2012 4th International Conference on Computer Engineering and Technology (ICCET 2012)
IPCSIT vol.40 (2012) © (2012) IACSIT Press, Singapore

A New Block-Matching Based Approach for Automatic 2D to 3D Conversion

Jian Chen, Yuesheng Zhu* and Xiyao Liu

Communication and Information Security Lab
Shenzhen Graduate School, Peking University, Shenzhen, China
Emal:10948156@sz.pku.edu.cn, zhuys@pkusz.edu.cn

Abstract. This paper presents a novel block-matching based approach for automatic 2D to 3D conversion, in which a depth map is generated by fusing three depth extraction modules with a main depth perception from motion-parallax, and two auxiliary depth perceptions from color and linear perspective information. An improved block-matching algorithm and a corresponding depth correction procedure are described in the depth from motion-parallax module. Also a joint bilateral filter is applied to diminish the block artifacts and the staircase edges. Experimental results have shown that the proposed approach can greatly reduce the computation burden with almost the same visual quality compared to a mainstream existing motion based method, and is more suitable for real-time applications.

Keywords: Block-matching, motion estimation, real-time application, colour information, 2D to 3D conversion

1. Introduction

Rapid development of 3D displays technologies and digital video processing has brought 3DTV into our life. As more facilities and devices are 3D capable, the demand for 3D video contents is increasing sharply. However, the tremendous amount of current and past media data is in 2D format and 3D stereo contents are still not rich enough now. Compared to the direct capture of 3D video contents, video conversion from 2D to 3D is a low-cost and backward compatible solution.

Depth Image Based Rendering (DIBR) [1] is an enabling technique used for rendering new virtual views of the 3D scene with 2D images and corresponding depth maps. The key issue in DIBR is how to generate the depth maps from 2D video contents. Various algorithms have been developed to generate depth information recently [2]-[7]. One of the dominant techniques for depth map generation is based on the concept of depth from motion-parallax [7], in which the relationship between the distance of moving objects from camera and the registered motion for them is used to estimate the depth information. Despite its simplicity and compatibility with H.264/AVC and other standards, the generated results are very rough. Some efforts have been made to improve this issue. Getting the depth information from motion-parallax fused with color segmentation is one of the important methods and has been adopted to get more realistic depth estimates in [8][9]. Although the Fusion with Color Segmentation (FCS) method can produce clear and good region boundaries and eliminate the block effect caused by block-matching, over segmentation would happen when lighting source changes and result in an inconsistent depth, also the computational burden limits its real-time applications. In addition, it is noted that in the widely used DIBR systems [10], a depth map pre-processed procedure is applied before 3D image warping to reduce artifacts and dis-occlusion in the rendered virtual images. However, it could blur the depth map edges.

* + Corresponding author. Tel.: + (86075526035352); fax: +(86075526035352).
E-mail address: zhuys@pkusz.edu.cn.

In this paper, a novel block-matching based algorithm for depth map generation is presented. Instead of using FCS method to produce reliable depth maps with accurate and clear edges, in the proposed approach, an improved block-matching algorithm and a corresponding depth correction procedure are developed to refine the depth map from the motion-parallax directly. Then the depth results from color and linear perspective are fused with the refined depth map and the joint bilateral filter is applied to produce a smooth and reliable depth map.

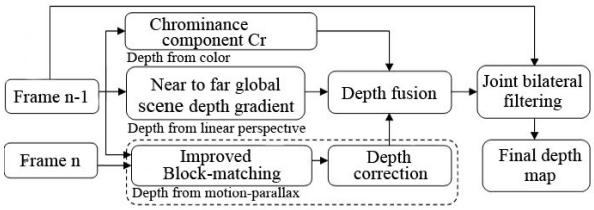

Fig.1: Block diagram of the proposed approach

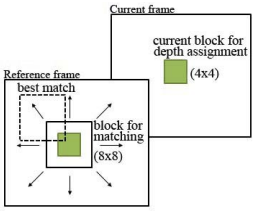

Fig.2: Improved motion estimation: different block sizes are used for matching and depth assignment.

The remainder of the paper is organized as follows. The proposed algorithm and the procedure of depth map generation are presented in Section 2. The simulation results are shown in Section 3. Finally, Conclusions are given in Section 4.

2. New Block-Matching Based Approach

The block diagram of proposed block-matching based approach is shown in Fig.1. A depth map is generated from three modules, Depth from Motion-parallax module (DM), Depth from Color module (DC), and Depth from Linear Perspective module (DLP). In the DM module, an improved block-matching algorithm is proposed to estimate the motion vectors on block level, and a depth correction procedure is applied to reduce the unreliable motion vectors and generates an initial depth map. Then the depth results from DC and DLP module are fused with the initial depth map to refine the detail depth value. Finally, a joint bilateral filter is adopted to eliminate the block effect and the staircase edges. The approach and the corresponding algorithms are described in detail as follows.

2.1. Depth from motion-parallax

2.2.1. Improved block-matching algorithm

It is noted that the motion vectors estimated from block matching may still be unreliable, factors like matching block size and uniform distribution of area luminance component would cause mismatch and produce unreliable motion vectors. In our proposed scheme, an improved block-matching motion estimation algorithm based on [7] is developed to estimate the depth information. To deal with the issue mentioned above, two improvements are made as follows.

Different block sizes are used for matching and depth assignment. The size of matching block is important for the estimated depth result. If the block size is too big, the estimated depth result will be very rough and lose a lot of detail depth information. Conversely, too small block size will decrease the matching accuracy and increase the number of unreliable motion vectors. To tackle this problem, different block sizes are used for matching and depth assignment in Fig.2. Current frame is divided into small blocks for depth assignment while the corresponding small blocks in reference frame are used as center and expanded to bigger blocks for matching. Then block-matching based motion estimation is performed to find the best match block and the motion vectors generated are used to assign depth for small blocks. In our simulation, the size of a small block is 4x4 and a big block is 8x8.

Adoption of color information in motion estimation. Most of the motion estimations are only performed in luminance domain, which may cause mismatch in the areas where the luminance components tends to distribute uniformly. Color Information has been proved very useful for the unreliable motion vector detection at the video decoders in [12]. In our method, color information is used in motion estimation to get more reliable motion vector. The evaluation metrics can be presented as follows:

$$E_{m,n} = \sum_{(i,j)\in b_{m,n}^{Y}} |r_{Y}(i,j)| + \alpha \times \sum_{(i,j)\in b_{m,n}^{Cb}} |r_{Cb}(i,j)| + \beta \times \sum_{(i,j)\in b_{m,n}^{Cr}} |r_{Cr}(i,j)| \quad (1)$$

Where $r_Y(i,j)$, $r_{Cb}(i,j)$, and $r_{Cr}(i,j)$ are the mean absolute differences (MAD) of the *Y, Cb* and *Cr* components of the 8x8 block ($b_{m,n}$), respectively. α and β are the weights used to emphasize the degree of color different. For the selection of α and β, we need to be careful not to overemphasize the color since the luminance is still the fundamental element for matching.

To reduce the computational complexity of motion estimation, diamond search algorithm is used in our application. The depth value *D(i , j)* is estimated by the magnitude of the motion vectors as follows:

$$D(i,j) = C\sqrt{MV(i,j)_x^2 + MV(i,j)_y^2} \quad (2)$$

Where $MV(i,j)_x$ and $MV(i,j)_y$ are the horizontal and vertical components of the motion vectors and *C* is a predefined constant. Fig.3 shows the depth results after the improvements.

2.2.2. Depth correction

Some unreliable depth blocks still exist in the initial depth map produced above. To refine the depth map, block-based mean filtering and median filtering are used to correct the unreliable depth value *(DV)*, in the block-based mean filter, the block *DV* is replaced by the mean *DV* of its neighboring blocks, while in the block-based median filter, the block is updated by the median DV of the blocks in the neighborhood.

Four thresholds (Th_1, Th_2, Th_3, and Th_4) are adopted in our scheme. Th_1 is the difference of the block *DV* and the mean *DV* of its neighboring blocks. In our experiments, it has been tested that if the difference is larger than 10, then either the *DV* is unreliable or there is a moving edge within the block. Th_2 is the sum of the pixel values in 4x4 blocks of the residue frame between the current frame and the reference frame. We use Th_2 to determine whether the block includes the edge of a moving object. In the residual frame blocks, moving objects edges usually contain more residual energy than static objects and the background [10]. With an appropriate threshold Th_2, most of the moving edges blocks can be detected. In our simulation, if Th_2 is larger than 700, then there is high probability that there is a moving edge within the block.

With Th_1 and Th_2, most of the unreliable depth blocks can be successfully detected. Then Th_3 and Th_4 are applied to determine whether the block is noise depth block or normal unreliable depth block. Noise depth block, which is usually either very big or very small in depth value, often appear randomly like image noises. In our experiment, let Th_3 be 240, and Th_4 be 20, the block with the DV larger than Th_3 or lower than Th_4 is considered as noise block and corrected by block-based median filtering. Otherwise, the block is normal unreliable block and updated by block-based mean filtering. Fig.3 shows the final depth after the depth correction.

2.2 Depth from Color

In [5], it has been shown that color component *Cr* can be used as proxy for depth despite the ambiguous and imprecise depth result. In our situation, *Cr* is not only used as a depth cue, but also used to refine the remaining block-effect of the depth map from motion parallax. To fuse depth map from motion parallax, *Cr* is mapping to a linear increasing gain from *Da* to *Db*, where *Da* and *Db* are the depth value interval of the final depth map from motion parallax.

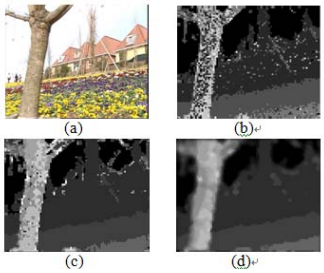

Fig.3: (a) The “flower” video. (b) The depth map estimate by block-matching algorithm. (c) The depth map after both the improvements is applied to the block-matching algorithm. (d) The final depth map after depth correction.

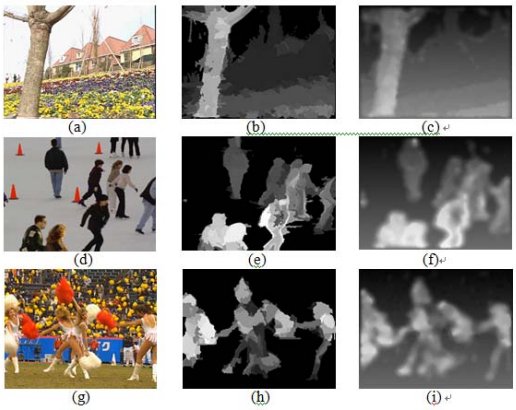

Fig.4: Experiment results. (a, d, g) 2D video sequence, (b, e, h) depth map estimated by FCS, (c, f, i) depth map estimated by our approach . h (d, e, f)

2.3 Depth from Linear Perspective

Research in [12] shows the result that depth from linear perspective can make the stereoscopic video more comfortable for human to watch. In our method, near to far global scene depth gradient is applied as the major cue as human visual perception tends to interpret the lower field of the image is closer than the higher field. The implementation is mainly from the work of [12].

2.4 Depth fusion and joint bilateral filtering

The three different depth maps are fuse according to W_m, Wc and W_l ($W_m + Wc + W_l = 1$) which are weighting factors for depth from motion-parallax, depth from color and depth from linear perspective, respectively. For selecting the values of W_m, W_c and W_l, we follow the principle that depth from motion is the main depth cue, depth from color and linear perspective are the auxiliary depth cues. Depending on different kinds of video, the weights may be changed a little. The fused depth map is then filtered by the joint bilateral filter [14].

3. Experimental Result

The performance of our proposed method is tested by using three 352x288 format 2D video sequences, “flower”, “ice” and “cheerleaders”, with a length of 12-second, 16-second and 12-second, respectively. The estimated depth maps generated by our approach and FCS [8] are shown in Fig.4. It is observed that both our algorithm and FCS can yield reliable depth estimates, but color segmentation algorithm usually has limitations like over segmentation when lighting source changes, which will generate inconsistent depth results. The visual quality of the final stereo video streams is subjectively evaluated by 10 volunteers participated. The volunteers were asked to watch stereoscopic videos in a random order and rate each video based on two factors, video quality and visual comfort. For the criteria of video quality, a five-segment scale assessment is used: (1) No 3D perception; (2) Poor 3D perception; (3) Fair 3D perception; (4) Good 3D perception; (5) Excellent 3D perception. For the criterion of visual comfort, another five-segment scale assessment is used: (1) Very uncomfortable with annoying artifact; (2) Uncomfortable with consistent artifact; (3) Mildly uncomfortable with tolerable artifact; (4) Comfortable with little artifact; (5) Very comfortable with no artifact. The artifact that makes the views uncomfortable includes image distortion caused by the inaccuracy of depth value and flickering artifact caused by inconsistence of the depth map.

The two factors are illustrated in Fig.5. For the video quality, FCS produce a slightly better protrusion effect than proposed algorithm, but for the visual comfort, the proposed algorithm performs better than FCS, and this is due to the inconsistence of the depth map caused by color segmentation algorithm. Both the algorithms work well for the sequences with regular motions, e.g., flower and ice sequence, but if the objects have complex self motions, e.g., cheerleaders sequence, both the algorithm produce different depth values for the same object, but as long as the depth values’ difference vary in a small range, the artifact and the visual discomfort caused can be negligible. Fig.6 shows the stereoscopic images produced by our approach.

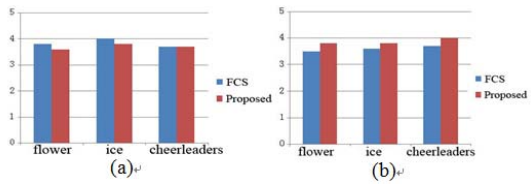

Fig.5: subjective tests results: (a) Image quality, (b) visual comfort.

Fig.6: Red-Cyan stereoscopic images of the test sequences: (a) flower, (b) ice, (c) cheerleaders.

The proposed algorithm achieves 0.82 fps while the FCS does 3.25 fps on *Intel® Core™ 2 Duo CPU T5670 @1.8GHz*. It is demonstrated that our algorithm has lower computational complexity, and is more suitable for real-time applications.

4. Conclusion

In this paper, a novel block-matching based algorithm for automatic 2D to 3D conversion is presented. An improved block-matching based motion estimation algorithm is proposed to get the main depth cue, in which different block sizes are used for block matching and depth assignment, and color information is adopted in motion estimation. With a depth correction procedure, the main depth cue from motion are fused with color cue and linear perspective cue to increase depth perception, and finally joint bilateral filtering is applied to refine the depth map. It has been shown that the proposed approach can greatly reduce the computation complexity with almost equally satisfactory quality compared to FCS and be more practical for real-time applications.

5. References

[1] C.Fehn, Depth-image-based rendering (DIBR), compression, and transmission for a new approachon 3D-TV, *SPIE*, 5291, pp. 93-104, 2004.

[2] P. Harman, J. Flack, S. Fox, and M. Dowley, Rapid 2D to3D conversion, *in Stereoscopic Displays and Virtual Reality Systems IX, vol. 4660 of Proceedings of SPIE*, pp. 78–86, SanJose, Calif, USA, January 2002.

[3] W. J. Tam, A. S. Yee, J. Ferreira, S. Tariq, and F. Speranza, Stereoscopic image rendering based on depth maps created from blur and edge information, *in Stereoscopic Displays and Virtual Reality Systems XII, vol. 5664 of Proceedings of SPIE* ,pp. 104–115, San Jose, Calif, USA, January 2005.

[4] Y.-M. Tsai, Y.-L. Chang, and L.-G. Chen, Block-based vanishing line and vanishing point detection for 3d scene recons truction, *Proceedings of the 2006 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS 2006)* , 2006.

[5] Wa James Tam, Carlos Vázquez, Filippo Speranza, Three-dimensional TV: A novel method for generating surrogate depth maps using colour information, *SPIE Electronics Imaging* 2009.

[6] I. Ideses, L. P. Yaroslavsky, and B. Fishbain, Real-time 2D to 3D video conversion, *Journal of Real-Time Image Processing,* vol. 2, no.1, pp. 3–9, 2007.

[7] Lai-Man Po, Xuyuan Xu, Automatic 2D-to-3D video conversion technique based on depth-from-motion and color segmentation, *IEEE International Conference on Signal Processing, ICSP 2010*.

[8] M Pourazad, Panos Nasiopoulos, Generating the Depth Map from the Motion Information of H.264-Encoded 2D Video Sequence, *EURASIP Journal on Image and Video Processing, Jan 1,* 2010.

[9] L.Zhang, W.J. Tam, Stereoscopic image generation based on depth images for 3DTV, *IEEE. Trans. Broadcasting 51 (2)* (2005) 191–199.

[10] A.M. Huang and T.Nguyen, Motion vector processing using the color information, *IEEE International Conference on Image Processing, ICIP,* 2009.

[11] Sung-Fang Tsai, Chao-Chung Cheng, Chung-Te Li, and Liang-Gee Chen A Real-Time 1080p 2D-to-3D Video Conversion System, *IEEE Transactions on Consumer Electronics*, Vol. 57, No. 2, May 201.

[12] J. Kopf, M. F. Cohen, D. Lischinski, and M. Uyttendaele, Joint bilateral up sampling, *ACM Trans. Graph.*, vol. 26,jul. 20.