Dense Stereo Matching Method Based on Local Affine Model

Jie Li
School of Electronic Information, Wuhan University, Wuhan, China
Email: jielonline@gmail.com

Wenxuan Shi, Dexiang Deng
School of Electronic Information, Wuhan University, Wuhan, China
Email: shiwxdsp@gmail.com, whuddx@gmail.com

Wenyan Jia
Department of Neurosurgery, University of Pittsburgh, Pittsburgh, PA, USA
Email: jiawenyan@gmail.com

Mingui Sun
Departments of Neurosurgery, Electrical engineering and Bioengineering, University of Pittsburgh, USA
Email: drsun@pitt.edu

***Abstract*—A new method for constructing an accurate disparity space image and performing an efficient cost aggregation in stereo matching based on local affine model is proposed in this paper. The key algorithm includes a new self-adapting dissimilarity measurement used for calculating the matching cost and a local affine model used in cost aggregation stage. Different from the traditional region-based methods, which try to change the matching window size or to calculate an adaptive weight to do the aggregation, the proposed method focuses on obtaining the efficient and accurate local affine model to aggregate the cost volume while preserving the disparity discontinuity. Moreover, the local affine model can be extended to the color space. Experimental results demonstrate that the proposed method is able to provide subpixel precision disparity maps compared with some state-of-the-art stereo matching methods.**

***Index Terms* — Local affine model, ZNCC, edge-preserving, occlusion handling**

I. INTRODUCTION

Binocular stereo vision infers 3D scene geometry from two images with different viewpoints. Dense stereo matching as the key in the binocular stereo vision is one of the most active research topics estimating disparity information from different views. However, dense stereo matching remains a difficult vision problem for the following reasons: (1) Noise. Unavoidable light variations, image blurring, and sensor noise exist in image formation cause the image polluted by the noise. (2) Textureless regions. These regions lack useful image information, which leads the intensity consistency constraint useless. (3) Disparity discontinuities. Since disparity discontinuities often appear at object boundaries, the smoothness constraint would be broken in these boundaries. However, accurate object boundaries detecting is difficult. (4) Occlusions. Some pixels appear in one view may be occluded in another view. How to give a reasonable disparity value to these pixels is worth considering.

A good stereo matching algorithm should be robust, which could filter the noise, handle occlusion, and smooth the disparity in the textureless region while preserving the disparity discontinuities at the boundaries. In general, dense stereo matching can be broadly classified into two categories, local and global [1], depending on whether they rely on local window-based computations or the minimization of a global energy function. The global methods consider the correlation of disparities in the neighborhood. The correlation is usually modeled as smoothness prior function. And the joint energy function is globally minimized to find the disparity. These approaches are related to Markov random field (MRF) models and energy minimization methods, such as graph cuts [2], belief propagation [3]. Global methods in general give impressive results than the local ones, especially when they are combined with segmentation [4] and plane fitting [5]. However, most global methods are computationally expensive and sometimes the parameters in the methods are hard to determine. In local-based methods, the disparity at a given point depends only on intensity values within a finite window. Despite of this drawback, the speed and parallelism advantage of local techniques keeps research in local techniques thriving. Moreover, the observation of [6] proves that local method is also able to achieve high quality result.

Many effective local algorithms for stereo matching have been reported on [1]. These traditional techniques assume that all pixels within a fixed window have the same disparity. Clearly, fixed window often doesn't work well for the whole image, especially on the boundaries. In general, a reliable window in local algorithm should be

Manuscript received October 12, 2012; revised November 17, 2012; accepted November 20, 2012.
Corresponding Author: Wenxuan Shi, shiwxdsp@gmail.com.

doi:10.4304/jcp.8.7.1696-1703

large enough to include enough intensity variation for reliable matching but small enough to cover only pixels with equal disparity. To properly deal with this dilemma, local methods generally adopt approaches by adaptive windows [6, 8], where different window sizes are applied to different pixels in the same image. Therefore, when calculating the matching cost for each region, the specific window has to be determined by examining the characteristics of edges, where sharp changes in intensity appear.

To obtain a high-quality disparity map, a number of local stereo matching methods [1] have been proposed by defining the support weight, which can implicitly measure the similarity of disparity values between two candidate pixels. These methods use the assumption that neighboring pixels with similar intensity/color should have similar disparities. By using this local support windows [6, 8, 9], the output at a pixel is a weighted average of nearby pixels, whose weights depend on the intensity/color similarities. The dilemma mentioned above is reduced effectively while the discriminative power of the dissimilarity measurement is increased. These methods, which smooth small fluctuations and preserve edges in a local window, could perform good results. Although some simplified adaptive weight methods [10, 11] have been proposed recently, the aggregation accuracy would degenerate. The proposed algorithm is organized as follows: First, an improved dissimilarity measurement is used to construct a disparity space image (DSI) by a left/right input. Then, in order to get a more reliable cost volume, a local affine model is used when the DSI satisfies the zero-mean normalized cross correlation (ZNCC) condition [12], which is determined by the neighborhood information. After that, Winner-Take-All (WTA)[1] is presented to obtain a disparity map. We can also get another disparity map using the other (right/left) input as the reference by the same method. After left-right checking [1], the occlusion regions and mismatching regions from two disparity maps (left and right as references, respectively) are determined. Finally, a proposed weighted median filter is performed to fill the occlusion regions and unstable regions.

II. ALGORITHM

Assume a rectified stereo pairs, and the local stereo algorithms generally perform the following four steps: matching cost computation, cost aggregation, disparity computation and disparity refinement [1]. Accordingly, our algorithm can be partitioned into four blocks. In this section, we give a detailed description of the building blocks outlined below (seen in Fig.1).

a. Matching Cost Computation

Typically, the matching cost, which constructs the DSI, is computed at each pixel for all disparities under consideration. Heiko *et al* [13] gave an evaluation for some popular matching cost methods, and the Census method gives the best result. However, Census method uses Census transformation to get a group of binary code flow, and it uses the color weight and distance weight to complete weight calculating. It uses the double channel method to accumulate the matching cost, which seems complicate. A sample linear combination performs very well in [14], which linearly combines absolute intensity differences (SAD) and a gradient for computing matching cost. We consider the difference between the color spaces, and use the modified self-adapting dissimilarity measurement to construct DSI. Assuming that C_{CSAD} is the absolute color differences and the color gradient difference is C_{CGRAD}. The matching cost $C(x,y,d)$ on pixel (x,y) in the disparity d can be defined as follows:

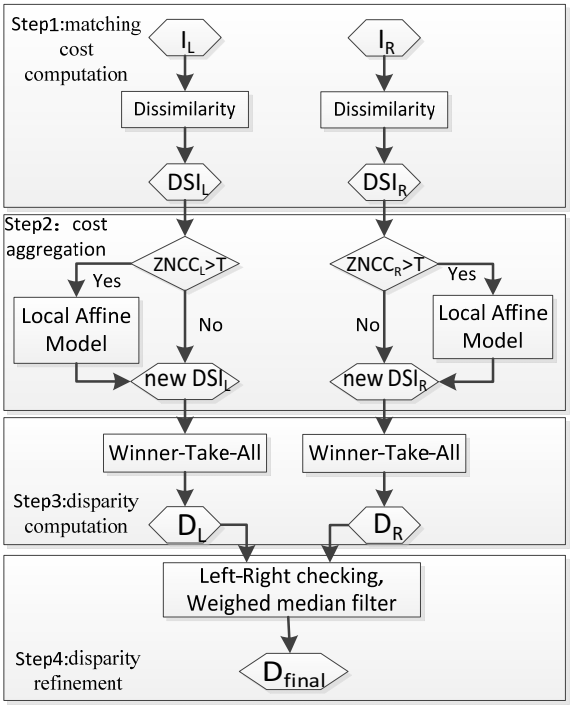


Figure1. Block diagram of the proposed method

$$C(x,y,d) = (1-\alpha)C_{CSAD}(x,y,d) + \alpha C_{CGRAD}(x,y,d) \quad ,$$

$$C_{CSAD}(x,y,d) = \sum_{(i,j)\in\Omega(x,y)} min\left(\frac{\left(\sum_{c\in\{r,g,b\}}|I_l^c(i,j)-I_r^c(i+d,j)|\right)}{3}, \tau_c\right) \quad ,$$

$$C_{CGRAD}(x,y,d) = min\left(C_{grad}^r(x,y,d), C_{grad}^g(x,y,d), C_{grad}^b(x,y,d), \tau_g\right) \ ,$$

$$C_{grad}^c(x,y,d) = \sum_{(i,j)\in\Omega(x,y)}|\nabla_x I_l^c(i,j) - \nabla_x I_r^c(i+d,j)| + \left|\nabla_y I_l^c(i,j) - \nabla_x I_r^c(i+d,j)\right|, \quad c=r,\ g,\ b. \tag{1}$$

where α balances the color and gradient terms. $\nabla_x I_l^c$ and $\nabla_y I_l^c$ represent the horizontal and vertical gradient in the c color channel. In order to decrease the influence of noises, a small surrounding window $\Omega(x,y)$ centered on (x,y) is added to calculating the matching cost. τ_c and τ_g

are truncation values. A DSI is constructed by all the *C(x,y,d)* calculated from (1) .

b. Cost Aggregation Based on Local Affine Model

The aggregation step aims to aggregate each pixel's matching cost over a weighted region to reduce the matching ambiguities and noises in the initial cost volume. The accuracy of local stereo matching methods is highly dependent on the cost aggregation schemes used.

As the introduction part said, an ideal cost aggregation strategy, to deal with disparity discontinuities and ambiguous regions (textureless areas, repetitive patterns, etc), should modify the support window at each position according to image content to include only those points with the same (unknown) disparity. The key of our stereo matching aggregation step is based on the local affine model, which has been proven useful in image matting [15, 16] and image denosing [17, 18]. Assume there are two small local matching regions Ω_P, Ω_Qcentered on pixel P and pixel Q in the two rectified images L and R, respectively. An affine relation between the intensities of the local matching regions is described as:

$$R_{k'} = M_p L_k + B_p \quad , k \in \Omega_P, k' \in \Omega_Q \quad , \tag{2}$$

where M_p and B_p are linear coefficients which are constant in Ω_p. Assume the left image is the reference image, we can get the disparity information from $D_k = R_{k'}$-L_k. Combing with (2), D_k can be obtained by:

$$D_k = R_{k'}\text{-}L_k = M_p L_k + B_p\text{-}L_k = A_p L_k + B_p \quad , \tag{3}$$

where $A_p = M_p - 1$. Equation (3) gives the relationship between the disparity map D and input image L in a local region Ω_p. Additionally, the local affine model satisfies $\nabla D_k = A_p \nabla L_k$, which ensures that the image D has an edge if image L has an edge. In this way, the disparity discontinuities are also the edges of the reference image.

In order to getting A_P and B_P , we minimize the difference between input p_k and the output D_k using square optimal estimator in a local window,

$$E(A_P, B_P) = \\ min\left(\sum_{k \in \Omega_p} ((A_\mathrm{P} I_k + B_\mathrm{P} - p_k)^2 + \varepsilon A_\mathrm{P}^2) \right) \quad , \tag{4}$$

where p_k is the pixel in the window Ω_p in an initial input disparity map denoted by $\widehat{D}$. ε is set to prevent A_p from being too large. Guided filter (GF) proposed by [17] uses the least-squares estimate to solve the function (4), then A_p and B_p are calculated as follows:

$$A_p = \frac{\frac{1}{|\Omega_\mathrm{p}|}\sum_{k \in \Omega_p} I_k p_k - \mu_p \overline{p_p}}{\sigma_k^2 + \epsilon} \quad , \\ B_p = \overline{p_p} - A_p \mu_p \quad , \tag{5}$$

where μ_p and σ_k^2 are the mean and variance of input image *I* in window Ω_p. $|\Omega_p|$ is the number of pixels in Ω_p. $\overline{\mathrm{p_p}}$ is the mean of pixels in Ω_p. Since p is included in many windows centered by its neighborhood, correspondingly, there are many A_p and B_p in different windows including p. Hence, GF simply average all possible value of D_k as:

$$D_k = \bar{A}_k I_k + \bar{B}_k \quad , \tag{6}$$

where $\bar{A}_k$ and $\bar{B}_k$ are defined as follows:

$$\bar{A}_k = \frac{1}{|\Omega_k|}\sum_{p \in \Omega_k} A_p \quad , \ \bar{B}_k = \frac{1}{|\Omega_k|}\sum_{p \in \Omega_k} B_p \quad , \tag{7}$$

where Ω_k centered at pixel k. GF makes a strong assumption that a single linear model and a solid window size are sufficient to model a local patch. If the window Ω_k in image *I* is in a flat patch, and the solution of (5) is: $A_k \approx 0$, and $B_k \approx \overline{p_k}$, then the output $\mathrm{D_k}$ is smooth clearly. If *I* changes a lot in Ω_k, then (5) is solved by $A_k \approx 1$ and $B_k \approx 0$.

However, when consider the case in Fig. 2, the texture in the object is much different from some neighborhood regions. GF simply aggregates all the multipoint estimates for p. This solution makes $D_k < \bar{A}_k \nabla L_k$, because $\bar{A}_k$is the output of an average filter which makes the gradients much smaller than those gradients of strong edges in the reference image *I*.

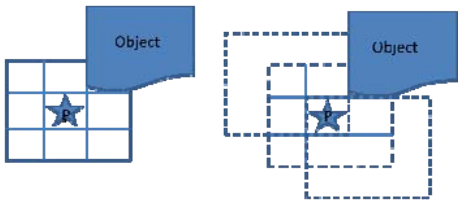


Figure 2. Variances appear in local windows

In fact, strong smoothing for pixel p is achieved only when it is contained in a window without involving pixels in the "Object" shown in Fig. 2. Therefore, it is necessary to check the validity before using this model (seen in (2)), since our local model is assumed to be built on all the local matching regions. The zero-mean normalized cross correlation (ZNCC) is used as a good measurement for calculating the affine similarity in the proposed method, since [14] mentions that ZNCC is insensitive to the affine transformations of image intensity values of two images. The maximal absolute value "1" indicates two signals are related by an affine transformation. Due to the same affine property, if the left image and the right image patch are local affine, it is the same to the left image and the disparity map. Hence, the correlation of point *(i,j)* in image I_l and the point *(i+d, j)* in image I_r is defined to be:

$$p(x,y)_{zncc} = \\ \frac{\sum_{(i,j)\in\Omega(x,y)}(I_l(i,j)-\bar{I}_l(x,y))(I_r(d+i,j)-\bar{I}_r(d+x,y))}{\sqrt{\sum_{((i,j)\in\Omega(x,y)}(I_l(i,j)-\bar{I}_l(x,y))^2(I_r(d+i,j)-\bar{I}_r(d+x,y))^2}} \quad , \tag{8}$$

where $\overline{I_l}(x,y)$ and $\overline{I_r}(d+x,y)$ are the means of pixel for a given windows centered at (x,y). We also use the multiple windows [19] to calculate all the ZNCC value including *p(x,y)*. Initialize a matrix Z, which has the same size with the window Ω_k. If the correlation is bellow some threshold, e.g. $\left|\left|p(i,j)_{zncc}\right|\right| < T$, we set *Z(i,j)=0*; otherwise *Z(i,j)=1*. The sum of pixels with Z(i,j)=1 in the reference image is *length(Z)*. Hence, the final possible value of D_k is defined as:

$$D_k(x,y) = \frac{1}{length(Z)}\sum_{(i,j)\in\Omega_N} Z(i,j)(A_k(i,j)I(x,y) + B_k(i,j)). \quad (9)$$

Alternatively, the linear model in (2) can be replaced by a vector model:

$$D_k(x,y) = \sum_{i\in\Omega_k}(\boldsymbol{A_k^T I_i} + B_k) \ , \quad (10)$$

where the black body $\boldsymbol{I_i}$ and $\boldsymbol{A_k}$ means a 3×1 vector, and $\boldsymbol{A_k^T I_i} = A_k^r I_i^r + A_k^g I_i^g + A_k^b I_i^b$.

In the proposed algorithm, (9) is used to aggregate the matching cost. Correspondingly, the DSI is updated by a new DSI.

c. Disparity Computation

Additionally, we compute the DSI for the right image as a reference image in a similar manner. When two disparity space images are obtained, the WTA is applied to form two disparity maps: D_R and D_L. However, a stable pixel should complete the mutual consistency checking, which requires the pixels in the left and right disparity maps, on the pixel grid, are perfectly consistent (same disparity value). This is performed by a subsequent mutual consistency checking (often called left- right checking) that divides all the pixels into stable or unstable pixels:

$$D_L(x_l) - D_R(x_l - D_L(x_l)) \le T_{lf} \ . \quad (11)$$

We mark pixels in the left disparity map as occluded pixels if the (11) does not hold. T_{lf} is a threshold.

d. Disparity Refinement

Suppose the unstable pixels are always the occluded pixels. The occluded pixels are assigned to the lowest disparity value of the spatially closest non-occluded pixels which lie on the same scanline, since the images are rectified precisely. This simple occlusion filling strategy can generate streak-like artifacts in the disparity map. According to the assumption, neighboring pixels with similar colors should have similar disparities. Correspondingly, we use a weighted median filter to fill all the pixels.

The relationship among reference image *I* and output disparity map *D* are given by (5) and (6). These can be represented in weight form like bilateral filter:

$$D_k = \frac{1}{|\Omega_k|^2}\sum_{i\in\Omega_k} W_{ik}(I)p_i \quad , \quad (12)$$

where $W_{ik}(I)$ is defined as :

$$W_{ik}(I) = \frac{1}{|\Omega_k|^2}\sum_{i\in\Omega_k}(1 + \frac{(I_i-\mu_k)(I_k-\mu_k)}{\sigma_k^2+\epsilon}) \ . \quad (13)$$

In disparity refinement step, when every pixel's weight in the certain window Ω_k is obtained, we use (12) to calculate the refined disparity value.

To reduce the discontinuities caused by the quantization in the disparity, a subpixel precision method is proposed, which can be achieved by:

$$d_f = d_c + 0.5 * \frac{C(x,y,d_c-1)-C(x,y,d_c+1)}{C(x,y,d_c-1)+C(x,y,d_c+1)-2C(x,y,d_c)} \ . \quad (14)$$

Since a disparity map has been produced from the aggregation step and the matching cost of the three candidate matching costs: $C(x,y,d_c-1)$, $C(x,y,d_c+1)$ and $C(x,y,d_c)$ are obtained from the aggregated disparity space image after the aggregation step, the enhanced disparity map can be easily acquired. Finally, we replace each disparity value with the average of those values over a 3×3 window.

III. EXPERIMENTAL RESULT

In this section, we evaluated our approach on the Middlebury stereo benchmark [20] using four stereo pairs named *Tsukuba*, *Venus, Teddy* and *Cones*. Experiments include the following parts: parameters setting, evaluation and comparing results with several classical local-based stereo matching methods.

The results are presented in Fig.3. Note that the parameters are constant across all datasets. In order to improve the stability, we normalize the pixel intensity in [0,1].

a. Parameter Setting

Cost computation: In our algorithm, the computation of matching cost is inherently parallel at each pixel and each disparity level. The parameters in (1), which depend on the size of the occluded regions and the signal-to-noise ratio, are set as follows: τ_c=0.0028, τ_g=0.016. The balance value α is 0.92. Ω is chosen to be a 3×3 window.

Cost aggregation: Although (10) can be used to get higher accuracy in some scenes, however, considering the computation time used in the color space is more than intensity space, and the average result in intensity is also a good, we use (9) to do the aggregation. The parameters are set as follows: Ω_k is a 15× 15 window; ϵ =0.0009.

Occlusion detection and filling: Because error in computation is existed, we choose T_{lf} to be 2 in (11). The $median\ filter$ step in (12) and (13), we choose Ω_k as a 5×5 window, $\sigma_k = 5, \epsilon$=0.0001;

b. Evaluation

A quantitative way is needed to evaluate the performance of a stereo algorithm. There are two kinds of quality measures [1], with respect to the ground truth data, to compute error statistics:

RMS error (root-mean-squared): It was measured in disparity pairs of the estimated disparity map $d_c(x, y)$ and the ground truth of disparity map $d_T(x, y)$:

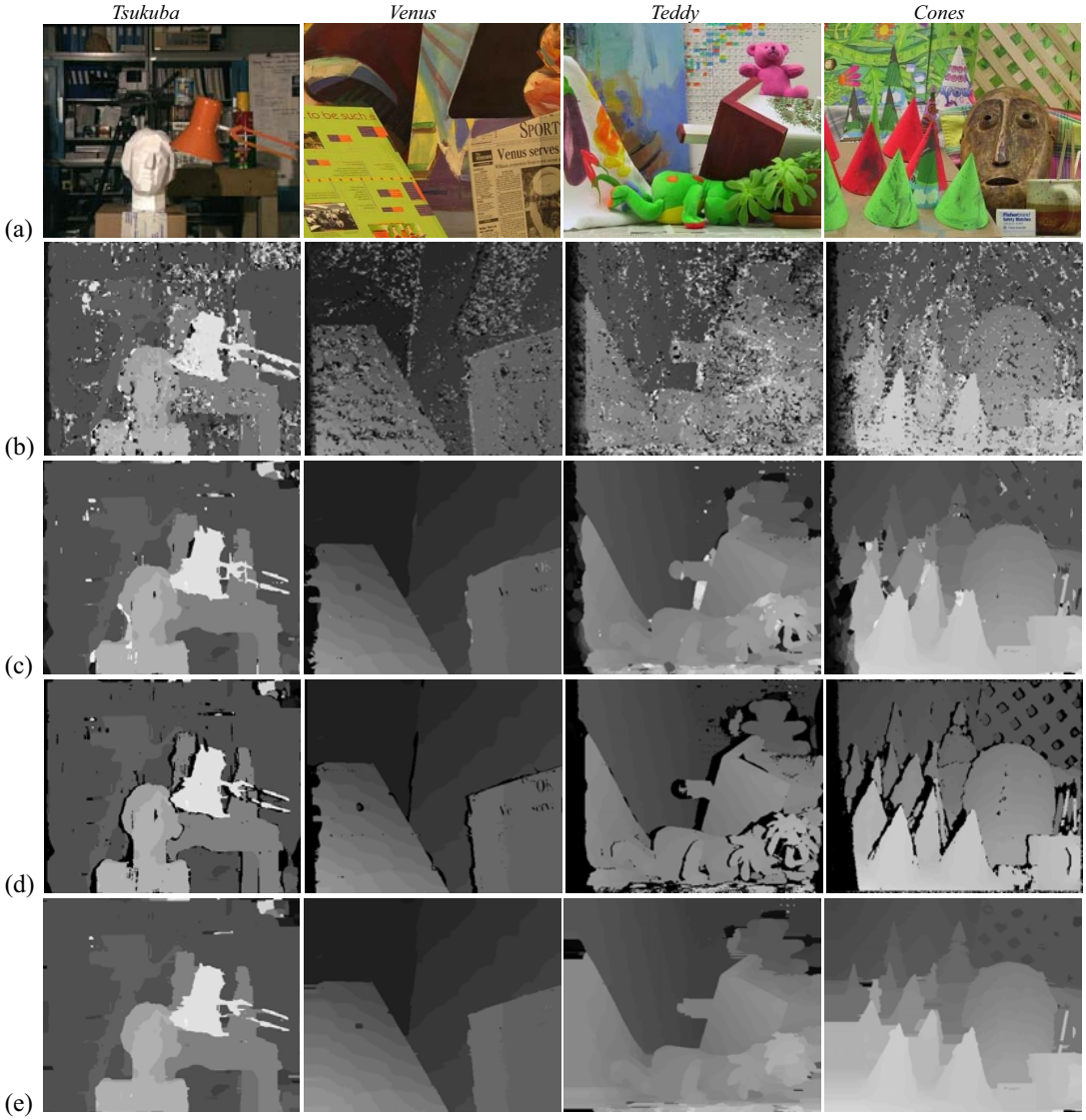
$$RMS = \sqrt{\frac{1}{N}\sum_{(x,y)} |d_c(x,y) - d_T(x,y)|^2} \quad , \qquad (15)$$

where N is the total number of matching pixels.

Percentage of bad matching pixels:

$$B = \frac{1}{N}\sum_{(x,y)}(|d_c(x,y) - d_T(x,y)| > \delta_d) \quad , \qquad (16)$$

where δ_d is the disparity error tolerance. And we used the percentage of bad matching pixels to evaluate the proposed algorithm on the Middlebury stereo benchmark with error threshold δ_d =0.5. In this paper, we show error percentages in three different regions: nonoccluded regions (***nocc)***, discontinuity regions (***disc)*** and either nonoccluded or half-occluded (denotes as ***all***). Then, we consider the average error in the whole disparity map. The disparity maps generated after each step of the proposed method using Middlebury data sets are presented in Fig. 3.



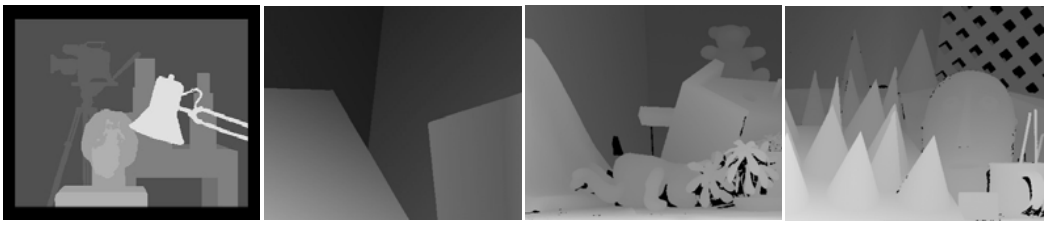


Figure 3. Intermediate results of the four different standard images by our algorithm compared to the ground truth. (a) Sample image of the stereo image pairs(the left image);(b) The WTA disparity maps after self-adapting dissimilarity measurement; (c)The WTA disparity maps after local affine model; (d) The disparity maps after applying left-right checking ;(e) The disparity maps after filling;(f) ground truth.

c. Comparing

We compared our results with the results of adaptive local method [6], fast biltalter method [21] and non-local aggregation method [22]. The results are given in TABLE I. The boldfaces mean the rank in all about 140 algorithms in the Middlebury stereo benchmark. We can observe that our method performs very well in *Tsukuba,* including high texture parts (such as the lamp stem in Fig. 3.(e)) and textureless parts(such as the box and desk). This is because our method has the property of edges-preserving, while smoothing the relatively flat regions.

Reliable local affine model prevents the edge smoothing. However, obtaining the accurate local affine model is difficult, even the ZNCC is used as the judgment criterion. The reason is that the inequation $\nabla D_k < \bar{A}_k \nabla L_k$ still works in few high texture regions (seen in Table I ***disc***) in the proposed algorithm. It leads to certain details losing, which can be seen in *Tsukuba* of Fig. 3(c), the lamp holder breaks somewhere. This type of errors will decrease if the local window is chosen smaller.

When we use the same parameters in *Venus* and change our local affine model window to 19×19, the final percentage of bad matching pixels in ***nonocc, all*** and ***disc*** regions would decrease to 6.26, 6.83 and 13.0, respectively. That's because *Venus* has large plane regions, and the windows in plane region should be large.

In the left-right checking stage, all the disparity of unstable pixels are set to 0. However, some of these unstable pixels may be correct disparities. Since the noise causes inconsistency in the stable pixels, some inliers are identified as outlier by mistake. Hence, the percentage of bad matching pixels increases (seen in Table II). After filling step, some steak-like artifact (seen in *Cones* of Fig. 3(e)) still appears. That is because several unmatched regions were filled by the occlude region method. How to increase the accuracy in unstable region detecting is worthy considered.

We also performed some experiments by other images. The disparity maps after each step are presented in Fig. 4. Experimental results show that our algorithm is a strong and effective method.

From the (7), (8) and (9), we found that all the summations are calculated by box filter. Therefore, these terms can be computed in ***O***(N) time.

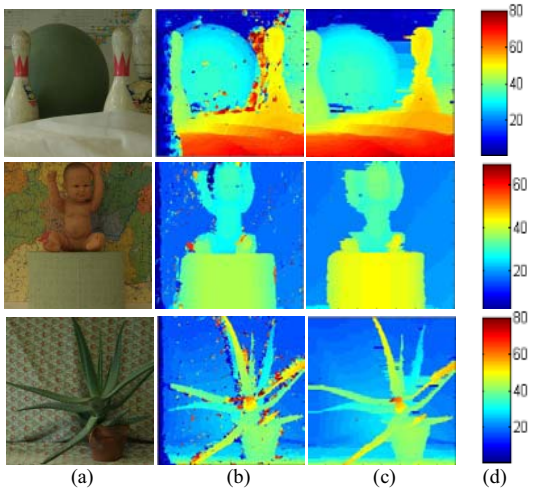


Figure 4. The intermediate results from proposed algorithm; (a) Sample image of the stereo image pairs (the left image); (b)~ (c)disparity maps after local affine model and filling;(d) color bar of the disparity map.

TABLE I Comparing results (δ_d=0.5)

Algorithm	**Avg. Error**	***Tsukuba***			***Venus***			***Teddy***			***Cones***		
		Nonocc	**all**	**disc**	**Nonocc**	**all**	**disc**	**Nonocc**	**all**	**disc**	**Nonocc**	**all**	**disc**
Adaptive local	18.1	18.1**(63)**	18.8**(61)**	18.6**(41)**	7.77	8.40	15.8	17.6	23.9	34.0	14.0	19.7	20.6
Fast biltalter	17.4	21.5**(85)**	22.4**(86)**	22.9**(88)**	5.71	6.66	14.9	16.2	23.3	32.1	9.10	15.8	18.1
Non-local	14.1	11.1**(25)**	11.6**(24)**	15.4**(13)**	8.64	9.11	14.0	13.5	19.9	25.3	9.40	15.0	16.6
Our method	15.0	10.2**(22)**	11.0**(21)**	20.0**(54)**	7.52	8.12	14.1	14.7	20.9	29.9	9.55	15.8	21.6

TABLE II The evaluation in each step (δ_d=0.5)

Algorithm	Avg. Error	*Tsukuba*			*Venus*			*Teddy*			*Cones*		
		Nonocc	all	disc	Nonocc	all	disc	Nonocc	all	disc	Nonocc	all	disc
Local affine	18.4	11.3	12.6	22.4	9.01	10.4	24.3	12.3	21.1	27.9	10.7	20.1	23.9
Left Right-checking	21.6	12.8	14.8	28.9	9.95	11.5	30.2	19.6	28.0	39.2	12.7	22.7	28.8
Filling	15.0	10.2	11.0	20.0	7.52	8.12	14.1	14.7	20.9	29.9	9.85	15.8	21.6

IV. CONCLUSION AND FUTURE WORK

A new local stereo matching method, which combines a new dissimilarity measurement and local affine cost aggregation model, is presented in this paper. In the post processing, two approaches are used, where occlusions are first detected, then they are filled in by neighboring areas. We did some tests on the Middlebury stereo benchmark, the experiment results showed that our proposed approach can produce good disparity maps. The computation time of the proposed algorithm is $\boldsymbol{O}$(N) while performing nice result in Middlebury data sets. However, various factors prevent the matching of pixel pairs in the real world, such as radiometric changing, camera device changing and illumination changing. Therefore, robust dissimilarity measurements and hardware implementation will be considered in order to overcome these problems in the future work.

ACKNOWLEDGMENT

This research was supported by the Chinese National Natural Science Foundation under grant No.61072135 and National Institutes of Health grants U01HL91736 and R01CA165255 of the United States.

REFERENCES

[1] D. Scharstein and R. Szeliski, “A taxonomy and evaluation of dense two-frame stereo correspondence algorithms”, *International journal of computer vision*, vol.47, no.1-3, pp.7-42, 2002.

[2] Y. Boykov, O. Veksler, and R. Zabih, “Fast Approximate Energy Minimization via Graph Cuts”, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.23, no.11, pp.1222-1239, 2001.

[3] J. Sun, Heung-Yeung Shum, and Nan-Ning Zheng, “Stereo Matching using Belief Propagation”, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.25, no.7, pp.787-800, 2003.

[4] J. Zeng, S. Yu, X. Fu, and C. Li, “A Line Segments Matching Method based on Epipolar-line Constraint and Line Segment Features”, *Journal of software*, vol.6, no.9, pp.1746-1754, 2011.

[5] Q. Yang, L. Wang, R. Yang, H. Stewenius, and D. Nister, “Stereo matching with color-weighted correlation, hierarchical belief propagation and occlusion handling”, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.31, no.3, pp.492-504, 2009.

[6] K. Yoon and I. Kweon, “Adaptive Support-Weight Approach for Correspondence Search”, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.28, no.4, pp.650-656, 2006.

[7] O. Veksler, “Stereo Correspondence with Compact Windows via Minimum Ratio Cycle”, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.24, no.12, pp.1654-1660, 2002.

[8] L. Li, C. Zhang, and H. Yan, “Stereo Matching Algorithm Based on a Generalized Bilateral Filter Model”, *Journal of software*, vol.6, no.10, pp.1906-1913, 2011.

[9] S. Paris and F. Durand, “A fast approximation of the bilateral filter using a signal processing approach”, *In Proceedings of European Conf. on Computer Vision*, pp.568-580, 2006.

[10] X. Sun, X. Mei, S. Jiao, M. Zhou, and H. Wang, “Stereo matching with reliable disparity propagation”, *In Proc. 3DIMPVT*, pp.132–139, 2011.

[11] Y. Wei, C. Tsuhan, F. Franz, and C. H. James, “High performance stereo vision designed for massively data parallel platforms”, *IEEE TCSVT*, vol.99, pp.1–11, 2010.

[12] Y. S. Heo, K. M. Lee, and S. U. Lee, “Robust Stereo Matching using Adaptive Normalized Cross Correlation”, *IEEE Trans. on Pattern Analysis and Machine Intelligence,* vol.33, no.4, pp.807-822, 2011.

[13] H. Hirschmüller and D. Scharstein, “Evaluation of Stereo Matching Costs on Images with Radiometric Differences”, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.31, no.9, pp.1582-1599, 2009.

[14] A. Klaus, M. Sormann, and K. Karner, “Segment-Based Stereo Matching Using Belief Propagation and a Self-Adapting Dissimilarity Measure”, *In Proceedings of International Conference on Pattern Recognition*, pp.15-18, 2006.

[15] A. Levin, A. Rav-Acha, and D. Lischinski, “Spectral Matting”, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.30, pp.1699-1712, 2008.

[16] A. Levin, D. Lischinski, and Y. Weiss, “A Closed Form Solution to Natural Image Matting”, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.30, pp.228-242, 2008.

[17] C. Rhemann, A. Hosni, M. Bleyer, C. Rother, and M. Gelautz, “Fast cost-volume filtering for visual correspondence and beyond”, *International Conf. on Computer Vision,* pp.3017-3024, 2011.

[18] K. He, J. Sun, and X. Tang, “Guided Image Filtering”, *European Conf. on Computer Vision*, pp.1-14, 2010.

[19] A. Fusiello, V. Roberto, and E. Trucco, “Efficient stereo with multiple windowing”, *In Proceedings of International Conf. on Computer Vision*, pp.858 – 863, 1997.

[20] http://vision.middlebury.edu/stereo.

[21] S. Mattoccia, S. Giardino, and A. Gambini, “Accurate and efficient cost aggregation strategy for stereo correspondence based on approximated joint bilateral filtering”, *In Proceedings of Asian Conf. on Computer Vision*, pp. 371-380, 2010.

[22] Q. Yang, “A non-local cost aggregation method for stereo matching”, *In Proceedings of International Conf. on Computer Vision*, pp.1402-1409, 2012.

Jie Li received the B.S. degree from Hubei university of technology in 2007 and received M.S. degrees from Wuhan University in 2009. She is currently pursuing the Ph.D. degree from the School of Electronic Information in Wuhan University, Wuhan, China. Her current research interests include 3D reconstruction, image processing and embedded system.

Wenxuan Shi received the B.S. and Ph.D. degree from Wuhan University in 2005 and 2011, separately. His current research interests are image processing, compressed sensing and embedded system.

Dexiang Deng was born in 1961. He is a Professor in School of Electronic Information, Wuhan University. His current research interests are spatial image processing and embedded system.

Wenyan Jia is a research assistant professor of neurological surgery in University of Pittsburgh.

Mingui Sun is a professor of neurological surgery, bioengineering and electrical engineering in University of Pittsburgh.