Principal Depth Estimation Using Cost-Volume Filtering for Fast Stereo Matching

Jaeseung Lim, Jinho Park, Hasil Park, Heegwang Kim and Joonki Paik*
Image Processing and Intelligent Systems Laboratory, Department of Image
Graduate School of Advanced Imaging Science, Multimedia, and Film
Chung-Ang University, Seoul, 06974, Korea
*wanicono@gmail.com, dkskzmffps@gmail.com, hahaha2470@gmail.com, heegwang27@gmail.com, paikj@cau.ac.kr**

Abstract

This paper presents an effective method to reduce computational cost of Cost-volume filtering for stereo matching. We first obtain initial depth values by using pixel-wise cost volume and estimate principal depth values for reducing computational cost. The proposed method can be applied to advanced driver assistance system, object recognition, and other algorithms that require 3-D depth information.

Keywords: Stereo matching, Cost-volume filtering, guided filter, ADAS

1. Introduction

Stereo matching obtains 3-D depth information by using a pair of images shows same scene with different viewpoints. These days, it is applied to advanced driver assistance system, recognition, tracking and others.

Stereo matching estimate optimal depth by comparing patch from reference image and corresponding patches from target image with search range. So its computational cost is in direct proportion to image size M , search range L and patch size N . If N is too small, noisy result occur in plane region of depth map. So stereo matching should uses proper N . In addition, stereo matching has edge-fatting problem that error regions is showed where depth discontinuities occur. To isolate this problem, Yoon and Kweon [6] shows good performance by using adaptive weight with patches.

However, those method can estimate accurate depth map but cause high computational cost. Therefore, there is a continuing need for stereo algorithm capable of small computational cost and estimating accurate depth map.

To solve those problem, Rhemann et Al. [1] proposed Cost-volume filtering (CVF) is not affected by N and shows good performance in edge-fatting problem by using coarse-to-fine method and guided filter [2]. However, it perform guided filtering L times which occupied most computational cost. So if L is bigger, computational time is too high to apply real time process.

Therefore, this paper proposes estimating principal depth values to overcome high computational time of CVF and using census transform [3] to prevent loss of detail.

2. Proposed method

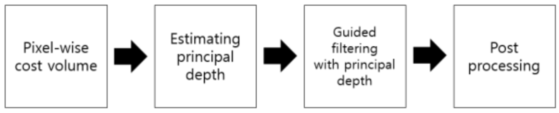


Fig. 1. Block diagram of the proposed method

The block diagram of the proposed method is shown in Fig. 1. First, the proposed method creates pixel-wise cost volume and obtain initial depth map from it. And using histogram of initial depth values to estimate principal depth values. Next step is guided filtering with principal depth values. Finally, streak filling and weighted median filter are used for post processing to fill the holes of occlusion region.

A. Pixel-wise cost-volume

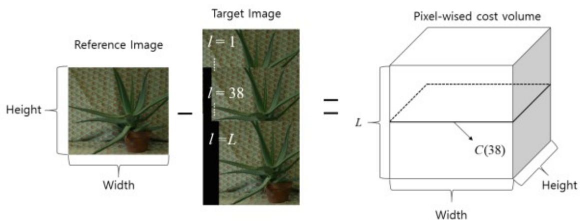


Fig. 2. Pixel-wise cost volume

Pixel-wise cost volume is generated by stacking L cost images and cost image $C(l)$ with depth value l is calculated by reference image I and l -shifted target image I'_l as shown in Fig. 2. $C(l)$ is defined as

$$C(l) = \alpha \cdot \min(\|I - I'_l\|, \gamma_1) + \beta \cdot \min(\mathrm{H}(I, I'_l), \gamma_2) + (1 - \alpha - \beta) \cdot \min(\|\nabla I - \nabla I'_l\|, \gamma_3) \quad (1)$$

where each term represent difference of color intensity, census transform and gradient, $\mathrm{H}(I, I_l')$ is hamming distance of census transformed images, ∇I and $\nabla I_l'$ are normalize gradient images between [0, 1], α and β balance the three terms and γ_1, γ_2 and γ_3 are truncation values.

B. Estimating principal depth

Initial depth value l_p and cost c_p of pixel p is defined as

$$l_p = \arg\min_l C_p(l) \tag{2}$$

$$c_p = \min_l C_p(l) \tag{3}$$

We choose l_p to estimate histogram of ground truth, if corresponding cost c_p is lower than threshold th_c . And principal depth values l' are estimated by top $L' = round(n\sqrt{L})$ ranked depth values by its frequency with predefined value n .

C. Guided filtering with principal depth

Guided filtering is performed to cost image $C(l')$ as

$$C_p'(l') = \sum_{q \in W_p} \omega_{q,p}(I) C_q(l') \tag{4}$$

where W_p is window of guided filter centered at p, q is random pixel in W_p , $\omega_{q,p}$ is filter weight and $C_p'(l')$ is filtered cost image.

And estimate final depth map by using Winner-Takes-All [4] and filtered cost images as

$$l_p' = \arg\min_{l'} C_p'(l') \tag{5}$$

To fill holes in occlusion region, streak filling and weighted median filter is used for post processing.

3. Experimental result

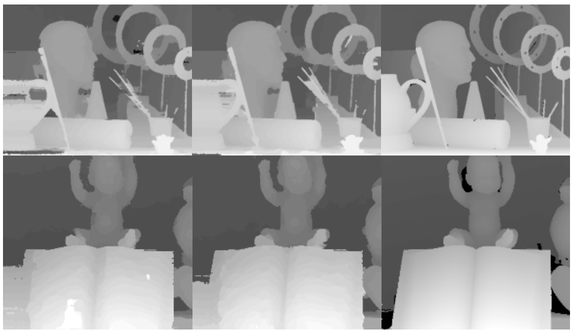
To evaluate the performance of the proposed method, we measure computational time and bad pixel error [4]. We use Middlebury benchmark [5] for experimental data. Tables 1, 2 and Fig.3 shows comparison result of the proposed method and cost volume filtering. The proposed method performs 2 times faster in computational time and 0.1% better in bad pixel error than CVF.

Table 1: Computational time (s)

	Baby2	Dolls	Aloe	Art	Average
CVF	176.8	267.6	225.7	280.1	237.5
Proposed	98.6	128.1	123.2	126.8	119.2

Table 2: Bad pixel error (%)

	Baby2	Dolls	Aloe	Art	Average
CVF	1.803	1.202	1.935	2.138	1.769
Proposed	1.554	1.181	1.927	1.959	1.656



(a) (b) (c)

Fig. 3. Estimated depth map: (a) CVF, (b) Proposed method, (c) Ground truth

4. Conclusion

This paper proposes reducing computational cost by estimating principal depth values. We estimate principal depth by initial depth with low cost and use census transform for preventing loss of details. Further work will be focused on post processing to improve hole filling algorithm in occlusion region.

Acknowledment

This work was supported by the second phase of the Brain Korea 21 Program in 2017.

References

[1] C. Rhemann, A. Hosni, M. Bleyer, C. Rother, and M. Gelautz, “Fast cost-volume filtering for visual correspondence and beyond,” *In proc. Computer Vision and Pattern Recognition*, pp. 3017-3024, June

[2] K. He, J. Sun, X. Tang, "Guided Image Filtering", *Proc. European Conf. Computer Vision*, 2010.

[3] R. Zabih, and J. Woodfill, “Non-parametric local transforms for computing visual correspondence,” *In proc. European conference on computer vision*, pp. 151-158, May 1994

[4] D. Scharstein, and R. Szeliski, “A taxonomy and evaluation of dense two-frame stereo correspondence algorithms,” *International journal of computer vision*, vol. 47, no.1-3, pp. 7-42, April 2002

[5] http://vision.middlebury.edu/stereo

[6] K. J. Yoon, and I. S. Kweon, “Adaptive support-weight approach for correspondence search,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 4, pp. 650-656, April 2006