Segmentation-aided local stereo matching

Baogui Yan Yisong Chen
Graphics Laboratory, EECS Department, Peking University
yanbg@graphics.pku.edu.cn chenys@graphics.pku.edu.cn

Abstract

This paper presents a new segmentation-aided checking method for distinguishing mismatch pixel within a complete stereo matching framework. Instead of employing traditional cross-checking method, our method successfully integrates segmentation information in the disparity refinement step. Segmentation-aided checking method follows the plane smoothness assumption, i.e. pixels within a plane should vary depth continuously. Conceptually, our method goes as follows: first, the pixels are clustered into a set of segments; second, planes are fitted for each segment; and finally, we choose the plane contain most stable pixel (i.e. passing through the Left-right Match Constraint) as the main plane of the segment and mark all the pixels in other planes as unstable. Compared to cross-checking method based on Left-right Match Constraint, our method is more effective especially in the repetitive structure areas and low texture areas. Compared to previous segment-based method that treat segments as match units, our method is more lightweight and easier to implement. Experimental results from our implementation evaluated on the Middlebury platform shows that proposed method is comparable to previous state-of-the-art local methods.

1 Introduction

Stereo matching is one of the most extensively studied problems in the field of computer vision. Most stereo matching methods usually take the following four steps. **Matching cost computation**: the matching costs are initialized for every pixel at each disparity level with its color/intensity information; **Cost aggregation**: the initial matching costs are aggregated within every pixel's support region. Cost aggregation methods are often based on the assumption that neighboring pixels with similar color/intensity share same disparity. **Disparity computation**: the disparity map is computed with local methods (e.g. winner-takes-all strategy) or optimized with a global energy function. **Disparity refinement**: the disparity map obtained with previous three steps may include occlusion and mismatch pixels. This step detects these outliers and then corrects them with various post-processing techniques.

Up to now, large numbers of stereo matching methods have been proposed to obtain dense disparity map. Most of them can be classified into two types: global methods and local methods[1]. Global methods usually treat the stereo matching problem as an optimization problem and minimize the energy function with graph cuts[18] or belief propagation[19]. Local methods, on the other hand, compute the disparity at a given pixel over a local support region. Local methods primarily rely on cost aggregation step (step.2). And recently, filtering the cost volume with edge-aware filter shows good matching accuracy. Yoon and Kweon[4] first demonstrated that bilateral filter[5] is very effective for preserving depth boundaries and can be used for cost aggregation. After that, a number of bilateral filter based cost aggregation methods have been developed, such as [6][8][9][11]. He *et al.* proposed a guided image filter, which can be used for image processing and stereo matching. This method shows leading performance and low computational expense (linear to the pixel number plus disparity level number). These edge-aware filter based methods successfully deal with the sharp edge retrieval, but likely produce mismatch in some repetitive structure areas and low texture areas.

By assuming that neighboring pixels with similar colors should have similar disparity, segment-based method can improve depth performance especially for large low texture areas. Previous methods usually treat segments as basic matching units and then solve an optimization problem by minimizes a global cost function. However, these methods don't allow large disparity variation inside the segment, which inevitably affect the disparity estimation.

In this paper, an approach for edge-aware filter based stereo matching is presented, with certain improvements in the disparity refinement step (step.4). The proposed method can successfully deal with above-mentioned trade-off. Specifically, the initial matching cost combines two individual costs: the absolute difference cost, and a gradient based cost (step.1). An aggregated cost volume is computed by minimum spanning tree using the initial matching cost (step.2). And then the winner-takes-all approach is applied to produce the disparity map (step.3). Finally, our segmentation-aided checking method is proposed for distinguishing the mismatch pixels and an efficient post-processing is applied to enhance the disparity map (step.4).

2 Matching Cost Computation

This step computes the initial matching cost for every individual pixel at all disparity level. The absolute difference on image color/intensity (AD) is a simple and widely used measure function. Given a pixel $p = (x, y)$ in the reference image (e.g. the left image) and a disparity level d, its correspondence pixel in the matching image (e.g. the right image) is $pd = (x - d, y)$. The cost term C_{AD} is shown as follows:

$$C_{AD}(p,d) = \frac{1}{3}\sum_{i=R,G,B} |I_i^{ref}(p) - I_i^{mat}(pd)| \qquad (1)$$

On the other hand, the image intensity derivatives in image x, y directions are extracted, and the absolute differences of each derivative value in the two directions are summed up as another cost measure i.e. C_{GRAD}.

$$C_{GRAD} = \sum_{x,y} |\nabla I^{ref}(p) - \nabla I^{mat}(pd)| \qquad (2)$$

Merging above two different costs derives the total matching cost C: absolute difference on image color/intensity and absolute difference on image gradients. An exponential function is preferred [2][3] for normalization and control on the influence of the outliers with λ:

$$\rho(C,\lambda) = 1 - exp(-\frac{C}{\lambda}) \qquad (3)$$

The joint matching cost value $C(p, d)$ is computed as follows:

$$C(p,d) = \rho(C_{AD}(p,d), \lambda_{AD}) + \rho(C_{GRAD}(p,d), \lambda_{GRAD}) \qquad (4)$$

Which turn out to improve following non-local cost aggregation result compared to simply sum the two cost terms up.

3 Cost Aggregation

This step aggregates each pixel's matching cost over a support region to reduce the matching ambiguities and noise in the initial cost volume. Here, we prefer Yang's non-local cost aggregation method [11].

Different from other edge-aware filter based cost aggregation methods that rely on local support region; Yang's method aggregates matching costs over the entire image with a minimum spanning tree (MST) structure.

The input image I is treated as a 4-connected, undirected grid graph $G = (V, E)$

where V denotes all the pixels in I and E is all the edges connect neighboring pixels. Let e denote an edge connecting neighboring pixel pair p_1 and p_2, the weight of e is

$$w(e) = w(p_1, p_2) = |I(p_1) - I(p_2)| \quad (5)$$

The minimum spanning tree T is derived from G by removing edges with bigger weights that are more likely cross the depth boundaries. Let p_1 and p_2 be any two pixels in T, and their distance $D(p_1, p_2)$ is defined by the sum of the weights along the shortest path. The similarity between p_1 and p_2 is denoted as follows:

$$S(p_1, p_2) = \exp(-\frac{D(p_1, p_2)}{\sigma}) \quad (6)$$

Where σ is used to adjust similarity between two pixels. Thus, the aggregation cost function is defined as follows:

$$C_d^A(p) = \sum_q S(p,q) C_d(\mathrm{q}) = \sum_q \exp(-\frac{D(p,q)}{\sigma}) C_d(\mathrm{q}) \quad (7)$$

Disparity is estimated by 'winner-takes-all method, i.e. by simply selecting the disparity level with the lowest cost.

4 Disparity Refinement with Segmentation-aided Checking Method

Edge-aware filter based methods work well in depth boundaries, but still produce mismatch in some repetitive structure areas and low texture areas.

Cross-checking method is widely used for distinguishing occlusion areas and easy to implement. A pixel $p = (x, y)$ is marked as stable if it satisfies the following constraint for the disparity maps D^{ref} and D^{mat} of the stereo image pair:

$$D^{ref}(p) = D^{mat}(p - [D^{ref}(p), 0]) \quad (8)$$

Though cross checking method works well in outlier detection, it can't deal with aforementioned problem. As shown in Figure 1, marking mismatch pixels as stable will ruin neighboring areas.

Repetitive structure | Cost Filter [10] | Non Local Filter[11] | Our method

Figure 1 Once mismatch pixels are marked as stable, propagation method will use these wrong disparities to computer neighboring unstable pixels' disparity.

Reexamining the mismatch area in Figure 1, we can find those areas and neighboring correct areas belong to same plane, while pixels' disparities of the same plane should vary continuously. Here, we give this idea a definition theoretically, and then use it to remove the mismatch pixels.

Plane Smoothness Assumption: Plane smoothness assumption is a kind of strong smoothness assumption. Pixels in the same plane are enforced to vary disparity continuously. Pixels that couldn't satisfy this assumption are marked as unstable.

Based on this assumption, we segment the image using mean-shift method[13] and treat each segment as a plane, as shown in Figure 2. For each segment, we try to fit planes with the pixels passing the cross-checking. Note that we enforce that the plane should contain sufficient stable pixels, and the disparity gradient shouldn't exceed given threshold. Finally, we choose the plane with most stable pixels as the principal plane and mark all the other pixels as unstable.

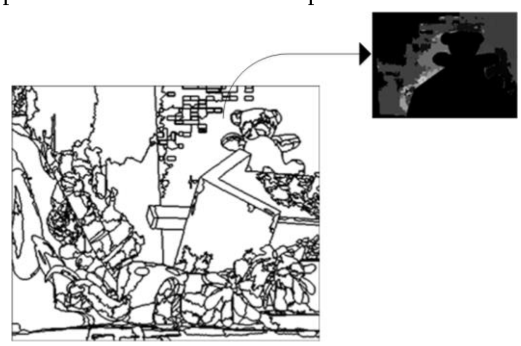


Figure 2 Teddy's segmentation map

Pixels passing the cross checking and segmentation-aided checking are shown as Figure 3. The repetitive structure areas are labeled correctly.

To handle the large unstable regions in Figure 3, we define a new cost value for each pixel p at each disparity d:

$$C_d^{new}(p) = \begin{cases} |d - D(p)|, & p \ is \ stable \\ 0, & else \end{cases} \tag{9}$$

Figure 3 Stable pixels map

Where D denote the disparity map. For all unstable pixels, their new cost values are assigned as zero at every disparity level, thus their disparities will completely depend on stable pixels.

And then we use the above-mentioned non-local cost aggregation method aggregating the new cost values $C_d^{new}(p)$ to propagate the disparity values from stable pixels to unstable pixels.

5 Experimental Results

We test our method with the Middlebury benchmark. The test platform is MacBook pro laptop computer with a 2.5 GHz Intel Core i5 CPU and 4 GB memory. We keep the parameters constant for all the data sets: $\lambda_{AD} = 10$, $\lambda_{GRAD} = 15$, $\sigma = 0.1$, $hs = 10$, $hr = 8$, $ms = 30$ (hs, hr, ms are parameters of mean shift algorithm).

For visual comparison, we present disparity results of Teddy and Baby1 dataset in Figure 7. The segmentation-aided method improves the performance in large low texture areas and around depth boundaries.

Our method ranks 21 in the Middlebury evaluation, and outperforms all the edge-aware filter based method, as shown in Figure 4. The disparity results are presented in Figure 5.

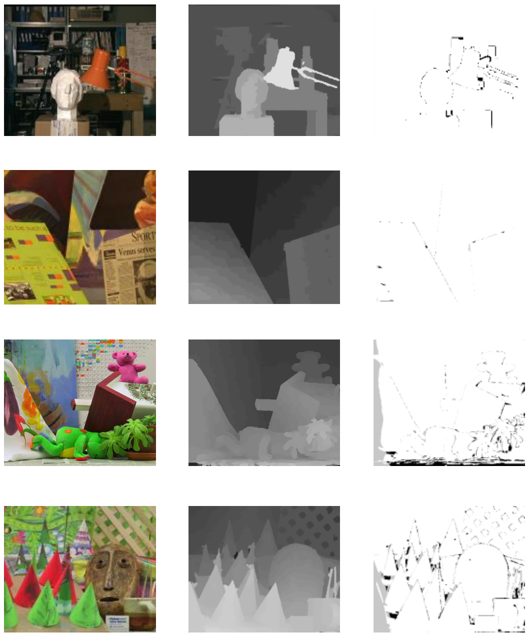


Figure 5 Left: Four stereo pairs of Middlebury evaluation platform. Middle: Final resulting disparity maps. Right: "bad" pixels of the produced disparity maps evaluated for the 1 pixel error threshold. Mismatched pixels are indicated by gray (occlusion) and black (non occlusion).

Figure 6 shows the comparison result of edge-aware filter based methods in recent years.

Algorithm	Avg. Rank	Avg. Erro	Tsukuba			Venus			Teddy			Cones		
			nonocc	all	disc	nonocc	all	disc	nonocc	all	disc	nonocc	all	disc
Our method	35.8	4.99	1.32	1.67	7.06	0.2	0.34	1.97	4.69	9.74	12.5	3.07	9.01	8.29
HEBF[15]	37.7	5.41	1.1	1.38	5.74	0.22	0.33	2.41	6.85	11.8	15.2	2.78	9.28	8.1
CostFilter[10]	43.6	5.55	1.51	1.85	7.61	0.2	0.39	2.42	6.16	11.8	16	27.1	8.24	7.66
NonLocalFilter[11]	46.8	5.48	1.47	1.85	7.88	0.25	0.42	2.6	6.01	11.6	14.3	2.87	8.45	8.1
AdaptWeight[2]	75.2	6.67	1.38	1.85	6.9	0.71	1.19	6.13	7.88	13.3	18.6	3.97	9.79	8.26

Figure 4 Results from the Middlebury evaluation platform for the 1 pixel threshold. Columns record from left to right: method; average rank; average percent of bad pixels; errors for the Tsukuba, Venus, Teddy, Cones stereo pairs;

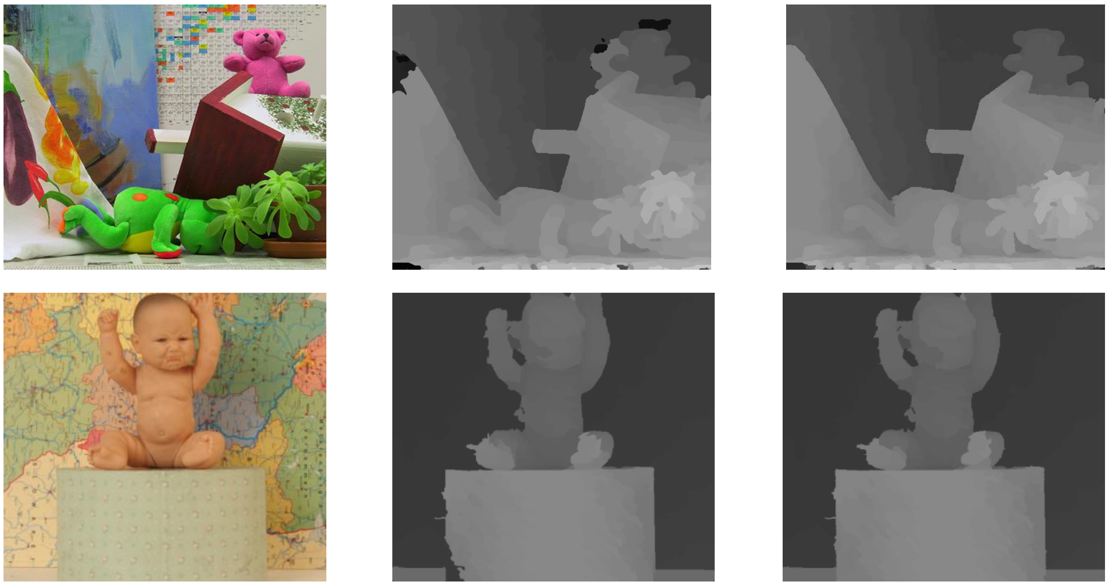
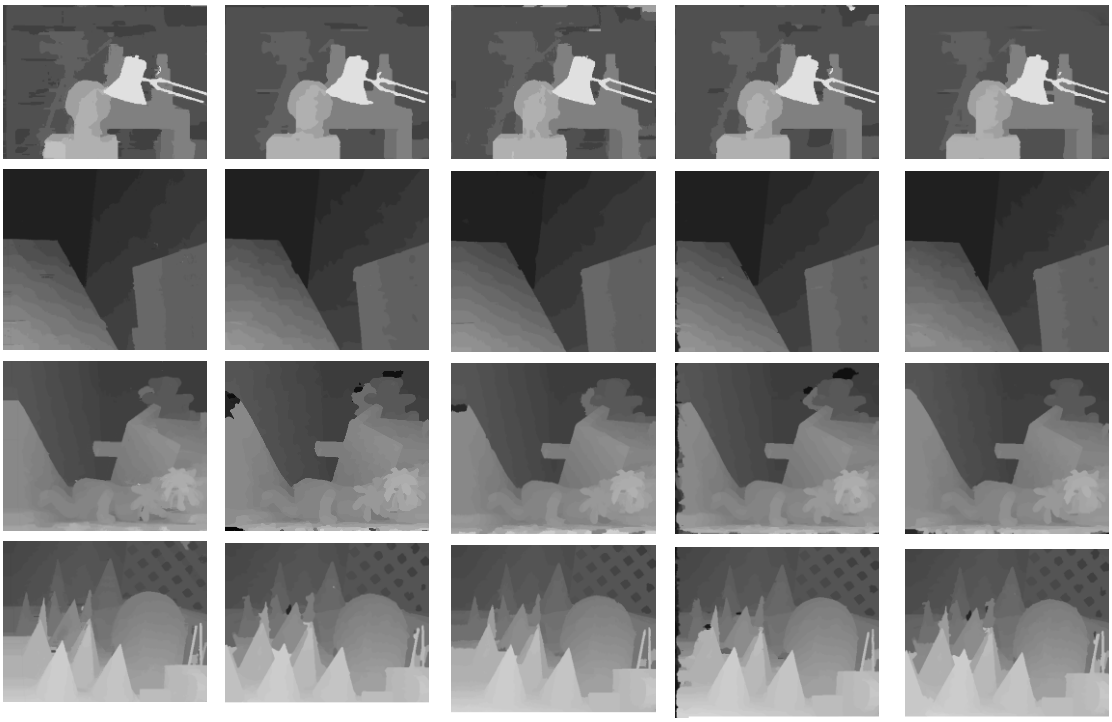


Figure 7 The disparity results of Teddy and Baby1 stereo pairs. From left to right: Original left images; disparity results without segmentation-aided checking; disparity results with segmentation-aided checking.



AdaptWeight[2] (6.67%)	NonLocalFilter[11] (5.48%)	CostFilter[10] (5.55%)	HEBF[15] (5.41%)	Our Method (4.99%)

Figure 6 Edge-aware filter based methods' experimental results on the Middlebury data sets. From top to bottom Tsukuba; Venus; Teddy; Cones; method and its average percent of bad pixels

6 Conclusion

This paper has presented a novel segmentation-aided checking method, which proves to be an effective supplement for edge-aware filter method. The proposed method significantly improves the disparity map's quality especially in repetitive structure areas and low texture areas. Our future work includes improving the sub-pixel performance, considering various segmentation methods and parallel implementation.

References

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

[2] Yoon K J, Kweon I S. Adaptive support-weight approach for correspondence search[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2006, 28(4): 650-656.

[3] Mei X, Sun X, Zhou M, et al. On building an accurate stereo matching system on graphics hardware[C]//Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on. IEEE, 2011: 467-474.

[4] Yoon K J, Kweon I S. Adaptive support-weight approach for correspondence search[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2006, 28(4): 650-656.

[5] Tomasi C, Manduchi R. Bilateral filtering for gray and color images[C]//Computer Vision, 1998. Sixth International Conference on. IEEE, 1998: 839-846.

[6] Paris S, Durand F. A fast approximation of the bilateral filter using a signal processing approach[M]//Computer Vision–ECCV 2006. Springer Berlin Heidelberg, 2006: 568-580.

[7] Porikli F. Constant time O (1) bilateral filtering[C]//Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on. IEEE, 2008: 1-8.

[8] Yang Q, Tan K H, Ahuja N. Real-time O (1) bilateral filtering[C]//Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009: 557-564.

[9] Yang Q, Wang S, Ahuja N. Svm for edge-preserving filtering[C]//Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. IEEE, 2010: 1775-1782.

[10] Rhemann C, Hosni A, Bleyer M, et al. Fast cost-volume filtering for visual correspondence and beyond[C]//Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. IEEE, 2011: 3017-3024.

[11] Yang Q. A non-local cost aggregation method for stereo matching[C]//Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE, 2012: 1402-1409.

[12] Fusiello A, Trucco E, Verri A. A compact algorithm for rectification of stereo pairs[J]. Machine Vision and Applications, 2000, 12(1): 16-22.

[13] Comaniciu D, Meer P. Mean shift: A robust approach toward feature space analysis[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2002, 24(5): 603-619.

[14] Wang Z F, Zheng Z G. A region based stereo matching algorithm using cooperative optimization[C]//Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on. IEEE, 2008: 1-8.

[15] Q. Yang. Hardware-efficient bilateral filtering for stereo matching. PAMI 2012.

[16] Scharstein D, Pal C. Learning conditional random fields for stereo//Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on.IEEE, 2007: 1-8

[17] Yang Q, Wang L, Yang R, et al. Stereo matching with color-weighted correlation, hierarchical belief propagation, and occlusion handling[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2009, 31(3): 492-504.

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

[19] Klaus A, Sormann M, Karner K. Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure[C]//Pattern Recognition, 2006. ICPR 2006. 18th International Conference on. IEEE, 2006, 3: 15-18.

[20] Levin A, Lischinski D, Weiss Y. A closed-form solution to natural image matting[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2008, 30(2): 228-242.

[21] He K, Sun J, Tang X. Guided image filtering[M]//Computer Vision–ECCV 2010. Springer Berlin Heidelberg, 2010: 1-14.