https://doi.org/10.2352/ISSN.2470-1173.2019.3.SDA-656

Saliency Map based Multi-View Rendering for Autostereoscopic Displays

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

Abstract

With the provision of motion parallax and viewing convenience, multi-view autostereoscopic displays have been popular in recent years. Obviously, increasing the number of views improves the quality of 3D images/videos and produces to better motion parallax. The tradeoff is the larger number of view images required to generate in real time leading to the need of the huge amount of computing resources for the systems. In fact, people often focuses on the distinctive objects in a scene. For achieving the same level of motion parallax, it can use more views to present distinctive objects and less views for the rest. As a result, fewer computing resources are required for rendering multi-view images. With exploiting this principle, a new multi-view rendering scheme based on visual saliency is proposed for the application of autostereoscopic displays. The new method uses saliency maps to extract distinctive regions with different saliency level in a scene and dynamically control the number of views to generate for them. Points in the distinctive regions with high saliency use more views, while points in the regions with low saliency use less views. By controlling the number of views in use for different salient regions, the proposed scheme can maintain low computation complexity without causing significant degradation in 3D experience. In this paper, a 2D+Z format based multi-view rendering system with the use of saliency maps is presented to illustrate the feasibility of the new scheme. Subjective assessment results demonstrate that the saliency map based multi-view system has slight degradation in 3D performance compared with true 28-view system and achieves 55% reduction in computation complexity.

Introduction

Multi-view autostereoscopic displays allow viewers to see 3D images without the glasses [1-2]. In addition to the viewing convenience, they produce effective depth cues, including stereo parallax, movement parallax, accommodation and convergence [3], to provide better 3D experience to viewers. By using more views, multi-view autostereoscopic displays have larger viewing zones and provide more viewers smoother motion parallax. Traditional autostereoscopic displays usually have less than 10 views. However, in recent years, more and more autostereoscopic displays with tens of views appear.

The images corresponding to different views are usually obtained from video camera acquisition [4]. Increasing the number of views in multi-view displays significantly increases the amount of data to be stored and transmitted in the display systems. To reduce the amount of data for transmission and storage, many image-based rendering techniques have been developed to use depth maps to deduce virtual intermediate views from one or a few images captured by cameras, instead of capturing all the view images by cameras [5]. Depth map can be obtained by using distance sensors or be estimated by using depth generation algorithms. It has been widely employed in the 3D industry for doing multi-view rendering. 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 [6].

In the real world, our visual environment contains much more information than we can perceive at once. Our visual attention usually unconsciously focuses on the most distinctive regions in a scene [7]. In other words, our visual attention guides the movement of the eyes, allowing an accurate inspection by the fovea of the chosen areas. The salient areas are commonly called Region-of-Interest in the image processing domain. Saliency maps are widely used to measure the visual importance of image pixels and plays important roles in many image processing applications [8]. Recently, it has been used in 3D rendering to improve the quality of experience [7].

Inspired by dynamic multi-view autostereoscopy [9], in this paper, we propose a novel multi-view rendering scheme for the application of autostereoscopic displays, which employs the information of saliency map to make a trade-off between 3D quality experienced and computational complexity in need. In the new scheme, the scene shown in the autostereoscopic display is divided into different regions according to the defined saliency levels. Different number of view images are generated for different regions. Regions with high saliency use more views and regions with low saliency uses fewer views. The proposed scheme maintains low computational complexity by controlling the number of views in use for different salient regions without causing significant degradation in the resultant 3D quality. In order to illustrate the feasibility of the scheme, a 2D+Z format based multi-view rendering method is presented for autostereoscopic displays.

The rest of the paper is organized as follows. Section II presents the method of saliency detection. Section III describes the system architecture of saliency map based multi-view rendering for autostereoscopic displays with 2D+Z format inputs. Experiment results are discussed in Section IV. Finally, the paper is concluded in Section V.

Saliency Detection

Visual saliency is a fundamental problem in neuroscience, psychology, and vision perception [10]. It measures the distinctiveness of a region from its neighboring regions. High salient regions convey more information than low salient ones. Visual saliency relies on four kinds of features for saliency detection [11]: local features, global features, visual forms and high-level factors. For instance, if a region is distinctive in local features such as intensity, color, texture and motion, it is considered as a high salient region. For a model of 3D saliency detection, depth factor has a vital role. A depth map contains depth information about each pixel of an image. The values of saliency are ranged from 0 to 255 in the saliency map. 0 denotes the lowest saliency, while 255 denotes the highest saliency.

Referring to the structure of the RGB-Signature algorithm [12],

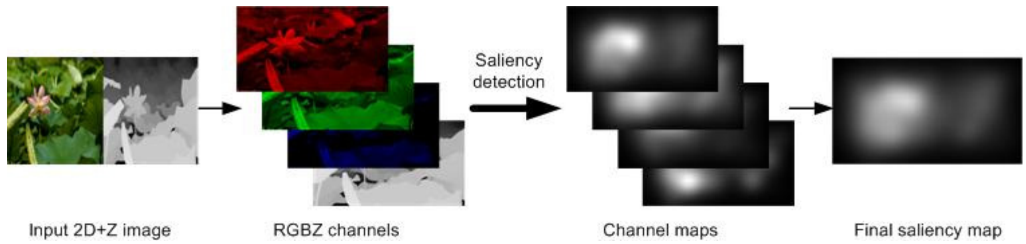

Figure 1. Saliency detection algorithm for 2D+Z images

a saliency detection algorithm with the 2D+Z information is designed for 3DTV applications, as shown in Fig. 1. The input 2D+Z image is decomposed into four channels, including three color channels and one depth channel. A saliency map is computed for each color channel independently by using the algorithm of graph-based visual saliency (GBVS) [13]. As a simple bottom-up visual saliency model, GBVS is very powerful for making reliable predictions on human fixation. The final saliency map is simply the average of the four-channel saliency maps. It is necessary to analyze the weighting factors for each channel carefully and use the most reliable weighting factors in saliency detection. This paper focuses on the way to make use of saliency information to implement multi-view rendering for the application of autostereoscopic displays. In this case, we just concern whether a simple saliency detection algorithm can help to improve 3D performance of the multi-view rendering with the given computation complexity.

Proposed Method

We proposed the concept of dynamic multi-view autostereoscopy in [9]. According to the concept, the view number for autostereoscopic displays can be changeable in different regions. The concept is utilized in this paper partly. We partition a scene into different regions by using saliency map. For more distinctive regions, more view images are used in multi-view rendering.

The process of saliency map based multi-view rendering for autostereoscopic displays with input format 2D+Z is shown in Fig.2. The first step in the process flow is pre-processing. It is to resize color images, depth maps and saliency maps to the dimension of the display screen. Other processing algorithms like color adjustment, image filtering and depth enhancement can be handled in this step. The second step is saliency partition, wherein the saliency map is divided into several regions according to the threshold values defined for different saliency ranges and is converted to a leveled saliency map with setting every point in each salient region to the saliency value defined for the corresponding saliency range. In this example, the leveled saliency map is shown in Fig. 2, which consists of three regions. The third step is RGB/Z segmentation. Based on the leveled saliency map, depth map and RGB image are segmented into different regions, corresponding to different salient ranges. The regions segmented in depth maps are corresponding to those segmented in RGB images. The most salient region for depth map and RGB image is shown in Fig. 2. The fourth step is to generate $\check{N}$ images for each region. Here, $\check{N}$ denotes the view number for a certain region, which is different from the denotation N, the view number used in conventional multi-view rendering methods. $\check{N}$ is changeable for different salient regions. Different number of views are generated for different regions according to the saliency ranges they belong. The fifth step is integration, which aims to put the image regions 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.

Saliency Partition

There are two key problems about saliency partition. One is how many regions the saliency map should be partitioned into. Another is how to set the ranges for saliency partition. They can affect the 3D performance and computational complexity of novel multi-view rendering method. The objective of this paper is to prove the feasibility of saliency map based multi-view rendering, so we just give an example with a simple partition way.

Based on the defined saliency ranges, the saliency map can be partitioned into different regions. As shown in Fig. 2, there are three regions partitioned in this example. Tab. 1 shows the three saliency ranges in use and the corresponding values after resetting. Resetting saliency is the operation to generate a leveled saliency map, which

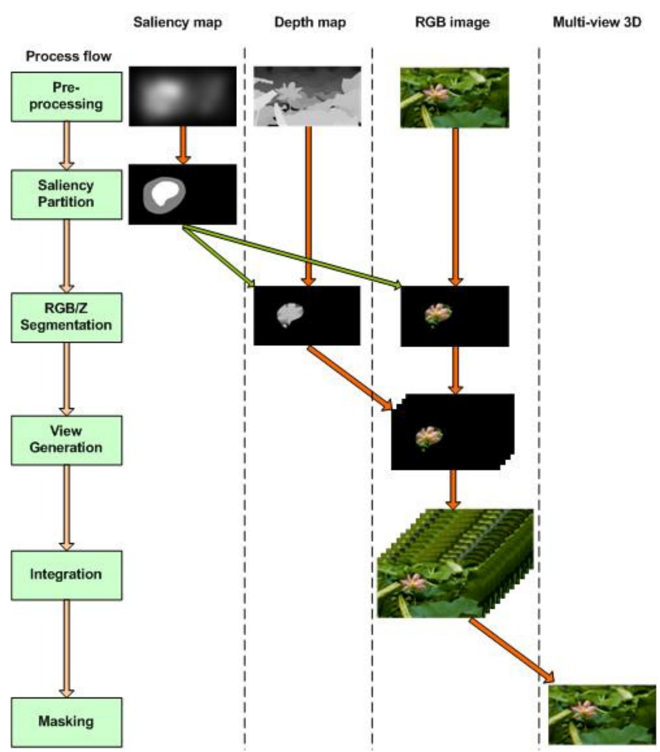

Figure 2. Saliency map based multi-view rendering for autostereoscopic display with 2D+Z format input

is used to segment depth map and RGB image in the next RGB/Z segmentation step.

RGB/Z Segmentation

Based on the leveled saliency map obtained in saliency partition step, depth maps and 2D RGB images are segmented into the regions corresponding to different salient ranges. In this example, depth map and RGB image are segmented into three regions. Fig. 3a-c show the three regions of depth map, while Fig. 3d-f show the three regions of RGB image corresponding to their depth regions.

View Generation

View generation is to render view images for different viewing zones. In the example shown in Fig 2, the full frame of depth map and 2D RGB image are segmented into three regions according to their pixel saliency and the defined saliency ranges. We suppose the multi-view autostereoscopic display with 28 viewing zones. That means N equals to 28. For each region, different numbers of views are generated. In this case, they are 7, 14 and 28 for the Region 3, 2 and 1, respectively. Fig. 4 shows the view distribution for these three regions. 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 Region 3. View 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25 and 27 are generated for Region 2 and View 1, 2, …, 28 are generated for Region 1. 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 regions. Depth image-based rendering has been widely employed for multi-view rendering. Since it is out of the scope of this paper, it will not be described further. At the end, the 28-view images are put together to form a full-frame multi-view 3D image by means of masking.

Table 1. Ranges of *saliency* for saliency partition

Region	Saliency range	Saliency after resetting
1	171~255	255
2	86~170	128
3	0~85	0

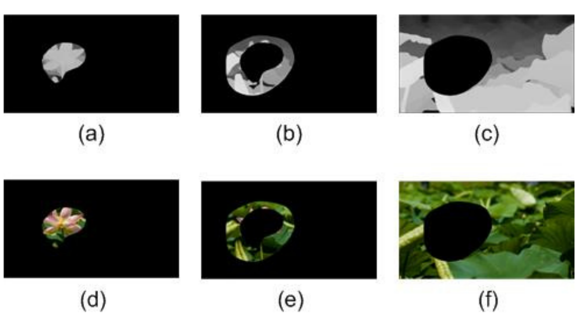

Figure 3. The three regions of depth map and RGB image after RGB/Z segmentation

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 view for 28 viewing zones corresponding to (a) Region 1, (b) Region 2 and (c) Region 3

Experimental Results

This section presents the subjective assessment experiments for evaluating the 3D performance of saliency map based multi-view rendering scheme and discusses its results. In the experiments, 20 2D+Z images are used to compare the performance of the new scheme and the true 28-view system. RGB and depth images are divided into 3 regions. In view generation, 7, 14 and 28 views are used for the regions from low saliency to high saliency. Double-stimulus continuous quality-scale (DSCQS) method [14] is used to assess the subjective quality of 3D multi-view images. In addition, the computational complexity of the new method is analyzed.

Subjective Assessment

Fifteen persons of ages ranged from twenty to forty-five years old participated in the assessment. Amongst them, four have experiences about 3D projects. There are totally 20 sets of multi-view images under test. Each set has two versions of images: one is processed by a true 28-view system and the other is processed by the new saliency based multi-view method. 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 new method. Fig. 5 shows the process of DSCQS. At the beginning of each test set, there would be a grey image displayed on the display to let the participants' eyes to rest before doing the test. In each test set, the two versions of test images are shown one after the other and repeats once. All the test images are displayed on a 50-inch 28-view lenticular autostereoscopic display.

Since, the subjective assessment aims at the comparisons of the two methods in 3D image quality, the following question is designed:

What is the level of the 3D effect of the image?

There are five levels from 1 to 5, representing bad, poor, fair, good and excellent. The objective of the question is to assess the 3D quality of the multi-view autostereoscopic display.

The mean scores of the test participants rated for the true 28-view system and the proposed saliency map based multi-view scheme are shown in Fig. 6. The ratings for the new scheme are slightly lower than that of the true 28-view system. Basically, the true 28-view system should have better performance in nature. However, the assessment results show that the new scheme achieves similar level of 3D effect as the true 28-view system.

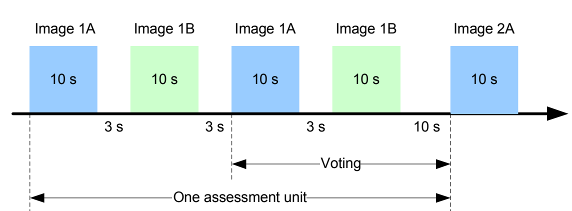

Figure 5. The test process of DSCQS

Computational Complexity

The computational complexity of the new scheme is based on the depth map of 2D+Z image. In order to determine the relationship between computational complexity and region number, we calculate the computational complexities for the partitioning of three salient regions using the 2D+Z images under test in the last section. The computational complexity is defined as below

$$C_P = \frac{1}{S}\sum_{sample=1}^{S}\sum_{region=1}^{R} pn \times vn$$

where S denotes the number of 2D+Z images used to average the computational complexity, R denotes the region number for dividing an image, pn and vn denote the pixel number in each extended region and the view number used in each region, respectively. R=1 means that the current scheme belongs to conventional 28-view scheme and there is no partitioning for all pixels. C_1, the computational complexity for the method when R=1 is easy to calculate, which is equal to 2160x3840x28 for a 4k 28-view display system.

Based on these denotations, the relative computational complexity is derived as C_P/C_1. Using the relative computational complexity, it is easy to determine whether the new scheme can reduce the computational complexity compared with the true 28-view scheme. By considering the 2D+Z images under test and the saliency ranges mentioned above, the overall relative computational complexity obtained is 44.7%. It is 55% less than that of the true 28-view system. Tab. 2 lists the average percentages of the pixels and computation amounts used in different salient regions. Note that the least salient regions contain around 55% pixels (more than half of the test image), while the most salient regions contain only 17% pixels. However, the computation amounts for the most salient regions is greater than that for the least salient regions. The results reveal that the saliency map based multi-view rendering scheme can save lots of computation resources compared with the conventional multi-view rendering scheme.

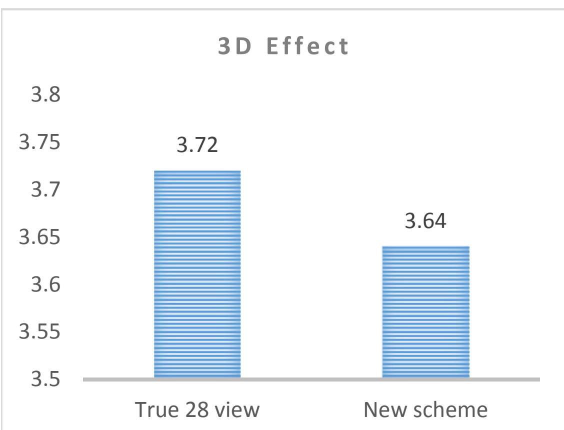

Figure 6. Subjective assessment results of systems with true 28 views and new scheme

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

Region	Original pixel	Computation amount
1	17.09	38.23
2	27.54	30.81
3	55.37	30.96

Conclusions

This paper proposes a saliency map based multi-view rendering scheme for autostereoscopic display systems. Based on saliency maps, 3D scenes shown in the autostereoscopic display are divided into different regions according to the defined saliency levels. By controlling the number of views used in presenting different regions, the proposed scheme can reduce the computational complexity without causing significant 3D quality degradation. A 2D+Z based multi-view system is presented to illustrate the feasibility to implement the proposed scheme. Subjective assessment results reveal that the 3D performance of the saliency map based multi-view scheme is slightly lower than that of the true 28-view method. More than 55% reduction in computational complexity is achieved.

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] A. Vetro and K. Müller, "Depth-based 3D video formats and coding technology," in Emerging Technologies for 3D Video: Creation, Coding, Transmission and Rendering, 1st ed. Chichester, ch. 8, UK: Wiley, 2013, pp. 139-161.

[6] 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.

[7] C. Chamaret, S. Godeffroy and etc., "Adaptive 3D Rendering based on Region-of-Interest," in Proceedings of SPIE 7524, Stereoscopic Displays and Applications XXI, 75240V, San Jose, USA, 2010.

[8] 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.

[9] Y. Jiao, M. C. Chan and M. P. C. Mok, "Dynamic Multi-view Autostereoscopy," submitted to EI 2019.

[10] 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.

[11] 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.

[12] 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.

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

[14] 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 system.

Mark, P.C. Mok received his Master and Bachelor 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.