A Journal of Software Engineering and Applications, **2012, 5, 144-148**
doi:10.4236/jsea.2012.512b028 Published Online December 2012 (http://www.scirp.org/journal/jsea)

A Fast Depth-Map Generation Algorithm based on Motion Search from 2D Video Contents

Weiwei Wang[1], Yuesheng Zhu[2]

[1]Communication and information Security Lab; [2]Shenzhen Graduate School, Peking University, China.
Email: cxjxfkl@126.com ,zhuys@pkusz.edu.cn

Received 2012

ABSTRACT

Generation of a depth-map from 2D video is the kernel of DIBR （Depth Image Based Rendering） in 2D-3D video conversion systems. However it occupies over most of the system resource where the motion search module takes up 90% time-consuming in typical motion estimation-based depth-map generation algorithms. In order to reduce the computational complexity, in this paper a new fast depth-map generation algorithm based on motion search is developed, in which a fast diamond search algorithm is adopted to decide whether a 16x16 or 4x4 block size is used based on Sobel operator in the motion search module to obtain a sub-depth-map. Then the sub-depth-map will be fused with the sub-depth-maps gotten from depth from color component Cr and depth from linear perspective modules to compensate and refine detail of the depth-map, finally obtain a better depth-map. The simulation results demonstrate that the new approach can greatly reduce over 50% computational complexity compared to other existing methods.

Keywords: Block-Matching; Depth-Map; Motion Search; DIBR

1. Introduction

Commercialization and industrialization of three-dimension televisions (3D TV) [1]not only depend on the development of 3D display as well as standardized technology, but also rely on a large amount of 3D video contents. Although 3D movie are on its way to develop, currently the 3D video contents are still not rich enough to satisfy the 3D-video market needs. In fact, the market is overwhelmed with 2D video. Converting 2D video into 3D video automatically and enabling the existing movies to be played on 3D displays becomes an important way to alleviate the shortage of 3D program. Also the 2D to 3D conversion technique can deliver the 3D videos effectively and efficiently. Therefore, the transition from 2D to 3D video is a low cost solution for the 3D industry compared with that captures 3D video directly.

There are some approaches for converting 2D video into 3D video [2-10]. Depth-map contains information relating to the distance of the scene objects from a viewpoint in a video content, and generating a depth-map effectively from 2D video is the kernel of DIBR in 2D-3D video conversion systems. The basic principle of DIBR [2,11] is to obtain a depth-map from 2D video and then synthesize the left and right views. The depth from motion （DFM）[5] is a kind of depth-map generation algorithm in which video is segmented first and the frame disparity is estimated to obtain the depth-map. But the DFM requires that moving objects must exist in successive frames. Fusion with color information can improve the depth-map quality [12] [13]. In the literature [12], the depth-map generated from motion-parallax is fused with color segmentation to obtain a clear and reliable depth-map, but its color segmentation algorithm introduces high computational complexity. In the literature [13] motion estimation is performed by using luminance and chrominance information in motion search module to yield a reliable depth-map and reduce the computation complexity. However, the computational complexity of the color information in motion search module is still high.

In this paper, a new fast depth-map generation algorithm based on motion search is developed, in which a fast diamond search algorithm is adopted to decide whether a 16 x 16 or 4x4 block size is used based on Sobel operator[14,15] in the motion search module without using color information to obtain a main sub-depth- map, and then the depth from color component Cr [4] and the depth from linear perspective [6] are used as auxiliary sub-depth-maps to fuse the main sub-depth- map. Finally the bilateral filter is adopted to eliminate the block effect and the staircase edges that remained in the fused depth-maps. The results show that with the proposed algorithm a smooth and reliable depth-map and a better visual 3D video can be obtained with low computational complexity compared to the methods in [12] and [13].

The remainder of the paper is organized as follows. The proposed depth-map generation algorithm is presented in section II. Experimental results are provided in section III. Finally, a conclusion is given in section IV.

2. The Proposed Depth-Map Generation Algorithm

The block diagram of proposed algorithm is shown in **Figure 1**.

In **Figure 1**, the final depth-map is fused with three sub-depth-maps, that is, depth from improved motion estimation, depth from color component Cr, and depth from linear perspective. In depth-map fusion, depth from color component Cr and depth from linear perspective are used as auxiliary sub-depth-map to compensate the main sub- depth-map gotten from improved motion estimation. And then a bilateral filter is adopted to eliminate the block effect and the staircase edges. In this section, the module of improved block-matching based depth from motion estimation is developed and described. The approach and the corresponding algorithms are described in detail as follows.

2.1. Depth from Improved Motion Estimation

In paper [7], the motion estimation is performed by using luminance information, which may cause mismatch in the areas where the luminance components tend to distribute uniformly. In paper [10], luminance and chrominance information is adopted in the motion vector processing which uses Y (luminance component), C_r (red component) and C_b (green component) to calculate the motion vectors. Our comparison results of time consumption in motion search module for four cases:(1) Y, (2) Y and C_b,(3) Y and C_r, (4) Y, C_r and C_b are shown in **Figure 2**.

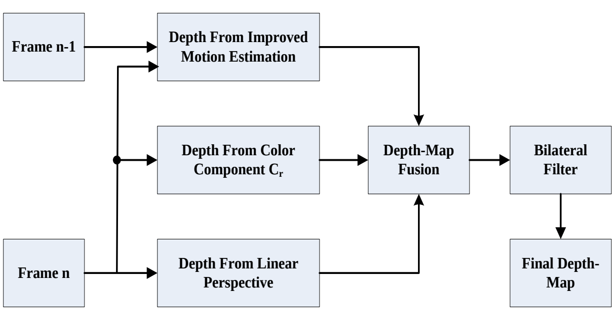

Figure 1. Block diagram of proposed algorithm.

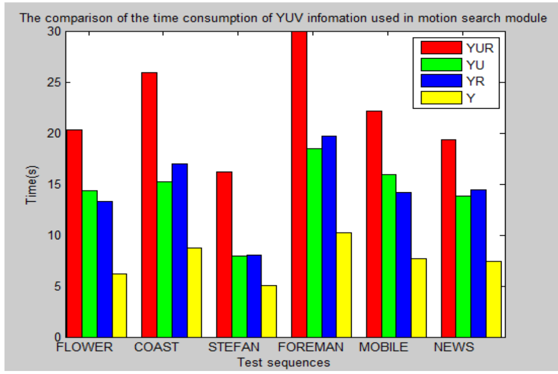

Figure 2. The time consumption in motion search module in four cases.

The results indicate that using color information in motion search module would increase the computational complexity. So, in the proposed module, the motion estimation is performed by using luminance information only in motion search module without using color information. But the depth from color component C_r is used as auxiliary sub-depth-map to fuse the sub-depth-map obtained from improved motion estimation.

In motion search module 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. The block-matching algorithm based motion estimation [7] utilizes the fact that objects with different motions usually have different depths. Near objects move faster than far objects and the relative motions are used to estimate the depth-map. The depth value D(i,j,k) are estimated by the magnitude of the motion vectors as follows:

$$D(i,j,k) = C\sqrt{MV(i,j,k)_x^2 + MV(i,j,k)_y^2} \quad (1)$$

where MV(i,j,k)x and MV(i,j,k)y are horizontal and vertical components of the motion vectors and C is a predefined constant. It is noted that the motion searching module takes as much as 40 percentage of the total time consumption. In order to reduce the computational complexity a fast diamond search algorithm is adopted in the new motion search module.

The common algorithms generate depth-map based on motion estimation in 4 x 4 block size. While, we observe that if the frame picture is homogenous enough, the depth value of blocks is close to their neighbors. We find it will save much computational complexity if we use 16x16 block size instead of 4 x 4 block size in homogeneous area. To evaluate the smoothness of a picture, statistical measurement such as standard deviation, variance, skewness and kurtosis [16] are used. Paper [14,15] use Sobel operator to create the edge maps of pictures for high efficiency of video coding process. In order to classify the homogeneity of a block, the amplitude of edge vector is defined by formula (2).

The vertical and horizontal directions are defined according to a luminance or chrominance pixel at position (i,j) with value Vij by formula (2)(3). Then the block homogeneity measurement H can be set by formula (5).

$$Amp(\overline{E}_{i,j}) = \left|Ex_{i,j}\right| + \left|Ey_{i,j}\right| \quad (2)$$

$$Ex_{i,j} = v_{i-1,j+1} + 2\times v_{i,j+1} + v_{i+1,j+1} - v_{i-1,j-1} - 2\times v_{i,j-1} - v_{i+1,j-1} \quad (3)$$

$$Ey_{i,j} = v_{i+1,j-1} + 2\times v_{i+1,j} + v_{i+1,j+1} - v_{i-1,j-1} - 2\times v_{i-1,j} - v_{i-1,j+1} \quad (4)$$

where $i \in 1,2,\ldots,R,\ j \in 1,2,\ldots,C,$

$$H_{r,c} = \begin{cases} 1, & \sum_{0\le i,j<N} Amp(\overline{E}_{i,j}) < Thd_H \\ 0, & otherwise \end{cases} \quad (5)$$

After the block size is assigned, the fast diamond search algorithm makes two round iterations to calculate the motion vectors [17]. In the first round, it takes a large diamond search pattern (Fig.3a) and 9 test points are compared to find the best points. The iteration process breaks out when the center point is just the most matching point. The second round aims to find the best point in a small diamond search pattern (Fig.3b). The first round takes at least 90% time according to related experiment. So we design a quick quit scheme to exit the iterative loop ahead of time if the SAD (Sum of Absolute Difference) of current point is less than the predefined threshold. A fast depth-map generation algorithm based on motion search is as follows.

Step1. Read the current macro block data (16x16 block size) and compute the homogeneity measurement H of the block by formulas (2)(3)(4)(5). If H equals to 0, the block is divided into 16 little 4 x 4 block size, and the motion search module will be based on the 4 x 4 block size, otherwise the large 16 x 16 block size will be used.

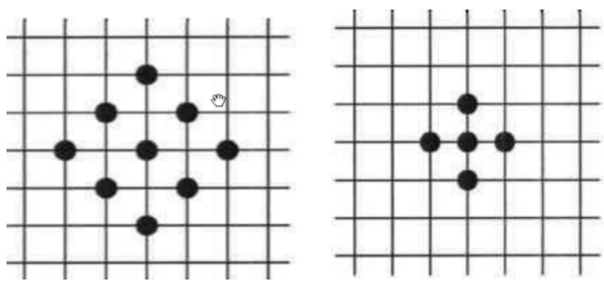

Figure 3. Diamond search pattern. (a) Large diamond search pattern. (b) Small diamond search pattern.

Step2. Compute the SAD of the 9 neighbor points of center point in large diamond pattern (Fig.3a). If the SAD value of current point is less than the given threshold T1, go to step 4; if it is larger than T1 but less than T2, go to step 3; otherwise, compute the minimum SAD value and loop over the 9 points.

Step3. Compute the SAD of the 5 neighbor points of center point in small diamond pattern (Fig.3b). The best point will be chosen as the matching point.

Step4. Compute the depth value of the block with formula (1).

2.2. Depth from Color Component Cr

In the research of [4], it has been proved that different objects have different hues in the 2D color video sequences, and each of the hues has its own associated grey level intensities in the Cr color component images. If we take the gray level intensities as indexes of depth, the depth of the boundaries of each object is different from that of its immediate surroundings. Therefore gray intensity images associated with the color component Cr of standard 2D-colour video sequences can be used as proxy depth-map. In our situation, the depth from color component Cr which is derived directly from the current frame of the 2D images is used as auxiliary sub-depth-map to fuse the sub-depth-map from improved motion estimation. The fused depth-map can increase the accuracy and detail of sub-depth-map. Others, for static scene we can get the different depth values using color component Cr while the other methods [7] obtain the same depth values. So depth from color component Cr can strengthen the layer of the stereo videos. The depth from color component Cr is shown in Fig. 4(b).

2.3. Depth from Linear Perspective

Research in [6] shows 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 auxiliary depth map as human visual perception tends to interpret most of the

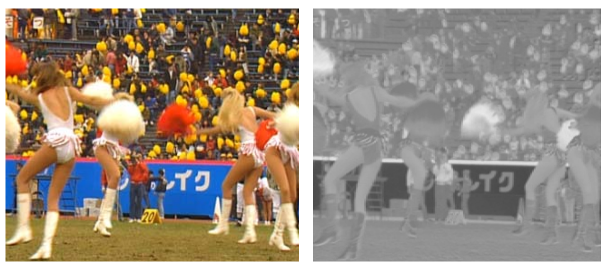

Figure 4. (a) The depth-map from color component Cr.(a) The original "cheerleader" video image.(b)The depth from color component Cr.

images represent scenes in which the bottom part is related to the ground and consequently close to us and the upper part represents the sky and consequently far from us.

2.4. Depth-Map Fusion

The final depth-map is fused with three sub-depth-maps, that is, depth from improved motion estimation, depth from color component Cr, and depth from linear perspective. In this paper a simple linear module is used to fuse the three kinds of sub-depth-maps. The fused depth-map can be described in the following equation:

$$D_{all} = D_m \times W_m + D_c \times W_c + D_l \times W_l \quad (6)$$

$$W_m + W_c + W_l = 1 \quad (7)$$

Where Dm, Dc and Dl are the values of sub-depth-maps estimated by motion estimation, color component Cr and linear perspective respectively. Dall is the values of the fused depth-map while Wm, Wc and Wl are the weights of them.

Because in depth-map fusion, depth from improved motion estimation is used as the main sub-depth-map, and depth from color component Cr and depth from linear perspective are used as the auxiliary sub-depth-maps, selecting the values of Wm, Wc and Wl we follow the principle that the value of Wm is larger than Wc and Wl.

3. Experimental Results

To evaluate our proposed algorithm, several test sequences are used to perform the [12] ，[13] and our method to make some comparisons on the efficiency. We set the threshold of Sobel to 5000 while the thresholds of motion search modules are 0.15 and 0.3. The test results are shown in Table 1.

Table 1. The test results of several sequence

Sequence	Time(s)		
	Po [12]	Chen [13]	Our method
FLOWER	4.12	8.12	2.92
COAST GUARD	3.45	8.23	1.98
CARPHONE	4.23	7.98	2.03
FOREMAN	4.47	8.43	2.65
MOBILE	5.01	8.90	3.21
CALENDAR	5.43	8.79	3.45

According to Table 1,we can easily find that the efficiency of our method is better than the algorithms in [12] and [13], about 31% of time saving than [12] and 65% to [13]. The promotion is especially remarkably in CAR PHONE and COASTGUARD for the large scale of smooth area in these two sequences. It is the contribution of Sobel operator which decides whether a 16x16 or 4x4 block size is used, and the quick quit scheme to exit the iterative loop ahead of time in the search module.

The sub-depth-map obtained from the improved motion estimation module is shown in Fig.5.(b). From the Fig.5.(b), we can see that there are many isolated points in this sub-depth-map. So this sub-depth-map will be fused with the sub-depth-maps gotten from depth from color component Cr and depth from linear perspective modules to eliminate these isolated points and compensate this sub-depth-map and get a better fused depth-map as shown in Fig.5.(c). Finally, a smooth and reliable depth-map shown in Fig.5.(d) is obtained by passing the bilinear filter.

4. Conclusions

In this paper, a fast depth-map generation algorithm based on motion search is proposed to enhance the efficiency of the generation of depth-map. In the new proposed modules, a fast diamond search algorithm is adopted to decide whether a 16x16 or 4x4 block size is used based on Sobel operator in the motion search module without using color information to obtain a sub-depth-map, and this sub-depth-map will be fused with the sub-depth-maps gotten from depth from color component Cr and depth from linear perspective modules respectively to compensate this sub-depth-map and obtain a improved fused depth-map. Finally, the bilateral

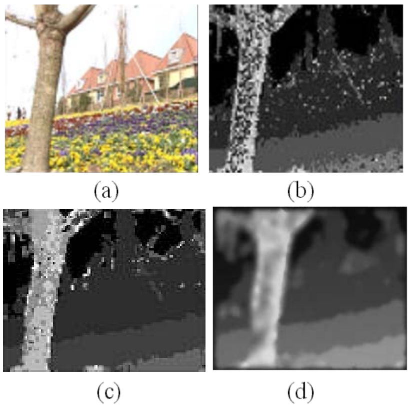

Figure 2. Depth-map from our proposed algorithm (a) The original "flower" video image. (b) The sub-depth-map estimate by improved motion estimation module. (c) The fused depth-maps after fused with three sub-depth-maps. (d) The final depth-map by passing bilinear filter.

filter is adopted to eliminate the block effect and the staircase edges that remained in the fused depth-map. The results show that with the proposed algorithm a smooth and reliable depth-map and a better visual 3D video can be obtained with over 50% reduction of computational complexity compared to the other methods.

REFERENCES

[1] M. Op de Beeck, and A. Redert, "Three Dimensional video for the Home," Proceedings of International Conference on Augmented, Virtual Environments and Three-Dimensional Image, May-June 2001, pp. 188-191.

[2] Fehn, C.: Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV. SPIE, vol.5291, no.2, pp. 93-104 2004.

[3] P. Harman, J. Flack, S. Fox, M. Dowley.: Rapid 2D to 3D Conversion. Proceedings of SPIE, vol. 4660, pp. 78-86, 2002

[4] Wa James Tam, Carlos Vázquez, Filippo Speranza.: Three-dimensional TV: A novel method for generating surrogate depth maps using color information. Proc. SPIE Electronics Imaging-stereoscopic Displays and Applications XX, 2009.

[5] D. Kim, D. Min, K. Sohn.: Stereoscopic video generation method using motion analysis. in Proceedings of the 3DTV Conference, pp. 1–4, Kos Island, May 2007.

[6] Sung-Fang Tsai, Chao-Chung Cheng, Chung-Te Li, Liang-Gee Chen.：A Real-Time 1080p 2D-to-3D Video Conversion System. IEEE Transactions on Consumer Electronics, Vol. 57, No. 2, pp. 803–804, May 2011.

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

[8] C. Tomasi, R. Manduchi.: Bilateral Filtering for Gray and Color Images. Proceedings of the IEEE International Conference on Computer Vision, Bombay, India, pp.839-846, Bombay January 1998.

[9] M. T. Pourazad, P. Nasiopoulos, and R. K. Ward.: An H.264-based scheme for 2D to 3D video conversion. IEEE Transactions on Consumer Electronics, vol.55, no.2, pp. 742–748, 2009.

[10] A.-M. Huang, T. Nguyen.: Motion vector processing using the color information. IEEE International Conference on Image Processing, ICIP, pp. 1605-1608, Cairo, 2009.

[11] W. J. Tam, f. Speranza, L. Zhang, R. Renaud, J. Chan, C. Vazquez.: Depth Image Based Rendering for Multiview Stereoscopic Displays: Role of Information at Object Boundaries. Three-Dimensional TV, Video, and Display IV, vol. 6016, pp. 75-85, 2005.

[12] L. Po, X. Xu, Y. Zhu, S. Zhang: Automatic 2D-to-3D video conversion technique based on depth-from-motion and color segmentation. IEEE International Conference on Signal Processing, ICSP, pp.1000-1003, 2010.

[13] J. Chen, Yuesheng Zhu and X. Liu.: A New Block-Matching Based Approach for Automatic 2D to 3D Conversion. The 4th International Conference on Computer Engineering and Technology, ICCET, pp. 109-113，Thailand,2012.

[14] D. Wu, S. Wu, K. Lim, F. Pan, Z. Li, X. Lin: Block INTER mode decision for fast encoding of H.264. IEEE International Conference on Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04). vol.3, 2004 pp. iii- 181.

[15] F. Pan, X. Lin, R. Susanto, K. P. Lim, Z. G. Li, G. N. Feng,D. J. Wu, and S. Wu.: Fast Mode Decision Algorithm for Intra Prediction in JVT. 7th JVT meeting, JVT-G013, Thailand, March 2003.

[16] K. R. Castleman.: Digital Image Processing. Prentice Hall Inc, 1996.

[17] Shan Zhu, Kai-Kuang Ma.: A new diamond search algorithm for fast block-matching motion estimation. IEEE Transactions on Image Processing, vol.9, no.2, pp.287-290, Feb 2000.

