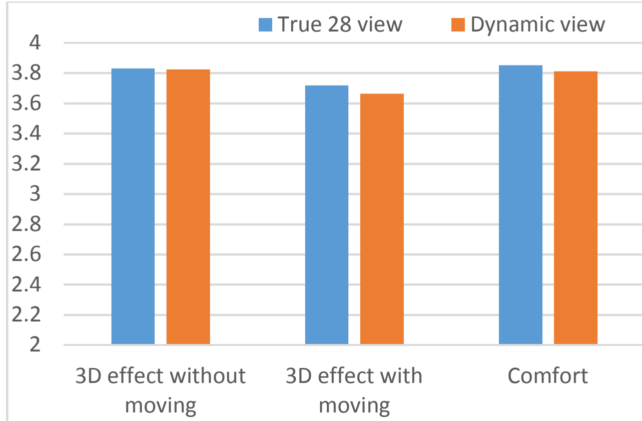


Figure 5. Subjective assessment results. Vertical axis: mean scores for 3D effect without head motion, 3D effect with slightly head motion and visual comfort, bars: the results with true 28-views and the results with dynamic views.

Table 2. Average percentage of the pixels and computation amount in the three segments

Segment	Pixels	Computation amount
1	38.7	17.4
2	30.6	27.5
3	30.7	55.1

Conclusion

This paper proposes a novel concept of dynamic multi-view autostereoscopy. By using fewer views to present small disparity parts on 3D scenes, the proposed scheme can greatly reduce the amount of view data in need to synthesize the multi-view 3D images and thus can save large amount of computing resources in implementing multi-view displays. In nature, the scheme can accomplish a uniformly distribution of disparity density for the points in the 3D scene. In this way, the computational complexity of multi-view system can be greatly reduced. This paper also presents a 2D+Z based multi-view autostereoscopic system to realize the dynamic multi-view stereo scheme. The subjective assessment results show that the proposed scheme just slightly degrades the 3D performance and can save up to 44.3% computational complexity when 3D scenes are divided into three segments.

References

[1] N. A. Dodgson, "Autostereoscopic 3D Displays," Computer, vol. 38, iss. 8, pp. 31-36, 2005.

[2] J. Konrad and M. Halle, "3-D Displays and Signal Processing," IEEE Signal Processing Magazine, vol. 24, iss. 6, pp. 97-111, 2007.

[3] Y. Takaki, K. Tanaka and J. Nakamura, "Super multi-view display with a lower resolution flat-panel display," Optics Express, vol. 19, no. 5, pp. 4129-4139, 2011.

[4] A. Kubota, A. Smolic and etc., "Multiview Imaging and 3DTV," IEEE Signal Processing Magazine, vol. 24, iss. 6, pp. 10-21, 2007.

[5] T. Hübner and R. Pajarola, "Single-pass Multi-view Volume Rendering," in IADIS International Conference on Computer Science and Information Systems: Computer Graphics and Visualization, Lisbon, Portugal, 2007.

[6] F. de Sorbier, Y. Nozick and H. Saito, "Multi-view Rendering using GPU for 3-D Displays," GSTF Journal on Computing, vol. 1 no.1, pp. 44-49, 2010.

[7] S. Zinger, D. Ruijters and etc., "View Interpolation for Medical Images on Autostereoscopic Displays," IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 1, pp. 128-137, 2012.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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 BS in safety engineering from Chinese University of Geosciences (1998), 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. In 2010, he joined in the Advanced Digital Systems Division (ADS) at the Hong Kong Applied Science and Technology Research Institute. As a senior engineer, his work has mainly focused on the algorithm researches, chip architecture design and system developments relative with autostereoscopic displays and advanced audio systems. His current research interests include 3D image enhancement, image-based rendering, crosstalk reduction in multi-view displays, adaptive noise cancellation in audio systems and 3D audio for VR.

Chan Man Chi received his Ph.D. degree in Information Engineering from The Chinese University of Hong Kong in 2000. He is graduated in Electronic Engineering from Hong Kong Polytechnic University at 1995 where he received his M.Phil. degree in 1997. He is currently a principal engineer at ASTRI and works in Advanced Digital Systems domain responsible for SoC firmware and application system development. Dr Chan has over ten years of software development experience in various systems including 3D autostereoscopic systems, wireless sensor networks, media processor SoCs and power line communications systems.

Mark, P.C. Mok received his Master and Bachelor's degrees in Electronic Engineering from The Hong Kong University of Science and Technology in 1997 and 2000 respectively. He has over 20 years working experience in IC design industry and currently a Senior Manager at ASTRI. He has been being the technical lead for 3D related R&D project for years and led the team to successfully deliver many glasses-free 3D display products in the market.

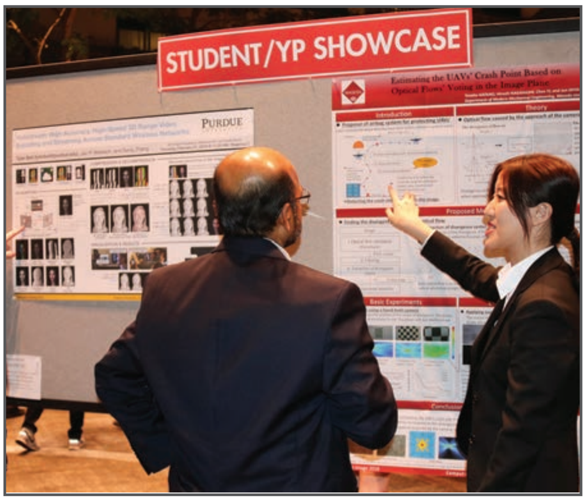
JOIN US AT THE NEXT EI!

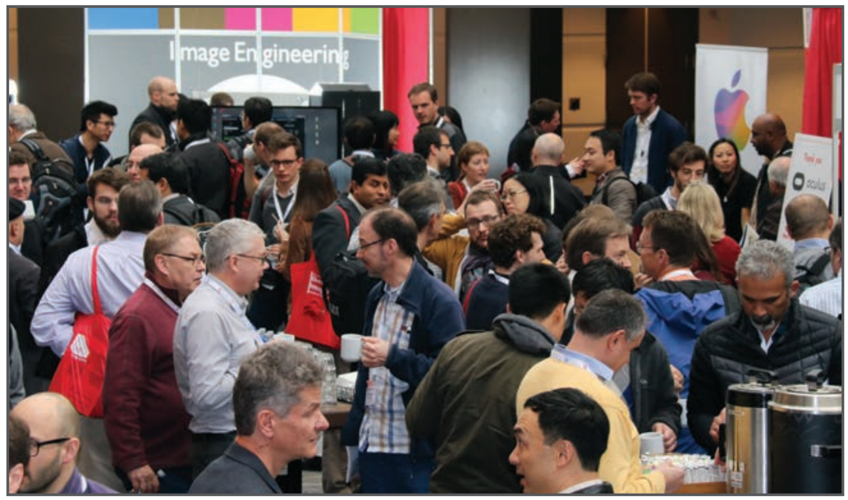
IS&T International Symposium on

Electronic Imaging

SCIENCE AND TECHNOLOGY

Imaging across applications . . . Where industry and academia meet!





• SHORT COURSES • EXHIBITS • DEMONSTRATION SESSION • PLENARY TALKS •
• INTERACTIVE PAPER SESSION • SPECIAL EVENTS • TECHNICAL SESSIONS •

www.electronicimaging.org

IS&T

imaging.org