An Automatic Depth Map Generation for 2D-to-3D Conversion

Hao Dong, Shouyi Yin, Weizhi Xu, Zhen Zhang, Rui Shi, Leibo Liu and Shaojun Wei
Institute of Microelectronics
Tsinghua University
Beijing, China
h-dong12@mails.tsinghua.edu.cn

Abstract—**In this paper, we propose an automatic algorithm which generates depth map of 2D image in good quality. The depth map is generated by fusing global depth gradient and local depth refinement. Image classification technique and edge-based depth cue are used to obtain the global depth gradient. In addition, image segmentation method and warm/cool color theory are adopted to generate local depth refinement. The proposed algorithm has advantages of good quality and low complexity. After the depth map is generated, the original 2D image or video can be converted to 3D for showing on stereoscopic display devices.**

Keywords—depth map generation; stereo vision; 2D-to-3D conversion; 3D TV;

I. Introduction

Three-dimentional television (3D-TV) is becoming more and more popular and attracting more and more viewers to experience vivid stereoscopic visual effects. Plenty of 3D contents are needed during the development of 3D-TV, however, the production of 3D contents of high quality are costly and time-consuming. To alleviate the problem of 3D material shortage, converting 2D contents to 3D contents is one way which can make full use of already existing 2D materials. The difference between 2D content and 3D content is the depth information which represents the relative distances of the objects in the 2D content. The generation of comfortable and natural depth is of great importance and the low computational complexity is also required.

This paper proposes a depth map generation algorithm which blends different depth cues and techniques. Both of them are low complexity. Combining these depth cues and techniques, our algorithm produces comfortable and natural depth maps from 2D materials.

II. Proposed Algorithom

A. Global depth gradient generation

The global depth gradient generation uses simple image classifcation technique and edge-based depth cue. To generate the global depth gradient, we first classify the scenes of images into two types, sky/ground type and normal type. Then we use different methods of the edge-based depth cue for different types.

The composition of a scene of sky/ground type is that the upper part is a sky and the lower part is the ground. We can judge a pixel whether belongs to a sky or the ground by the pixel value of HSI-color space [1]. That is

$$Sky(x,y)=1, \quad if \quad (100<H(x,y)<180)\,\&\&(100<I(x,y)<255) \tag{1}$$

$$Ground(x,y)=1, \quad if \quad (50<H(x,y)<100)\,\&\&(80<S(x,y)<255) \tag{2}$$

H(x,y), S(x,y) and I(x,y) are the hue, saturation and intensity of pixel (x,y). Sky(x,y) means the pixel belonging to a sky and Ground(x,y) means the pixel belonging to the ground.

$$Amounts=\sum_{x=1}^{X}\sum_{y=1}^{Y}(Sky(x,y)\,||\,Ground(x,y)) \tag{3}$$

$$Scene=\begin{cases} sky\,/\,ground\ type, & Amounts>threshold \\ normal\ type, & Amounts<threshold \end{cases} \tag{4}$$

X and Y denote the height and width of the scene. If Amounts, the total amounts of pixels belonging to sky or ground of a scene, is larger than a threshold, the scene of the image can be classified into sky/ground type. Otherwise, it is classified into normal type.

For the scene of sky/ground type, the upper part of the scene is a sky and the lower part of the scene is the ground so the depth gradient is roughly near-to-far from bottom-to-top. We make use of the cumulative horizontal edge histogram of the scene to assign the global depth gradient [2]. Because the horizontal edge histogram represents the horizontal complexity and the ground often contains more complex edges, there is a distinct depth change between the sky and the ground.

For the scene of normal type, the area containing more details often draw much attention of viewer. So it is reasonable to produce a closer depth of the area than other areas. The massive edges often represent more details and the area which contains most edges can be assigned a depth closer to the

This work is supported in part by the China Major S&T Project (No.2013ZX01033001-001-003), the International S&T Cooperation Project of China grant (No. 2012DFA11170), the Tsinghua Indigenous Research Project (No.20111080997) and the NNSF of China grant (No.61274131).

viewer than other areas. We simply divide the scene into blocks and find out the two blocks which contains the most edges. If the two blocks are adjacent and the difference between each amount of edges of each block is within a threshold, the two blocks can be merged to one region. Otherwise, the block which contains the most edges is used for the following computation. We obtain the centre point of the block or the region and regard it as the nearest point. Then the global depth gradient can be assigned according to the distance between each point of the scene and the centre point. The detailed procedure for global depth gradient generation of normal type is shown in Fig1.

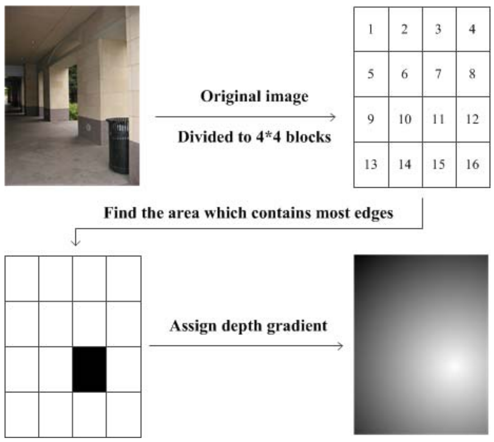

Fig. 1. Global depth gradient generation of normal type

B. Local depth refinment

Local depth refinement makes use of warm/cool color theory and simple image segmentation technique. Warm/cool color theory is that the pixel of a warm color is nearer to the viewer than the pixel of a cool color. So the pixel of redder color is assigned a depth closer to the viewer. We generate the local depth component D_{Cr} by simply using the Cr component of the YCbCr-color space.

We adopt a simple segmentation by using the higher 4-bit of Y component to combine the similar pixels together [3]. And Y_{seg} is used to represent the higher 4-bit of Y component. Because the higher luminance often gives viewer a closer feeling, we generate the local depth component D_{y-seg} by simply using the Y_{seg} .

Finally, the two local depth components are fused with the global depth gradient as the following equation:

$$D_{fused} = \alpha * D_{global} + \beta * D_{Cr} + (1-\alpha-\beta) * D_{y-seg} \quad (5)$$

where D_{fused} denotes the fused depth and D_{global} denotes the global depth gradient. α and β denote weights for the D_{global} and the D_{Cr} respectively and they are both less than 1. In practice, fixed weights are used to avoid high computational complexity and temporal flicker.

C. 3D image generation

We adopt 2D Gaussian filter to improve the quality of depth map and Depth-Image-Based Rendering technique to generate 3D images from the original 2D content and the depth map [4].

III. EXPERIMENTAL RESULTS

We tried various kinds of contents and found the depth maps are in good quality. Fig. 2 shows data on two different types of images of the original image, global depth gradient and final depth map.

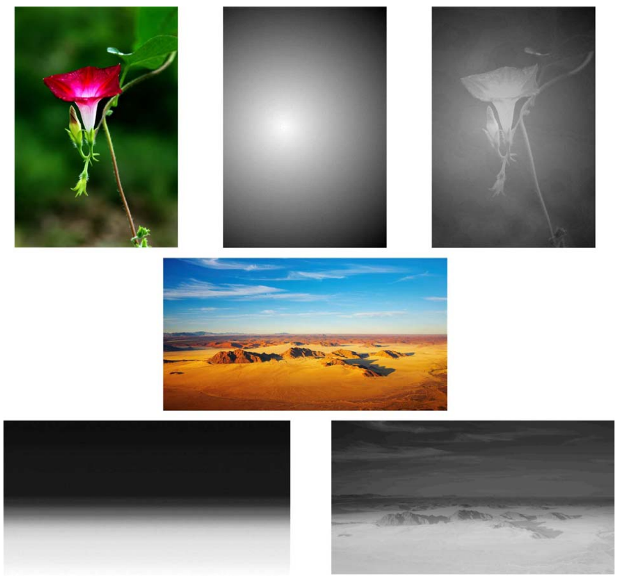

Fig. 2. Original 2D image, global depth gradient and final depth map of normal type (first row), original 2D image of sky/ground type (second row) and global depth gradient and final depth map of the image (bottom).

IV. CONCLUSION

The paper proposes a low-complexity depth fusion algorithm for 2D-to-3D conversion. We make use of different methods of edge-based depth cue according to different image types to generate the global depth gradient. To refine the global depth gradient, simple image segmentation technique and warm/cool color theory are used. The experiment shows that the generated depth maps are comfortable and natural.

REFERENCES

[1] Y. S. Huang, F. H. Cheng, and Y. H. Liang, "Creating Depth Map from 2D Scene Classification," 3rd International Conference on Innovative Computing Information and Control (ICICIC '08), pp. 69–69, June 2008.

[2] S.F. Tsai, C.C. Cheng, C.T. Li, and L.G. Chen, "A real-time 1080p 2D-to-3D video conversion system", IEEE International Conference on Consumer Electronics (ICCE), pp. 803-804, January 2011.

[3] C. A. Chien, C. Y. Chang, J. S. Lee, J. H. Chang and J. I. Guo, "Low complexity 3D depth map generation for stereo applications," IEEE International Conference on Consumer Electronics (ICCE), pp. 185–186, January 2011.

[4] L. Zhang, C. Vázquez, S. Knorr, "3D-TV Content Creation: Automatic 2D-to-3D Video Conversion," IEEE Transactions on Broadcasting, vol. 57, no. 2, pp. 372–383, June 2011.