Depth Map Generation from a Single Image Using Local Depth Hypothesis

Na-Eun Yang, *Student Member*, *IEEE*, Ji Won Lee, *Student Member*, *IEEE*, and Rae-Hong Park, *Senior Member*, *IEEE*

***Abstract*—This paper proposes an interactive method of depth map generation from a single image for 2D-to-3D conversion. First, the proposed method groups an input image into similar regions to preserve details and segments the image into salient regions with user interaction. Then, the proposed method generates local depth hypothesis using structural information of the input image and salient regions. Finally, the proposed method generates depth map using local depth hypothesis and grouped regions. Experimental results show that the proposed method gives more natural depth map in terms of human perception. It can be applied to depth image based rendering such as stereo view rendering.**

I. INTRODUCTION

Recently, commercial three dimensional (3D) markets have grown incredibly. However, because of a limited number of available 3D contents, 2D-to-3D conversion is proposed as an alternative to meet consumer's demand of 3D contents. Most of the 2D-to-3D conversion techniques are time-consuming processes with a lot of human intervention required. There are some automatic methods but they still need to be improved. Therefore, the 2D-to-3D conversion with less human intervention is needed.

Human perceives depth from various heuristic, monocular depth cues [1]–[3]: focus/defocus, relative height/size, texture gradient, structural feature from occlusion, geometry, and so on. These monocular depth cues make people perceive depth from a single-view image. Based on these facts, many studies have been done [1]–[3]. In a single image that has various monocular cues, it is difficult for these studies to estimate depth map naturally from monocular cues alone. We propose a depth map generation method from a single image for 2D-to-3D conversion using local depth hypothesis. The proposed method is a semi-automatic and simple method with a little intervention from a user.

The rest of the paper is organized as follows. Section II describes the proposed method of depth map generation. Experimental results of the proposed method are given and discussed in Section III. Finally, in Section IV conclusions and future research directions are given.

II. PROPOSED DEPTH MAP GENERATION METHOD

We propose an interactive depth map generation algorithm using local depth hypothesis. Fig. 1 shows a block diagram of the proposed depth map generation method. It consists of four parts: scene grouping, local depth hypothesis generation, depth assignment, and depth map refinement. Let I be an input image and M denote a user input, which indicates how to segment I into several salient regions S. H_{local} is the local depth hypothesis and G represents the grouped image using a graph-based segmentation algorithm [4]. D_{init} signifies the initial depth map and D_{final} denotes a refined final depth map. Description of each part is given in the following.

A. Scene grouping

The local depth hypothesis H_{local} represents abrupt depth discontinuity between salient regions. However, it is not enough to show details of depth variation in each salient region, and depth discontinuities can exist between objects in each salient region. So we need a segmentation to represent detailed depth discontinuities in the salient region. It is assumed that regions of similar intensity are likely to have similar depth. We use a graph-based segmentation algorithm [4] in grouping similar regions in order to improve salient segmentation and assign the same depth value at the next stage. Fig. 2(a) shows the scene grouped result G of input image with a larger number of detail regions compared to salient segmentation S in Fig. 2(c).

B. Depth hypothesis generation

Depth hypothesis is generated with structural information of an input image and a little user interaction. First, a user defines salient regions of an input image as shown in Fig. 2(b). First, an input image is segmented into two or three salient regions using a graph-cut algorithm [5]. One of segmented regions (gray region) represents a main object whereas the others (white region and black region) are background, as shown in Fig. 2(c). Then, to make the local depth hypotheses of each region, we detect lines from an edge map of the input image. Lines corresponding to each salient region determine a set of vanishing point, V [6]. Fig. 2(d) shows a detection result of the vanishing point of the gray region in Fig. 2(c).

In depth map, depth is expressed in grayscale. The vanishing point represents the farthest point. Depth values of regions are gradually increased from the vanishing point, where the brighter the region is, the closer it is located to a camera. If the vanishing point is inside the input image, the depth hypothesis will be determined by the chessboard

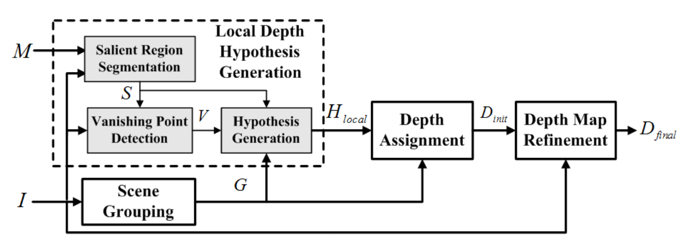

Fig. 1. Block diagram of the proposed method.

This work was supported in part by the Second Brain Korea 21 Project.

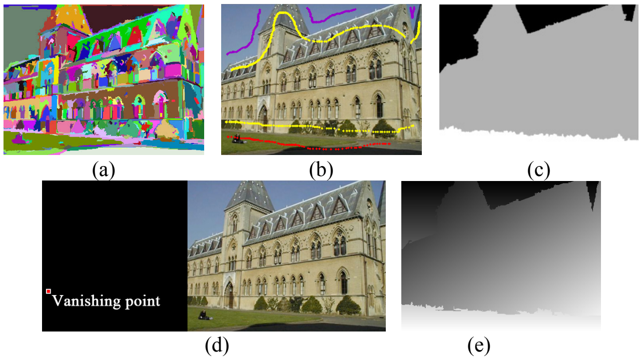

Fig. 2. Results of scene grouping and depth hypothesis generation. (a) scene grouping result, (b) original image with user input, (c) salient region segmented image, (d) detected vanishing point, (e) local depth hypothesis.

distance [7] from the vanishing point. Otherwise, the depth hypothesis is generated by the combination of four basic hypotheses: bottom to top (Fig. 3(a)), top to bottom (Fig. 3(b)), right to left (Fig. 3(c)), and left to right (Fig. 3(d)). If some salient regions have no detected line, the depth hypothesis of each region will be determined by default as bottom to top (Fig. 3(a)). With the salient region segmented image (Fig. 2(c)), we combine two types of depth hypotheses (Figs. 3(a) and 3(c)) to reflect the general tendency of depth discontinuity with salient features preserved. Its result, i.e., the local depth hypothesis H_{local} is shown in Fig. 2(e).

C. *Depth assignment and refinement*

We assign a depth value to each scene group using the local depth hypothesis. The depth value of a given group is assigned by local depth hypothesis and the average depth value in the local depth hypothesis as shown in Fig. 2(e).

In the initial depth map D_{init}, each region can have a depth value different from those of neighboring pixels though they have similar depth values that belong to the same object in the original image. If one region with the same depth in a real scene is divided into several sub-regions with different depth values, it can produce unnatural artifacts. So, the proposed method refines the initial depth map, in which a cross-bilateral filter is used for depth smoothing [2]. The input image is used as a reference image for cross-bilateral filtering. This process can preserve discontinuities in depth, while smoothing regions of similar intensity at the same time. Fig. 4(c) shows the final result of the proposed method.

III. EXPERIMENTAL RESULTS AND DISCUSSIONS

Fig. 4 shows simulation results of the proposed method compared with two existing 2D-to-3D conversion methods [2], [3]. Fig. 4(a) shows the depth map by Cheng *et al*.'s algorithm [2] which uses a single (right to left) depth hypothesis. Fig. 4(b) shows the depth map by Han and Hong's method [3], in which a depth hypothesis is estimated from Gaussian distribution with a vanishing point and height depth cue. Because both methods are global approaches with a single depth hypothesis, the results cannot accurately reflect the local depth discontinuity. In the results of Cheng *et al*.'s algorithm, the right part of the sky appears closer than building. And two

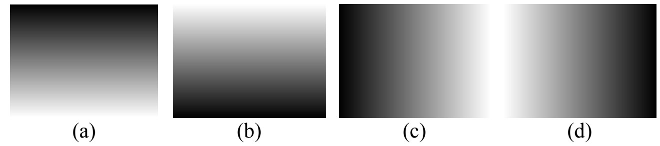

Fig. 3. Four basic hypotheses. (a) bottom to top, (b) top to bottom, (c) right to left, (d) left to right.

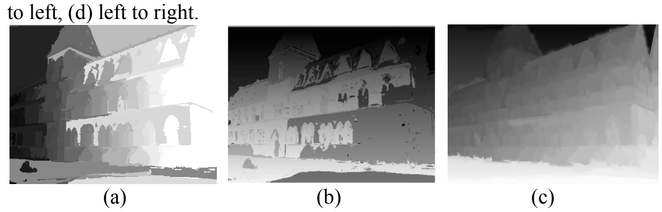

Fig. 4. Final depth map. (a) Cheng *et al.*'s method [2], (b) Han and Hong's method [3], (c) proposed method.

existing methods give a big depth difference between the right side of outer wall of the building and the left side. However, our method can effectively reflect the local depth transition by the salient segmentation. Therefore, the sky appears farther than building and the depth changes gradually in the outer wall. With simple user interaction, we can generate a depth map that preserves both the global and local transitions of depth. The proposed method can be applied to key frame interpolation for 2D-to-3D video conversion [8].

IV. CONCLUSIONS

This paper proposes a depth map generation method from a single image for 2D-to-3D conversion with user interaction. The proposed method combines depth hypotheses with the salient segmented image, and refines the initial depth map using a cross-bilateral filter. The proposed depth map maintains salient depth values and local transition of depth. It can generate natural depth map from the viewpoint of human perception. Future research will focus on reducing human intervention, so that ultimately, the proposed depth map generation method can be automated.

REFERENCES

[1] Y. J. Jung, A. Baik, J. Kim, and D. Park, "A novel 2D-to-3D conversion technique based on relative height depth cue," in *Proc. Stereoscopic Displays and Applications XX*, vol. 7237, doi: 10.1117/12.806058, San Jose, CA, Jan. 2009.

[2] C.-C. Cheng, C.-T. Li, and L.-G. Chen, "A novel 2D-to-3D conversion system using edge information," *IEEE Trans. Consumer Electronics*, vol. 56, no. 3, pp. 1739–1745, Aug. 2010.

[3] K. Han and K. Hong, "Geometric and texture cue based depth-map estimation for 2D to 3D image conversion," in *Proc. 2011 IEEE Int. Conf. Consumer Electronics*, pp. 651–652, Las Vegas, NV, Jan. 2011.

[4] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *Int. J. Computer Vision*, vol. 59, no. 2, pp. 167–181, Sep. 2004.

[5] Y. Li, J. Sun, C.-K. Tang, and H.-Y. Shum, "Lazy snapping," *ACM Trans. Graphics*, vol. 23, no. 3, pp. 303–308, Aug. 2004.

[6] V. Cantoni, L. Lombardi, M. Porta, and N. Sicard, "Vanishing point detection: Representation analysis and new approaches," in *Proc. 11th Int. Conf. Image Anal. and Process.*, pp. 90–94, Palermo, Italy, Sept. 2001.

[7] R. C. Gonzalez and R. E. Woods, *Digital Image Processing, Third edition*. Upper Saddle River, NJ: Pearson Education Inc., 2010.

[8] W.-N. Lie, C.-Y. Chen, and W.-C. Chen, "2D to 3D video conversion with key-frame depth propagation and trilateral filtering," *Electron. Lett.*, vol. 47, no. 5, pp. 319–321, Mar. 2011.