Stereo Matching by Adaptive Weighting Selection Based Cost Aggregation

Lingfeng Xu[†], Oscar C. Au[†], Wenxiu Sun[†], Lu Fang[‡], Ketan Tang[†], Jiali Li[†], Yuanfang Guo[†]
[†]The Hong Kong University of Science and Technology
[‡]University of Science and Technology of China
Email: [†]{lingfengxu, eeau, eeshine, tkt, jiali, eeandyguo}@ust.hk, [‡]fanglu@ustc.edu.cn

***Abstract*—Cost aggregation is the most essential step for dense stereo correspondence searching, which measures the similarity between pixels in the stereo images. In this paper, based on the analysis of the optimal adaptive weight, we propose a novel support aggregation strategy by adaptive weighting selection. The proposed method calculates the aggregation cost by the joint optimization of both left and right matching cost. By assigning more reasonable weighting coefficients, we exclude the occlusion pixels while preserving sufficient support region for accurate matching. The proposed optimal strategy can be integrated by any other adaptive weighting based cost aggregation method to generate more reasonable similarity measurement. Experimental results show that, compare with traditional methods, our algorithm can reduce the foreground fatten phenomenon while increasing the accuracy in the high texture regions.**

I. Introduction

Stereo matching is one of the most active research topics in computer vision. By capturing images / videos by stereo cameras, we can rectify the stereo images [1], and estimate the depth maps by disparity estimation [2]–[5]. The depth maps can be applied into view synthesis [6], object tracking, image based rendering, etc.

According to Daniel's paper [3], stereo matching generally performs (subsets of) four steps: 1. Matching cost computation; 2. Cost (support) aggregation; 3. Disparity computation / optimization; 4. Disparity refinement. The first step measures the similarities between two pixels, where one pixel is from left image and another one is from the right image. In order to include more texture and reduce the affection of the noise and color inconsistence between the stereo images, the similarity measurements are aggregated in a local region, such like square windows, shiftable windows, or weighted window. After that, the disparity of each pixel can be estimated by local methods like Winner Take All (WTA) approach or some other global methods such as Graph cuts [7] and belief propagation, which include the smoothness constraint between neighboring pixels. After the three steps, an initial depth map can be generated with some outliers in the occlusion regions and textureless region. Some post-processing methods are proposed to refine the depth maps by cross checking, median filter, plane fitting, etc.

Among the four steps, cost aggregation is the most essential part for dense stereo matching, which measures the similarity of pixels among the stereo images. In order to obtain high quality matching cost, quite a number of aggregation algorithms are proposed, such like square windows, shiftable windows [8], windows with adaptive size [9], and windows with adaptive weighting coefficients [4], [10], [11]. Because of the performance and accuracy, the most common used methods utilize the windows with adaptive weighting coefficients for cost aggregation:

$$E(p,d) = \frac{\sum_{q\in\Omega(p)} W(p,q)DSI(q,d)}{\sum_{q\in\Omega(p)} W(p,q)} \quad (1)$$

As shown in (1), the main idea of adaptive weighting methods is to find a good support region for each pixel. Here $DSI(q,d)$ is the disparity space image [3] which measures the matching cost of pixel p in current image and pixel $(p-d)$ in the reference image. The support region should contain the pixels within the same object of center pixel by assigning large weighting coefficients, and exclude the pixels in different depth level by assigning small weighting coefficients. Meanwhile, the support regions should also get rid of occlusion regions which only exist in one image.

Many prior researches focused on how to assign large weight to pixels in the same object and small weight in different object. The weighting coefficients are calculated based on color similarity and spatial distance, such like segmentation based method [12], [13], geodesic distance based method [10] , and soft segmentation based method [4], etc. Those algorithms have a good performance to separate the foreground and background. However, all of them have problems to exclude the pixels in the occlusion regions. Wang and Rachna's papers [12], [13] did not exclude the occlusion pixels, which produced more foreground fatten phenomenon. Yoon and Hosni's papers [4], [10] excluded the occlusion region by multiplying the left and right weighting coefficients. However, it produced noise in the high texture region because when multiplying two weighting coefficients, the support regions were much reduced and could not include sufficient texture information for correspondence matching.

In our paper, we focus on the problem of how to find more reasonable weighting coefficients: at one hand, occlusion pixels should be excluded; at the other hand, we want to include as many support regions as possible. The rest of the paper is organized as follows: section II will explain the detailed algorithm of proposed aggregation method. Experimental and comparison results will be shown in section III, followed by the conclusions and future work in section IV.

978-1-4673-5762-3/13/$31.00 ©2013 IEEE

II. Proposed Algorithms

The target of cost aggregation is to aggregate the matching cost within a 'good' support, where the support should contain the pixels in the same depth and exclude the pixels in the different depth and occlusion regions. In this section, we first propose the optimal model for adaptive weighting coefficients based on the assumption that disparity maps are given. Then, we analyze the common used existing methods and explain the problems they have. Finally, we will propose the adaptive weighting selection based cost aggregation which tries to approximate the optimal weighting, together with a fast algorithm for weighting cost computation.

A. The optimal adaptive weighting model

As mentioned above, the weighting coefficients are highly related with the depth value of each pixel. Based on the assumption that the ground truth depth maps are given, we propose the 'optimal' adaptive weighting model:

$$\begin{aligned} &W(p,q) = [1 - occ(q)] \cdot \exp[-\alpha \cdot DS(p,q)] \\ &where \\ &occ(q) = \begin{cases} 1 & \text{if pixel } q \text{ is occluded} \\ 0 & \text{otherwise} \end{cases} \\ &DS(p,q) = \| d(p) - d(q) \| \end{aligned} \tag{2}$$

Here p is the center pixel of the window. And we want to calculate the weighting coefficient for each pixel q around p in the window. Based on the assumption that the ground truth disparity is given, we can indicate whether pixel q is occluded or not, then assign zero weight to those occlusion pixels. Meanwhile, by measuring the disparity similarity of pixel p and q by $DS(p,q)$, pixels share the same depth of center pixel will be assigned large weights while other pixels will have small weights.

B. Analysis of existing adaptive weighting methods

The proposed optimal adaptive weighting model can generate accurate weights for pixels according to their disparities. However, before stereo matching, there are no depth maps. a lot of researches tried to utilize the color and spatial information [4], [10], [12], [13] to approximate the depth correlation between pixels. Those approximations are reasonable in some ways: pixels with similar color are usually within the same object; and adjacent pixels usually have large correlation so large weights are given, vise versa.

However, the problems of the approximations also exist: pixels with different color could also be in the same object, and occlusion region can not be indicated directly because of the lack of the depth information. In Wang and Rachna's papers [12], [13], they did not consider the occlusion problem. The weighting coefficients are calculated by one image without considering another image. When calculating the left disparity map, the weighting coefficients are calculate based on the color and spatial correlation of the left image:

$$W(p,q) = Cr_{_L}(p,q) = exp[-\alpha \cdot CS_L(p,q) - \beta \cdot dis_L(p,q)] \tag{3}$$

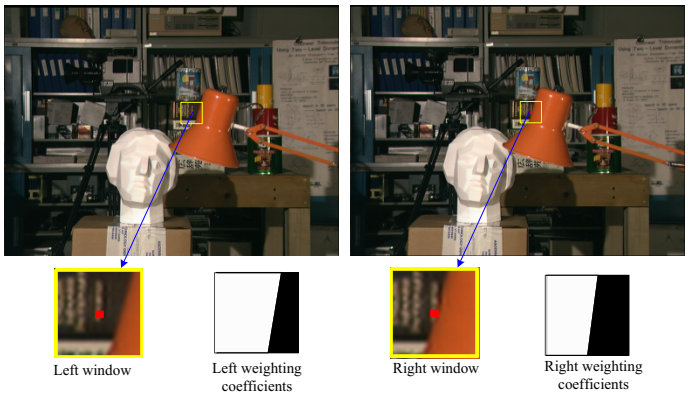


Fig. 1: The weighting coefficients for left & right window

$Cr_{_L}(p,q)$ measures the correlation of pixel p and q by their color similarity $CS(p,q)$ and spatial distance $dis(p,q)$ in the left image. In most of the case, the two terms color similarity and spatial distance can well approximate the disparity similarity term $DS(p,q)$. However, this method did not deal with the occlusion problems. As shown in Fig.1, the center pixels of the two blocks are corresponding pixels. White pixels in the weighting coefficients window have large weight, vice versa. In [12], [13], when estimating the left disparity map, only left weighting coefficients window are chosen for the weighted SAD by Eqn.1. When the window contains the occlusion region, the matching cost is very large even if the two windows are well matched. And the window with larger disparity in the right image will usually have smaller cost because of the similarity of background region. So in this case, even the center pixel is not occluded, the foreground disparity still gives a smaller matching cost, which causes the foreground fatten phenomenon around the occlusion region.

Due to the foreground fatten phenomenon, most of the high performance cost aggregation algorithms [4], [10], [14] exclude the occlusion regions by multiplying left and right weighting coefficients.

$$W(p,q) = Cr_{_L}(p,q) \cdot Cr_{_R}(p,q) \tag{4}$$

In one hand, when multiplying the two weighting coefficient windows, only the pixels having large weights in both of the views can survive and remain as the support pixels for cost aggregation. The occlusion pixels will be excluded by the multiplication. In the other hand, some of the 'good' support pixels are also removed. In the high texture regions, both of the weighting coefficients windows contain very few support regions. And the support regions are further removed because of the multiplication. Wrong disparities will be produced because too few support pixels do not contain enough texture information for correspondence search. Another problem with multiplication based algorithm is the computation complexity. For each disparity, the weighting coefficients need to be recalculated, which increase the computation complexity. And fast recursive based technologies can not be applied in this case because of the inconsistence of the weighting coefficients.

C. Adaptive Weighting Selection

In order to exclude the occlusion pixels as well as preserve sufficient support pixels, we propose the following aggregation algorithm based on adaptive weighting selection:

$$E_L^{Cr_L}(p,d) = \frac{\sum_{q\in\Omega(p)} Cr_L(p,q)DSI_L(q,d)}{\sum_{q\in\Omega(p)} Cr_L(p,q)} \quad (5)$$

$$E_L^{Cr_R}(p,d) = \frac{\sum_{q\in\Omega(p)} Cr_R(p-d,q-d)DSI_L(q,d)}{\sum_{q\in\Omega(p)} Cr_R(p-d,q-d)} \quad (6)$$

$$E_L(p,d) = \min(E_L^{Cr_L}(p,d), E_L^{Cr_R}(p,d)) \quad (7)$$

$E_L^{Cr_L}$ aggregates the cost for left image with the left correlation window which is defined in Eqn. 3. And $E_L^{Cr_R}$ measures the left cost by the right correlation window. The final matching cost for left image is selected by the minimum of the two matching costs by different weighting coefficients. The idea is that, occlusion regions only exist in one of the image. In other words, at least one of the images does not contain the occlusion regions and can be used to generate a good support for matching. Taking Fig. 1 for example, the left weighting coefficients window assigns large weight to the occlusion region, so the aggregated cost is large. However, the right weighting coefficients window does not contain the occlusion region and can produce a small matching cost. In this case, right weighting coefficients window is chosen. Compare with multiplication based algorithm in Eqn.4, our algorithm can not only exclude the occlusion regions, but also preserve a larger support region. It reduces the foreground fatten phenomenon and is more accurate in the high texture regions. Meanwhile, fast recursive algorithm can also used to reduce the complexity of the aggregation.

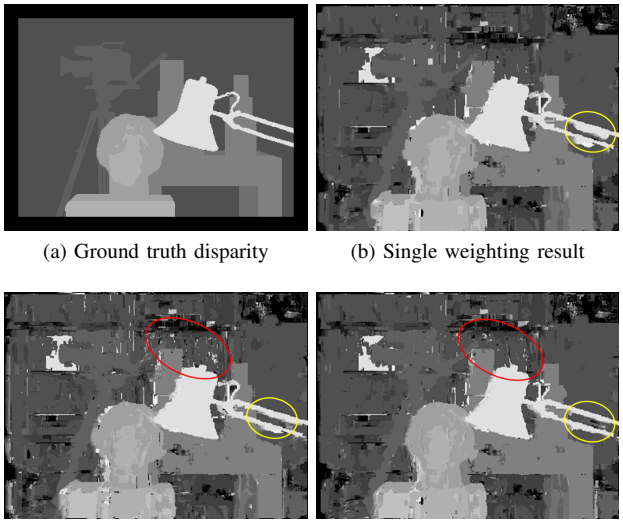
Fig. 2 shows the cost aggregation results for three different approaches: single weighting algorithm, multiplication weighting algorithm and our proposed weighting selection method. The disparity maps are calculated by winner take all without taking any disparity refinement. As shown in Fig. 2b, the regions marked by the yellow circles contains the foreground fatten phenomenon, while 2c and 2d have more accurate results. Meanwhile, compare with 2c, our result in 2d are more accurate in the high texture regions (marked by the red circles). Here the correlation is measured by segmentation label. $label_L(p)$ is the segmentation label for pixel p in the left image. If pixel q is in the same segment of center pixel p, then the weight is 1, otherwise the weight is 0. In our algorithm, mean-shift based segmentation [15] is utilized.

$$Cr_L(p,q) = (label_L(p) == label_L(q)) \quad (8)$$

$$Cr_R(p,q) = (label_R(p) == label_R(q)) \quad (9)$$

D. Symmetric Cost Mapping

In our proposed weighting selection algorithm, for each pixel, we need to aggregate twice: one by left weighting coefficients and another one by right weight coefficients. In this part, we will prove the symmetric properties of the matching cost between left image and right image. Half of the complexity can be reduced by proposed symmetric cost mapping.



(a) Ground truth disparity (b) Single weighting result

(c) Multiplication weighting result (d) Weighting selection result

Fig. 2: WTA result for 'tsukba'(without disparity refinement)

During the stereo matching, we first calculate left weighting cost which is selected by $E_L^{Cr_L}(p,d)$ and $E_L^{Cr_R}(p,d)$. After that, the right weighting cost is derived by weighting selection of $E_R^{Cr_L}(p,d)$ and $E_R^{Cr_R}(p,d)$. The definition of the right weighting cost is followed:

$$E_R^{Cr_L}(p,d) = \frac{\sum_{q\in\Omega(p)} Cr_L(p+d,q+d)DSI_R(q,d)}{\sum_{q\in\Omega(p)} Cr_L(p+d,q+d)} \quad (10)$$

$$E_R^{Cr_R}(p,d) = \frac{\sum_{q\in\Omega(p)} Cr_R(p,q)DSI_R(q,d)}{\sum_{q\in\Omega(p)} Cr_R(p,q)} \quad (11)$$

Property 1: $E_R^{Cr_L}(p,d) = E_L^{Cr_L}(p+d,d)$

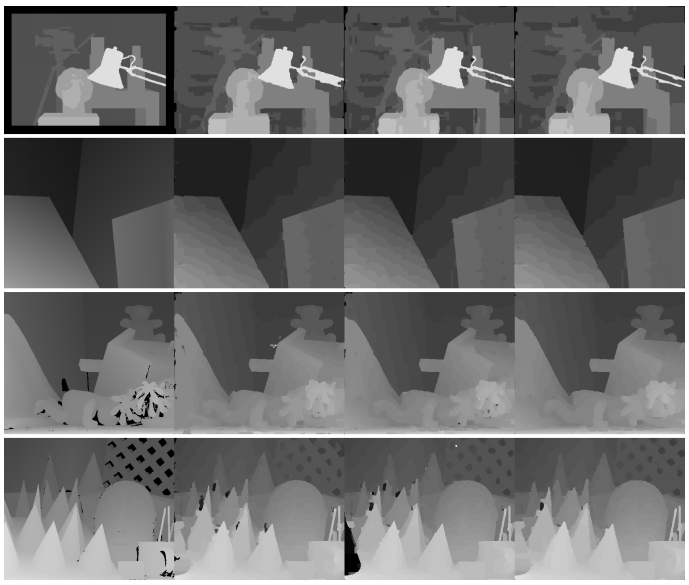
The property can be proved easily from the definition. Property 1 means that, we do not need to perform another cost aggregation to calculate $E_R^{Cr_L}(p,d)$. In fact, it can be mapped by $E_L^{Cr_L}(p,d)$ directly. Another property can be derived by similar proof.

Property 2: $E_L^{Cr_R}(p,d) = E_R^{Cr_R}(p-d,d)$

In the implementation, the cost $E_L^{Cr_L}(p,d)$ and $E_R^{Cr_R}(p,d)$ are first computed by cost aggregation. Then $E_L^{Cr_R}(p,d)$ and $E_R^{Cr_L}(p,d)$ are generated by cost mapping with neglectable computation complexity. Compare with brute force method which needs four aggregations, our algorithm only calculates the aggregation cost twice and another two are derived by proposed symmetric cost mapping. Half of the complexity are reduced. Therefore, by the proposed weighting selection strategy, a more accurate support is utilized to exclude the occlusion problem as well as preserve enough texture for correspondence search. Meanwhile, the complexity of the algorithm is reduced by half by proposed cost mapping.

TABLE I: Performance Comparison of the Proposed Method

	Tsukuba			Venus			Teddy			Cones		
	nonocc	all	disc.	nonocc	all	disc.	nonocc	all	disc.	nonocc	all	disc.
Selection based	**1.12**	**1.47**	**5.98**	**0.14**	**0.30**	**1.58**	**4.75**	**8.70**	**12.8**	**2.66**	**9.06**	**7.54**
Single based	1.85	2.37	9.91	0.34	0.56	3.48	6.08	11.5	16.2	3.75	10.4	9.96
multiplication based	1.36	1.64	6.93	0.18	0.34	1.87	5.33	9.45	14.7	3.17	9.68	8.98



(a)Ground truth (b)Single based (c)Multiplication based (d) Selection based

Fig. 3: Results for Tsukuba, Sawtooth, Venus, and Map image

III. Experimental results

In this section, we present the experimental results of proposed algorithm on the Middlebury benchmarks. Graph cuts [7] and plane fitting [16] are utilized to generate the final disparity maps after cost aggregation. And the disparity similarity image DSI is calculated by truncated absolute difference which is more robust to the noise. The results are compared with existing methods: single weighting based aggregation [12] and multiplication based aggregation [10]. (The weighting coefficients are modified into segmentation based weight shown in Eqn.8 and Eqn.9 to ensure to fairness of the comparison.) As shown in Fig. 3, significant improvements can be clearly noticed in the occlusion regions: the single weighting method have the problem of foreground dilation. Compare with multiplication based algorithm, some wrong disparities in the high texture regions are removed.

Tab. I shows the error percentages with regards to the ground truth. Three different regions are measured: non-occluded region, all region and dis-occlusion region. Our selection based method outperforms the existing single weighting and multiplication algorithms.

IV. Conclusion

In this paper, we propose a novel cost aggregation algorithm which based on adaptive weighting selection. Compare with existing methods, our method can not only exclude the occlusion regions but also preserve a sufficient support region. More accurate results can be derived. Meanwhile, we also prove the symmetric property of the aggregation cost and reduce the complexity by half. In the future, we will investigate the recursive algorithm to further speed up the algorithm.

V. Acknowledgement

This work has been supported in part by the Research Grants Council (GRF Project no. 610210) and the Hong Kong Applied Science and Technology Research Institute Project (ASTR52-20A00210/11PN)

References

[1] L. Xu, O. Au, W. Sun, Y. Li, S. H. Chui, and C. W. Kwok, "Image rectification for single camera stereo system," in *Image Processing. IEEE International Conference on*, 2011.

[2] Y. Li, O. Au, L. Xu, W. Sun, S.-H. Chui, and C.-W. Kwok, "A convex-optimization approach to dense stereo matching," in *Image Processing, IEEE International Conference on*, 2011.

[3] D. Scharstein, R. Szeliski, and R. Zabih, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," in *Stereo and Multi-Baseline Vision. IEEE Workshop on*, 2001.

[4] Y. Kuk-Jin and K. In-So, "Adaptive support-weight approach for correspondence search," in *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 2006.

[5] H. Li and G. Chen, "Segment-based stereo matching using graph cuts," in *Computer Vision and Pattern Recognition. IEEE Computer Society Conference on*, 2004.

[6] W. Sun, O. Au, L. Xu, S. H. Chui, C. W. Kwok, and Y. Li, "Error compensation and reliability based view synthesis," in *Acoustics, Speech and Signal Processing. IEEE International Conference on*, 2011.

[7] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," in *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 2001.

[8] A. F. Bobick and S. S. Intille, "Large occlusion stereo," in *International Journal of Computer Vision*, 1999.

[9] M. Okutomi and T. Kanade, "A locally adaptive window for signal matching," in *International Journal of Computer Vision*, 1992.

[10] A. Hosni, M. Bleyer, M. Gelautz, and C. Rhemann, "Local stereo matching using geodesic support weights," *Image Processing, IEEE International Conference on*, 2009.

[11] F. Tombari, S. Mattoccia, L. Di Stefano, and E. Addimanda, "Near real-time stereo based on effective cost aggregation," *Pattern Recognition. International Conference on*, 2008.

[12] D. Wang and K. B. Lim, "A new segment-based stereo matching using graph cuts," in *Computer Science and Information Technology. IEEE International Conference on*, 2010.

[13] Rachna, H. Singh, and A. K. Verma, "Segment controlled window shape to compute disparity map from stereo images," *IJCA Special Issue on Electronics, Information and Communication Engineering*, 2011.

[14] Q. Yang, L. Wang, R. Yang, H. Stewenius, and D. Nister, "Stereo matching with color-weighted correlation, hierarchical belief propagation, and occlusion handling," in *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 2009.

[15] D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis," in *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 2002.

[16] L. Xu, O. Au, W. Sun, Y. Li, and J. LI, "Hybrid plane fitting for depth estimation," in *APSIPA Annual Summit and Conference*, 2012.