High-Accuracy Stereo Matching Based on Adaptive Ground Control Points

Chenbo Shi, *Member, IEEE*, Guijin Wang, *Member, IEEE*, Xuanwu Yin, Xiaokang Pei, Bei He, and Xinggang Lin

***Abstract*— This paper proposes a novel high-accuracy stereo matching scheme based on adaptive ground control points (AdaptGCP). Different from traditional fixed GCP-based methods, we consider color dissimilarity, spatial relation, and the pixel-matching reliability to select GCP adaptively in each local support window. To minimize the global energy, we propose a practical solution, named as alternating updating scheme of disparity and confidence map, which can effectively eliminate the redundant and interfering information of unreliable pixels. The disparity values of those unreliable pixels are reassigned with the information provided by local plane model, which is fitted with GCPs. Then, the confidence map is updated according to the disparity reassignment and the left—right consistency. Finally, the disparity map is refined by multistep filers. Quantitative evaluations demonstrate the effectiveness of our AdaptGCP scheme for regularizing the ill-posed matching problem. The top ranks on Middlebury benchmark with different error thresholds show that our algorithm achieves the state-of-the-art performance among the latest stereo matching algorithms. This paper provides a new insight toward high-accuracy stereo matching.**

***Index Terms*— Stereo matching, ground control points, confidence map, alternating updating, weighted median filter.**

I. Introduction

STEREO matching is vital for many applications such as 3D reconstruction, robot navigation, object segmentation, detection and tracking, etc. Generally, stereo matching combines the information of the same scene from several different viewpoints, and estimates the dense depth map using the disparity of objects in the image pairs.

The major problem in stereo matching is the ambiguity in occluded and textureless regions. The ambiguous pixels cannot be matched with their corresponding ones in another image based on their own information. We have to resort to the messages from those reliable neighbor pixels to estimate their disparities. According to the scheme of message merging, stereo matching algorithms can be divided into local and global approaches [1]. In local schemes, the disparity at a given pixel only depends on the data cost within a finite window. Most early local based methods, such as line scanning [2] and adaptive windows [3], improved the performance by posing space constraints to neighboring pixels. However, it was difficult to handle textureless regions and occlusion due to the limited window size. In recent years, weighted support window were proposed to evaluate the importance of each pixel [4]. Similar to the bilateral filter, Yoon [4] weighted each pixel by measuring the space and color similarity between pixels. Defined as the shortest path connecting two pixels in the color volume, geodesic distance was used by Hosni [5] as the support weight. These methods were able to obtain the disparity in textureless regions and kept the discontinuity property on the edge, however with huge computational cost. Gupta and Cho [6] proposed binary window technique to choose the pixels in the support window for speeding up. But their model was not suitable for slant planes and was sensitive to the algorithm parameters. Confidence-based support window (CSW) proposed by Shi [7] selected reliable pixels for the local disparity plane fitting and achieved good results, but still left some errors in ambiguous areas.

Global methods typically make explicit smoothness assumptions and treat the disparity assignment as a global optimization problem in Markov Random Field (MRF). Traditional global optimization algorithms were proven to be suitable to compute the global minimum, such as dynamic programming (DP) [8], Belief Propagation (BP) [9] and Graph Cuts (GC) [10]. However they performed poorly on the discontinuous edges. Recently better results were achieved by combining new constraint hypothesis such as color segmentation [11], [12] matting [13], [14] and weighed support window [15]. But most global optimization algorithms need a complex global model and were sensitive to algorithm parameters. In recent years, Ground Control Points (GCP) based approaches also got good results and attracted a lot of attention. Dense depth map could be estimated by these sparse reliably matched GCPs. These GCP based algorithms were reviewed in detail in Section I-A.

In this paper, we propose a novel high-accuracy stereo matching approach with confidence-based adaptive GCP selection. Different from general energy optimization, we formulate an energy function that incorporates the coarse disparity map and confidence map, and then present a practical alternating scheme to update the disparity map and confidence map. A Confidence-based Support Window (CSW) is defined for each pixel with the reliability of disparity value, color similarity, and spatial constraint. And then matching involves an iteration process of two steps, 1) select high-confidence

Manuscript received October 20, 2012; revised October 7, 2013 and December 23, 2014; accepted January 6, 2015. Date of publication January 15, 2015; date of current version March 3, 2015. This work was supported by the National Natural Science Foundation of China under Grant 61271390 and Grant 61327902. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Béatrice Pesquet-Popescu.

The authors are with the Department of Electronic Engineering, Tsinghua University, Beijing 100084, China (e-mail: shichenbo@gmail.com; wangguijin@tsinghua.edu.cn; 11235houston@gmail.com; greatpxk@sina.com; b-he08@mails.tsinghua.edu.cn; xglin@tsinghua.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2015.2393054

1057-7149 © 2015 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

pixels to conduct local plane fitting and then re-assign the disparity of all points in the CSW using the fitted plane. 2) Update the confidence map by the re-assigned disparity map and the left-right consistence to reselect GCP set. The iteration will terminate until the confidence map converges. In the iteration process, adaptive GCPs can reduce the sensitiveness of the disparity map to input parameters. After the iteration, multi-step filters are applied to refine the disparity map. Extensive experiments demonstrate the effectiveness and superiority of our proposed scheme.

The rest of the paper is organized as follows. In Section I-A, some related work is introduced. In Section II, we briefly overview the motivation and theory of our stereo matching algorithm and present the proposed framework. Section III describes the detailed implementation of our algorithm. Section IV presents the experimental results on the public well-know benchmark and various datasets, and investigates the overall performance in each module and over various parameters. The last section presents our conclusion and some potential improvements as future work.

A. Related Work

The notion of Ground Control Points (GCPs) was firstly introduced by Bobick and Intille [16]. GCPs were sparse points that can be matched reliably. The disparity values of the other points are estimated by merging the messages from GCPs. Thus, a semi-dense disparity map could be obtained directly [17]. Wei [18] combined GCPs messages with image segmentation. To deal with the half-occlusion problem, Xu and Jia [19] defined the Outlier Confidence by the probability of pixel occlusion. With the outlier confidence, they defined a penalty term for occlusion and added it to the energy function to be optimized. Good results were achieved in both occluded and non-occluded areas. Sun et al. [20] proposed an algorithm of reliable pixels' message propagation by line segments along 1-D direction. Different from [17], [18], their algorithm only propagated the reliable seed points' messages through the scanning lines. Single-direction propagation was used to avoid the highly complex and unstable color segmentation. However, owning to the uncertainty of the endpoints of the line segments, line segments could be affected by the stripe defect problem. Wang [21] used MRF as the global optimization model with GCP as inputs. Their energy function consisted of three terms: the energy constraint with GCP, the smoothness constraint between neighboring pixels and the data term measured by absolute difference (AD) between the left view and the right view. The top rank in the Middlebury website of their method proved that GCP was a good expression of regional discrimination. GCP-based algorithms could strengthen the useful information of the scene.

II. Theory and Framework

The concept of confidence has been adopted in previous research works [3], [19], [21], where the main goal of applying the confidence is to choose initial pixels for subsequent processing. In this work, the confidence is firstly derived from the disparity cost and will be utilized throughout the optimizing process.

This section firstly gives some notions, then formulates the GCP based energy function, and finally presents our new alternating updating framework.

A. Definition

1) Confidence Definition: According to [3], the estimated disparity at pixel p can be written as the sum of the ground truth d_p^* and the noise N_p:

$$d_p = d_p^* + N_p. \tag{1}$$

Different from [3], we assume N_p as a zero-mean uniform distribution $U(-\omega_p, \omega_p)$, where ω_p indicates the range of the possible candidates. Generally, the matching cost $Cost(p, d)$ at pixel p should be smaller when $d = d_p^*$. ω_p can be expressed by the signal to noise ratio (SNR),

$$\omega_p \sim \frac{\sum_{d \neq d_p^*} \frac{1}{Cost(p,d)}}{\frac{1}{Cost(p,d_p^*)}} = \sum_{d \neq d_p^*} \frac{Cost(p, d_p^*)}{Cost(p, d)} \tag{2}$$

On the other hand, by the smoothness assumption in disparity map, the disparity of the p-th pixel can be estimated with its neighbors. ω_p can also be affected by the disparity variance in the neighborhood, which is expressed as

$$\omega_p \sim \left(\frac{1}{n} \sum_{q \in A(p)} (d_q - d_q^*)^2 \right)^{1/2}, \tag{3}$$

where $A(p)$ is the neighborhood of p. The confidence of disparity at pixel p is defined as the probability of estimated value being equal to the true value. Thus the relationship between the confidence and the noise is expressed as follows,

$$f(p) \propto \frac{1}{\max(\omega_p, 1)}. \tag{4}$$

Accordingly, the confidence can be expressed in either the matching cost data space or the disparity space. In the initialization, the confidence of the p-th pixel is evaluated by the ratio between the maximum and secondary value, called Peak Ratio (PKR), of the matching cost. While during the updating stage, the new disparity value is estimated by the reliable disparities in the neighborhood, and the confidence is described by the consistency of them.

According to the definition above, the confidence is closely related to disparity matching cost, which is a vector of dimension $D_{range} = D_{max} - D_{min}$, with the search range $D_{min} \sim D_{max}$, as shown in the Fig. 1 (Left). Generally, $D_{min} = 1$. In practice, the cost vector is always reduced to two dimensions as (d_m, f), where d_m is the disparity achieving the lowest matching cost, and f is the associated confidence.

The pixel with high confident disparity are selected as Ground Control Points (GCPs). Messages are propagated from GCPs to the low confident pixels and high-accuracy depth estimation can be achieved.

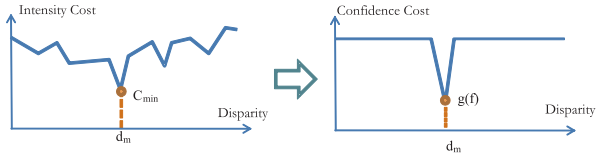


Fig. 1. The dimension reduction of the intensity data cost by the confidence.

2) Local Plane Model of Smoothness Assumption: In general MRF based global algorithms, the smoothness assumption keeps the smooth property in the same object and the discontinuity between different objects. It can be expressed as the disparity constraint between pixels in neighborhood as follows:

$$V_{p,q}(d_p, d_q) = V\{d_q, d_p + F_p(q)\} = V\{\Delta d_{pq}, F_p(q)\}, \quad (5)$$

where $F_p(\cdot)$ denotes the smoothness assumption model at p and $\Delta d_{pq} = d_p - d_q$ denotes the disparity difference between p and q. In general, $F_p(q)$ is a function related to the location between p and q. If $F_p(q) \equiv 0$, the smoothness model is the front parallel plane, which is simple but not suitable for obviously slant disparity planes. In this paper, the local plane model [22] is used, which is defined as

$$F_p(q) = A_p(q_x - p_x) + B_p(q_y - p_y), \quad (6)$$

where A_p and B_p are the plane parameters. Similar to [22], the continuity of the plane parameters between adjacent pixels is taken into account.

B. Energy Function

Different from the previous color image based MRF formulation, we initialize the global energy with the coarse disparity map and confidence map, and then iteratively minimize the energy function in a coarse-to-fine way. The coarse disparity d^{cor} and confidence f^{cor} are formulated as $d^{cor} = \{d_L{}^{cor}, d_R{}^{cor}\}$ and $f^{cor} = \{f_L{}^{cor}, f_R{}^{cor}\}$, respectively. The maximum a-posteriori estimation of disparity map is derived as follows,

$$\arg\max_d p(d|d^{cor}, f^{cor}) = \arg\max_d \frac{p(d^{cor}, f^{cor}|d)p(d)}{p(d^{cor}, f^{cor})}. \quad (7)$$

Since $p(d^{cor}, f^{cor})$ is fixed, (7) is reduced to

$$\arg\max_d p(d|d^{cor}, f^{cor}) = \arg\max_d p(d^{cor}, f^{cor}|d)p(d). \quad (8)$$

The first item $p(d^{cor}, f^{cor}|d)$, called data term, represents the maximum likelihood estimation given observation d^{cor} and f^{cor}. Assumed that the disparity of each pixel has independent identical distribution, we express the data term of the whole image as

$$p(d^{cor}, f^{cor}|d) \propto \exp\left(-\sum_{p\in S} Data(d_p)\right), \quad (9)$$

where S represents the set of pixels in the whole image. The data cost $Data(\cdot)$ of pixel p at disparity d_p can be computed with p's coarse disparity d_s^{cor} and confidence f_p^{cor},

$$Data(d_p) = g(f_p^{cor} \cdot T(d_{L,p}^{cor} - d_{R,p+d_p}^{cor})). \quad (10)$$

The function $g(\cdot)$ is formulated as

$$g(x) = \exp(-\beta \cdot x), \quad (11)$$

where β is a constant to balance the penalty for the disparity shifting from the coarse value. $T(\Delta d)$ is a truncation function of disparity difference with the form

$$T(\Delta d) = \begin{cases} \alpha K & if\ |\Delta d| < T_D \\ K & otherwise, \end{cases} \quad (12)$$

where K is the unit cost, α a constant greater than 1 (Typical $5 \sim 10$) and T_D is a threshold.

The second item in the right-hand side of equation (8), called smoothness term, represents the model inherent in the disparity map. The smoothness term $p_p(d)$, at pixel p, is the accumulation of the potentials $V_{p,q}$ between p and its neighborhood q

$$p(d) \propto \exp\left(-\sum_{\{p,q\}\in N} V_{\{p,q\}}(d_p, d_q)\right), \quad (13)$$

The potential $V_{p,q}$ in equation (13) evaluates the smoothness cost between pixel p and q, expressed as

$$V_{\{p,q\}}(d_p, d_q) = g(u_{\{p,q\}} \cdot T(\Delta d)), \quad (14)$$

where $u_{\{p,q\}}$ is the color similarity measured by

$$u_{\{p,q\}} = \begin{cases} f_p & if\ |I(p) - I(q)| < T_c \\ 0 & otherwise, \end{cases} \quad (15)$$

where T_c is a constant threshold and $T(\Delta d)$ is the same as equation (12). Δd is the disparity smooth constraint with the form

$$\Delta d = (d_p + F_p(q) - d_q) = F_p(q) - \Delta d_{pq}, \quad (16)$$

where $F_p(q)$ represents an offset given by the local plane model. If q does not fit to the smooth model, i.e., Δd in (16) is too large, $T(\Delta d)$ is large accordingly. This means the contribution of pixel q to p gets smaller by (13) and (14). Thus the total smoothness energy is formulated as the sum of each pixel's energy

$$\begin{aligned} &\sum_{\{p,q\}\in N} V_{\{p,q\}}(d_p, d_q) \\ &\quad = \sum_{\{p,q\}\in N} g(u_{\{p,q\}} \cdot T(d_p, d_q)) \\ &\quad = \sum_{p\in S}\left(g\left(\sum_{q\in A(p)} f_q \cdot T(F_p(q) - \Delta d_{pq})\right)\right). \end{aligned} \quad (17)$$

The neighborhood N in (13) and (17) is defined/interpreted by $A(p)$. $A(p)$ is not the typical 4-connected or 8-connected neighborhood, instead it refers to the point set satisfying the color similarity condition $u_{\{p,q\}}$ defined in (15). The total matching energy is

$$\begin{aligned} E = E_{data} + E_{smooth} = \sum_{p\in S}\Bigg(&g\left(f_p^{cor} \cdot T(d_{L,p}^{cor} - d_{R,p+d_p}^{cor})\right) \\ &+ g\left(\sum_{q\in A(p)} f_q \cdot T(F_p(q) - \Delta d_{pq})\right)\Bigg) \end{aligned} \quad (18)$$

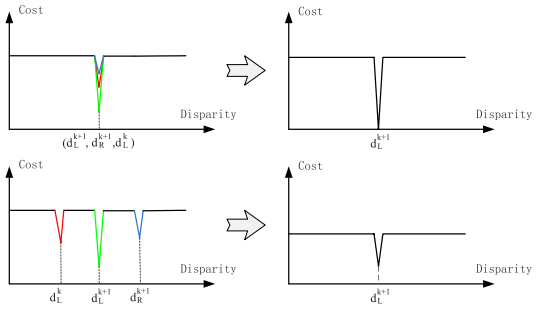


Fig. 2. Minimize the energy of the fused messages by the dimension reduction. Top: the estimated disparity at the same value; Bottom: the disparities at different values.

This is the target function to be optimized. We will introduce our optimization strategy in next section.

C. Adaptive GCP Optimization Framework

In the last section, we formulate the energy minimization problem utilizing the coarse disparity and confidence maps. Different from the belief propagation in grayscale images [9], we propose a novel optimization based on adaptive GCP.

The smoothness term in equation (17) transfers the disparity messages from each pixel q in $A(p)$ to pixel p. The message here is different from those based on the matching cost. It is a vector of dimension D_{range}, with the form of $m_q = [0, \ldots, 0, f_q, 0, \ldots, 0]$. Only the value at d_q-th dimension is f_q, while other values are all set to 0.

The optimization includes the following steps:

1. Initialize the disparity maps and confidence maps by the matching cost from input color images; For each pixel, the initial disparity is chosen to be the d giving the minimum cost and the initial confidence is obtained by the PKR of the cost function.

2. In the k-th iteration, firstly the disparity smoothness model $F_p^k(q)$ is calculated based on the GCP set in the neighborhood $A(p)$ of a pixel p according to the confidence constraint. In this work, a local plane is used for the smoothness model (please see Section III-B). The smoothness messages are obtained as

$$m_p^* = \sum_{q \in A(p)} m_q^k \cdot T(F_p^k(q) - \Delta d_{pq}), \tag{19}$$

where m_p^* is the fused message.

3. Combine the smoothness messages with the data item and minimize the total energy. The minimization is equivalent to maximize the combined message as follows,

$$\begin{cases} m_p^{k*} = m_p^k \cdot T(d_{L,p}^{cor} - d_{R,p+d_p}^{cor}) + m_p^* \\ d_p^{k+1} = \arg\max_d (m_p^{k*}(d)) \\ m_p^{k+1} = \begin{cases} \dfrac{m_p^{k*}(d)}{\|m_p^{k*}\|} & if\ d = d_p^{k+1} \\ 0 & otherwise \end{cases} \end{cases} \tag{20}$$

where $m_p^*(d)$ denotes the value at d-th dimension and m_p^{k+1} is the message of pixel p for the next iteration.The energy at pixel p, decreasing in confidence message, has multiple pulses as shown in the left figure of Fig. 2. Denote the smoothness item by (d_L^{k+1}, d_L^k) and the data item by (d_L^{k+1}, d_R^{k+1})

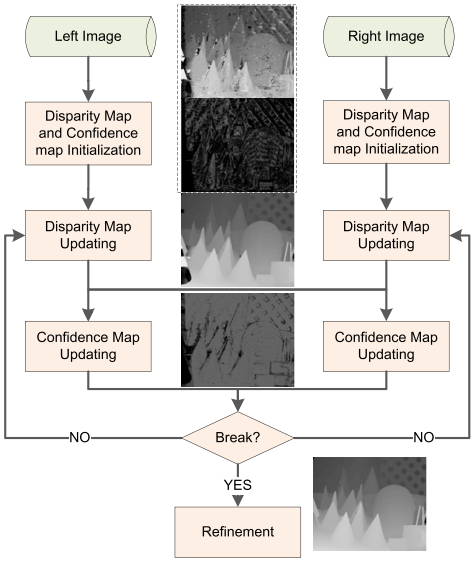


Fig. 3. The flow chart of the proposed algorithm.

in equation (10). Similar to the description in II-A, the disparity value yielding the minimum cost will be chosen as the new candidate and the PKR of the message changes because of the cost on different dimensions. And thus, if the fused messages accumulates at the same disparity (Top-right in Fig. 2), the confidence should be increased, otherwise the confidence will be decreased (Bottom-right in Fig. 2). In practice, the confidence is updated according to the consistence of the multi-estimated disparity values (please see Section III-C). Then the GCP set will be re-selected based on the newly updated confidence map that is more reliable. The adaptive GCPs propagate more precise and reliable information to other pixels.

Once most of the confidence map pixels converge to higher values, the optimization will terminate. Based on the above processing, the algorithm framework is proposed as shown in Fig. 3, which including 4 main parts.

1) Initialization of the Disparity and the Confidence Maps: This module aims to get the initial disparity map and confidence map according to the input image pair. Our framework is compatible with various matching cost functions, such as SAD, CENSUS [23], Adaptive window [4], Geodesic [5], etc. The confidence is defined according to the feature of each cost function. Different representations of confidence is evaluated in [24]. Considering the features of different cost functions, we design a new initialization method combining AD with CENSUS dissimilarity measure in Section III-A.

2) Disparity Map Updating: In this module, we apply local plane fitting for each Confidence Support Window (CSW) [7]. Smoothness messages are merged and are used to reassign the low confident pixels by combination with the data cost. The parameters (CSW window size and color similarity) are adaptively set to achieve better results during the alternating updating process. More details are described in Section III-B.

3) Confidence Map Updating: Confidence map updating aims to minimize the cost energy in each iteration. In the initialization, the confidence of each pixel is evaluated by the

TABLE I
PARAMETER SYMBOLS LIST

Parameter Description	Symbol
Size of AD window	w_{AD}
Size of CENSUS window	w_{CEN}
AD Confidence Threshold	T_{AD}
CENSUS Confidence Threshold	T_{CEN}
Size of Median filter	w_m
Size of Weighted-Median filter	w_{wm}
Color Threshold in LAB Space	T_s
Size of Confidence Support Window	w_b
Inlier Threshold of RANSAC	T_{RAN}
Tolerances of Disparity difference	${T_D}^{LRC}, {T_D}^{Update}$
Confidence Updating Step	c_f
Confidence Updating Positive Factor	λ

PKR of the matching cost. During updating stage, the variance of confidence is reflected as the fluctuation of neighbors, referring to equation (3). Each pixel should converge to the ground truth by received voting information from reliable neighbors. If the disparity of one pixel remains unchanged between iterations, its confidence will be increased and updated.

4) Refinement of the Disparity Map: Disparity Refinement is one of the important steps in stereo matching. In order to deal with the missing disparities on object boundaries and occluded regions, a multi-step refinement is applied, including left-right check and weighted median filter referred in [25].

III. IMPLEMENTATION DETAILS

Table I shows some parameters and their symbols referred in this paper. The parameters of window size are denoted by w and the parameters of color and confidence threshold are denoted by T.

A. Auto-AD & Census Initialization

According to [26], color space can give better results in measuring distortion regions while CENSUS is more effective for grey-scale images. Wegner [27] proposed a new cost function using the product of AD and CENSUS. Sun [20] used the sum of AD and CENCUS as the integrated cost data. However the measure criterions of AD and CENSUS are different. The directly fused function will easily result in reduction of entire dissimilarity. In our initialization, we present a new hybrid SAD and CENSUS, called Auto-AD&CENSUS, as our matching cost to obtain a valid set of GCPs. First the matching cost data is calculated by SAD in the small windows w_a and CENSUS in the large windows w_c. Then confidence maps f_a and f_c are obtained as equation (4). Since the SAD measuring has better similarity than CENSUS in silent regions, it will be given priority in the selection. The disparity map and the confidence map are initialized as follows.

$$d(p) = \begin{cases} d_a(p) & if\ f_a(p) > T_a \\ d_c(p) & otherwise \end{cases}$$

$$f(p) = \begin{cases} \max(f_a(p), f_c(p)) & if\ d_a(p) = d_c(p) \\ f_a(p) & elseif\ f_a(p) > T_a \\ f_c(p) & elseif\ f_c(p) > T_c \\ \min(f_a(p), f_c(p)) & otherwise, \end{cases} \quad (21)$$

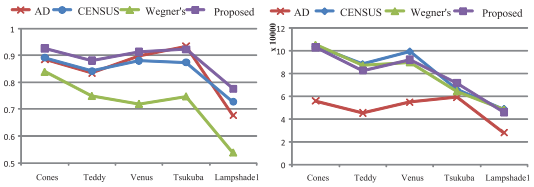


Fig. 4. The quantity and inlier of GCPs corresponding with different initializations. Comparing with AD, CENSUS, Wegner [27] and Proposed method, the curves show Inlier (left) and absolute number(right) in 5 test images.

where the items with subscript a and c represent SAD and CENSUS respectively. Fig. 4 shows the quantity and inlier of the correct GCPs by four different algorithms when the AD and CENSUS window size is 3 and 11. The curve of AD shows better inlier precision but less correct points while CENSUS is on the contrary. Auto-AD&CENSUS improves inlier rate by more than 3% over the CENSUS while retaining the absolute number. The better initial disparity map can bring benefit to the following updating process.

B. Disparity Map Updating

Because of the ambiguity in smooth and occluded regions, there are many error pixels in the initial disparity map. The updating process aims to re-assign the disparity value of those pixels. The CSW algorithm presented in [7] which takes into account both the color similarity and confidence is applied in this section.

The pixels of low confidence need to be reassigned, and they are chosen as the center of the support windows. Suppose that R_p is the support window centered at p of a constant size. Firstly, we generate a binary mask S_p for R_p,

$$S_p(q) = \begin{cases} 1 & if\ d_{Lab}(p,q) < T_s \\ 0 & otherwise, \end{cases} \quad (22)$$

where $q \in R_p$ and $d_{Lab}(p,q)$ is the Euclidean color distance. Applying morphology filtering to S_p, we can obtain a more reliable set S'_p, in which all the pixels are thought to be similar to p in color. With S'_p, the CSW E_p is obtained by picking up all GCP pixels whose confidences are larger than a threshold T_{conf} (we set $T_{conf} = 2c_f$, where c_f is the confidence updating step, see below).

$$E_p(q) = \begin{cases} 1 & if\ Conf(q) > T_{conf}, S'_p(q) = 1 \\ 0 & otherwise. \end{cases} \quad (23)$$

Then we use the pixels in E_p to calculate optimal local plane model to re-assign the disparity for each pixel in S'_p. To reduce the effect of outliers, RANSAC [28] is employed in local plane fitting as in equation (6). The disparity will be reassigned only when the RANSAC succeeds.

Since the continuity of the disparity surface of the same object, the pixels that need to be reassigned in a CSW are expected to have similar disparity plane parameters. Observing this, we do not need to choose a CSW for every pixel in current iteration. To speed up, we choose sparse window centers that can form a 'cover' for the whole image. Specifically, the CSWs are chosen such that they overlap at least half size of

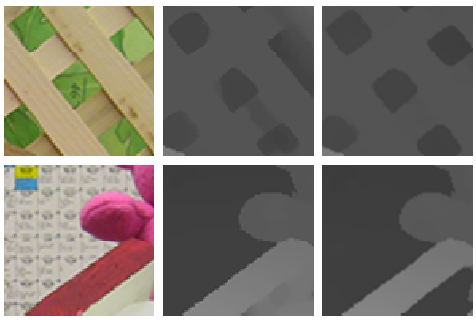


Fig. 5. Comparing results in different size of support window. (a) the original image, (b) CSW with window size of 67, (c) CSW with window size of 47; the first and second frames show better results under $w_b = 67$ and $w_b = 47$, respectively.

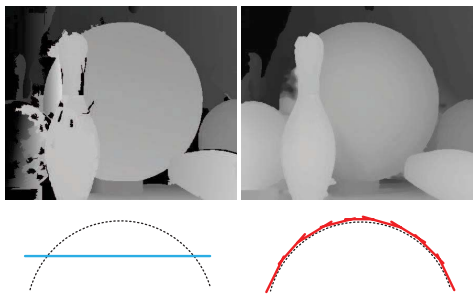


Fig. 6. Sub-precision Results on sphere surface compared with [29]. Left: the result of [29], Right: the result of this paper. The performance of this paper is more accurate than [29] on the ball surface.

the window. The multi-reassigned pixel will choose the disparity value that is closest to the support window center. More details are referred in our previous work [7].

The local plane fitting at different regions require different window sizes. In Fig. 5, for example, the proposed algorithm gives the most accurate depth estimation of the fence on the Cones data when $w_b = 47$, while on Teddy data, the most accurate depth estimation of the roof is achieved when $w_b = 67$. To make the messages propagate more efficiently between distant pixels, we adjust the parameters adaptively during the disparity updating process.

$$\begin{cases} T_s^k = \min(T_{\max\_s}, T_s^{k-1} + 0.1) \\ w_b^k = \min(w_{\max\_b}, w_b^{k-1} + 1) \end{cases} \tag{24}$$

where k is the iteration number. T_{max_s} and w_{max_b} refer to the upper limit of the color and space dissimilarity. Since the selected CSWs are kept to overlap, the suitable plane models of adjacent CSWs keep the smoothness of disparity map. Similar to soft-segmentation [11], the proposed approach achieves better smoothness than segment-based algorithms. Fig. 6 shows disparity results on a sphere surface computed by our proposed algorithm and [29]. The result of our proposed approach achieves sub-pixel precision on the bowling and retains better smoothness than the segment-based method presented in [29], verifying that the local plane model is a good approximation for the curved surface.

C. Confidence Map Updating

The confidence map is mostly determined by the variance of disparity neighbors between iterations, therefore we update the confidence map after the disparity map is updated. As shown in Fig. 2, the smoothness message is described by d_L^{k+1} and d_L^k. The new disparity d_L^{k+1} is evaluated by the neighborhood. The variance of the disparity neighbors could be described by divergence of disparities between adjacent iterations. The confidence adjustment of the updating term is as follows.

$$\Delta f_1^{k+1}(p) = \begin{cases} \lambda c_f & |d_L^{k+1}(p) - d_L^k(p)| < T_D^{Update} \\ -c_f & otherwise, \end{cases} \tag{25}$$

where c_f and λ refer to the update step and a constant factor. If the disparity value of the same pixel remains unchanged between iterations, the confidence will be increased, and vice versa. The data item is denoted by d_L^{k+1} and d_R^{k+1}, which is similar to the left-right consistence. The confidence updating of the data item is simplified as

$$\Delta f_2^{k+1}(p) = \begin{cases} -c_f & |d_L^{k+1}(p) - d_R^{k+1}(p')| > T_D^{LRC} \\ 0 & otherwise. \end{cases} \tag{26}$$

If the pixel doesn't satisfy the left-right consistence, the confidence is decreased by a constant. Otherwise, the confidence keeps the same. The updated confidence is the combination of the two incremental terms above.

$$f^{k+1} = f^k + \Delta f_1^{k+1} + \Delta f_2^{k+1}. \tag{27}$$

The adaptive GCP set is re-selected by the updated confidence map. In related works [19], neighbors are determined by the color or distance similarity and are not changed. Some segment-based methods [11], [30] arrange the size of the window via reselect referred pixels. Compared with those algorithms, our adaptive GCP scheme selects more reliable neighbors. Moreover, the confidence map updating can make the approach converge more quickly.

D. Multi-Step Refinement

After the matching process, filters are often used to remove some mismatches and fill the holes. Considering that the discontinuous edges always come along with the color differences, we applied a color-weighted median filter [25]. We designed a multi-step combining refinement process which is able to get refinement effect in different regions, with steps described as follows:

- *Left-Right Consistency Check Refinement:* when the difference of disparity value between the left and right pixel is greater than T_s, then take the smaller disparity value to eliminate errors resulting from occlusion;
- *Conventional Median Filter:* a 5×5 median filter is used to deal with discrete noise and holes in the disparity map;
- *In-situ Median Filter:* a 7×7 weighted median filter is used to refine the discontinuous edges and smooth areas. Different from [25], the filter operation is done in-situ which is more effective for occluded area.

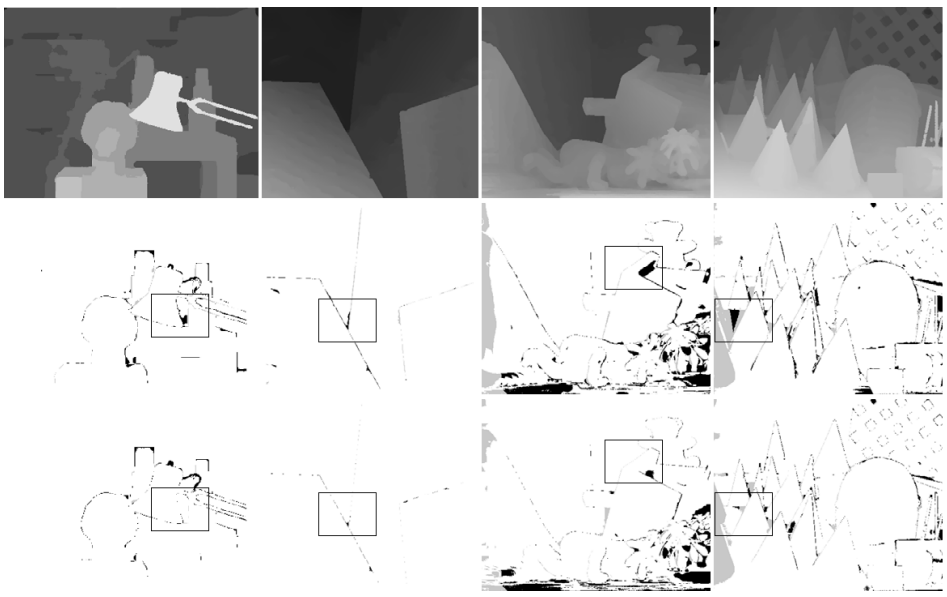


Fig. 7. The disparity maps and the error maps on Middlebury dataset. Top: the final disparity maps obtained by our algorithm. Middle: the corresponding error maps of CSW results without updating [7]. Bottom: the corresponding error maps of final results.

TABLE II
THE VALUES OF WINDOW PARAMETERS

w_{AD}	w_{CEN}	w_b	w_m	w_{wm}
3	11	67	5	7

TABLE III
THE VALUES OF COLOR AND CONFIDENCE THRESHOLDS

T_s	T_{AD}	T_{CEN}	T_{RAN}	$T_D{}^{LRC}$	c_f	λ
2.0	30	40	0.75	$0.75scale$	5	4

IV. EXPERIMENTS

In this section, we evaluate our scheme on various datasets. The proposed method is implemented using C++ under Windows XP OS with Intel CPU (E8400), 3.0GHz and 2G RAM. Tables II and III show all the parameters used in our experiment for all test images. In this section, experiments are presented in 5 parts. The results and ranks on Middlebury website are shown in the first part. After that, experiments with different parameters indicate the wide adaptiveness of our algorithm. The third part gives the performance in terms of sub-pixel precision. In the remaining two parts, more results on different datasets are shown. Experimental results verify the effectiveness of our algorithm and the acceptable computational cost.

A. Ranks on Middlebury Benchmark

Fig. 7 shows comparing results on Middlebury benchmark. Tsukuba image is suitable for testing algorithms on forward parallel plane objects. Since we integrated the advantages of CSW and adaptive GCP, the performance is more robust. As shown in Fig. 7, even in regions such as lamp shade and bracket, the disparity values are estimated accurately and the region behind the bookrack is smoother. Slant regions under different levels are verified in Venus image, including texture-less regions simultaneously. Due to the consideration of slant plane model and the updating scheme, more accurate results are achieved by proposed method. Additionally, the multi-step refinement would reduce errors along discontinuous edges. As AD and CENSUS are combined in our algorithm, high-accuracy confidence map is obtained in Teddy image. For example, the complex region of the flowerpot is estimated effectively, i.e. leaves under different levels could be separated correctly. Disparity results in the left border of the image are improved though errors in the bottom slant plane arise slightly. Cones image contains several complex scenes, such as occluded regions, hole regions, and non-texture regions. Our algorithm achieves more accurate results on most of those regions. Though regions out of the boundary could not be matched reliably, disparity values are correctly estimated since the model of these regions coincides with the correct matched regions. Our algorithm achieves excellent results since several texture-less regions along the right edge are estimated via AD and CENSUS followed by refinement on discontinuity edges.

Global optimization based methods rank top on the benchmark. Our results get rank 1 when the error threshold is 1.0 as shown in Fig. 8. For Cones image, our results get rank 1 on all the three error thresholds. Especially on the non occlusion region, the error rate decreases to 1.81 while ObjectStereo [31] takes the second place by 2.2%.

B. Convergence and Parameters

1) Convergence Performance: The confidence updating scheme brings lots of benefits to the convergence of the algorithm. Fig. 9 presents the updating process on an artificial image which includes a slant textureless plane and texture front parallel background. The initial disparity map has many error values because of the color ambiguities. Only those pixels on the boundary of the slant plane have correct

Error Threshold = 1 (Error Threshold...) — Sort by nonocc / Sort by all / Sort by disc

Algorithm	Avg. Rank	Tsukuba ground truth nonocc	Tsukuba all	Tsukuba disc	Venus ground truth nonocc	Venus all	Venus disc	Teddy ground truth nonocc	Teddy all	Teddy disc	Cones ground truth nonocc	Cones all	Cones disc	Average Percent Bad Pixels
YOUR METHOD	6.3	1.03 13	1.29 5	5.60 15	0.10 3	**0.14** 1	1.30 4	4.63 13	6.47 5	12.5 14	**1.81** 1	**5.70** 1	**5.33** 1	3.83
ADCensus [94]	9.3	1.07 17	1.48 14	5.73 20	0.09 2	0.25 8	1.15 3	4.10 8	6.22 3	10.9 7	2.42 10	7.25 8	6.95 11	3.97
CoopRegion [41]	11.3	0.87 4	**1.16** 1	4.61 4	0.11 5	0.21 4	1.54 8	5.16 20	8.31 13	13.0 17	2.79 23	7.18 7	8.01 29	4.41
AdaptingBP [17]	11.4	1.11 21	1.37 8	5.79 22	0.10 4	0.21 5	1.44 6	4.22 10	7.06 8	11.8 11	2.48 12	7.92 15	7.32 15	4.23
RVbased [116]	14.7	0.95 9	1.42 12	4.98 9	0.11 7	0.29 12	**1.07** 1	5.98 27	11.6 38	15.4 34	2.35 8	7.61 9	6.81 10	4.88
DoubleBP [35]	15.1	0.88 6	1.29 4	4.76 7	0.13 9	0.45 29	1.87 14	3.53 6	8.30 12	9.63 4	2.90 29	8.78 38	7.79 23	4.19
RDP [102]	15.3	0.97 10	1.39 10	5.00 10	0.21 26	0.38 20	1.89 15	4.84 14	9.94 23	12.6 15	2.53 13	7.69 11	7.38 16	4.57
OutlierConf [42]	16.0	0.88 5	1.43 13	4.74 6	0.18 18	0.26 10	2.40 26	5.01 16	9.12 19	12.8 16	2.78 22	8.57 29	6.99 12	4.60
SurfaceStereo [79]	21.3	1.28 35	1.65 23	6.78 42	0.19 20	0.28 11	2.61 37	3.12 3	**5.10** 1	**8.65** 1	2.89 28	7.95 17	8.26 37	4.06

Fig. 8. The print screen of ranks on Middlebury website corresponding with error threshold = 1.0.

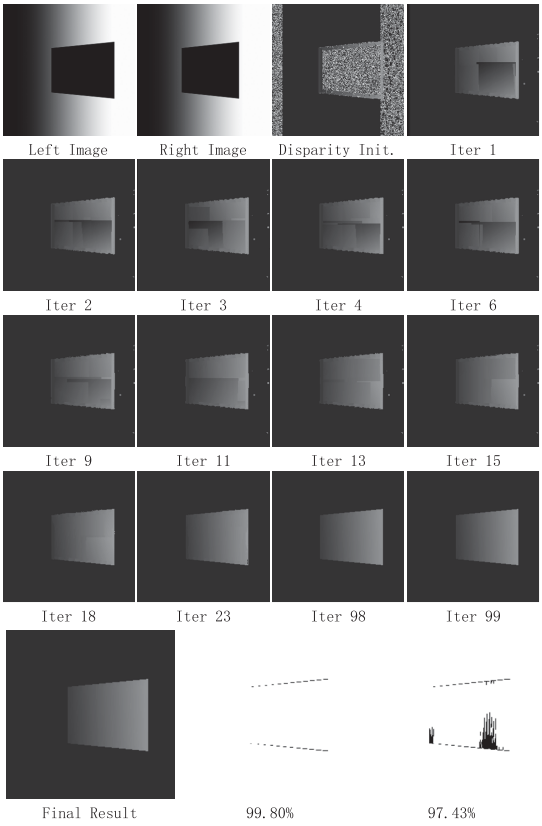


Fig. 9. Convergent process on Slant image [32] with textureless foreground and front parallel background. The left two frames is the input images, and after that is the iteration 1, 2, 3, 4, 6, 9, 11, 13, 15, 18, 23, 98, 99 in order. The last row shows the results and error map corresponding with error threshold = 1.0 and 0.5.

disparity values. In the first several iterations, the disparities are drastically changed and the GCP set cannot be selected correctly. The slant plane is separated into several models. Since the confidence updating depresses the difference between messages and keeps the correspondence of maximal probability model, the disparity map is gradually stable after about 10 iterations. It can be seen that the process converges after 20 iterations. The final result shows high accuracy at both pixel and sub-pixel precision.

We also verify the convergence in the actual scene. Taking Cones for example, the number of unconfident pixels and

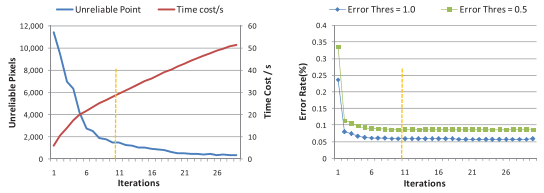


Fig. 10. Convergence performance on Cones Data. Left: The number of unreliable pixels and the accumulated time cost with the iterations. Right: The error rate with the iterations for pixel and sub-pixel accuracy.

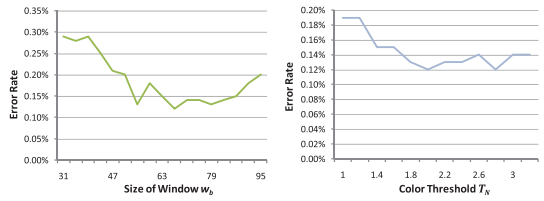


Fig. 11. The error rate under different parameters on Venus images. Left: the error rate changing with the size of support window when $T_N = 2.0$. Right: the error rate changing with color thresholds when $w_b = 67$.

the error rate versus iteration times are listed in Fig. 10. In the initialization, 12,000 pixels are used as center points of CSW and the process costs 7∼10s. Along with iterations, the number of unconfident pixels decreases quickly and it costs only 1s for each iteration. The iterative process terminates when it reaches the maximum number of iterations or the number of unconfident pixels and the error rates are lower than corresponding thresholds. It can be drawn out that the error rate decreases faster than unconfident pixels. As shown by the yellow dashed line, the error rate converges to a stable state after 14 iterations. The total processing time of our algorithm on Cones image is less than 50s.

2) Initial Parameters: We also discuss the influence of different parameters, such as the color similarity threshold T_s and the window size of CSW w_b. These two parameters influence the matching results by adjusting the neighborhood of each pixel. Fig. 11 gives the matching results on Venus data with different parameters. The error curve is shown in the left figure with w_b changing from 31 to 95 when the color threshold takes a typical value $T_s = 2.0$. The right figure shows the error curve with T_s changing from 1.0 to 3.0 with the window size $w_b = 67$. The error rate are restricted

TABLE IV
THE TOP RANKS ON WEBSITE [37] UNDER DIFFERENT ERROR THRESHOLDS

Algorithms	Avg.Rank and Actual Ranking					Overall Ranking
	$T_E = 0.5$	$T_E = 0.75$	$T_E = 1.0$	$T_E = 1.5$	$T_E = 2.0$	
Proposed	10.1(2)	15.5(3)	**6.30(1)**	**8.0(1)**	**7.80(1)**	**9.54(1.60)**
SubPixSearch[34]	**4.20(1)**	**4.80(1)**	21.4(10)	9.80(2)	13.5(3)	10.7(3.40)
SubPixDoubleBP[35]	18.0(7)	16.9(4)	21.3(10)	14.4(8)	17.4(8)	17.6(7.40)
CoopRegion[36]	26.3(15)	27.3(13)	11.3(3)	13.0(6)	14.8(4)	18.5(8.20)
LLR[38]	23.9(10)	19.8(7)	22.5(12)	20.8(11)	18.5(9)	21.1(9.80)
GC+SegmBorder[39]	11.3(4)	17.8(5)	31.0(19)	23.9(13)	22.6(11)	21.3(10.4)
ADCensus[33]	39.0(30)	33.5(20)	9.30(2)	9.90(3)	10.2(2)	20.4(11.4)
PMBP[40]	10.1(2)	15.3(2)	30.8(17)	30.4(18)	30.4(18)	23.4(11.4)
AdaptingBP[29]	34.8(26)	30.4(18)	11.4(4)	12.8(4)	16.3(7)	21.1(11.8)
Undr+OvrSeg[41]	23.7(12)	20.8(8)	31.9(20)	27.6(16)	28.2(16)	26.4(14.4)

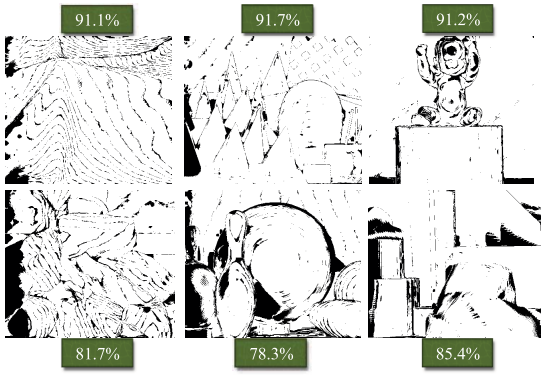


Fig. 12. Sub-pixel error map and precision on partial data in Middlebury dataset. From left to right in order, there are Cloth1, Cones, Baby1, Rocks2, Bowling2 and Lampshade2.

between 0.32% $\sim$ 0.13% and 0.23% $\sim$ 0.14% in these two cases, which shows that the proposed approach is insensitive to input parameters. The performance benefits from the adaptive updating framework with better flexibility and robustness.

C. Sub-Pixel Precision

Compared with other sub-pixel algorithms, We achieve the top results under different error thresholds in Table IV. Our algorithm ranked as top 3 with all error thresholds. In details, our results get rank 1 when the error threshold is 1.0, 1.5, or 2.0, and get rank 2 and rank 3 when the error threshold is 0.5 or 0.75. The average rank of our algorithm is 1.6. Mei [33] get the best results when the error threshold is 1.0 and 2.0. However, the performance decreases severely when the error thresholds are 0.5 and 0.75. Mizukami et al. [34] improves the sub-pixel precision based on [33] and achieves higher ranks. The average rank of some typical methods, such as Yang et al. [35], Wang [36] and Klaus [29], is 7.4, 8.2 and 11.8 respectively. Our algorithm is competitive to these start-of-the-art algorithms.

Fig. 12 shows the sub-pixel results on other examples by our algorithm. The correct rate of all pixels is labeled in Fig. 12 for each image with the error threshold set to 0.5. The error maps show good performance on the continuous smooth surface, such as the cloth and the bowling. Most of the estimated disparities in the textureless regions of Lampshade2 image are correct. The discontinuous edges in the Baby1 and Cones images are also kept well.

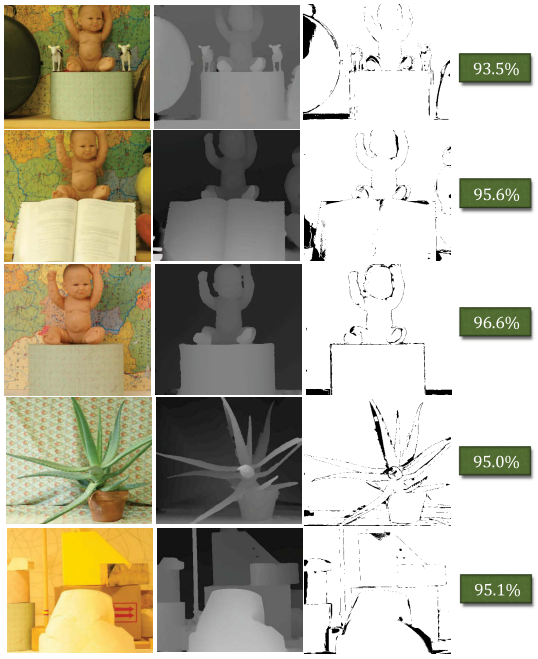


Fig. 13. Result and precision on Baby1, Baby2, Baby3, Aloe and Lampshade1 images. From left to right in order, there are original images, disparity maps, error maps and precision (error threshold is 1.0).

D. Overall Middlebury Dataset Results

Fig. 13 shows more results under different scenes. The correct precision is close and stable. Fig. 14 compares our proposed algorithm and Bleyer's method [40] by measuring the error rates on the non-occlusion regions of another 30 image pairs in the Middlebury dataset. Note that the same parameters are used for different image pairs in our algorithm. In Bleyer [40], AD measuring function, linear truncation as the smoothness term, and Simple Tree based dynamic programming are utilized. The Blue, red and green curves represent results via different similarity measurements of gray, RGB and LUV color spaces respectively in [40]. The orange curve is the performance of the proposed method. As shown, the average error rate of our algorithm is the lowest. The average error rate of the occlusion regions is 7.74%, lower than 13.7% of Bleyer's. Our algorithm all outperforms Bleyer's except for Cloth2 and monopoly images. For "Baby1" $\sim$ "Baby3" and "Cloth1," "Cloth3," "Cloth4" images, we achieve error rate less than 4%, which is more accurate. Comparing with the best

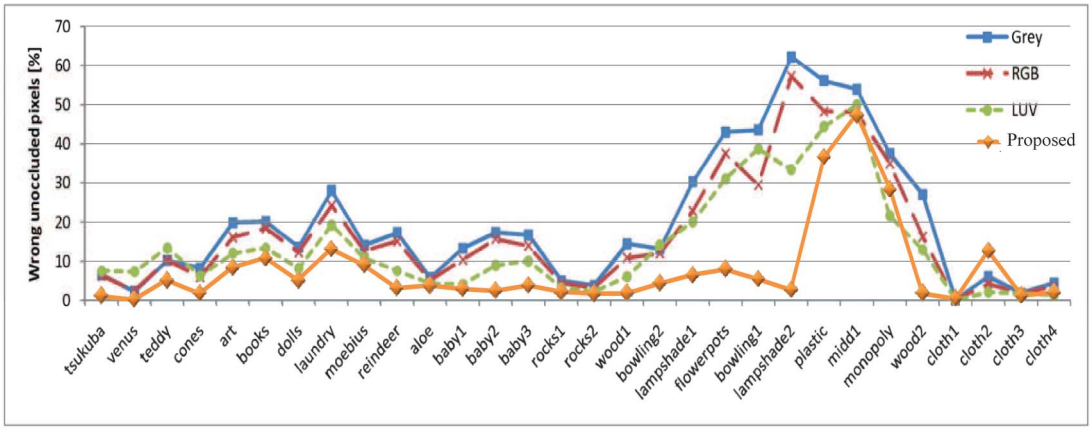


Fig. 14. Error rate in Non-occluded regions on 30 images in Middlebury dataset by proposed algorithm and [40]. Blue, red and green curves represent results on Gray , RGB and LUV space by [41]. The orange curve is the result by this paper. The average error rate reduces about 6% overall.

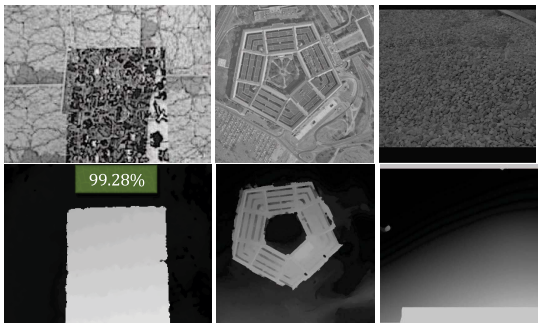


Fig. 15. Results from proposed algorithm on gray images. From left to right in order, there are Map, Pentagon and Rocks images.

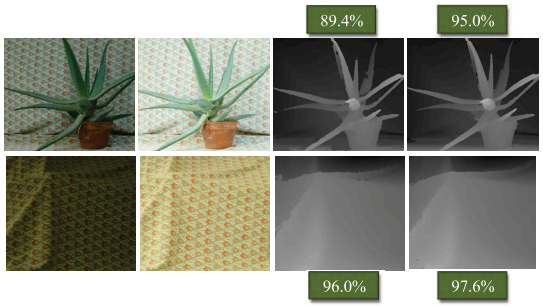


Fig. 16. Results under different illumination and exposure. From left to right, there are left images, right images, corresponding results and results under the same illumination. The first row includes illum2-Exp0 and illum2-Exp1 Aloe images. The second row includes illum1-Exp1 and illum2-Exp2 Cloth1 images.

result from [40], error rates of "lampshade1," "lampshade2," "Bowling1," "Bowling2," "Flowerpots," and "wood2" decrease more than 15%. Additionally, our results are robust besides Plastic, Midd1, and Monopoly images. As these images have texture-less regions including wall and curve, even human could not figure out the true depths, it is understandable that current matching methods do not work well on such regions. The average error rate of non-occluded regions excluding the above three is less than 4.4%. Further more, the average error rates of all regions are 10.9% and 21.1% while the error thresholds are 1.0 and 0.5 respectively for all the 30 images.

E. Other Discussions

1) Gray-Scale Images: In [40], methods based on color similarity could not achieve good results because of the low distinctiveness of color texture. With little modification, our scheme can be extended to gray images straightforward. We utilize CENSUS to make our algorithm adaptive to gray images. Our algorithm selects similarity color pixels strictly. The initial color threshold and updating step of gray images are 1/3 of the ones of colorized images while other parameters remain the same. The result of Map image is shown in the first column in Fig. 15. Results based on segmentation are poor due to the lack of color information. The error rates are 99.28% and 97% without and with sub-pixel accuracy respectively due to jagged errors introduced by non-discriminative gray levels. The result of the pentagon from CMU dataset is given in the second column. We can see that the middle pentagon region is extracted accurately and hollows are achieved among pentagon regions. Additionally, some high buildings outside the pentagon could be figured out from the depth map. As the road piled with stones in the image is noisy, initial disparity values are cluttered. The result in the right column verifies the effectiveness and smoothness of our algorithm.

2) Differernt Illuminance: Due to the difference between equipments, the binocular camera would capture images under different exposure sensibility. The single camera confronts problems of different exposure, additive noise and luminance. Measurements of MI, NCC and CENSUS presented by [42] could only solve these problems partly and need further improvements. Since the color similarity of the same image is used to describe smoothness of neighborhood, the aberration of different images would not affect the remaining models. Additionally, initial depth and confidence map based on AD and CENSUS can depress noise, different luminance and insensitive exposure.

Results on images with different exposure and luminance in the Middlebury dataset are shown in Fig. 16. The first row shows two Aleo images under different exposures, followed by their corresponding disparity results. Disparity results are given in the right first row. Although there exist some errors along the boundary of leaves, accurate estimations for the

flowerpot, forward leaves and cloth can be achieved. Since the AD measurement fails and disturbs under different exposure, initial results are calculated by the CENSUS measurement which introduces a few errors along boundaries. The second row shows Cloth images with corresponding disparity results. Disparities are estimated well in smooth regions while inaccurately along the boundaries. The average accurate rate is 94%, while the average accurate rate is 97.2% under the same exposure. In conclusion, our algorithm can solve the problems of different luminance, exposure, and color.

V. Conclusion

Global based algorithms outperform than others in stereo matching research. Among which, color segment-based and GCP-based methods play important roles. This paper presents a novel energy function and the global model is optimized with adaptive GCPs. Different from previous methods, we fuse the color segmentation in the updating process and propagate not only disparity messages, but also confidence messages. According to the adaptive GCP model, an alternating updating framework utilizing disparity map and confidence map is proposed. The method selects adaptive neighborhoods and minimize the cost energy to help the correct message propagation. In addition, an automatic AD and CENSUS selection is presented to strengthen the initialization and a multi-step refinement is proposed to improve the performance on discontinuous edges. The top ranks achieved by our algorithm on the standard images on Middlebury web demonstrate that our algorithm is one of the most effective global-based algorithms. The results under different error thresholds prove that the method is more robust than others. We also test our algorithm on more than 30 image pairs in Middlebury datasets. The average error rate in non-occluded area is less than 8%. Beyond that, our algorithm achieves good performance under gray-scale and different illuminations. Our future work is to speed up the processing and improve the performance on objects with similar color at different depth.

References

[1] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int. J. Comput. Vis.*, vol. 47, nos. 1–3, pp. 7–42, 2002.

[2] S. Birchfield and C. Tomasi, "A pixel dissimilarity measure that is insensitive to image sampling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 4, pp. 401–406, Apr. 1998.

[3] T. Kanade and M. Okutomi, "A stereo matching algorithm with an adaptive window: Theory and experiment," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 9, pp. 920–932, Sep. 1994.

[4] K.-J. Yoon and I. S. Kweon, "Adaptive support-weight approach for correspondence search," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 4, pp. 650–656, Apr. 2006.

[5] A. Hosni, M. Bleyer, M. Gelautz, and C. Rhemann, "Local stereo matching using geodesic support weights," in *Proc. 16th IEEE Int. Conf. Image Process. (ICIP)*, Nov. 2009, pp. 2093–2096.

[6] R. K. Gupta and S. Y. Cho, "Real-time stereo matching using adaptive binary window," in *Proc. 3DPVT*, 2010, pp. 1–8.

[7] C. Shi, G. Wang, X. Pei, B. He, and X. Lin, "Stereo matching using local plane fitting in confidence-based support window," *IEICE Trans. Inf. Syst.*, vol. E95.D, no. 2, pp. 699–702, 2012.

[8] O. Veksler, "Stereo correspondence by dynamic programming on a tree," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2005, pp. 384–390.

[9] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient belief propagation for early vision," *Int. J. Comput. Vis.*, vol. 70, no. 1, pp. 41–54, 2006.

[10] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 11, pp. 1222–1239, Nov. 2001.

[11] M. Bleyer, C. Rother, and P. Kohli, "Surface stereo with soft segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2010, pp. 1570–1577.

[12] B. He, G. Wang, C. Shi, X. Yin, B. Liu, and X. Lin, "High-accuracy and quick matting based on sample-pair refinement and local optimization," *IEICE Trans. Inf. Syst.*, vol. E96-D, no. 9, pp. 2096–2106, 2013.

[13] M. Bleyer, M. Gelautz, C. Rother, and C. Rhemann, "A stereo approach that handles the matting problem via image warping," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vols. 1–4. Jun. 2009, pp. 501–508.

[14] B. He, G. Wang, and C. Zhang, "Iterative transductive learning for automatic image segmentation and matting with RGB-D data," *J. Vis. Commun. Image Represent.*, vol. 25, no. 5, pp. 1031–1043, 2014.

[15] 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. Pattern Anal. Mach. Intell.*, vol. 31, no. 3, pp. 492–504, Mar. 2009.

[16] A. F. Bobick and S. S. Intille, "Large occlusion stereo," *Int. J. Comput. Vis.*, vol. 33, no. 3, pp. 181–200, 1999.

[17] M. Lhuillier and Q. Long, "Match propagation for image-based modeling and rendering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 8, pp. 1140–1146, Aug. 2002.

[18] Y. Wei and L. Quan, "Region-based progressive stereo matching," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1. Jun./Jul. 2004, pp. I-106–I-113.

[19] L. Xu and J. Jia, "Stereo matching: An outlier confidence approach," in *Proc. 10th Eur. Conf. Comput. Vis. (ECCV)*, 2008, pp. 775–787.

[20] X. Sun, X. Mei, S. Jiao, M. Zhou, and H. Wang, "Stereo matching with reliable disparity propagation," in *Proc. 1st Joint 3DIM/3DPVT Conf. 3D Imag., Modeling, Process., Vis., Transmiss.*, May 2011, pp. 132–139.

[21] L. Wang and R. Yang, "Global stereo matching leveraged by sparse ground control points," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 3033–3040.

[22] M. Bleyer, C. Rhemann, and C. Rother, "PatchMatch stereo-stereo matching with slanted support windows," in *Proc. Brit. Mach. Vis. Conf.*, 2011, pp. 1–11.

[23] R. Zabih and J. Woodfill, "Non-parametric local transforms for computing visual correspondence," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 1994, pp. 151–158.

[24] X. Hu and P. Mordohai, "Evaluation of stereo confidence indoors and outdoors," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 1466–1473.

[25] C. Rhemann, A. Hosni, M. Bleyer, C. Rother, and M. Gelautz, "Fast cost-volume filtering for visual correspondence and beyond," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2011, pp. 3017–3024.

[26] M. Bleyer and S. Chambon, "Does color really help in dense stereo matching?" in *Proc. 3DPVT10*, 2010, pp. 1–8.

[27] K. Wegner and O. Stankiewicz, "Similarity measures for depth estimation," in *Proc. 3DTV Conf., True Vis.-Capture, Transmiss., Display 3D Video*, May 2009, pp. 1–4.

[28] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[29] A. Klaus, M. Sormann, and K. Karner, "Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure," in *Proc. 18th Int. Conf. Pattern Recognit.*, vol. 3. 2006, pp. 15–18.

[30] Y. Taguchi, B. Wilburn, and C. L. Zitnick, "Stereo reconstruction with mixed pixels using adaptive over-segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2008, pp. 1–8.

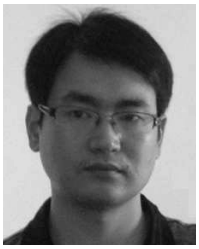
[31] M. Bleyer, C. Rother, P. Kohli, D. Scharstein, and S. Sinha, "Object stereo—Joint stereo matching and object segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2011, pp. 3081–3088.

[32] J. Sun, Y. Li, S. B. Kang, and H.-Y. Shum, "Symmetric stereo matching for occlusion handling," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2005, pp. 399–406.

[33] X. Mei, C. Cui, X. Sun, M. Zhou, Q. Wang, and H. Wang, "On building an accurate stereo matching system on graphics hardware," in *Proc. GPUCV*, 2011, 467–474.

[34] Y. Mizukami, K. Okada, A. Nomura, S. Nakanishi, and K. Tadamura, "Sub-pixel disparity search for binocular stereo vision," in *Proc. 1st Int. Conf. Pattern Recognit.*, Nov. 2012, pp. 364–367.

[35] Q. Yang, R. Yang, J. Davis, and D. Nister, "Spatial-depth super resolution for range images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.
[36] Z.-F. Wang and Z.-G. Zheng, "A region based stereo matching algorithm using cooperative optimization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2008, pp. 1–8.
[37] D. Scharstein and R. Szeliski. (2002). *Middlebury*. [Online]. Available: http://www.cs.unc.edu/
[38] S. Zhu, L. Zhang, and H. Jin, "A locally linear regression model for boundary preserving regularization in stereo matching," in *Proc. 12th Eur. Conf. Comput. Vis.*, 2012, pp. 101–115.
[39] F. Besse, C. Rother, A. Fitzgibbon, and J. Kautz, "PMBP: PatchMatch belief propagation for correspondence field estimation," in *Proc. BMVC*, 2012.
[40] M. Bleyer, S. Chambon, U. Poppe, and M. Gelautz, "Evaluation of different methods for using colour information in global stereo matching approaches," *Remote Sens. Spatial Inf. Sci.*, vol. 37, pp. 415–422, Jul. 2008.
[41] M. Bleyer and M. Gelautz, "Simple but effective tree structures for dynamic programming-based stereo matching," in *Proc. Int. Conf. Comput. Vis. Theory Appl. VISAPP*, 2008, pp. 415–422.
[42] H. Hirschmuller and D. Scharstein, "Evaluation of cost functions for stereo matching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2007, pp. 1–8.



Chenbo Shi (M'13) was born in 1984. He received the B.S. and Ph.D. degrees from the Department of Electronics Engineering, Tsinghua University, Beijing, China, in 2005 and 2012, respectively. From 2008 to 2012, he authored over 10 international journal and conference papers. He is currently a reviewer for several international journals and conferences. He is a Post-Doctoral Researcher with Tsinghua University. His research interests are focused on image stitching, stereo matching, matting, object detection, and tracking.

Guijin Wang (M'08) received the B.S. and Ph.D. (Hons.) degrees in electrical engineering from Tsinghua University, Beijing, China, in 1998 and 2003, respectively. From 2003 to 2006, he was a Researcher with Sony Information Technologies Laboratories. Since 2006, he has been with the Department of Electronics Engineering, Tsinghua University, as an Associate Professor. He authored over 60 interna-tional journal and conference papers, and holds 10 patents with numerous pending. He was the Session Chair of the IEEE CCNC'06. His research interests focus on wireless multimedia, depth imaging, pose recognition, intelligent human–machine UI, intelligent surveillance, industry inspection, and online learning.

Xuanwu Yin was born in Changchun, China, in 1987. He received the B.S. degree from Tsinghua University, Beijing, China, in 2011, where he is currently pursuing the Ph.D. degree in electronics engineering. His research interests include the applications of image processing and pattern recognition in image/video matting, registration, and 3D reconstruction.

Xiaokang Pei was born in 1988. He received the B.S. and M.S. degrees in signal and information processing from the Department of Electronics Engineering, Tsinghua University, Beijing, China, in 2009 and 2012, respectively. His research interests are focused on stereo matching, 3D reconstruction, image and video processing, and computational photography.

Bei He was born in Anhui, China, in 1987. He received the B.S. degree from the Nanjing University of Science and Technology, Nanjing, China, in 2008. He is currently pursuing the Ph.D. degree in electronics engineering with Tsinghua University, Beijing, China. His research interests include in the applications of image processing and pattern recognition in image/video matting, registration, and mosaicing.

Xinggang Lin received the B.S. degree in electronics engineering from Tsinghua University, Beijing, China, in 1970, and the M.S. and Ph.D. degrees in information science from Kyoto University, Kyoto, Japan, in 1982 and 1986, respectively. He joined the Department of Electronics Engineering, Tsinghua University, in 1986, where he has been a Full Professor since 1990. He received the Great Contribution Award from the Ministry of Science and Technology of China, and the Promotion Awards of Science and Technology from Beijing Municipality. He was the General Co-Chair of the Second IEEE Pacific-Rim Conference on Multimedia, an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, and a Technical/Organizing Committee Member of many international conferences. He is a fellow of the China Institute of Communications. He authored over 140 referred conference and journal papers in diversified research fields.