Tree Structural Watershed for Stereo Matching

Xiao Tan
SEIT of UNSW Canberra
Canberra, ACT 2610, Australia
xiao.tan@csiro.au

Changming Sun
CSIRO Mathematics, Informatics and Statistics
11 Julius Avenue
North Ryde, NSW 2113, Australia
changming.sun@csiro.au

Xavier Sirault
CSIRO Plant Industry
Clunies Ross Street,
Canberra, ACT 2601, Australia
xavier.sirault@csiro.au

Robert Furbank
CSIRO Plant Industry
Clunies Ross Street,
Canberra, ACT 2601, Australia
robert.furbank@csiro.au

Tuan D. Pham
Aizu Research Cluster for Medical Engineering and Informatics
The University of Aizu,
Aizu-Wakamatsu
Fukushima 965-8580, Japan
tdpham@u-aizu.ac.jp

ABSTRACT

We present a new method for dense stereo matching based on a tree structural cost volume watershed (TSCVW) and a region combination (RC) process. Given a cost volume as the data cost and an initial segmentation result, the proposed TSCVW method reliably estimates the disparities in a segment by using energy optimization to control plane segmentation and plane fitting. Then the disparities in the incorrectly fitted and occluded regions are refined using our RC process. Experimental results show that our method is very robust to different initial segmentation results and the shape of a segment. The comparison between our algorithm and the current state-of-the-art algorithms on the Middlebury website shows that our algorithm is very competitive.

Categories and Subject Descriptors

I.4.8 [**Scene Analysis**]: Stereo

General Terms

Algorithms

Keywords

Watershed optimization, Region combination, Stereo, Disparity calculation

1. INTRODUCTION

1.1 Previous Work

IVCNZ '12 November 26 - 28 2012, Dunedin, New Zealand

Dense stereo matching is one of the key issues in the area of computer vision. A good 3D reconstruction result requires a very accurate disparity map to be obtained from dense matching. However, errors may occur in the process of stereo matching due to many factors such as noise, distortion, lighting reflection, lack of texture, and occlusion. Stereo matching algorithms should be robust to these factors and be able to provide reliable results. A survey of stereo vision problems and reviews of different types of stereo matching algorithms can be found in [12].

In recent researches, nearly all the top performing algorithms employ segmentation information. Some methods use the segmentation information as a hard constraint [19, 10], and others use it as a soft constraint [15, 20, 4]. These methods give a good estimation to the disparities in the large textureless regions. Moreover, depth discontinuities can be preserved, because depth discontinuities usually occur at the boundaries of the obtained segments. Most of the segment based methods typically have two main steps, namely image segmentation and disparity assignment [18, 1, 17, 10, 23]. In the first step, color segmentation approaches are used in many researches [17, 6, 3]. However, it is hard to obtain a good segmentation for dense stereo matching. If a segment is too small especially if it is in a textureless region, the accuracy of the estimation for the disparities of pixels in the segment may be reduced. However, if a segment is too large, the pixels on different objects with different disparities may be contained in one segment (as indicated by the black circles in Fig. 1(a)). In the second step, the disparities of pixels in a segment are formulated by a disparity model such as a plane [10, 20, 19], a B-spline surface [11, 4], or a single value [23]. The main problem in the segment based algorithms is the use of a predefined model to formulate the disparities of pixels in a segment where the disparities may not follow a specific model due to a false segmentation or due to the curved shape of a surface on an object (as in the region of the white circle in Fig. 1(b)).

Yuichi *et al.* presented a joint method which updates the segmentation and depth iteratively in [16]. Their method uses a Gaussian function to model the color and spatial distribution of the depth of a segment. In their method, the segment boundaries can be finally located at the depth boundaries. However, as over-segmentation is used, it fails at the pixels where the surface has a large slope. Some researches [3, 5] start at an initial over-segmentation and group the

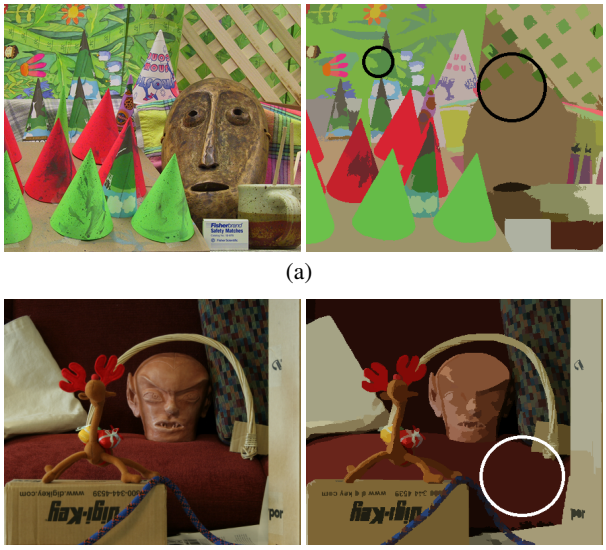

Figure 1: (a) Left image and the segmentation result of the "Cones" dataset. Notice the two objects are incorrectly grouped into one segment. (b) Left image and the segmentation result of the "Reindeer" dataset. Notice the segment of a curved surface on the cushion.

small segments into a larger segment. But the initial segmentation must meet the requirement that no multiple objects are grouped into one segment. Otherwise, the result will be even worse after grouping. As no preliminary information about the scene is available, this requirement cannot be guaranteed.

1.2 Our Approach

In this paper, we use planes to formulate the disparities of regions. Compared with traditional segment based algorithms [20, 19, 10], our scheme splits a stereo image into segments which can be better formulated by planes and then performs plane fitting for such segments. Our scheme starts with an initial segmentation which is obtained from a color segmentation algorithm. Then, segmentation classification is performed. After that, the segment optimization using our new tree structural cost volume watershed (TSCVW) algorithm is performed in the disparity space to give a more reliable plane fitting. As bad pixels may be grouped in the textureless and occluded segments, this will make the fitting results for such regions unreliable. However, no specific scheme has been proposed to handle these unreliable fitting results in previous researches. In our research, a region combination (RC) algorithm detects these regions using a consistency measurement and then gives accurate disparity estimations for these segments.

The rest of the paper is organized as follow. An overview of our algorithm is given in Section 2. Detailed descriptions to our algorithm are given in Sections 3, 4, and 5. Section 6 presents the experimental results. The conclusion is given in Section 7.

2. OVERVIEW OF PROPOSED ALGORITHM

The flow chart of our proposed algorithm is shown in Fig. 2. The blocks of the pre-processing module include cost volume calculation, initial segmentation, pixel classification, and segment classification. The segment optimization includes our tree structural cost volume watershed and an energy based optimization. In disparity refinement, the last stage of our algorithm, a region combination step and belief propagation (BP) are used to further improve the results.

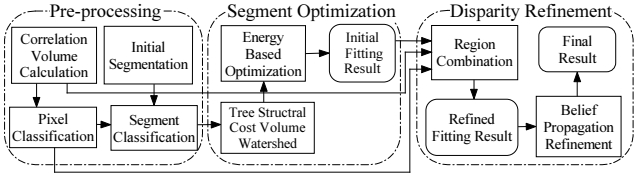

Figure 2: Flow chart of our proposed algorithm.

3. PRE-PROCESSING

Different cost volume calculation methods can be used for obtaining the data cost, such as using the color-weighted measurement [21] and the dissimilarity measurement [2]. Here, we use the cross region based algorithm as described in [22]. The initial segmentation result is obtained from a mean-shift algorithm [6].

In the first step of the pixel classification, visibility check [15] is used to classify pixels into occluded or unoccluded pixels; and then the unoccluded pixels are further classified into reliable or unreliable pixels by a confidence check. Before performing the visibility check, we run BP refinement on the cost volume of the right image and then apply the winner-take-all (WTA) method to obtain a disparity map for the right image. Given the cost values of a pixel p at different disparities, we check its reliability by:

$$\frac{\exp\left(-C_p\left(d_1\right)\right)}{\sum_d \exp\left(-C_p\left(d\right)\right)} > \exp\left(\frac{T_r}{|d_1 - d_2|}\right) \tag{1}$$

where d_1 and d_2 are the best and second best disparity values among all the disparities within disparity search range; $C_p\left(d\right)$ is the value of the cost volume of pixel p at disparity d. T_r is a parameter determined empirically, which is set to -1.7 in our algorithm. The left hand side of Eq. (1), the normalized probability of the best disparity, measures the distinctiveness of the best disparity over all possible disparities. The right hand side of Eq. (1), a threshold controlled by the distance between the best disparity and second best disparity, follows the observation that the pixel with a higher value of $|d_1 - d_2|$ is less likely to be declared as a reliable pixel.

Segment classification is obtained using the percentage of reliable pixels in the segment. If the percentage of reliable pixels is higher than a threshold (e.g., 90%), it is declared as a reliable segment; otherwise it is an unreliable segment.

4. SEGMENT OPTIMIZATION

We perform plane fitting in the disparity space where a specific plane can be represented by parameters a, b, and c:

$$f = ax + by + c \tag{2}$$

where f is the disparity value; a and b control the horizontal and vertical slope of the plane; c is the disparity offset. In our algorithm, we use different schemes to estimate the plane parameters for different types of segments.

4.1 Tree Structural Cost Volume Watershed

This scheme is used to split the reliable segments which contain curved surfaces or surfaces of different objects into subsegments. In each level of the TSCVW, we use the cost volume watershed to

split the segment into two subsegments; and then assign the plane parameters to the subsegments by minimizing an energy function.

The seed regions of the cost volume watershed are generated by a two steps sequential RANSAC algorithm [8] on the disparity values of the reliable pixels in a segment. In the first step of each sequence, we randomly select points in the reliable pixels and calculate the plane parameters. Inlier/outlier identification is done by checking if the distance from a pixel to the plane is smaller than a threshold T_R which is set to 0.5, the maximum rounding error. The parameters of the plane having the maximum number of inliers is chosen as a set of possible plane parameters (PPP). In the second step, for the outliers identified in the first step, we compare its distance to the plane parameterized by the obtained PPP. If the distance is smaller than T_p, the point will be regarded as locating on this plane and will be removed from outliers to inliers. Then, the maximum connected region of inlier pixels is selected to be a seed region corresponding to the obtained PPP in step one. Outliers in the second step are then fed to the next sequence of RANSAC to find another PPP. If the segment contains multiple planes or objects, these outliers will typically be the pixels of another plane or object. The sequential RANSAC stops when a predefined number, N_r, of PPP is reached or no outliers are left.

In our TSCVW, we set N_r to 2. Segments may contain objects having curved surfaces. In such a case, the bad pixels near the ridges may cause large disparity jump. We handle this by using jump cost to penalize jumps and encourage placing the boundary of two planes near the projection of their intersection line. Denote the obtained two seed regions by s_1 and s_2, and denote their corresponding PPPs by ς_1 and ς_2. Let $f_\varsigma(p)$ be the disparity of a plane, ς. The jump cost of the two PPPs at point p is defined as:

$$J_p\left(\varsigma_1,\varsigma_2\right)=\left|f_{\varsigma_1}(p)-f_{\varsigma_2}(p)\right| \tag{3}$$

Let ψ_i be the set of immersed pixels by the plane parameterized by ς_i. Denote the data cost of a pixel p at the disparity f_p be $D_p\big(f(p)\big)$. We define the level of a pixel p at the boundary of an immersed region as follow:

$$\Delta_{\varsigma_i}(p)=D_p\big(f_{\varsigma_i}(p)\big)+\left(J_p\left(\varsigma_1,\varsigma_2\right)-\min_{q\in N(p)\cap\psi_i}J_q\left(\varsigma_1,\varsigma_2\right)\right) \tag{4}$$

where $N\left(p\right)$ is the set of 8-neighboring pixels of pixel p, i is equal to 1 or 2. As $f_\varsigma(p)$ is usually a continuous label, the subpixel estimation is obtained by linear interpolation between two nearest integer labels.

The cost volume watershed starts by setting ψ_1 and ψ_2 to s_1 and s_2 respectively. For a specific water level, the boundary pixel whose level is smaller than the water level will be immersed. As the water level gradually increases, all pixels in the segment will finally be immersed by ψ_1 or ψ_2, and a ridge will be obtained.

Subsegments obtained as described above may over fit noisy pixels or bad pixels. Thus, a plane may be incorrectly segmented into two subsegments. To handle this, we assign the parameters for two subsegments by minimizing an energy function after each step. The energy function is defined as,

$$\begin{aligned}E\left(\varsigma_{\psi_1},\varsigma_{\psi_2}\right)&=E_{\text{plane}}\left(\varsigma_{\psi_1},\varsigma_{\psi_2}\right)\\&\quad+E_{\text{smooth}}\left(\varsigma_{\psi_1},\varsigma_{\psi_2}\right)+E_{\text{mdl}}\left(\varsigma_{\psi_1},\varsigma_{\psi_2}\right)\end{aligned} \tag{5}$$

where $(\varsigma_{\psi_1},\varsigma_{\psi_2})$ are the plane parameters assigned to the two subsegments ψ_1 and ψ_2 respectively. The assignment is one of the three parameter combinations: $(\varsigma_1,\varsigma_1)$, $(\varsigma_2,\varsigma_2)$, and $(\varsigma_1,\varsigma_2)$. The assignment, $(\varsigma_1,\varsigma_2)$, indicates the plane is splitted into two subsegments. The first term is the plane fitting term that encourages the assignment to follow the plane fitting result obtained from the cost volume watershed. The plane fitting term is defined as,

$$E_{\text{plane}}\left(\varsigma_{\psi_1},\varsigma_{\psi_2}\right)=\sum_{p\in s_1}D(f_{\varsigma_{\psi_1}}(p))+\sum_{p\in s_2}D(f_{\varsigma_{\psi_2}}(p)) \tag{6}$$

The second or, the smoothness term, which encourages a smooth boundary between ψ_1 and ψ_2, is defined as,

$$E_{\text{smooth}}\left(\varsigma_{\psi_1},\varsigma_{\psi_2}\right)=\sum_{p\in B(\psi_1,\psi_2)}J_p\left(\varsigma_1,\varsigma_2\right) \tag{7}$$

where $B\left(\psi_1,\psi_2\right)$ is the set of pixels on the boundary between ψ_1 and ψ_2. The third term, the MDL term [4], is introduced to prevent over fitting to small group noisy pixels. It is defined as,

$$E_{\text{mdl}}\left(\varsigma_{\psi_1},\varsigma_{\psi_2}\right)=\begin{cases}\lambda_m & \varsigma_{\psi_1}\neq\varsigma_{\psi_2}\\ 0 & \varsigma_{\psi_1}=\varsigma_{\psi_2}\end{cases} \tag{8}$$

If the assignment of ψ_1 equals to that of ψ_2, it means the segment being processed will not be segmented further. Thus, the TSCVW for this segment stops. If the two assignments are not equal, in the next level of TSCVW we will process the obtained two subsegments using the same scheme as described above.

4.2 Plane Parameter Estimation for Unreliable Segments

As the ratio of the unreliable pixels in the unreliable segment is comparably high, the plane parameters obtained by only using the pixels in the unreliable segment are usually incorrect. Thus, for the unreliable segment, we find a plane with small disparity jumps at the segment boundaries. The plane is found by minimizing the following energy function,

$$E\left(\varsigma\right)=E_\psi\left(\varsigma\right)+\sum_{\varsigma_k}E_{\text{smooth}}\left(\varsigma,\varsigma_k\right),\varsigma\in S_\varsigma \tag{9}$$

where ψ is an unreliable segment; ς is the plane parameter assigned to ψ. S_ς is the set of all plane parameters obtained by using sequential RANSAC on all pixels in ψ. ς_k is the plane parameter of a neighboring segment. The smoothness term $E_{\text{smooth}}\left(\varsigma,\varsigma_k\right)$ is similarly definitied as in Eq. (7). The data term $E_\psi\left(\varsigma\right)$ is

$$E_\psi\left(\varsigma\right)=\sum_{p\in\psi}D(f_\varsigma(p)) \tag{10}$$

We optimize the energy function iteratively using a greedy search. In each iteration, we check all PPPs of each segment and find the one minimizing Eq. (9) and change assignment simultaneously at the end of iteration.

5. DISPARITY REFINEMENT

5.1 Region Combination

Our fitting refinement scheme is based on a region combination process which is motivated by the fact that most occluded regions and unreliable fitted regions are parts of another plane. In this scheme we are trying to find the corresponding planes for the occluded or unreliable fitted regions.

We call a pixel as a bad fitting pixel if it is in an occluded region or its plane fitting result has a large bias with the disparity obtained from the data cost by applying WTA. The segment with a high ratio of bad fitting pixels is more likely to be unreliable. For a given pixel p, we check if it is a bad fitting pixel by,

$$B\left(p\right)=\begin{cases}1 & \text{if }\left|f_{\text{best}}(p)-f_\varsigma(p)\right|>0.5\text{ or }O\left(p\right)=1\\ 0 & \text{otherwise}\end{cases} \tag{11}$$

where $f_{\text{best}}(p)$ is the disparity obtained from the data cost by applying WTA, and $f_{\varsigma}(p)$ is the disparity obtained from plane fitting results. For a pixel p, if the bias between $f_{\text{best}}(p)$ and $f_{\varsigma}(p)$ is larger than the maximum rounding error 0.5 or it is detected as an occluded pixel in the pixel classification step, $O(p) = 1$, it will be declared as a bad fitting pixel, $B(p) = 1$. For each segment, if its bad fitting pixels is higher than a threshold, it will be regarded as a part of another segment. The threshold for determining if a segment should be combined with other segment is given by a determination function. The determination function can be obtained by training or set empirically by users. In our algorithm, the determination function is defined empirically as,

$$R(\psi) = 0.1\log_2(\text{Card}(\psi)/5) \tag{12}$$

where $\text{Card}(\psi)$ is the number of pixels in a segment ψ. The value of the determination function reduces with the size of the segment, which follows the fact that a smaller segment is more likely to be a part of another segment. We test different types of determination functions, such as linear, quadratic, and logarithmic functions. The logarithmic function gives the best result.

In the region combination process, for a segment ψ, we find its corresponding segment, ψ_C, by:

$$\psi_C = \arg\min_{\psi_N}(F_{\psi}(\varsigma_{\psi_N})) \tag{13}$$

where $F_{\psi}(\varsigma_{\psi_N})$ is defined as

$$F_{\psi}(\varsigma_{\psi_N}) = w_{\text{cost}}(\varsigma_{\psi_N}) + w_{\text{p}}(\varsigma_{\psi_N}) + w_{\text{c}}(\varsigma_{\psi_N}) \tag{14}$$

$N(\psi)$ is the set of neighboring segments which will be defined later; ς_{ψ_N} is the plane parameters of segment ψ_N. $w_{\text{cost}}(\varsigma_{\psi_N})$, the data term, is defined as the average over the data cost of all unoccluded pixels in segment ψ. If pixels in ψ are all occluded pixels, we do not use the data term for this segment. According to the Gestalt grouping principle [21, 14], if a segment is closer to ψ in the spatial distance or in the color distance, it is very likely that they should be grouped together. The strength of grouping by proximity is measured by a spatial term, $w_{\text{p}}(\varsigma_{\psi_N})$, and a color term, $w_{\text{c}}(\varsigma_{\psi_N})$, which are defined as $\frac{\Delta_{\text{p}}}{\gamma_p}$ and $\frac{\Delta_{\text{c}}}{\gamma_c}$ respectively. γ_{p} and γ_{c} control the weights of the spatial and color terms. The distance between two segments in the image space, Δ_{p}, is defined as the minimum distance between two pixels which belong to the two segments respectively. The color difference, Δ_{c}, is defined as the Euclidean distance of the colors of the two segments in the RGB space. The color of a segment is defined as the average color of all pixels in the segment.

The set of neighboring segments includes all segment whose spatial distance to ψ is smaller than 2. Some segments may be occluded by foreground objects and their corresponding segments are therefore far from them. To handle this, we increase spatial search range to find the correct corresponding segment. For the segments whose spatial distance to ψ within the range $(2, N_{\text{s}})$, we use the method proposed in [5] to check if it is connected with ψ in 3D space. The segments passing the 3D connectivity check may be occluded by foreground object, so we put them into the set $N(\psi)$.

As some segments belonging to the same plane should be grouped together, we perform region combination iteratively. In iterations, we record the minimum value of $F_{\psi}(\varsigma_{\psi_N})$ of a segment in Eq. (14). For each iteration, if the new assignment of ψ_C gives a smaller value of $F_{\psi}(\varsigma_{\psi_N})$ than the recorded minimum value, we take the new assignment and record the new minimum value. The region combination stops, until no assignment changes.

5.2 BP Refinement

Some small groups of pixels which are segmented incorrectly may be regarded as noise and still remain in the final fitting result. To handle this, we use the hierarchical belief propagation (BP) as proposed in [7] to further improve our result. Our result is formulated by,

$$U_p(f_p) = \begin{cases} 0 & \text{if } f_p = f_- \text{ or } f_p = f_+ \\ \lambda_{\text{fit}} & \text{otherwise} \end{cases} \tag{15}$$

where f_- ($f_- < f$) and f_+ ($f_+ > f$) are the two nearest integer labels of f_p which is a continuous label obtained from the fitting result of a pixel. λ_{fit} is a penalty to the disparities violating the plane fitting result. The energy function which the hierarchical BP will optimize is given by,

$$E_{\text{BP}}(f_p) = \sum_p \lambda_{\text{BP}} D_p(f_p) + \sum_p U_p(f_p) + \sum_{(p,q)\in N} V(f_p, f_q) \tag{16}$$

where V is: $V(f_p, f_q) = \min(|f_p - f_q|, d_{\text{BP}})$. As each pixel is connected with 4-neighboring pixels in V, we set λ_{fit} to 4 times of d_{BP}, the truncation in V, to make large disparity jumps introduced by any bad fitting result to be corrected by V.

6. EXPERIMENTAL RESULTS

In order to test the effectiveness of our new method, we implemented our algorithm using Visual C++ 2008 and run the algorithm on images from the Middlebury website [13] and other sources. In this section, we firstly test the performance of our TSCVW algorithm on different initial segmentation results. After that we give the intermediate and final results for several stereo datasets.

6.1 Performance of TSCVW

In this section, we test the proposed TSCVW algorithm using different initial segmentation results. In initial segmentation results, the neighboring objects having similar color or texture are grouped into one segment. Traditional segmentation based stereo matching algorithms usually fail in such segments. Our algorithm solves this by using TSCVW to discriminate different objects in such segments. We test our TSCVW with three different initial segmentation results from over to under segmentations using the mean-shift segmentation algorithm [6]. The results in Fig. 3 show that our TSCVW algorithm performs very well with different initial segmentation results, especially in segments which contain multiple objects.

6.2 Comparison of Results

In this section, we first test the accuracy of our algorithm using the Middlebury stereo dataset. We keep the parameters constant and compare our results with that of other algorithms. The parameter values used in the mean-shift segmentation and BP are the default parameters. As mentioned above, our method is robust to the initial segmentation results. The parameters used in our algorithm are set to constant values of $T_p = 1.5$, $\lambda_m = 1200$, $N_s = N_d/8$, $\gamma_{\text{p}} = 80$, and $\gamma_{\text{c}} = 33$. The running time on the "Tsukuba" dataset is approximately 95 seconds on a 3.0 GHz PC with 4GB RAM. The TSCVW takes about 80 seconds, the region combination process takes about 5 seconds, and the BP refinement takes about a further 10 seconds. The quantitative evaluation for the Middlebury images pair is listed in Table 1, where the error threshold is one pixel of disparity. Our algorithm is ranked at the 9th among all the algorithms submitted. When the error threshold is two pixels of disparity, our algorithm is ranked at the 5th. And it is ranked 2nd among all segment based methods.

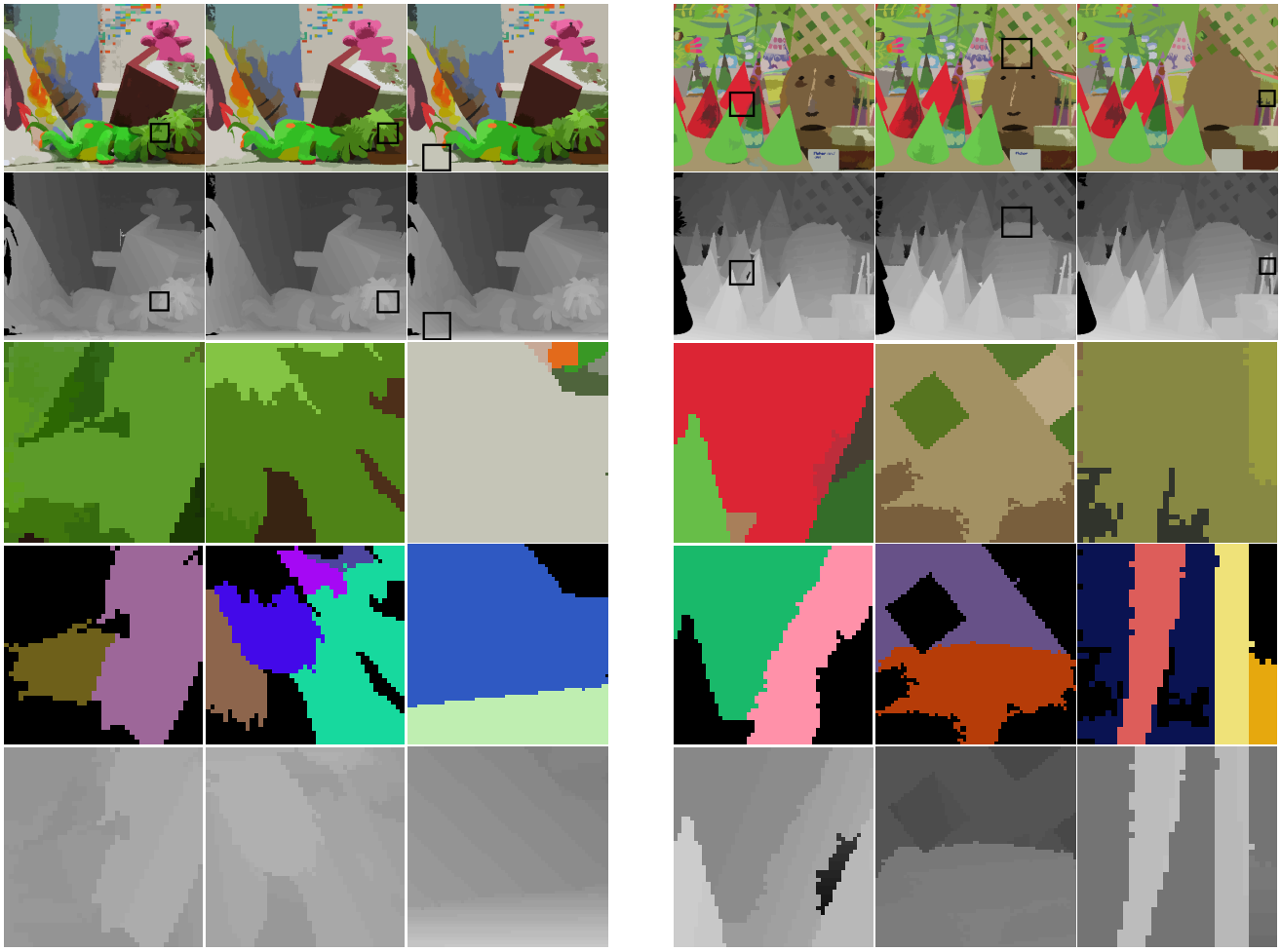

Figure 3: Results of TSCVW for the "Teddy" and "Cones" datasets respectively based on different initial segmentation results; First row: different initial segmentation results from over to under segmentations; Second row: disparity maps obtained from TSCVW; Third row: close view of a part of the initial segmentation results; Fourth row: segmentation results obtained from TSCVW; Last row: close view of the disparity maps obtained from TSCVW.

7. CONCLUSIONS

We further test our algorithm on more stereo datasets. To test the robustness of our algorithm, the parameters used in these datasets are the same as in the previous experiment. The next two datasets for further experiment are also chosen from the Middlebury datasets [9]. The last two are a close view of a rock and a bottom view of a tower. The results are shown in Fig. 4. The experiments show that our method gives a great improvement to the disparity map obtained directly from the data cost.

In this paper, we proposed a new dense stereo matching algorithm. The algorithm incorporates our TSCVW and RC schemes together with BP. Our TSCVW algorithm which performs plane segmentation and plane fitting simultaneously gives an accurate estimation to disparities of the pixels in segments which may contain multiple objects or contain curved surfaces. The proposed RC scheme gives an accurate estimation to the disparities in the occluded region and the unreliable fitting region by finding their corresponding segments. The experiment results show the effectiveness of our method. The comparison between our algorithm and the current state-of-the-art algorithms has been carried out and good results obtained. We believe our work is an attempt on the implementation of watershed on the cost volume and energy optimization and it would inspire further research in this area.

8. ACKNOWLEDGEMENT

We thank Chao Zhang for his precious comments. Tan was partially supported by the China Scholarship Council. Sun was partially supported by the CSIRO's Transformational Biology Capability Platform.

9. REFERENCES

[1] S. Baker, R. Szeliski, and P. Anandan. A layered approach to stereo reconstruction. *CVPR*, pages 434–441, 1998.

[2] S. Birchfield and C. Tomasi. A pixel dissimilarity measure that is insensitive to image sampling. *PAMI*, 20(4):401–406, 1998.

[3] M. Bleyer and M. Gelautz. A layered stereo algorithm using image segmentation and global visibility constraints. *ICIP*, 5:2997–3000, 2004.

[4] M. Bleyer, C. Rother, and P. Kohli. Surface stereo with soft segmentation. *CVPR*, pages 1570–1577, 2010.

Table 1: The result of our algorithm and other segment based algorithms on the Middlebury stereo datasets.

Algorithms	Tsukuba			Venus			Teddy			Cones		
	nonocc	all	disc	nonocc	all	disc	nonocc	all	disc	nonocc	all	disc
AdaptingBP	1.11	1.37	5.79	0.10	0.21	1.44	4.22	7.06	11.80	2.48	7.92	7.32
OurMethod	1.08	1.55	5.57	0.19	0.39	1.83	4.18	5.96	10.70	3.42	8.80	9.20
SurfaceStereo	1.28	1.65	6.78	0.19	0.28	2.61	3.12	5.10	8.65	2.89	7.95	8.26
PlaneFitBP	0.97	1.83	5.26	0.17	0.51	1.71	6.65	12.10	14.70	4.17	10.70	10.60
OverSegmBP	1.69	1.97	8.47	0.50	0.68	4.69	6.74	11.90	15.80	3.19	8.81	8.89

Figure 4: Performance of the TSCVW on the surface boundary. Top row: Left images of datasets; Middle row: Initial disparity maps obtained by applying winner-take-all to the data costs; Last row: Results obtained from our algorithm.

[5] M. Bleyer, C. Rother, P. Kohli, D. Scharstein, and S. Sinha. Object stereo - joint stereo matching and object segmentation. *CVPR*, pages 3081–3088, 2011.

[6] C. Dorin and M. Peter. Mean shift: A robust approach toward feature space analysis. *PAMI*, 24(5):603–619, 2002.

[7] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient belief propagation for early vision. *IJCV*, 70(1):41–54, 2006.

[8] M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.

[9] H. Hirschmuller and D. Scharstein. Evaluation of cost functions for stereo matching. In *CVPR*, pages 1–8, 2007.

[10] A. Klaus, M. Sormann, and K. Karner. Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure. *ICPR*, 3:15–18, 2006.

[11] M. H. Lin and C. Tomasi. Surfaces with occlusions from layered stereo. *PAMI*, 26(8):1073–1078, 2004.

[12] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *IJCV*, 47(1):7–42, 2002.

[13] D. Scharstein and R. Szeliski. http://www.vision.middlebury.edu/stereo/, 2011.

[14] J. Shi and J. Malik. Normalized cuts and image segmentation. *PAMI*, 22(8):888–905, 2000.

[15] J. Sun, Y. Li, S. B. Kang, and H. Y. Shum. Symmetric stereo matching for occlusion handling. *CVPR*, 2:399–407, 2005.

[16] Y. Taguchi, B. Wilburn, and C. L. Zitnick. Stereo reconstruction with mixed pixels using adaptive over-segmentation. In *CVPR*, pages 1–8, 2008.

[17] H. Tao, H. S. Sawhney, and R. Kumar. A global matching framework for stereo computation. *ICCV*, 1:532–539, 2001.

[18] J. Y. A. Wang and E. H. Adelson. Representing moving images with layers. *ITIP*, 3(5):625–638, 1994.

[19] Z. F. Wang and Z. G. Zheng. A region based stereo matching algorithm using cooperative optimization. *CVPR*, pages 1–8, 2008.

[20] Q. Yang, L. Wang, R. Yang, H. Stewénius, and D. Nistér. Stereo matching with color-weighted correlation, hierarchical belief propagation, and occlusion handling. *PAMI*, 31(3):492–504, 2008.

[21] K. J. Yoon and I. S. Kweon. Adaptive support-weight approach for correspondence search. *PAMI*, 28(4):650–656, 2006.

[22] K. Zhang, J. Lu, and G. Lafruit. Cross-based local stereo matching using orthogonal integral images. *CSVT*, 19(7):1073–1079, 2009.

[23] C. L. Zitnick and S. B. Kang. Stereo for image-based rendering using image over-segmentation. *IJCV*, 75(1):49–65, 2007.