Disparity Map Refinement Method Using Coarse-to-Fine Image Segmentation

Jinyoung Jang*, Dong-Won Shin†, Yo-Sung Ho†
*Korea Communication Agency (KCA), Naju, Republic of Korea
E-mail: and0162@kca.kr(jyjang@gist.ac.kr)
†Gwangju Institute of Science and Technology (GIST), Gwangju, Republic of Korea
E-mail: {dongwonshin, hoyo}@gist.ac.kr

***Abstract*— Stereo matching methods are used mainly to obtain the depth information in stereo images. Since stereo images are taken from two different viewpoints for the same scene, we have occlusion regions that are visible in only one image, but not in the other image. Because the occlusion region lacks information on corresponding points, disparity errors occur in the disparity map. In order to correct the occlusion region, we propose a disparity map refinement method based on coarse-to-fine (CTF) image segmentation. Experimental results show the proposed method generates a higher quality depth information than other methods; in particular the proposed method reduces disparity errors in discontinuity regions.**

I. INTRODUCTION

Realistic content refers to next generation content that offers a real-world experience based on information and communications technology (ICT). 3D imaging technology is one of realistic content technologies and the most widely researched area. In order to use 3D image in various application fields, it is necessary to acquire depth information of either two or multi-view images. In general, depth information can be obtained through a passive method that uses only the 2D features of an image, such as stereo matching, or an active method using a device directly, such as a depth camera.

A stereo matching algorithm is mainly used to obtain the depth information in stereo images. And it is the process to find the corresponding points in stereo images and generate the disparity map related to the depth information. This stereo matching algorithm generally performs four steps: cost computation, cost aggregation, disparity selection, and disparity refinement [1].

In this paper, we propose a post-processing method in the stereo matching to improve accuracy of the disparity map. To obtain the initial disparity map, we use the truncated absolute difference of color and gradient (TAD C+G) cost function [2] and guided filter [3]. Also, the cross-consistency checking method is used to detect the occlusion region. In order to correct the occlusion region, we perform a disparity map refinement method based on coarse-to-fine (CTF) image segmentation.

II. INITIAL DISPARITY MAP ESTIMATION

A. Cost Computation Using TAD C+C Cost Function

For a stereo color pair, the cost volume C is constructed by applying cost computation. The cost volume represents matching costs for each pixel at all possible disparity levels.

For a single pixel $i = (x_i, y_i)$ and disparity level l, its cost function $C(i, l)$, which is TAD C+G, can be expressed by (1). Two terms, color and gradient term, exist in this function.

$$\begin{aligned} C(i,l) = (1-\alpha)\cdot \min(\|I_L(i) - I_R(i_l)\|, \tau_1) \\ +\alpha \cdot \min(\|\nabla_x I_L(i) - \nabla_x I_R(i_l)\|, \tau_2) \end{aligned} \quad (1)$$

where I_L and I_R denote the left and right color images, respectively; i_l is the corresponding pixel of a pixel i with a disparity level l (i.e., $i_l = (x_i - l, y_i)$); α is a parameter that balances the color and gradient terms; and τ_1 and τ_2 are truncation values to prevent the matching cost from increasing because of the occlusion region.

B. Cost Volume Filtering Using a Guided Filter

After a cost volume is constructed, the disparity level of the lowest cost is chosen at each pixel. The result is noisy because the pixel-to-pixel comparison is too sensitive to noise. Therefore, each slice of the cost volume must be filtered. To filter the cost volume, we use the guided filter because it has an edge-preserving property and low complexity.

The filtered cost volume C' is expressed as follows.

$$C'(i,l) = \sum_{j\in\omega_k} W(i,j)C(j,l) \quad (2)$$

where i and j are pixel indexes. W is the guided filter weight function and C is the cost volume at neighboring pixels in the window ω_k. The guided filter weight function is expressed as (3).

$$W(i,j) = \frac{1}{|\omega_k|^2} \sum_{k:(i,j)\in\omega_k} \left(1 + \frac{(I(i)-\mu_k)(I(j)-\mu_k)}{\sigma_k^2 + \epsilon}\right) \quad (3)$$

where μ_k and σ_k^2 are the mean and variance of guidance image I which is the left image. ϵ is a regularization parameter.

C. Disparity Selection by WAT Method

Once the cost volume is filtered, the disparity level of any pixel i is simply chosen in a winner-take-all (WAT) method. The use of WAT method is to find the minimum cost value. The initial disparity $d(i)$ is defined in (4).

$$d(i) = \arg\min_{l \in \mathcal{L}} C'(i, l) \quad (4)$$

And then we can obtain the initial disparity map. Fig. 1 represents initial disparity maps of the Middlebury stereo image dataset.

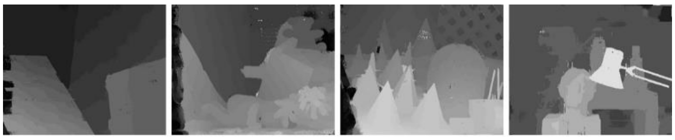


Fig. 1 Initial disparity maps.

III. Disparity Map Refinement

To solve the inaccuracy of the initial disparity map, we should perform disparity refinement as a post-processing step. There are two steps in disparity refinement process. First step is occlusion detection and second step is occlusion correction. We use the cross-consistency checking method to detect occlusion regions and disparity errors. The cross-consistency checking method is expressed as (5).

$$c(x) = \begin{cases} 0, & if\ d_l(x) \neq d_r(x - d_l(x)) \quad occlusion\ region \\ 1, & otherwise \quad non-occlusion\ region \end{cases} \quad (5)$$

where $d_l(x)$ is the disparity value of the x point in the left disparity map and $d_r(x - d_l(x))$ is the disparity value of the $x - d_l(x)$ point in the right disparity map.

The goal of the proposed method is to improve the accuracy of the disparity map using a new occlusion correction method. Fig. 2 shows the overall framework of the proposed method. Our method is categorized in the following four steps: 1) coarse-to-fine image segmentation; 2) image labeling and histogram calculation; 3) calculation of the candidate disparity value; 4) decision of the final disparity value.

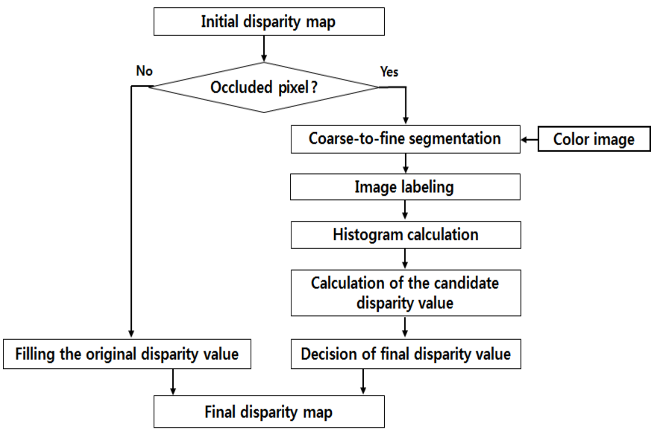


Fig. 2 Overall framework of the proposed method.

A. Coarse-to-fine Image Segmentation

Our proposed method uses a new coarse-to-fine paradigm. As illustrated in Fig. 3, multiple-level coarse-to-fine segmentations are conducted on the initial disparity maps using a color image. The initial disparity map is segmented starting from the coarsest to the finest segmentation level.

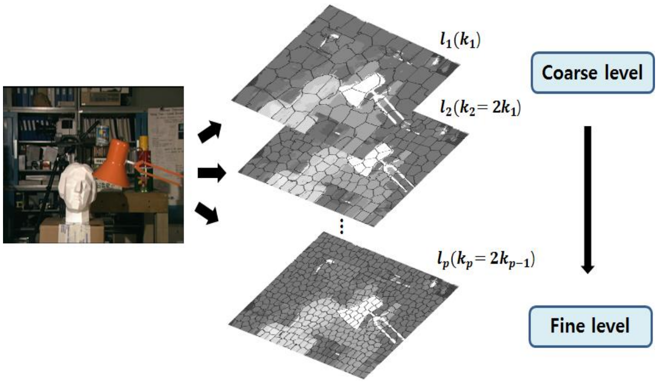


Fig. 3 Coarse-to-fine image segmentation.

Segmentation level l_p is determined by the number of superpixels k. First, in the coarsest segmentation level l_1, the disparity map is segmented according to the initial number of superpixels k_1. Then, in other segmentation levels, the disparity map is segmented by doubling the number of superpixels from the previous level.

In each segmentation level l_p, we segment the initial disparity map using the simple linear iterative clustering (SLIC) method [4].

The SLIC is a method of clustering all the pixels in an image with several superpixels using the color and distance information between the pixel and cluster. Simplicity and fast computational speed are the advantages that can be gained from this method. This is performed using cluster information only in a certain region (i.e., without using all cluster information of an image).

As shown in Fig. 4, the SLIC clusters the pixels in the image with the number of superpixels k.

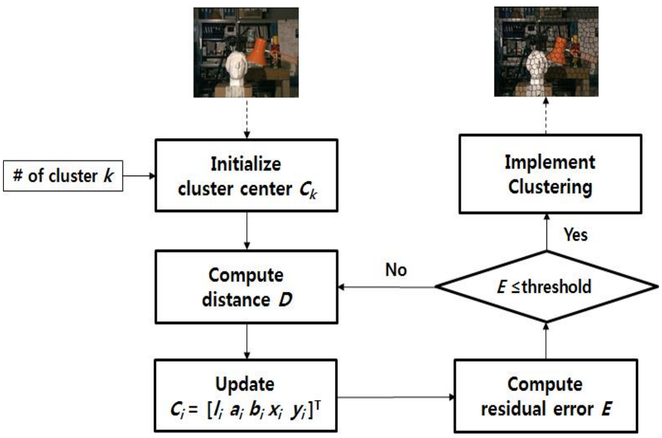


Fig. 4 Overall framework of the SLIC method.

B. Image Labeling and Histogram Calculation

After the initial disparity map is segmented at each segmentation level, each segmented region is labeled. The histogram then is used to obtain the mode of disparity value in each labeled region as shown in Fig. 5.

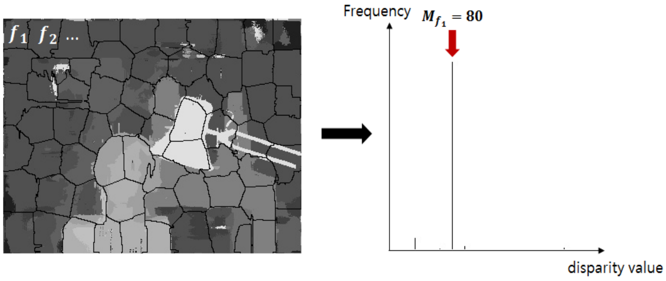


Fig. 5 Image labeling and histogram.

The mode M_{f_k} of the disparity value is expressed as follows.

$$M_{f_k} = \underset{d}{\operatorname{argmax}}\, c(d) \tag{6}$$

where $c(d)$ is the count of the disparity value and f_k is the label for $k = \{1,2,\dots,K\}$.

Image labeling and histogram calculation are repeated at each segmentation level.

C. Calculation of the candidate disparity value

To obtain the candidate disparity map, the mode of disparity value is filled in the occlusion region. The mode of disparity value is the candidate value $d^{l_p}(i)$. In the non-occlusion region, the original disparity value is filled. If the process is repeated at each segmentation level, the candidate disparity maps can be obtained at each segmentation level.

The candidate disparity $d^{l_p}(i)$ is expressed as (7). Fig. 6 shows the results of the candidate disparity maps at each segmentation level.

$$d^{l_p}(i) = M_{f_k}^{l_p}(i) \qquad if\ occlusion\ region \tag{7}$$

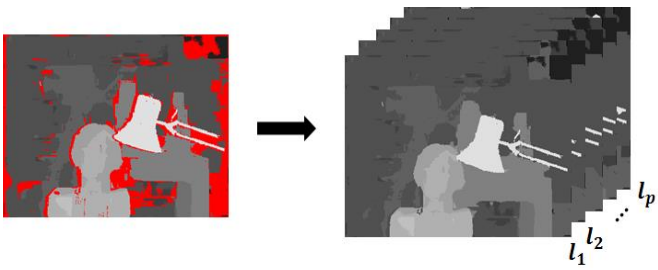


Fig. 6 Candidate disparity maps.

D. Decision of the final disparity value

In the occlusion region, the final disparity value is determined to be the minimum value of the average difference among the candidate disparity values. The final disparity value in the occlusion region is expressed as (8).

$$d'(i) = \min\left|d^{l_p}(i) - \mu(i)\right| \qquad if\ occlusion\ region \tag{8}$$

where $\mu(i)$ is the average of the candidate disparity values of all segmentation levels at pixel i. $\mu(i)$ is defined in (9), where p is the number of segmentation levels.

$$\mu(i) = \frac{1}{P}\sum_{p=1}^{P} d^{l_p}(i), \qquad p \in \{1,2,\dots,P\} \tag{9}$$

Fig. 7 shows that the final disparity map is determined based on the candidate disparity maps.

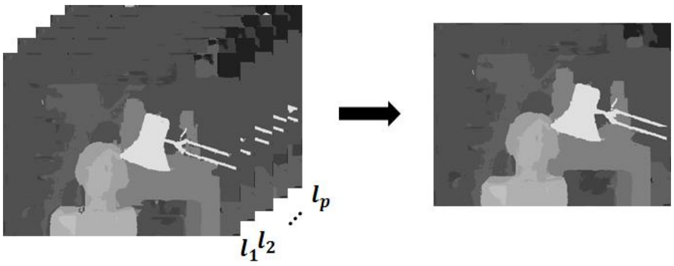
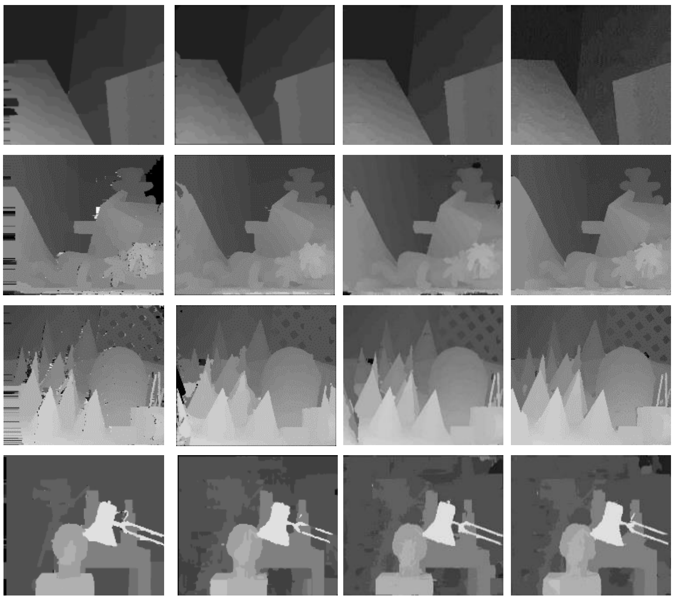


Fig. 7 Determination of final disparity map.

IV. Experiment Results

Our experiment on the proposed method includes four test images obtained from the Middluebury website: *Venus, Teddy, Cones, and Tsukuba*.

Fig. 8 shows the final disparity maps derived from the proposed method and other stereo matching algorithms.



(a) GC+occ (b) CCH+SegAggr (c) Chang's method (d) Proposed method

Fig. 8 Final disparity maps of proposed method and conventional methods.

In Fig. 8, it can verified that the proposed method come up with better results than the conventional stereo matching algorithms.

To evaluate the performance of the stereo matching algorithm, we require a quantitative means to estimate the quality of the computed correspondences. Therefore, we compute the bad pixel rate based on known ground truth data. The bad pixel rate B is defined by (10) [1].

$$B = \frac{1}{N} \sum_{(x,y)} (|d_C(x,y) - d_r(x,y)| > \delta_d) \tag{10}$$

where $d_C(x,y)$ and $d_r(x,y)$ are the computed disparity and ground truth maps, respectively; N is the total number of pixels; and δ_d denotes disparity error tolerance. In our experiments, we use $\delta_d = 1.0$.

We measure the bad pixel rate in three kinds of regions: all regions, non-occluded regions, and discontinuity regions, denoted as “All”, “Nonocc.”, and “Disc.”, respectively.

Table 1 lists the final results of the proposed method. The proposed method uses coarse-to-fine image segmentation for occlusion correction. We also compare the proposed method with other stereo matching algorithms. Results show that the overall bad pixel rates of the proposed method are lower than the rates in other stereo matching algorithms.

TABLE I
FINAL COMPARISON RESULTS

Algorithm		GC+occ [5]	CCH+SegAggr [6]	Chang's method [7]	Proposed method
Venus	All	2.19%	0.94%	1.06%	1.04%
	Nonocc.	1.64%	0.41%	0.43%	0.63%
	Disc.	6.75%	3.97%	2.90%	4.23%
Teddy	All	17.40%	14.30%	13.81%	10.46%
	Nonocc.	11.20%	8.08%	7.10%	4.84%
	Disc.	19.80%	19.80%	20.03%	15.86%
Cones	All	12.40%	12.90%	12.08%	10.71%
	Nonocc.	5.36%	7.07%	4.42%	3.70%
	Disc.	13.00%	16.30%	11.99%	10.38%
Tsukuba	All	2.01%	2.11%	2.60%	2.07%
	Nonocc.	1.19%	1.74%	2.30%	1.59%
	Disc.	6.24%	9.23%	9.57%	8.07%
Average		8.27%	8.07%	7.36%	6.13%

In Fig. 9, we show the average of bad pixel rate by the region types that are all, non-occlusion, and discontinuity regions. In the three different regions, the proposed method outperforms other stereo matching algorithms. In particular, our proposed method performs best in discontinuity regions than in other regions.

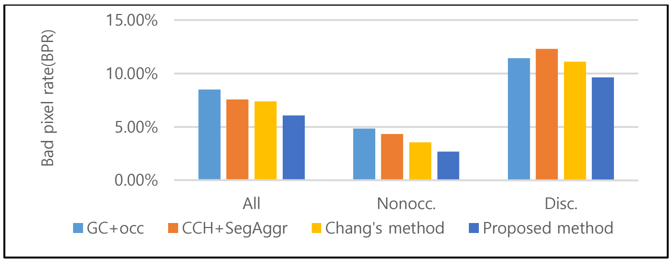


Fig. 9 Average of BPR by the region types.

V. CONCLUSIONS

In this paper, we propose a disparity refinement method based on coarse-to-fine image segmentation. In particular, we concentrate on the means by which to enhance the quality of a disparity map. First, we obtain an initial disparity map using the TAD C+G cost function and guided filter. Second, we propose a new post-processing method of stereo matching. We use a cross-consistency checking method to detect the occlusion and disparity errors. We then correct the occlusion and error pixels using coarse-to-fine image segmentation. We employ a SLIC superpixel segmentation algorithm for this purpose. Experimental results show that the proposed method performs better than the conventional method. In particular, the proposed method using coarse-to-fine image segmentation reduces disparity errors in discontinuity regions.

ACKNOWLEDGMENT

This work was supported by the 'Civil-Military Technology Cooperation Program' grant funded by the Korea government.

REFERENCES

[1] D. Scharstein, R. Szeliski and R. Zabih, “A taxonomy and evaluation of dense two-frame stereo correspondence algorithms,” *Stereo and Multi-Baseline Vision, 2001. (SMBV 2001). Proceedings. IEEE Workshop on. IEEE*, pp. 131-140, 2001.

[2] C. Rhemann, A. Hosni, M. Bleyer, C. Rother and M. Gelautz, “Fast cost-volume filtering for visual correspondence and beyond,” *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3017-3025, 2011.

[3] K. He, J. Sun and X. Tang, “Guided image filtering,” *European conference on computer vision*. Springer Berlin Heidelberg, pp. 1-14, 2010.

[4] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua and S. Süsstrunk, “SLIC superpixels compared to state-of-the-art superpixel methods,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 34, no. 11, pp. 2274-2282, 2012.

[5] V. Kolmogorov and R. Zabih, “Computing visual correspondence with occlusions using graph cuts,” *IEEE International Conference on Computer Vision*, pp. 508-515, 2001.

[6] T. Liu, P. Zhang, L. Luo, “Dense stereo correspondence with contrast context histogram, segmentation-based two-pass aggregation and occlusion handling,” Advances in Image and Video Technology, pp. 449–461, 2009.

[7] Y. J. Chang and Y. S. Ho, “Disparity map enhancement in pixel based stereo matching method using distance transform”, *Journal of Visual Communication and Image Representation*, vol. 40, no. 12, pp. 118-127, 2016.