Stereo matching with space-constrained cost aggregation and segmentation-based disparity refinement

Yi Peng, Ge Li*, Ronggang Wang, Wenmin Wang

School of Electronic and Computer Engineering, Shenzhen Graduate School, Peking University, Xili University Town, Shenzhen, China, 518055

ABSTRACT

Stereo matching is a fundamental topic in computer vision. Usually, stereo matching is mainly composed of four stages: cost computation, cost aggregation, disparity optimization and disparity refinement. In this paper, we propose a novel stereo matching method with space-constrained cost aggregation and segmentation-based disparity refinement. State-of-the-art methods are used for cost aggregation and disparity optimization stages. Three technical contributions are given in this paper. First, applying space-constrained cross-region in cost aggregation stage; second, utilizing both color and disparity information in image segmentation; third, using image segmentation and occlusion region detection to aid disparity refinement. The performance of our platform ranks second in the Middlebury evaluation.

Keywords: Stereo matching, space-constrained cost aggregation, segmentation, belief propagation, occlusion detection

1. INTRODUCTION

Stereo matching has been one of the most active research areas in computer vision. According to Scharstein and Szeliski [1], methods for stereo matching can be classified as either global methods or local methods. Local methods usually run faster and can be easily adopted in practical applications. For global methods such as graph cut [11] and belief propagation [2], disparity optimization is applied on the basis of energy minimization. Thus, global methods are relatively slower but tend to get a better disparity result. However, some global methods can achieve comparable running speeds with those of local methods in recent works [2]. Global approaches incorporate explicit smoothness assumptions and determine all disparities simultaneously by applying energy minimization techniques. Also, the data term or initial disparity is important for global methods, where data term is usually calculated based on the support/aggregation region. The region is likely to have similar disparity values, which are helpful for the results.

Most stereo matching methods can be divided into four stages [1]: cost computation, cost aggregation, disparity optimization, and disparity refinement. Cost aggregation is the most critical stage for the accuracy of stereo matching. In [7], Tombari et al. proposed a segment support region for the cost aggregation. Yoon et al. proposed an adaptive weight method [8] for the cost aggregation. Hosni et al. proposed a geodesic weight in [9]. Tombari et al. presented a color segmentation-based cost aggregation [19] with the assumption that pixels inside the same segment are likely to have similar disparity values. Zhang et al. proposed a cross-based method [10] with which the aggregation region could be an adaptive window. After that, they proposed a new cross-region [5] for cost aggregation which is more robust than the previous one [10]. As a result, its area is closer to the outline of objects and has good adaptability for the objects with complex shapes. However, the traditional cross-region based aggregation method is still not robust enough, since regions near object edges are not well structured. As a result, the cross shape is very sensitive to the color threshold, which makes the method not robust enough. To solve those difficulties, we propose a robust cross-region based method, called space-constraint cross-based cost aggregation in this paper.

In order to improve accuracy, many segmentation based disparity refinement methods have been proposed in recent years. Klaus et al. [13] used segmentation with slant to estimate the disparity plane. Yang et al. [14] applied segmentation with iterative refinement. Bleyer et al. [18] used segmentation with model fitting. However, color-based segmentation method may not be suitable for disparity estimation, since in order to separate the discrete disparity areas of the image, one must apply over-segmentation, which will weaken the function of segmentation. In [3], Comaniciu and

Three-Dimensional Image Processing, Measurement (3DIPM), and Applications 2015, edited by Robert Sitnik, William Puech, Proc. of SPIE-IS&T Electronic Imaging, Vol. 9393, 939309 · © 2015 SPIE-IS&T
CCC code: 0277-786X/15/$18 · doi: 10.1117/12.2083741

Meer proposed a mean-shift approach which can greatly improve the accuracy of the contour of the layer. The superpixel [4] approach was also proposed using space constraint, which segments the object boundary with very high accuracy only when it sets excessive superpixel resulting in small layers. However, a good segmentation method should arrange pixels with continuous disparity in the same layer. Therefore, Zhou and Boulanger proposed a segmentation method [15] based on disparity, with which the segmentation was spread according to neighboring pixels' disparity. Even though this method is simple, it is too sensitive to the initial disparity error. To overcome the above shortcomings, we propose an innovative segmentation method by utilizing both color and disparity information in this paper.

On the other hand, the initial disparity of occluded region tends to be wrong. Zitnick [16] and Grammalidis [17] proposed two methods for detection of the occluded region based on disparity space, but their methods are complicated and difficult to implement. In this paper, we use a robust and fast occlusion detection method based on our proposed segmentation result. With this method, to process a picture with $n*m$ size, the time complexity is reduced to $O(n*m)$.

Simulation results based on Middlebury dataset show that the proposed stereo matching method achieves second place among all 140 submitted methods.

2. PROPOSED ALGORITHM

Following Scharstein and Szeliski's taxonomy [1], our proposed method is made up of four steps: cost initialization, cost aggregation, disparity computation and refinement, which is shown in Figure 1. Detailed descriptions of these individual steps are as follows.

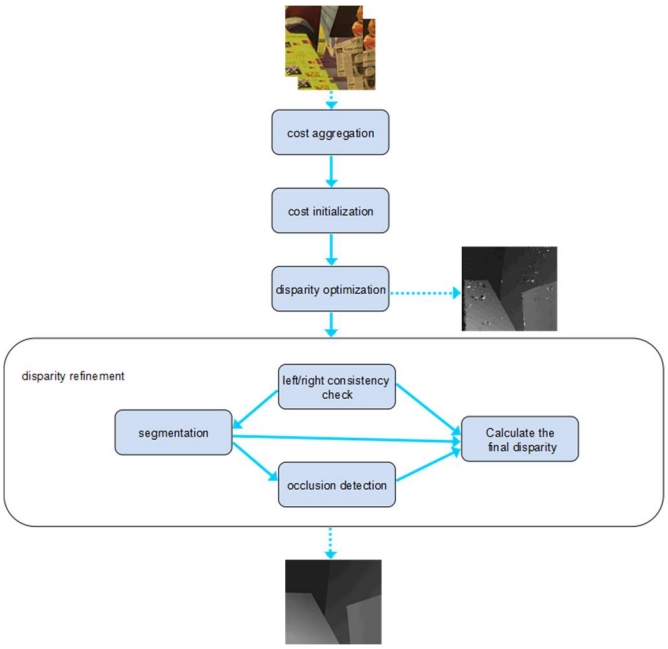


Figure 1. Flow chart of the proposed stereo matching method. (Solid blue arrows indicate the order of execution, and dotted blue arrows indicate input and output images)

2.1 Cost aggregation

A region based method on cross skeletons was first proposed by Zhang et al. [10, 5]. The cross region is constructed under the following two rules:

$$O_{lab}(p,q) < k_1 \text{ and } O_{lab}(q, q+(1,0)) < k_1, \quad (l_1 < L_1) \tag{1}$$

$$O_{lab}(p,q) < k_2, \quad (L_1 \le l_1 \le L_2) \tag{2}$$

Where l_1 is Euclidean distance of p to the search point q, and $O_{lab}(p,q)$ is Euclidean distance between point p and q in the lab color space. k_1, k_2, L_1, L_2 are user-defined parameters, where $k_1 > k_2$, $L_2 > L_1$. Rule 1 restricts the color difference between p and q. Rule 2 means when the arm length is more than L_1 but less than L_2, a much tight threshold value k_2 is applied for $O_{lab}(p,q)$ to warrant that the arm can only extends in a region with similar color pattern. A cross shape is decided by the color threshold k_1 and k_2. It is difficult to find a unified set of parameters simultaneously, which can accommodate multiple color changes over different pictures. As a result, a cross shape may be hard to match the contour of an object.

A traditional cross–based method applies some kinds of color segmentation. For each search pixel p the method calculates an area around p based on color and distance information. Recently, Achanta et al. proposed a space constraint segmentation method, named superpixel [4]. This method can greatly improve the stability and accuracy of the color segmentation by making segmented region closer to the contour of object. Lu et al. [20] also used this method to improve the original PatchMatch. Inspired by this method, we propose a robust space-constraint cross-based cost aggregation method to improve traditional cross-based method.

We define sample pixels as pixels sub-sampled with constant distance from one picture both in horizontal and vertical directions. As examples in Figure 2, yellow pixels are sample pixels. To construct one cross, besides the two rules in traditional cross-based method, we introduce a third rule as follows,

$$O_{lab}(p,q) < t * O_{lab}(q,e_i), \quad (\rho_1 * l_1 < l_2 < \rho_2 * l_1) \tag{3}$$

Where e_i is sample pixel in the neighborhood of p that satisfies the condition of $(\rho_1 * l_1 < l_2 < \rho_2 * l_1)$, l_1 is the distance from p to the search pixel q, and l_2 is the distance from e_i to the q. $O_{lab}(p,q)$ is the Euclidean distance between pixel p and q in the lab color space. ρ_1, ρ_2 are the user-defined parameters, where $\rho_2 > \rho_1$.

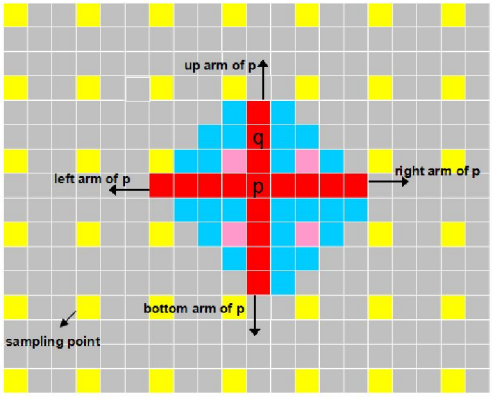


Figure 2. Space-constraint Cross Construction: an upright cross is constructed for each pixel. The support region of pixel p is modeled by merging the horizontal arms of the pixels (q for example) lying on the vertical arms of pixel p.

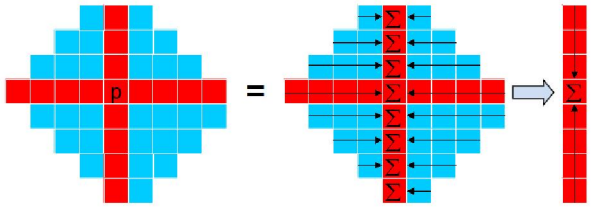


Figure 3. Cross Aggregation: the cost in the support region is aggregated with two passes along the horizontal and vertical directions [5].

Though we involve more thresholds, the color thresholds of k_1 and k_2 are relaxed with our method. In traditional method, along with k_1 and k_2 increase, the cross region tends to overlap with the border of the object. By applying our proposed space constraints of rule 3, this phenomenon can be avoided. Figure 4 gives cross region examples constructed by traditional method and the proposed method respectively. The color thresholds of traditional method are set to optimal values, in contrast to constant values in our method. As a result, the cross region constructed by the proposed method is closer to the object contour.

traditional method proposed method

Figure 4. Cross regions constructed by traditional method and the proposed method (the black pixel is the cross center, and the cross region is composed of white pixels and the black pixel).

2.2 Cost initialization

The cost function is a weighted sum of three terms. The first term is census [5] (C_{census}). The second term is the absolute value of the difference in color spaces (C_{AD}). The third term is a bidirectional gradient with grayscale image (C_{GD}).

$$C(x, y, d) = aC_{census} + bC_{AD} + gC_{GD} \tag{4}$$

Where x and y is the pixel coordinate, d is the pixel disparity, a, b and g is user defined parameters.

2.3 Disparity optimization

As we know, global method can greatly improve the disparity results. However it is a NP-hard problem. Graph cut [11] and belief propagation [2] are the two common approximation algorithms, and belief propagation is faster. Belief propagation can also achieve better results as shown in [13]. The confidence of B and the energy function E can be simply expressed as [2],

$$B = e^{-E} \tag{5}$$

Here, the maximum confidence of B is equivalent to the minimum of E. The disparity d_p of pixel p is expressed as an energy function.

2.4 Disparity refinement

After the previous steps, the disparity plane is still not accurate enough. As shown in figure 1, most of the inaccurate disparity is distributed around object boundaries. By distinguishing these areas, we can refine disparity plane by the proposed segmentation method.

2.4.1 Segmentation

A good segmentation method should arrange pixels with continuous disparity in the same layer and pixels with discontinuous disparity in the different layers. We propose an algorithm by combining color and disparity information. The proposed method is consisted of two steps. Firstly, traditional meanshift [3] and superpixel [4] methods are employed to segment the left image of the stereo pair into multiple layers. Then, the layers are merged with our proposed color based cluster and disparity based cluster methods as shown in Figure 5. Our segmentation method can obtain better disparity plane compared with meanshift and superpixel methods as shown in Figure 6.

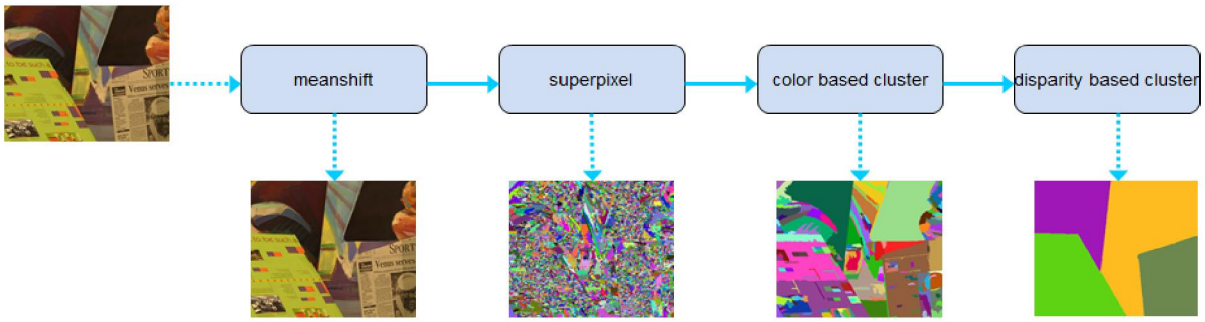


Figure 5. Flow chart of the segmentation method. (Solid blue arrows indicate the order of execution, and dotted blue arrows indicate input and output images)

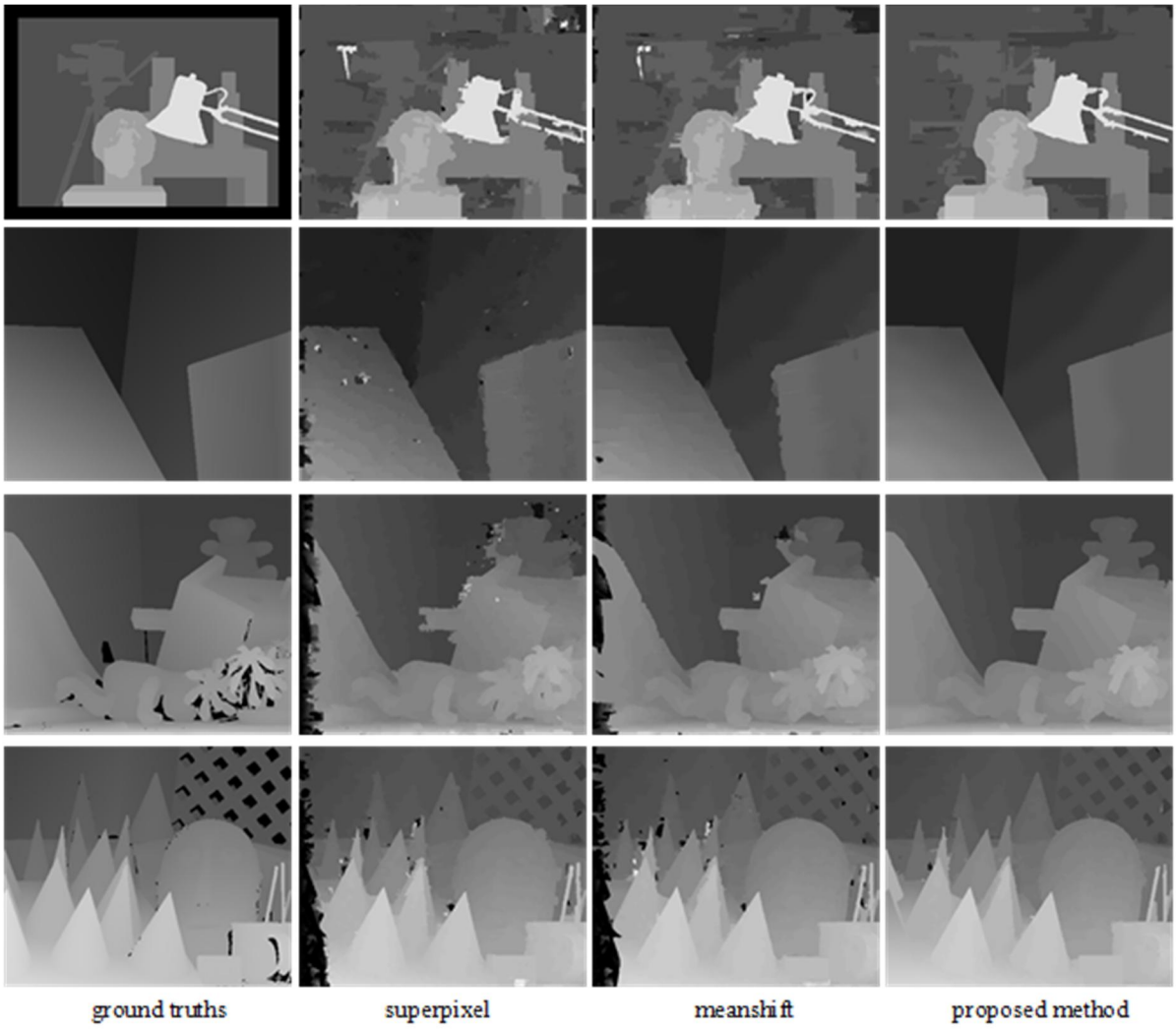


Figure 6. Comparison of the disparity plane results among several different segmentation methods.

2.4.1.1 Color based cluster

If squared color distance of two adjacent layers is close enough as in equation (6), we cluster them to improve the stability of the block.

$$MeanEuclid(l,s) < \frac{\tau}{\sqrt{vol(l)+vol(s)}} \tag{6}$$

where $vol(l)$ is the number of pixels in layer l and $vol(s)$ is the number of pixels in layer s. $MeanEuclid(l,s)$ is Euclidean distance between layer l and s in the mean lab color space. τ is the user-defined parameter.

2.4.1.2 Disparity based cluster

To make the segmentation more suitable for disparity estimation, disparity continuous layers should be combined into a continuous layer. We first select reliable pixels with traditional left/right consistency check method [12], and then these reliable pixels are used to cluster the layers by the following process.

(1) Some layers may have very few reliable pixels and they have not enough disparity information to perform segmentation accurately. To solve this issue, we cluster these layers with their adjacent layer according to color information, when the following conditions are met,

$$rp[l] < p[l]/\delta \tag{7}$$

where $rp[l]$ is the number of reliable pixels and $p[l]$ is the number of pixels in layer l. δ is the user-defined parameter.

(2) However, the disparities refined by above process may still be unreliable and they will affect the accuracy of segmentation. Here we use a robust method to arrange disparity continuous regions in the same layer. For a layer l adjacent to layer s, the neighboring pixels pair of u_i and w_i on the boundaries are processed as follows:

1) Averaging the disparities of reliable pixels in the same layer over a square window centered on the pixels *u* and *w*, to get $v_u(i)$ and $v_w(i)$ respectively. A variable is defined by,

$$th[u_i][w_i] = max|v_u(i) - v_w(i)| \quad i \in \omega_{u,w} \tag{8}$$

where $\omega_{u,w}$ is the collection of all neighboring pixels in l and s. When $th[u_i][w_i] < r$, the layer l and layer s are merged. *r* is the user-defined parameter.

2) Averaging the disparities of all reliable pixels in the layer l and s, to get d_l and d_s respectively. Another variable is defined by,

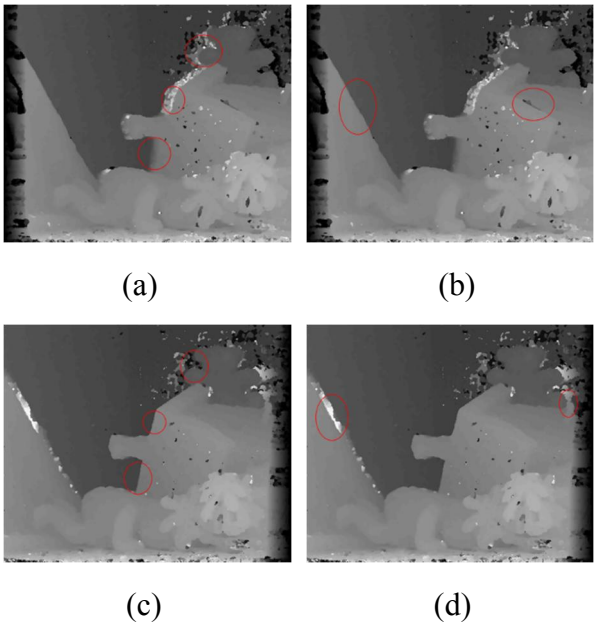
$$ds[l][s] = |d_l - d_s| \tag{9}$$

When $ds[l][s] < \sigma$, layer l is clustered with layer s. σ is the user-defined parameter.

2.4.2 Occlusion detection

The above proposed segmentation method is utilized to detect occlusions. As shown in Figure 7, for the left image, according to the characteristics of human eyes, the occlusion region is on the right of each layer.

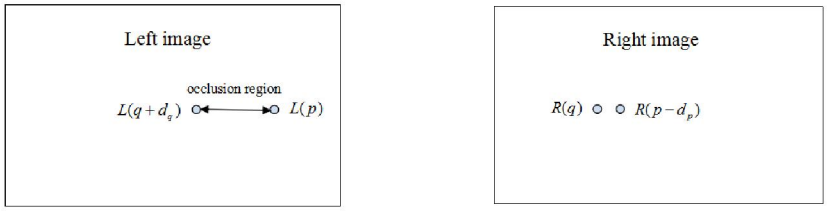
As shown in Figure 8. For the left first reliable pixel $L(p)$ in each row of each layer, we find its corresponding pixel $R(p - d_p)$ in the right image according to its disparity d_p. Then we search k pixels in the left of $R(p - d_p - 1)$ and select the reliable pixel $R(q)$ with the largest disparity. Here k is set to 5. Different value of k has different effects as shown in Figure 9. We get the disparity d_q of $R(q)$ and calculate its corresponding pixel $L(q + d_q)$ in the left image. The pixels between $L(p)$ and $L(q + d_q)$ in the left image are the occluded pixels.



(a) (b)

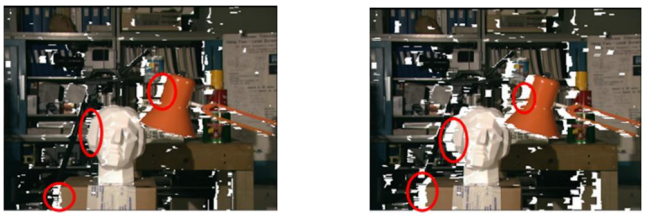
(c) (d)

Figure 7. On the left image. Initial disparity on the right region of each layer is not accurate (a) and on the right region of each layer is accurate (b). On the right image situation is reversed (c and d).



(a) (b)

Figure 8. Occlusion reasoning. In the picture we marked out the key points which computing needs.



(a) result of $k = 6$ (b) result of $k = 10$

Figure 9. The occlusion detection experiment on the images and white areas are occluded regions.

2.4.3 Calculate the final disparity

First, we estimate the disparity bidirectional gradients of the reliable pixels. Then, to estimate of the pixel p, we use the reliable pixels in a rectangular region centered by p and in the same layer with p. Multiple candidate disparities for pixel p are calculated by,

$$dx = (x[p] - x[q_i]) * derivationX[q_i] \tag{10}$$

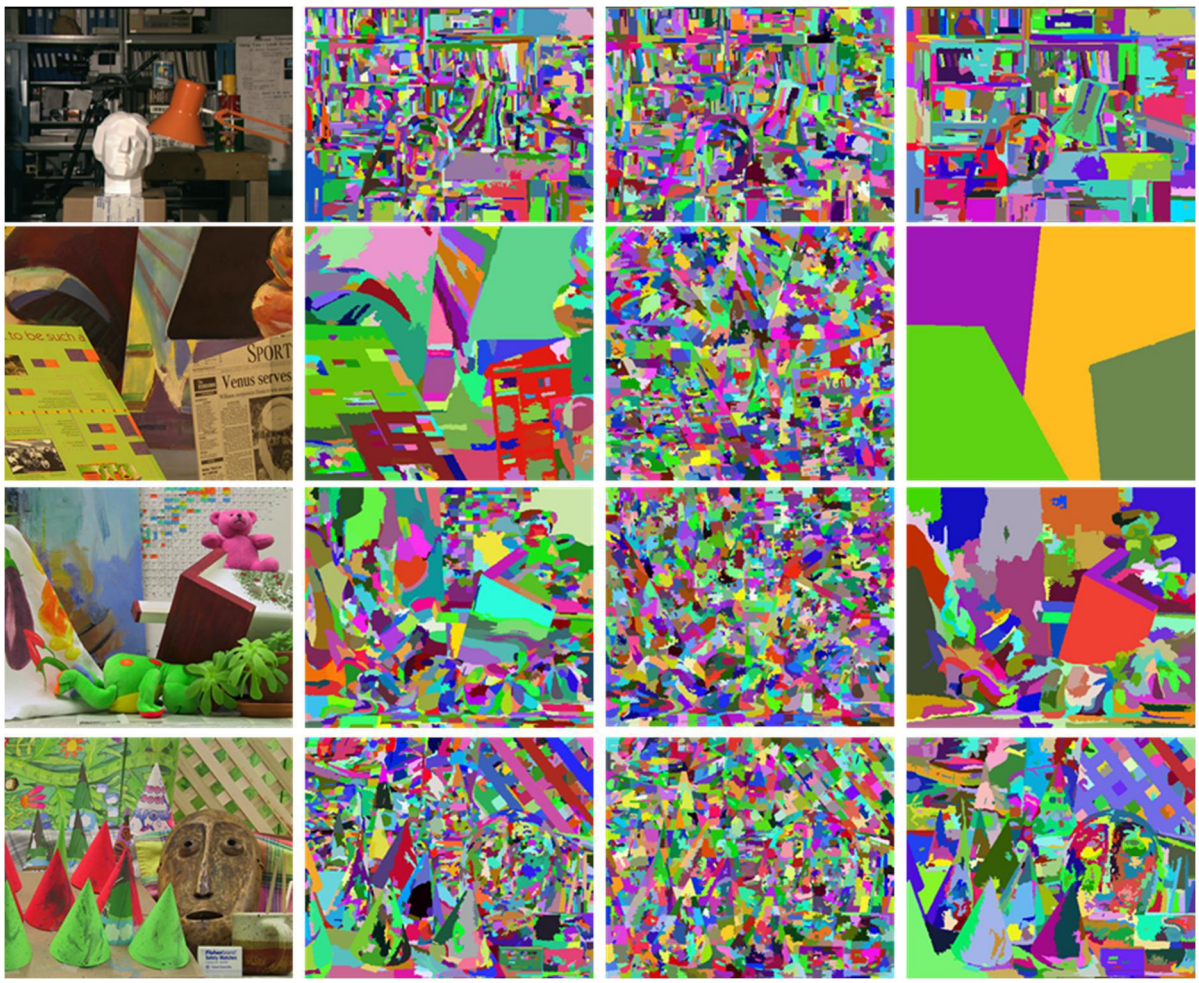
$$dy = (y[p] - y[q_i]) * derivationY[q_i] \tag{11}$$

$$d(p_i) = d(q_i) + dx + dy \tag{12}$$

where q_i is the reliable pixel, $d(q_i)$ is the disparity of q_i, $derivationX[q_i]$ is the X direction gradient, and $derivationY[q_i]$ is the Y direction gradient. Finally, we sort all the candidate disparities and take the median of them as the disparity of p.

3. EXPERIMENTAL RESULTS

We first compare performance of the proposed segmentation method with traditional meanshift [3] and superpixel [4] methods in Figure 10. The proposed segmentation method achieves the best performance. Then, we test the performance of the proposed method in the Middlebury test bed [6], and measure the percentage of bad matching pixels (where the absolute disparity error is larger than 1 pixel) with three subsets of an image: nonocc (the pixels in the non-occluded region), all (all the pixels), and disc (the visible pixels near the occluded regions). The parameters of our method set in the testing are given in Table 1. The test results are shown in Figure 11, the proposed stereo matching method ranks the second place among all the 140 submitted methods.



reference images mean shift superpixel proposed method

Figure 10. Comparison of the segmentation results of several different segmentation methods.

Table 1. The parameters of our proposed method.

k_1	k_2	l_1	l_2	ρ_1
15	5	1.5	3	10
ρ_2	a	b	g	t
100	20	74	6	1.1
τ	δ	r	σ	
78	10	0. 02	0.01	

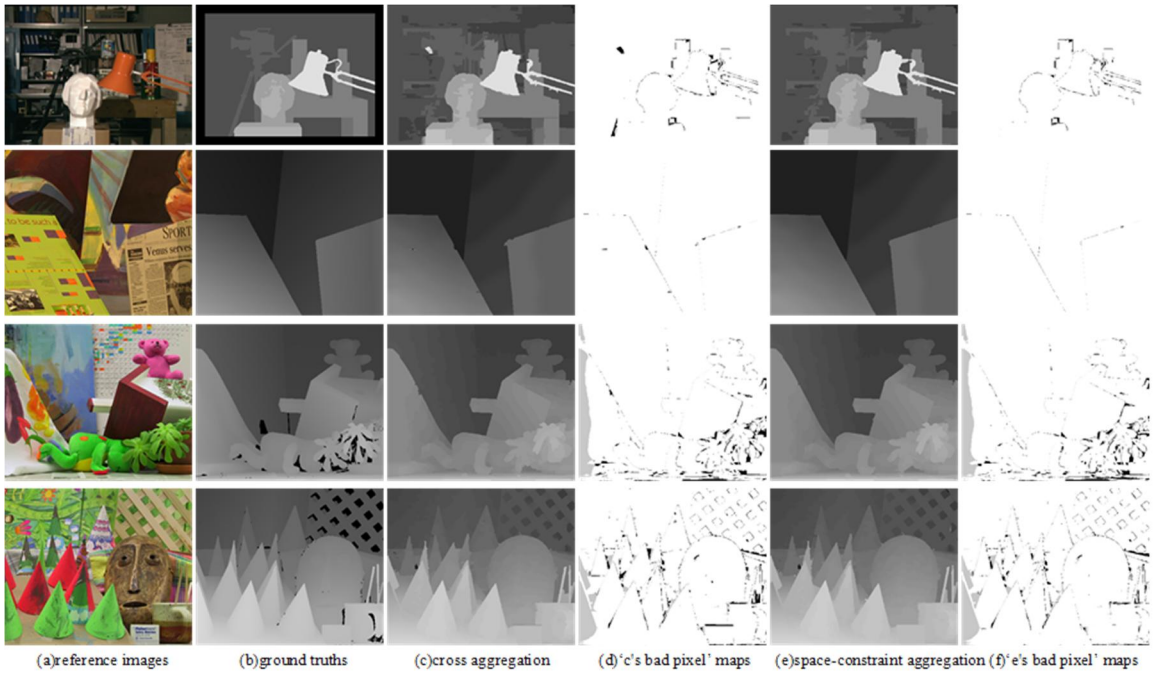


Figure 11. Results using the Middlebury datasets: Tsukuba, Venus, Teddy and Cones. Pixels with a disparity error greater than one pixel are displayed in the 'bad pixel' maps, where miss matches in non-occluded areas are indicated in black, and in occluded areas are in gray.

ACKNOWLEDGMENTS

This work was partly supported by a grant of National Science Foundation of China 61370115, Shenzhen Peacock Plan and Shenzhen fundamental research project JCYJ20120614150301623.

REFERENCES

[1] Scharstein D and Szeliski R. "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms [J]," International journal of computer vision, pages 47(1-3): 7-42 (2002).

[2] P. F. Felzenszwalb and D. P. Huttenlocher. "Efficient belief propagation for early vision," In CVPR, pages I: 261 – 268 (2004).

[3] D. Comaniciu and P. Meer. "Mean shift: A robust approach toward feature space analysis," IEEE:PAMI, pages 24(5):603 – 619 (2002).

[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 Trans. Pattern Anal. Mach. Intell., pages 34(11), pages 2274-2282, (2012).

[5] X. Mei, X. Sun, M. Zhou, S. Jiao, H. Wang, and X. Zhang. "On building an accurate stereo matching system on graphics hardware," GPUCV (2011).

[6] D. Scharstein and R. Szeliski. Middlebury stereo evaluation -version 2, 2010.http://vision.middlebury.edu/

[7] F. Tombari, S. Mattoccia, and L. D. Stefano. "Segmentationbased adaptive support for accurate stereo correspondence," In Proc. PSIVT, pages 427-438 (2007).

[8] K. J. Yoon and I. S. Kweon. "Adaptive support-weight approach for correspondence search," IEEE TPAMI, pages 28(4):650-656 (2006).

[9] A. Hosni, M. Bleyer, M. Gelautz, and C. Rheman. "Local stereo matching using geodesic support weights," In Proc. ICIP, pages 2093-2096 (2009).

[10] K. Zhang, J. Lu, and G. Lafruit. "Cross-based local stereo matching using orthogonal integral images," IEEE TCSVT, pages 19(7):1073-1079 (2009).

[11] Y. Boykov, O. Veksler, R. Zabih. "Fast approximate energy minimization via graph cuts," Pattern Analysis and Machine Intelligence, IEEE Transactions on, pages 23(11):1222-1239 (2001).

[12] P. Fua, "A parallel stereo algorithm that produces dense depth maps and preserves image features," Machine Vision and Applications, pages 6:35–49 (1993).

[13] A. Klaus, M. Sormann, K. Karner "Segment-Based Stereo Matching Using Belief Propagation and a Self-Adapting Dissimilarity Measure," Pattern Recognition, pages 3:15-18(2006).

[14] Q. Yang, L. Wang, R. Yang, H. Stewenius, and D. Nister. "Stereo matching with color-weighted correlation, hierachical belief propagation and occlusion handling," In CVPR (2006).

[15] X. Zhou, P. Boulanger. "New eye contact correction using radial basis function for wide baseline video conference system," In Pacific-Rim Conference on Multimedia, pages 68 – 79 (2012).

[16] C. L. Zitnick and T. Kanade. "A Cooperative Algorithm for Stereo Matching and Occlusion Detection," IEEE Trans. Pattern Analysis and Machine Intelligence, pages 22(7): 675-684(2000).

[17] N. Grammalidis and M. Strintzis. "Disparity and occlusion estimation in multiocular systems and their coding for the communication of multiview image sequences," IEEE Trans. Circuits Syst. Video Technol. Pages 8(3):328-344 (1998).

[18] M. Bleyer, C. Rother, and P. Kohli. "Surface stereo with soft segmentation," in Proc. IEEE Comput. Soc. Conf. CVPR, pages 1570 – 1577(2010).

[19] F. Tombari, S. Mattoccia, L.di Stefano, and E. Addimanda. "Near Real-Time Stereo Based on Effective Cost Aggregation," Proc. 19th Int'l Conf. Pattern Recognition, pages 1-4 (2008).

[20] Jiangbo Lu, Hongsheng Yang, Dongbo Min, and Minh N. Do. "Patch match filter: Efficient edge-aware filtering meets randomized search for fast correspondence field estimation," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition pages 1854 – 1861(2013).