Automatic Depth Map Generation from a Single Image Using Segment-Adaptive Depth Merging

Sung In Cho, Kyoungrok Cho, and Young Hwan Kim
Department of Electrical Engineering, Pohang University of Science and Technology (POSTECH)
csi2267@postech.ac.kr

***Abstract*-- This paper proposes an advanced approach to learning-based depth map generation for automatic 2D-to-3D conversion, which utilizes 3D histogram-based segmentation for a segment-adaptive depth merging. By using a segment-adaptive depth merging, the proposed method successfully enhances the accuracy of the conventional learning-based depth map generation.**

I. INTRODUCTION

With recent advances in 3D display technology, 3D-capable hardware device has come into wide use. However, the availability of 3D content production cannot meet the demands for 3D display. For alleviating this problem, 2D-to-3D conversion has been extensively exploited because it allows to produce 3D contents from enormous 2D contents.

A typical 2D-to-3D conversion process consists of two steps: depth estimation for a given 2D single image and depth-based rendering (DIBR) in order to produce a stereo-image pair from a given 2D single image [1]. Whereas DIBR has a typical process and generally produces good results, the depth estimation from a single image is a challengeable task. Hence, many studies for the depth estimation have been done. Among them, learning-based depth estimation [1], [2] successfully improved the quality of the depth map. However, this approach provides only the rough depth information because it produces a depth map by fusing photometrically similar training images with a given image. In addition, the accuracy of this approach can be significantly lowered when the number of training images is not enough to estimate a depth map. Hence, if additional information is utilized to refine the resultant depth map of learning-based method, the quality of the depth map can be significantly improved.

In this paper, we propose a new method for the depth map generation that utilizes 3D histogram-based segmentation to refine the depth map produced by the learning-based method. For the refinement, the proposed method selectively merges depth values of a segment depending on color consistency and geometrical information of the corresponding segment. In addition, a segment-based cross bilateral filtering is used for improving the quality of the depth map further.

II. PROPOSED DEPTH MAP GENERATION

As shown in Fig. 1, the proposed method consists of two main steps which are the learning-based initial depth map generation and the refinement of the initial depth map. The operation of the method is described in detail below.

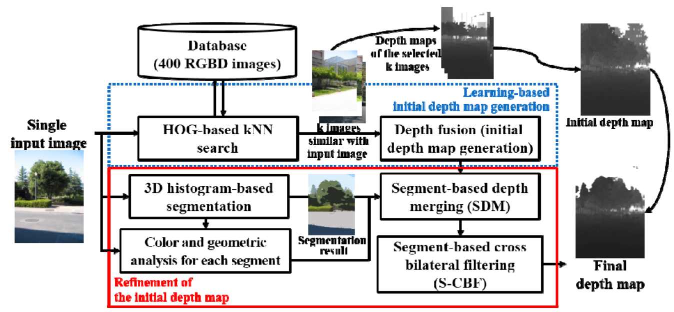


Fig. 1. Block diagram of the proposed method.

A. Learning-based initial depth map generation

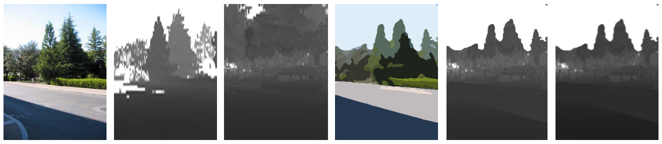
For the initial depth map generation, we adopt the method of learning-based depth map generation (LDM) [1]. In this method, 400 RGBD (RGB+depth) images in the Make3D dataset [3] was used as the database. Because RGB and depth images have different resolutions, we have re-sized these images to 240×320 pixels. For a given image, k images which have a high similarity with a given image, are selected from the database. The similarity is estimated by histograms of oriented gradients (HOGs). Then, the initial depth map (Fig 2(c)) is produced by performing median filtering across the ground-truth depth maps of the selected k images.

B. Refinement of the initial depth map

As shown in Fig 2(c), an initial depth map cannot provide the accurate depth map because it was generated by the fusing the photometrically similar images with a given image. Hence, the initial depth map is refined by utilizing the result of segmentation. 3D histogram-based segmentation [5] which is a simple approach to segmentation, is used for producing the label information of each segment. In this method, a given RGB image is transformed into CIELab color space and color quantization is performed. Then, 3D histogram is constructed on quantized color space. From the 3D histogram, peak bins are extracted and clustering is performed by using the extracted peak bins as cluster centroids. Finally, labels for each segment are generated as shown in Fig 2(d).

By using this segment information denoted by the label, the initial depth map can be refined by following two steps: segment-based depth merging (SDM) and segment-based cross bilateral filtering (S-CBF). Before SDM, we analyze color consistency and geometrical information of a segment to selectively apply the SDM to each segment. First, the color consistency and smoothness of a segment are estimated by the color variance (σ^2_C) of a segment and the ratio of the number of edge pixels (R_E) in a segment, respectively. σ^2_C is defined as follows:

This research was supported by IDEC and the Basic Science Research Program through IT Consilience Creative Program of MKE and NIPA (C1515-1121-0003).



(a) (b) (c) (d) (e) (f)

Fig. 2. Result of each procedure in the proposed method. (a) original image, (b) ground-truth depth map (c) initial depth map, (d) segmentation result, (e) result of SDM, (f) final depth map of the proposed method.

$$\sigma_C^2 = \sum_{X} \left(L[X]-\mu_L\right)^2 + \left(a[X]-\mu_a\right)^2 + \left(b[X]-\mu_b\right)^2 / 3N(X) \quad (1)$$

where L, a, and b are the color components of pixels in a segment, μ_L, μ_a, and μ_b are the corresponding mean values. **X** and $N(\mathbf{X})$ denote a coordinate vector of pixels and the number of pixels in a segment. For the R_E calculation, the Laplacian filtering is performed on an intensity image and the pixel having the resultant value higher than 20 is classified into the edge pixel. For the segment having a low R_E and σ^2_C, depth value is determined by averaging the initial depth values of a segment as shown in Fig. 2(e). The threshold values for R_E and σ^2_C are empirically set to 0.3 and 10^5, respectively. In addition to this, if a segment is in the top of an image (>1/5 of height) and has a high luminance ($\mu_L > 70$) with bluish color ($\mu_b < -5$), we classify this segment into a distant sky and assign the maximum depth value. For the segment in the bottom of an image (<1/5 of height), we do not apply SDM assuming that d_{Ini}^E generally provides an accurate result for this segment. This is because most images have a ground region at the bottom of an image. The assumptions we used for the geometrical analysis are widely used for real 2D-to-3D applications [1].

After SDM, S-CBF is applied on the result of SDM (d_{SDM}^E) to improve the quality of depth map and to generate the final depth value (Fig. 2(f)) as follows:

$$d^E[X] = \sum_{Y} d^E_{SDM} \cdot e^{\left(-\|X-Y\|^2/2\sigma_s^2\right)} \cdot e^{\left(-\|I[X]-I[Y]\|^2/2\sigma_r^2\right)} / N(X) \quad (2)$$

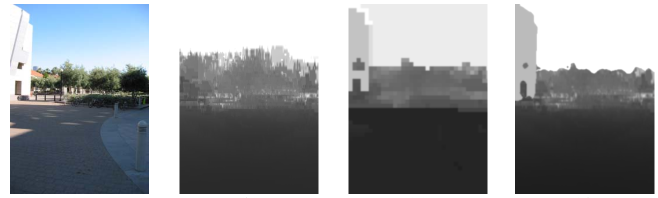
where **Y** denote the coordinate vectors of neighboring pixels. $I[\cdot]$ denote the intensity of a pixel $(\cdot)$. σ_s^2 and σ_r^2 are the variances of filters for spatial and intensity domains, respectively, and are set to one. In (3), we double the σ_r^2 value for a pixel grouped into the same segment with the current pixel. By this, we strongly smooth the depth values located in same segment.

III. SIMULATION RESULTS

For the simulations, the Make3D dataset [3] which consists of 134 test images and 400 training images was used. Normalized cross covariance (*NCC*) was used for the evaluation of quality of depth map. *NCC* is defined as follows [1]:

$$NCC = \sum_{X} \left(d^{GT}[X]-\mu^{GT}\right)\left(d^E[X]-\mu^E\right) / N\sigma_d^{GT}\sigma_d^E \quad (3)$$

where N denote the number of pixels in a given image. σ_d^{GT} and σ_d^E are the standard deviations of ground truth and result depth maps, respectively, while μ^{GT} and μ^E are the corresponding mean values. d^{GT} denotes the ground-truth depth value. For the benchmark methods, we used the typical learning based method (LDM) [1] and Cheng's method (CDM) [6]. CDM is the segmentation-based depth map generation method. For the CDM, the same initial depth map with the proposed method is used.



(a) (b) (c) (d)

Fig. 3. Depth map results of benchmark and proposed methods. (a) original image, (b) depth map by LDM, (c) depth map by CDM, (d) depth map by the proposed method.

TABLE I

NCC VALUES OF THE LDM, CDM, AND PROPOSED METHOD FOR THE 134 TEST IMAGES IN THE MAKE3D DATASET

Methods	LDM	CDM	Proposed
Avg. *NCC*	0.646	0.679	0.693
Std. *NCC*	0.187	0.189	0.173

As shown in Table 1, the proposed method provided the highest average *NCC* value among the benchmark methods. Specifically, the proposed method could provide the most robust results of depth map by exhibiting the smallest standard deviation among the benchmark methods. In addition, in Fig. 3, the proposed method exhibited the best quality of depth map by effectively merging depth values for some segments.

IV. CONCLUSION

In this paper, we proposed a novel depth map generation method that successfully enhances the quality of conventional learning-based depth map generation by performing the segment-based depth merging (SDM) and the segment-based cross bilateral filtering (S-CBF). Specifically, SDM is selectively applied for each segment depending on color and geometrical information of its corresponding segment to optimize the improvement in the quality of depth map, which is obtained by SDM. In the simulations, the proposed method increased the average *NCC* value up to 0.048 higher than that of the benchmark methods.

REFERENCES

[1] J. Konrad, M. Wang, P. Ishwar, C. Wu, and D. Mukherjee, "Learning-Based, Automatic 2D-to-3D Image and Video Conversion," *IEEE Trans. Image Process.*, vol. 22, no. 9, pp. 3485-3496, Sep. 2013.

[2] K. Karsch, C. Liu, and S. B. Kang, "Depth extraction from video using non-parametric sampling," in *Proc. Eur. Conf. Comput. Vis.*, pp. 775-788, 2012.

[3] (2012). *Make3D* Available: make3d.cs.cornell.edu/data.html

[4] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in Proc. *IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 886-893, Jun. 2005.

[5] S. I. Cho, S. J. Kang, and Y. H. Kim, "Human perception-based image segmentation using optimising of colour quantisation," *IET Image Proc.*, 2014. *Accepted.*

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