11.4: A Quality-Scalable Depth-Aware Video Processing System

Chao-Chung Cheng, Chung-Te Li, Yi-Min Tsai, Liang-Gee Chen
Graduate Institution of Electronics Engineering, National Taiwan University, Taipei, Taiwan

Abstract

The three-dimensional (3D) displays provide a dramatic improvement of visual quality over the 2D displays. To produce high quality three-dimensional contents, depth map generation is the major issue in content creation. In this paper, we propose a quality-scalable depth-aware video processing system which retrieves depth maps using hybrid depth cues from single view video content. The system uses a hardware-oriented algorithm and generates spectacular and comfortable results. With the help of the depth map, 2D videos can be converted to 3D format and can be displayed in 3D/Stereoscopic displays by depth image-based rendering technique. Moreover, for conventional 2D displays, we also propose a 2D video depth perception enhancement application. With the depth-dependent parameter adjustment, the stereo effect of the 2D content is also enhanced.

1. Introduction

3D video signal processing has become a trend in the visual processing field. As 3D display technology matures, people aspire to experience the videos that closer to reality. Emerging 3D displays generate better stereo effects than conventional 2D displays. However, for existing 2D content, depth information is not recorded. The lack of an effective 3D content generation approach is a dilemma for 3D industry. Therefore, depth-aware video processing becomes an important issue.

The main concept of depth generation is to retrieve the third dimensional information from multiple depth cues of existing 2D video content. When watching real world, human brain integrates various heuristic depth cues to generate depth perception. The major depth perceptions are binocular cue from two eyes and various monocular cues from single eye. Various depth generation techniques have been proposed in the previous works [1][2][3][8] [9][15] [16]. Therefore, in order to create high quality depth map from single view video, both binocular and monocular cues need to be explored. In this paper, we propose a depth-aware video processing system. The proposed system can provides depth map and multi-view video for 3D/stereoscopic displays and also can enhance the depth perception on conventional 2D displays.

2. Proposed Video Processing System

The system we proposed has three major cores, depth generation, depth-aware 2D video enhancement, and multi-view depth image-based rendering (DIBR)[7], as shown in **Figure 1**. The detail of each part is explained in the following sub-sections.

2.1. Depth Generation

The depth generation module is the major part in the system. The module combines multiple depth cues. They are depth from motion parallax (DMP), depth from geometrical perspective (DGP), and depth from relative position (DRP). These cues are integrated by priority depth fusion method as shown in **Figure 2**.

In binocular depth cue, multiple frames in temporal domain are used to approximate two views in spatial domain[5]. Firstly, the camera motion of the consecutive frames with respect to the current frame is analyzed by global motion estimation. The frame with most suitable baseline is selected and is warped to form a parallel view configuration with the current frame to compensate the camera motion.

Therefore, the warped frame and the current frame can be approximated as binocular images. The disparity in spatial domain can be seen as motion parallax in temporal domain. Motion parallax is computed by block-based matching with spatial smoothness consideration. The spatial smoothness is applied to generate smooth motion vector using local optimization method. Other global optimization method, such as graph cut and belief propagation may generate better results with high computation complexity. The energy function is shown in the following equation.

$$\text{Energy} = \alpha \, \text{SAD} + \beta * \text{Distance}(MV_{block} - MV_{neighbor})$$

Forward-backward raster scan order is applied in our implementation to generate smooth motion vectors with the definition of energy function. Then the depth was estimated by $\lambda\sqrt{MVx^2 + MVy^2}$ .Where λ is a scaling factor of motion vector to disparity vector.

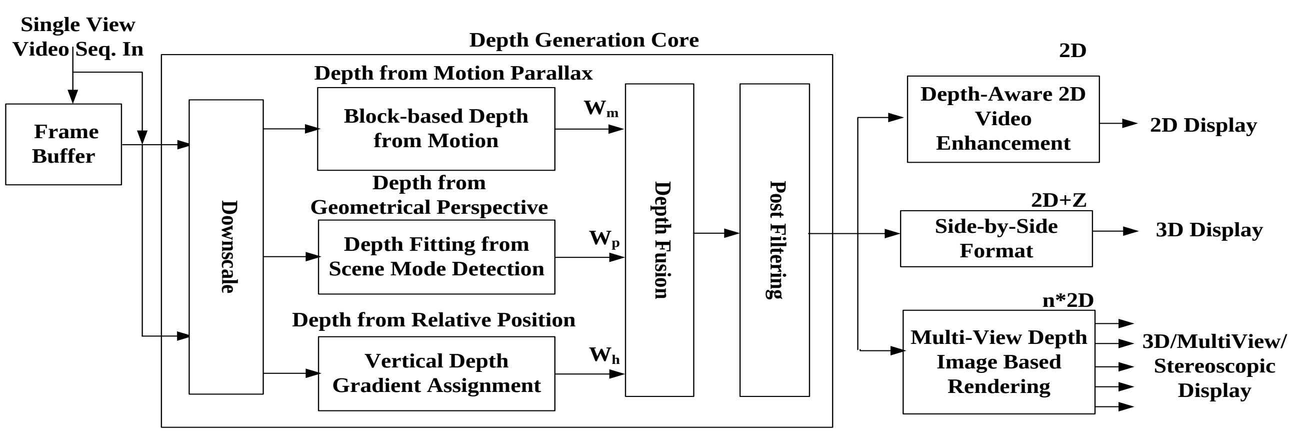

Figure 1. The block diagram of proposed depth-aware video processing system.

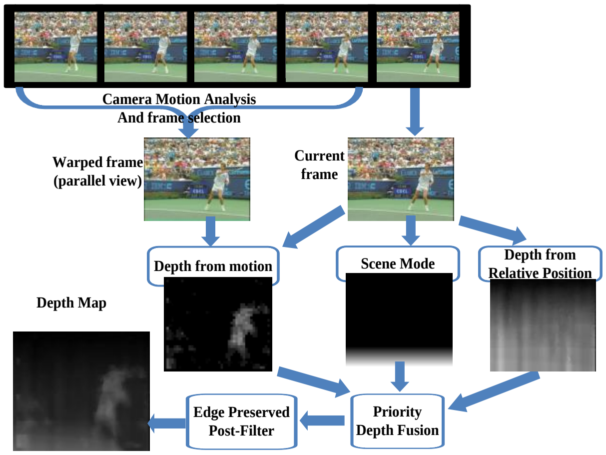

Figure 2. Data flow of the proposed depth generation algorithm**.**

In the scene mode module, depth is assigned by scene line structure analysis [4]. Two major structures, horizontal line and vanishing line, are implemented. Depending on the line structure of the video frames, two depth types of scene will be selected and be assigned. The temporal depth smoothness is also considered to keep the video from depth jerk.

The line-based DRP detects the horizontal edge of the video frames. The stepping depth map is assigned through the vertical scan line. The depth cue generates comfortable stable initial depth map when the video frames have standing objects and humans.

The priority depth fusion[9] is applied to integrate multiple depth cues according the characteristics of video content. When the video have large global motion, the weighting of the depth from motion parallax have large value. Otherwise, the depth from geometrical perspective and depth from relative height will have larger weighting. Finally, the fused depth map is filtered by the cross bilateral filter[14] to remove the block artifact. The filtered depth map also has comfortable visual quality because bilateral filtering generates smooth depth map inside the smooth region with similar pixel values and preserves edge on object boundary.

2.2. Depth-Aware 2-D Video Enhancement

For conventional 2D display, we propose an application to use depth information to enhance the depth perception of 2D video content. The depth-aware video enhancement makes the video more vivid and more realistic. Using the characteristics of human visual perception in [11][12][13]. Three cues, contrast, saturation, and edge, are used to enhance the depth perception according to relative depth range. The flow is shown in **Figure 3**.

2.3. Depth Image-Based Rendering

For 3D/Multi-view/Stereoscopic display, the depth map is used for rendering multiple view angles by the following equation.

$$x_i = x_c \pm \left(\frac{t_x}{2}\frac{f}{Z}\right)$$

Where the x_i is the horizontal coordinate of the interpolated view, and x_c is the horizontal coordinate of the intermediate view. Z is depth value of current pixel, f is camera focal length and t_x is eye distance. The edge dependent interpolation method is applied to preserve edge information of the interpolated area. In [6][7], we interpolates small hole along the edge using the modified ELA algorithm, thus edge information of interpolated area is preserved.

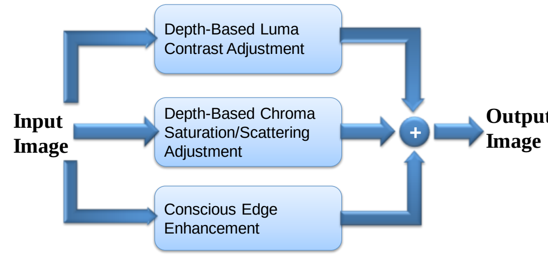

Figure 3. Flow of 2D Image Enhancement on Depth Perception

3. Experimental Result

We show the depth map generated by our proposed algorithm first, then the depth-aware enhancement for 2D display. Finally, we will show the multi-view images and generated stereoscopic images in red-cyan at the end of this section.

3.1. Depth Map

We show the result of original sequences and corresponding depth maps in **Figure 4**. By fusion of depth map from motion and scene, both static objects and moving objects generated depth.

Figure 4. Original images and depth maps

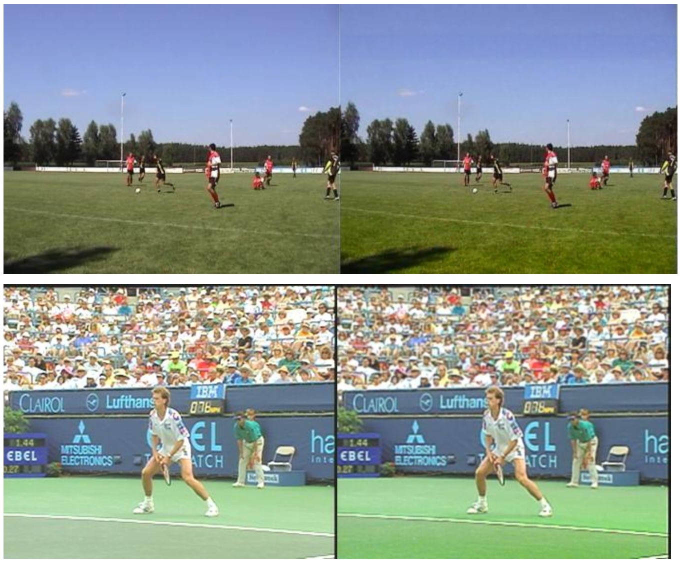

Figure 5. Original image(left) versus depth perception enhanced image(right).

3.2. Depth-Aware 2-D Video Enhancement

In this part, we show original sequences and enhanced ones as shown in **Figure 5**. Contrast, saturation, and edge are adjusted by a function of depth value. From the example image, we can see the near region have enhanced stereo perception.

3.3. Depth Image-Based View Rendering

Figure 6 and **Figure 7** show multi-view and red-cyan images by multi-view DIBR. Through this technique, the depth information can be used to generate arbitrary views from original sequences. From the red rectangle of **Figure 6**, we can see the disparity of human and background. The holes of rendered images are carefully filled by using edge-dependent algorithm in our previous work [6][7].

4. Discussion and Conclusion

Compare with the previous works, the proposed work have three major advantages as shown in Table 1. Firstly, the depth from multiple cues can provide extra depth information to generate more realistic effects. The depth from relative position generates stepping depth gradient over the object segment, the depth from motion parallax enhances the depth when video have camera motion, and the depth from geometrical perspective enhances the stereo effect of scene structure. Secondly, the process of proposed methods using block-based DMP and line-based DRP cooperate with bilateral filter and is more suitable for hardware implementation. As for [2], the computation time of the segmentation-based algorithm is not constant. Finally, the proposed method is quality-scalable. Depending on the applications, different block sizes can be selected or cooperated with multi-scale subsample. For example, large block size is suitable for 2D enhancement because 2D enhancement applications do not require high resolution depth maps. Larger block size will result in lower depth detail. Moreover, the system we proposed can be used both in depth perception enhancement of conventional 2D display and depth-based processing for modern 2D+Z and N*2D Display. It is suitable to be integrated into 3D display and broadcasting systems.

Table 1. Comparison with previous works

Comparison	[1]	[2]	[3]	Proposed
Depth from motion parallax	O	O	X	O
Camera motion compensation	X	O	n/a	O
Depth from geometrical perspective	X	X	X	O
Depth from relative height	X	X	O	O+ (Relative depth on gradient direction)
Constant Latency	X	X	X	O
Quality scalable	X	X	X	O

5. Acknowledgements

The project is supported in part by the Himax Technology Inc.

6. References

[1] Ianir Ideses, Leonid P. Yaroslavsky, and Barak Fishbain, “Real-time 2D to 3D video conversion,” Journal of Real-Time Image Processing, 2007

[2] Donghyun Kim, Dongbo Min, and Kwanghoon Sohn, "Stereoscopic Video Generation Method Using Motion Analysis," IEEE Transactions on Broadcasting 2008

[3] Yong Ju Jung, Aron Baik, Jiwon Kim, and Dusik Park, "A novel 2D-to-3D conversion technique based on relative height depth cue" in SPIE Electronics Imaging, Stereoscopic Displays and Applications XX, 2009

[4] Chao-Chung Cheng, Chung-Te Li, Po-Sen Huang, Tsung-Kai Lin, Yi-Min Tsai, and Liang-Gee Chen, "A Block-based 2D-to-3D Conversion System with Bilateral Filter," in IEEE Int. Conf. on Consumer Electronics, 2009

[5] Chao-Chung Cheng, Chung-Te Li, Yi-Min Tsai, and Liang-Gee Chen, "Hybrid Depth Cueing for 2D-To-3D Conversion System," in SPIE Electronics Imaging, Stereoscopic Displays and Applications XX, 2009

[6] W.-Y. Chen, Y.-L. Chang, and L.-G. Chen, “Real-time depth image based rendering hardware accelerator for advanced three dimensional television system,” IEEE Int. Conf. on Multimedia and Expo., 2006.

[7] W.-Y. Chen and Y.-L. Chang and S.-F. Lin and L.-F. Ding and L.-G. Chen, "Efficient Depth Image Based Rendering with Edge Dependent Depth Filter and Interpolation", in IEEE Int. Conf. on Multimedia and Expo. (ICME), 2005

[8] Jing-Ying Chang, Chao-Chung Cheng, Shao-Yi Chien, and Liang-Gee Chen, "Relative Depth Layer Extraction For Monoscopic Video By Use Of Multidimensional Filter," IEEE Int. Conf. on Multimedia and Expo. (ICME) 2006

[9] H.Murata et al.: ”Conversion of Two-Dimensional Images to Three Dimensions”, SID Digest of Technical Papers, 39.4, pp859-862 (1995)

[10] Y.-L. Chang, J.-Y. Chang, Y.-M. Tsai, C.-L. Lee, and L.-G. Chen, “Priority Depth Fusion for the 2D-to-3D Conversion System,” in SPIE Electronics Imaging, Three-Dimensional Image Capture and Applications 2008

[11] Ichihara, S., Kitagawa, N.,& Akutsu H.”Contrast and depth perception: Effects of texture contrast and area contrast,” Perception, vol. 36,pp. 686 – 695, 2007

[12] Monitoring Solar Radiation and Its Transmission Through the Atmosphere, Brooks, 2006

[13] Mather., G. “The use of image blur as a depth cue,” Perception, vol. 26, pp.1147 – 1158. 1997

[14] C. Tomasi and R. Manduchi, “Bilateral filtering for gray and color images,” Proceedings of the IEEE International Conference on Computer Vision, pages 839-846, January 1998

[15] H. Murata et al.: ”A Real-Time 2-D to 3-D Image Conversion Technique Using Computed Image Depth”, SID Digest of Technical Papers, 32.2, pp919-922 (1998)

[16] T. Iinuma, H. Murata, S. Yamashita, K. Oyamada, "Natural Stereo Depth Creation Methodology for a Real-time 2D-to-3D Image Conversion," SID Symposium Digest of Technical Papers, 2000

Figure 6. Multi-view images rendered by DIBR (left-view, central-view, and right-view images)

Figure 7. Red-Cyan images of four sequences.