SEGMENTATION BASED DISPARITY ESTIMATION USING COLOR AND DEPTH INFORMATION

Sang Yoon Park, Sang Hwa Lee, and Nam Ik Cho

School of Electrical Engineering and Computer Science,
Seoul National University, Seoul, 151-744 Korea
Emails: sanguni@ispl.snu.ac.kr, lsh@ipl.snu.ac.kr, nicho@snu.ac.kr

ABSTRACT

The well-known cooperative stereo uses two dimensional rectangular window for a local block matching, and three dimensional box-shaped volume for a global optimization procedure. In many cases, appropriate selections of these matching regions can provide satisfactory matching results. This paper presents a new method for iteratively modifying sizes and shapes of matching regions based on color and depth information. This algorithm computes the aggregated matching costs with two ideas. The first idea is to select matching regions based on object boundaries to avoid projective distortion. This provides the reliable matching scores as well as the prevention of the foreground fattening phenomenon. The second idea is to iteratively modify the segmentation map by merging the regions where the disparities are likely to be the same. Experimental results show that the proposed algorithm provides more accurate disparity map than other algorithms. Especially, the computed disparity map shows the advantage of our algorithm in disparity discontinuity regions.

1. INTRODUCTION

Dense disparity map has been required for many stereo matching applications including image based rendering, depth-keying, 3D object modeling, etc. Specifically, these applications require the disparity map which varies smoothly on object surfaces and changes sharply at its boundaries. Unfortunately, it is very difficult to satisfy these contradictory conditions. Although appropriate selection of the window may sometimes lead to the acceptable results in area-based stereo matching such as normalized correlation (NC) or sum of absolute difference (SAD), it would generate undesirable problems as follows. While adaptive window [1] smartly derives a statistical model of the uncertainty of disparity estimation within the 2D window, high computational complexity is required. The shiftable window [2] can simply detect and delineate object boundaries. However, a rectangular window of fixed size is disadvantageous, and sometimes generates blocky artifacts. The segmentation based methods proposed in [3, 4] may effectively reduce the foreground fattening error (or called boundary overreach in [2]). However, the color-based segmentation needs to be substituted for the depth-based segmentation. Global optimization method [5] based on Markov Random Fields (MRF) requires a computational overload for high dimensional optimization process in spite of reasonable results.

The well-known cooperative stereo [6] iteratively updates the initial matching costs obtained by the local similarity measure. The update functions diffuse the reliable matching scores into the neighbors, and inhibit values along the view-directions of the left or right eye. A box-shaped 3D support volume is employed to diffuse the correct matches into neighbors under the continuity constraint. The cooperative stereo is one of the smart matching algorithms in handling the ambiguous matching regions including textureless, noisy regions, etc. However, it generates undesirable problems in depth discontinuity regions. This paper proposes a new method to iteratively select matching regions[1] based on color and depth information, while conventional methods employ only color information. The proposed method computes the aggregated matching costs with two ideas. The first idea is to select matching regions based on object boundaries to avoid projective distortion. We employ the color image segmentation algorithm for this purpose. This provides the reliable matching scores as well as the prevention of foreground fattening phenomenon. However, the segments with the same disparities should be merged under the smoothness assumption. Thus, the second idea is to iteratively merge the regions where the disparities are likely to be the same. Experimental results show that the proposed algorithm provides more accurate disparity map than other algorithms with little additional computational overload. Especially, the computed disparity map shows the advantage of our algorithm in disparity discontinuity regions.

2. MOTIVATION

The cooperative stereo iteratively updates the initial 3D volumetric space constructed by the local matching scores like SAD. For the efficiency, the shapes and sizes of matching regions are fixed through the whole matching process. However, the fixed matching region around the interested pixel may bring about the disadvantages such as the foreground fattening phenomenon or projective distortion. To improve the quality of the disparity map, the modifications of the matching regions need to be considered. It would be ideal if the matching region could be adjusted to include only the pixels having the same disparities. Unfortunately, it would be probably impossible because of the absence of true disparity map. On the other hand, the ideal selections of the matching regions can provide us with the closest result to the ground truth. We perform the following procedures to solve this chicken-and-egg problem. Instead of the true disparity map, in our implementation, the matching regions are selected based on the object boundaries enforced by the color segmentation cue. It is based on the hypothesis that the segments with the similar colors may also have the sim-

[1] In this paper, let us denote the 2D windows and 3D boxes by "matching regions".

0-7803-8554-3/04/$20.00 ©2004 IEEE.

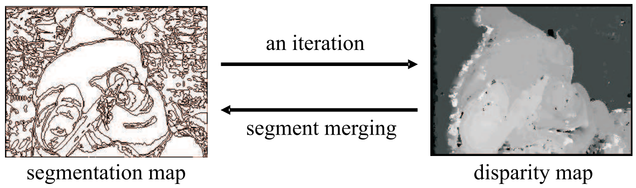


Fig. 1. Basic operation of the proposed method.

ilar disparities. However, this hypothesis is not true in general. Hence, it is necessary to merge the regions where the disparities are likely to be the same, which results in the depth-based segmentation. Note that we can obtain estimated disparity maps from the intermediate steps, i.e., the initial local matching or iteration of cooperative stereo, although true disparity map is not given. Segment merging plays a important role in that the propose method considers the color and depth information at the same time, and can also improve the results significantly. Following section describes the proposed method in detail.

3. ALGORITHMS

3.1. Basic framework

The cooperative stereo is competent to handle the ambiguous matching regions including textureless, noisy regions, etc. However, it generates undesirable problems in depth discontinuity regions. Therefore, the basic framework of the cooperative stereo needs to be modified to overcome these disadvantages. As shown in Fig. 1, basic framework of the proposed method is the iteration of cooperative stereo followed by segment merging. Specifically, the iteration is performed using the arbitrarily shaped matching regions based on the color segmentation cue. Then, the intermediate disparity map is determined to maximize the match value[2]. A new segmentation map is obtained using the latest disparity map and segmentation map. At first glance, the iteration seems to be performed using only color information. However, it is actually based on both color and depth information.

3.2. Initial matching

We select 2D local windows based on boundaries which are given by the color segmentation cue. For the color segmentation, we employ the method of [7] based on the mean shift algorithm. A segmentation threshold is meaningful, since boundaries determined by the color segmentation are regarded as the depth discontinuity in the initial matching. The low threshold may well be preferred for the avoidance of skipping true object boundaries [4]. If a 3D volumetric space for each pixel (x, y) and disparity d and the color segmentation map of reference image are prepared, the initial match value $P_0(x, y, d)$ is assigned to each pixel as

$$P_0(x, y, d) = F\Big(\frac{1}{N(x, y)} \sum_{i,j \in W'} \big|I_r(x+i-d, y+j) - I_l(x+i, y+j)\big|\Big), \tag{1}$$

[2]This step is often called *winner-takes-all*.

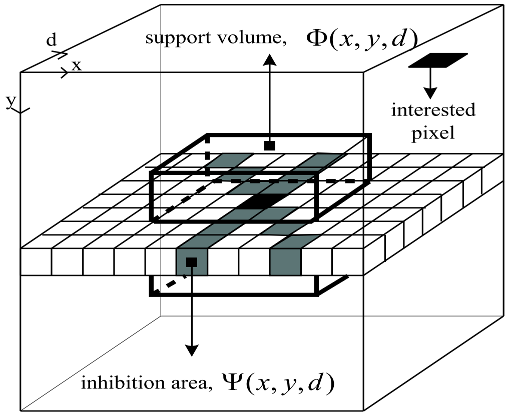


Fig. 2. Support and inhibition of the cooperative stereo.

where $N(x, y)$ is the number of samples within $W'(x, y)$, and $F(x)$ is the function that enforces the match values between 0 and 1, which is simply defined as

$$F(x) = \frac{1}{1 + \lambda x}, \tag{2}$$

where λ is usually set for 1. Let $W(x, y)$ be the conventional rectangular window of fixed size, and then we define the window $W'(x, y)$ as

$$W'(x, y) = \big\{(x', y') | [C_0(x', y') = C_0(x, y)] \cap [(x', y') \in W(x, y)]\big\}, \tag{3}$$

where $C_0(x, y)$ represents the color segmentation map whose pixel values are constant in the same segment. The modification of 2D window as eq. (3) can provide the reliable match for object boundary, and also conserve the contour of a slim object. The disparity estimation using eq. (3) tends to be poor if the number of pixels satisfying eq. (3) is less than about 20% of that of rectangular window. In this case, the $W'(x, y)$ is allowed to contain the pixels of neighboring regions, or be simply substituted for the $W(x, y)$.

3.3. Cooperation

The cooperative stereo [6] adds a stabilizing scheme to the $P_0(x, y, d)$ obtained by eq. (1) for global optimization. Its iterative algorithm refines the match values by using continuity among neighboring pixels and uniqueness of the disparity per pixel. Fig. 2 illustrates the support volume $\Phi(x, y, d)$ and inhibition area $\Psi(x, y, d)$ in 3D volumetric space. This method is also similar to the disparity estimations based on the Markov Random Fields (MRF) [5] in that they diffuse the match values among neighboring pixels. An iterative update function based on the continuity and uniqueness constraints is used to stabilize the match values as follows [6].

$$P_{n+1}(x, y, d) = P_0(x, y, d) \cdot \Bigg[\frac{S_n(x, y, d)}{\sum_{(x', y', d') \in \Psi} S_n(x', y', d')}\Bigg]^2. \tag{4}$$

In the above equation, the initial match value $P_0(x, y, d)$ is used to avoid the convergence to unexpected match value as iteration number n increases. The match value is inhibited by the sum of the match values within $\Psi(x, y, d)$, and locally supported as follows.

$$S_n(x,y,d) = \frac{1}{N(x,y,d)} \sum_{(x',y',d')\in\Phi'} P_n(x+x', y+y', d+d'), \tag{5}$$

where $\Phi'(x, y, d)$ is defined as

$$\Phi'(x,y,d) = \big\{(x',y',d')|[C_n(x',y') = C_n(x,y)] \cap [(x',y',d') \in \Phi(x,y,d)]\big\}. \tag{6}$$

Eq. (6) is a 3D extension of eq. (3), and similar to the definition of [3] except for the concept of local patch. After the match values are converged, we finally select the disparity corresponding to the maximum match value for each pixel, as given by

$$d(x,y) = arg\ \max_{d} P_n(x,y,d). \tag{7}$$

Subpixel accuracy can also be computed after completing the cooperative algorithm as

$$d' = d + \frac{1}{2} \cdot \frac{P_n(x,y,d-1) - P_n(x,y,d+1)}{P_n(x,y,d-1) - 2P_n(x,y,d) + P_n(x,y,d+1)}. \tag{8}$$

3.4. Merging of the segments

The segment merging is performed after the initial matching or iteration of cooperative stereo. Two segments need to be merged in the case when the disparities of adjacent segments are the same because of following two reasons. First, the wise selection of the window around an interested pixel gathers as many pixels as possible which have same disparities in the image. In this case, no matter how many pixels the window includes, the projective distortion will not occur. Second, the mergence of segments can provide more reliable matching costs since the intensity variation as well as the number of samples increases. The major disparity is defined as the disparity which occupies the majority in the segment based on the current disparity map. Adjacent segments of C_n need to be merged in the case when their major disparities are similar as follows.

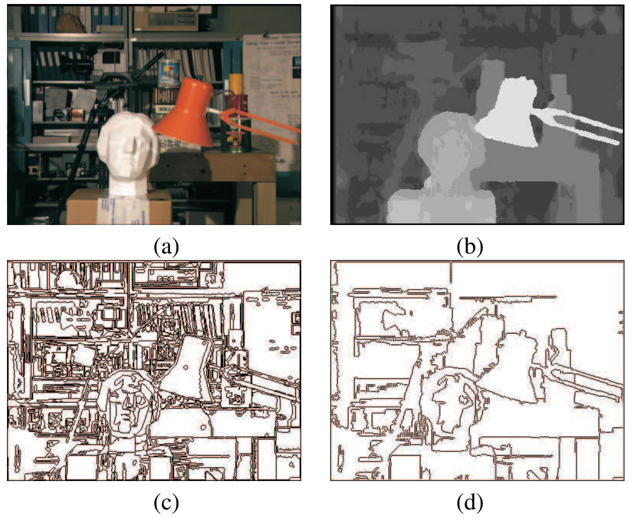
$$|d_1 - d_2| + \lambda \cdot |c_1 - c_2| \ < \ threshold, \tag{9}$$

where d_1, d_2, and c_1, c_2 represent major disparities and colors of two adjacent segments, respectively. As shown in (9), the similarity of color is also added to merging condition with the weight, λ. Since segment merging is inserted into iterations of the cooperative stereo, it is also performed iteratively. If the number of segments is not reduced any more after several iterations, this procedure may be skipped.

3.5. Summary of the proposed method

The proposed method is summarized as follows.

1. Prepare a 3D volumetric space and color segmentation map of reference image.
2. Calculate the initial match value $P_0(x, y, d)$ using eq. (1) based on the modified window.
3. Determine the disparity map to maximize the match value.
4. Perform the segment merging using the result of step 3.
5. Update the match value $P_n(x, y, d)$ using eqs. (4) and (5) based on the modified support region.
6. Determine the disparity map to maximize the match value.
7. Perform the segment merging using the result of step 6.
8. If the $P_n(x, y, d)$ is converged, compute the sub-pixel accuracy, else go to step 5.



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

Fig. 3. Simulation results on *head* image of Tsukuba university: (a) original left image, (b) disparity map by the proposed method, (c) initial region boundaries obtained by the method of [7], (d) converged region boundaries.

4. EXPERIMENTAL RESULTS

To show the improvement of the proposed method, we have applied our algorithm to real stereoscopic imagery. Fig. 3 shows the simulation results on famous *head* image of Tsukuba university. The sizes of 2D window and 3D support region are 5×5 and $7 \times 7 \times 3$, respectively. The computational complexity is very high when directly aggregating all pixels within a matching re-

Table 1. Error rates (%) of the computed disparity maps are evaluated at the webpage http://www.middlebury.edu/stereo in all (unoccluded), textureless, and discontinuity regions.

image		*head*	*sawtooth*	*venus*	*map*
[6]	all	3.49	2.03	2.57	0.22
	untex.	3.65	2.29	3.52	·
	disc.	14.77	13.41	26.38	2.37
without segment merging	all	1.75	1.18	0.86	0.69
	untex.	0.48	1.09	0.58	·
	disc.	9.22	6.93	7.61	8.80
proposed	all	1.40	0.89	0.80	0.44
	untex.	0.45	0.23	0.53	·
	disc.	7.57	5.40	6.63	6.20

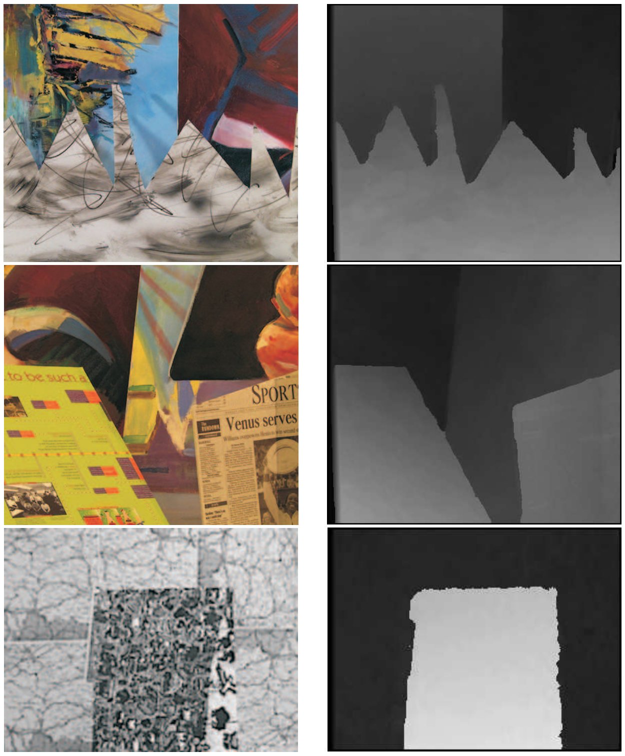


Fig. 4. Results of the stereo image pairs provided at the webpage http://www.middlebury.edu/stereo. Left column shows the reference images of *sawtooth*, *venus*, and *map*, and right column shows their disparity maps by proposed method.

gion. Thus, the well-known box filtering approach is first applied, and then pixel values of uninterest are subtracted from the result. Actually, the number of pixels which should be excluded is no more than 15% after the convergence of the segment merging, and thus an iteration can be computed in a few seconds. Figs. 3 (a) and (b) show the original left image and final disparity map, respectively. As shown in the figure, the proposed method provides an accurate disparity map. More specifically, object boundaries and thin structures as well as smooth surface are correctly recovered. Fig. 3 (c) shows the initial region boundaries which are obtained by the method of [7]. In case of the *head* image, oversegmentation hardly skips the depth discontinuities to our wishes. Fig. 3 (d) shows the converged boundaries after 7 iterations, and the number of segments is reduced to about 6% of the Fig. 3 (c). The webpage http://www.middlebury.edu/stereo provides the quantitative evaluations of various stereo matching algorithms for four stereo image pairs (*head, sawtooth, venus, and map*). The detailed descriptions about the evaluating methods can be shown in [8]. Tab. 1 shows matching results for the four stereo pairs in all (unoccluded), textureless, and discontinuity regions. As shown in the table, only use of color information without segment merging generates better results than the method of [6] which don't use any information. However, if color and depth are used at the same time, we can obtain more accurate results than other two cases. Especially, the results show the advantage of our algorithm in disparity discontinuity regions. Our method ranks as the No. 4 by results of Tab. 1 on that webpage (May 18, 2004). Fig. 4 shows the reference images and final disparity maps of the other 3 stereo image pairs.

5. CONCLUSION

This paper has presented a new method which can iteratively modify matching regions. While the conventional method only uses color segmentation cue, the proposed method uses both color and depth information. This algorithm computes the aggregated matching costs with two ideas. The first idea is to select matching regions based on object boundaries, and the second is to iteratively merge the regions where the disparities are likely to be the same. The proposed idea is very simple, and its final disparity map also shows both the smoothness on the object surface and sharpness of the boundaries. The proposed method has been evaluated at http://www.middlebury.edu/stereo, and the results show that our new method can compute the disparities more accurately than other approaches.

6. REFERENCES

[1] T. Kanade and M. Okutomi, "A stereo matching algorithm with an adaptive window: theory and experiment," *IEEE Trans. on PAMI*, vol. 16, no. 9, pp. 920-932, Sept. 1994.

[2] M. Okutomi, Y. Katayama, and S. Oka, "A simple stereo algorithm to recover precise object boundaries and smooth surfaces," *Int'l J. Computer Vision*, vol. 47, no. 1, pp. 261-273, Apr. 2002.

[3] Y. Zhang and C. Kambhamettu, "Stereo matching with segmentation-based cooperation," In *Proc. ECCV*, pp. 556-571, 2002.

[4] H. Tao, H. S. Sawhney, and R. Kumar, "A global matching framework for stereo computation," In *Proc. ICCV*, pp. 532-539, 2001.

[5] S. H. Lee and Y. Kanatsugu, "MAP-based stochastic diffusion for stereo matching and line fields estimation," *Int'l J. Computer Vision*, vol. 47, no. 1, pp. 195-218, Apr. 2002.

[6] C. L. Zitnick and T. Kanade, "A cooperative algorithm for stereo matching and occlusion detection," *IEEE Trans. on PAMI*, vol. 22, no. 7, pp. 675-684, July 2000.

[7] D. Comaniciu and P. Meer, "Robust analysis of feature spaces: color image segmentation," In *Proc. CVPR*, pp. 750-755, 1997.

[8] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int'l J. Computer Vision*, vol. 47, no. 1, pp. 7-42, Apr. 2002.