Segment-Tree Stereo Matching Algorithm Based on Improved Matching Costs

Cheng Han, Shiyu Lu, Longbin Jin, Shan Jiang and Hua Li

Abstract— **Matching cost aggregation is one of the oldest and still popular methods for stereo correspondence. While effective and efficient, cost aggregation methods typically aggregate the matching cost by summing/averaging over a user-specified, local support region. This is obviously only locally-optimal, and the computational complexity of the full-kernel implementation usually depends on the region size. In order to improve aggregation accuracy, we propose a segment-tree stereo matching method with improved matching costs. A reasonable weight for the matching process is assigned by introducing color and gradient multi-dimensional information components in order to overcome inaccuracies of weak texture regions. Next, similar regions are merged, whereby pixel points belonging to the same parallax consistency are merged with the corresponding generation tree. Finally, depth and color information are used for the tree reconstruction, while a color-depth weight is adopted in order to enhance the structure of the tree. Performance evaluation on 19 Middlebury data sets shows that the proposed method is comparable to previous state-of-the-art aggregation methods in disparity accuracy and processing speed.**

Index Terms—**Segment-tree, matching cost, color-depth weight, stereo matching**

I. Introduction

Extracting three-dimensional information from images is key in the field of machine vision. In particular, stereo matching is a major research topic within image extraction techniques, and has been widely applied in intelligent vehicles [1], robots [2], navigation [3], medical diagnosis [4], and drones [5]. For binocular parallax matching, the disparity value is generally calculated using information extracted from the two left and two right frames taken at the same horizontal line within the scene. A study by Scharstein divided the binocular matching algorithm into a global and local algorithm [6] in order to optimize the balance between matching accuracy and efficiency. Global algorithms have been able to improve results by minimizing the output of the energy equations, which tends to be computationally complex. Local algorithms, in contrast, are more efficient, yet they are less accurate.

Yang proposed a non-local cost-aggregation algorithm based on bilateral filtering [7]. The algorithm pre-structures the entire reference picture into a minimum spanning tree (MST) model, and then performs tree-based filtering [8] in order to decrease the complexity. However, there is a weight unevenness distribution problem in the cost calculation process. Mei proposed a segment-tree (ST) stereo matching algorithm, an MST variant based on image segmentation, whereby segment information is introduced into the non-local cost-aggregation framework [9]. More specifically, a two-step filtering method updates the cost at the second step of filtering. Weight information is adjusted using the depth information. In addition, Zhang [10] proposed a cross-scale framework to improve existing local and non-local aggregation algorithms. In particular, for ST-based algorithms, the spanning tree structure is directly determined using the image segment strategy. Furthermore, Ma proposed the application of a constant time weighted median filter in stereo matching, effectively improving the accuracy of image edges [11]. A three-dimensional cost-aggregation stereo matching algorithm was proposed by Li for multiple spanning trees [12]. In particular, the MST does not exhibit smoothing and mismatching due to the application of three-dimensional coordinates to define disparity maps combined with multiple minimum spanning trees and block matching [13]. Hamzah proposed a stereo matching algorithm for iterative guided filtering and image segmentation and combined numerous algorithms to determine the optimal matching effect [14]. However, the original proposed segment strategy neglects the parallax consistency assumption, which inevitably leads to the roughness of the initial segment result [15].

In the current study, we first calculate the inappropriate phenomenon of weight within the weak texture region, and introduce the color and gradient multi-dimensional information components for the construction of the cost function. Moreover, a new grouping strategy is proposed for the algorithm framework based on the segment tree. Experimental results demonstrate that the proposed method obtains optimal matching results while maintaining the time efficiency of the algorithm.

II. Improved Method

A. Improved initial matching cost function

The Absolute Deviation Gradient (AD-Gradient) similarity measure is generally adopted for the matching cost function of the traditional spanning tree structure. This measure is the first step of the original spanning tree algorithm. Assume that $C_d^{AD}(p)$ represents the cost of the

Manuscript received July 2, 2019; revised June 6, 2020. This work is supported by the Youth Science Funds of the National Natural Science Foundation of China (61602058), the Key Science and Technology Program of Jilin Province, China (20170203003GX, 20170203004GX, 20180201-069GX) and the Technology Development Plan of Jilin Province, China (20190103031JH, 20190201255JC).

Cheng Han, Shiyu Lu and Longbin Jin are with the School of Computer Science and Technology, Changchun University of Science and Technology.(e-mail: hancheng@cust.edu.cn, rain20500@126.coma, and 785120668@qq.com)

color component of pixel point p at disparity d, and $C_d^{GradX}(p)$ represents the cost of the horizontal direction gradient component of pixel point p at disparity d. Let α denote the user-defined adjustment factor, which acts as a contraint for the color and horizontal gradient components. In addition, τ_{col} and τ_{grdx} represent the truncation thresholds used to limit the negative effects of outliers for the color and gradient components, respectively. $I^{left}(p)$ represents the pixel point p of the left image. $I_i^{right}(p')$ represents the pixel point p' of the right image. The total cost $C_d(p)$, and the costs of the color and horizontal gradient components can be expressed as follows:

$$C_d(p) = \alpha \cdot C_d^{AD}(p) + (1-\alpha) \cdot C_d^{GradX}(p), \quad (1)$$

$$C_d^{AD}(p) = \min\left(\frac{1}{3}\sum_{i=R,G,B}\left|I_i^{left}(p) - I_i^{right}(p')\right|, \tau_{col}\right), \quad (2)$$

$$C_d^{GradX}(p) = \min\left(\left|\nabla_x I^{left}(p) - \nabla_x I^{right}(p')\right|, \tau_{grdx}\right). \quad (3)$$

Taking into account the presence of variations in the vertical direction of the scene, combined with the advantages of the AD-Gradient method, the matching function can be improved with the addition of a new weight component in the cost function. Let $C_d^{GradY}(p)$ represent the gradient cost amount in the vertical direction, and α and β denote the user-defined adjustment factors used to constrain the component items. The cost of the vertical direction gradient component can be determined as follows:

$$C_d(p) = (1-\alpha-\beta) \cdot C_d^{AD}(p) + \alpha \cdot C_d^{GradX}(p) + \beta \cdot C_d^{GradY}(p) \quad (4)$$

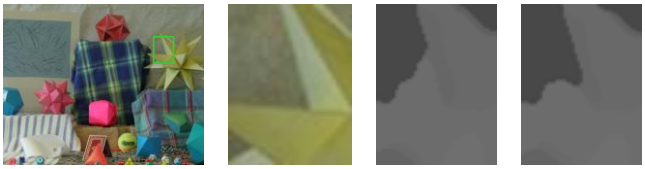
$$C_d^{GradY}(p) = \min\left(\left|\nabla_y I^{left}(p) - \nabla_y I^{right}(p')\right|, \tau_{grdy}\right). \quad (5)$$

Fig. 1 presents an experimental comparison of the traditional and proposed cost functions applied to the Moebius test image. The disparity map results of Fig. 1(c) (original cost function) and Fig. 1(d) (proposed vertical gradient cost function) demonstrate that the proposed method performs better than the original cost function in terms of the processing of vertical edge details within the image.

However, the cost of pixels within insufficient texture regions is close to zero. Thus, such pixels make a minimal contribution to the cost aggregation process compared with regions where the texture is sufficient. In order to overcome the defects in the weak texture region, a transformation can be performed based on the cost and gradient functions as follows:

$$C_d'(p) = \frac{1}{1+e^{-C_d(p)}}. \quad (6)$$

The optimized matching cost function converts the cost of pixels that may exhibit zero values in the traditional cost function into positive non-zero terms. Equation (6) is a monotonically increasing Sigmoid function. Via this transformation, the contribution values of pixel points in the weak texture region can be raised, thereby allowing for the differentiation of different pixel points in the weak texture region. Fig. 2 compares the disparity map between the stereo matching algorithm and the original algorithm following the initial cost function transformation. A comparison of the disparity map results of Fig. 2(b) (original cost function) and Fig. 2(c) (improved cost function) demonstrates that the proposed method has a minimal effect on the strong texture region, yet it can effectively deal with the weak texture region, thus improving the matching result.



(a) (b) (c) (d)

Fig. 1. (a) Moebius test image, (b) magnified area of test image, (c) contrast disparity map of the original cost function method, and (d) contrast disparity map of the improved method.

B. Segment tree grouping strategy

The cost aggregation of the segment tree is based on the construction process of the minimum spanning tree, where all pixels in the image are connected by edges, and the sum of the weights of the edges between connected pixels is minimized. This configuration regards the reference image as an undirected connected graph, denoted as $G = (V, E)$. The term V represents a set of pixel points within the image, while E represents a set of edges for each pair of adjacent pixels within the image. Let s and r be a pair of adjacent pixels, then $I(s)$ and $I(r)$ are the color information of pixel points s and r. Its weight w_e is determined as follows:

$$w_e = w(s,r) = \left|I(s) - I(r)\right|. \quad (7)$$

a) (b) (c)

Fig. 2. Comparison of disparity maps obtained from the cost function following optimization. (a) Lampshade test image, (b) original cost function method, and (d) improved method.

Yang proposed taking the spanning tree as the minimum spanning tree as edges with smaller weights are less likely to spin deep edge regions. In order to obtain accurate aggregation results, such edges need to be selected during the process of the spanning tree construction to determine the spanning tree with the smallest sum of edge weights. In the minimum spanning tree structure, for any two pixel points p and q, only one path existing connects the two pixels. The distance $D(p,q)$ between the two is determined by the sum of the edge weights along the path. The similarity measure functions for two pixel points p and q can then be calculated using the distance value as follows:

$$D(p,q) = \sum_{e \in E} w(p,q), \quad (8)$$

$$S(p,q) = \exp\left(-\frac{D(p,q)}{\sigma}\right). \quad (9)$$

where σ is a user-defined adjustment factor. By

definition, if the colors between two points are closer and the distance $D(p,q)$ is greater, the similarity between the two nodes will be lower. In the tree structure, pixels closer to point p or pixels similar to point q will obtain a greater weight.

The segment tree structure proposed by Mei [9] can be divided into three steps, as demonstrated in Fig. 3. Note that for simplicity, only six nodes are used in the figure. The three steps can be described as follows:

1) Initialization (Fig. 3(a)): The edges in edge set E are sorted into ascending order according to the defined weights, and each node in point set V is regarded as a graph subtree.

2) Grouping (Fig. 3(b)): The edge set E is fully scanned, and the two subtrees are connected to form a larger new subtree, whereby each subtree is the MST of the corresponding segment. Let v_p and v_q denote the nodes that are connected by edge $e_j \in E$, if v_p and v_q belong to different subtrees, the weights w_{e_j} of the edges connected between the subtrees satisfy the criteria proposed in the literature [16]. More specifically, if $Int(T_p)$ and $|T_p|$ represent the maximum edge weight and region size in subtree T_p, respectively, and k is a constant parameter specified by the user, then w_{e_j} satisfies:

$$w_{e_j} \le \min\left(Int(T_p) + \frac{k}{|T_p|}, Int(T_q) + \frac{k}{|T_q|} \right). \tag{10}$$

3) Link (Fig. 3(c)): Following the grouping step, a linking step is performed whereby subtrees T_p and T_q are merged into a complete ST structure.

The ST method can be thought of as a segment algorithm, as it creates an MST for the entire graph, as well as for each segment. It can be inferred that the ST structure directly determines the final result of the cost aggregation. This means that the grouping step plays an important role in building the ST structure. However, the original segment algorithm usually performs insufficiently due to several limitations. First, k is not able to efficiently control the expected region size, resulting in a poor segment effect for the original segment algorithm. The larger the value of k, the looser the threshold at the beginning of the grouping and the insufficient segment of the depth discontinuity. Conversely, for smaller values of k, the initial grouping is more efficient, yet it will lead to excessive segments. In this case, small segments with too low a pixel value cannot obtain the required support weights from adjacent regions. Second, increases in the size of region $|T|$ result in a sharp decrease in the $k/|T|$ term, as well as the ability to distinguish the same type of region. Meanwhile, when $|T|$ increases, the execution of grouping decisions is greatly constrained. The original grouping method easily divides areas belonging to the parallax into different regions. Based on the above limitations, a grouping strategy to construct a new ST structure is proposed. In particular, we let $k = 0.06$. The proposed optimized method can provide a more constrained threshold at the beginning of grouping, as well as a relaxation threshold when merging regions. This forces the merging of similar regions to better satisfy the different consistency assumption as follows:

$$w_{e_j} \le \min\left(\begin{matrix} Int(T_p) + k \cdot \log(1+|T_p|), \\ Int(T_q) + k \cdot \log(1+|T_q|) \end{matrix} \right). \tag{11}$$

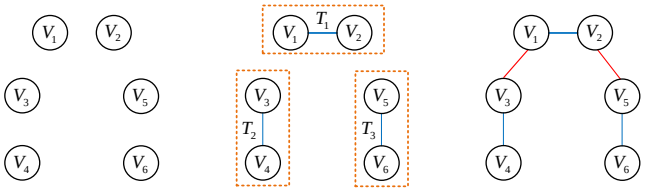


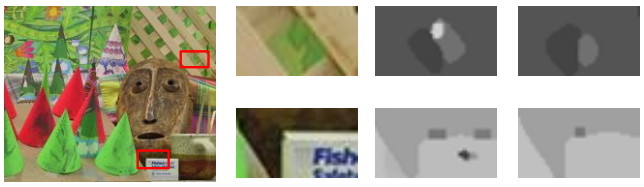
Fig. 3. Flow of building a segment tree proposed by Mei. (a) Initialization, (b) grouping, and (c) link.

C. Enhanced segment based on color and depth information

The original MST algorithm only uses the color information in the edge weight function to construct the tree. This leads to the lack of 3D cues in the matching process, leading to a greater number of mismatches in the disparity map. Based on this, we fuse the depth and color information to reconstruct the tree. Moreover, the algorithm is segmented a second time by color and depth weights to enhance the structure of the tree. Studies have demonstrated that the segment effect can be effectively improved [17] when two clues are used as features. More specifically, the weight of edge e connected by pixels s and r can be updated. Δ_I normalize I to the range of [0,1], Δ_D normalize I to the range of [0,1]. We introduce the weight term λ to balance the proportion of color and depth. Depth image D then be determined via Equation (12):

$$w_e = \lambda \frac{|I(s)-I(r)|}{\Delta_I} + (1-\lambda)\frac{|D(s)-D(r)|}{\Delta_D}. \tag{12}$$

Fig. 4 displays the disparity map results calculated based on the color and depth information method. By comparing the disparity map results of the initial method (Fig. 4(c)) and the method using the depth information component (Fig. 4(d)), it can be seen that our proposed method is more reliable when processing the details in the depth discontinuous region.



(a) (b) (c) (d)

Fig. 4. Disparity map based on color and depth information. (a) Cones test image, (b) intercepting enlarged area, (c) initial method, (d) method using depth information component.

III. RESULTS AND DISCUSSION

A. *Experimental results and performance analysis*

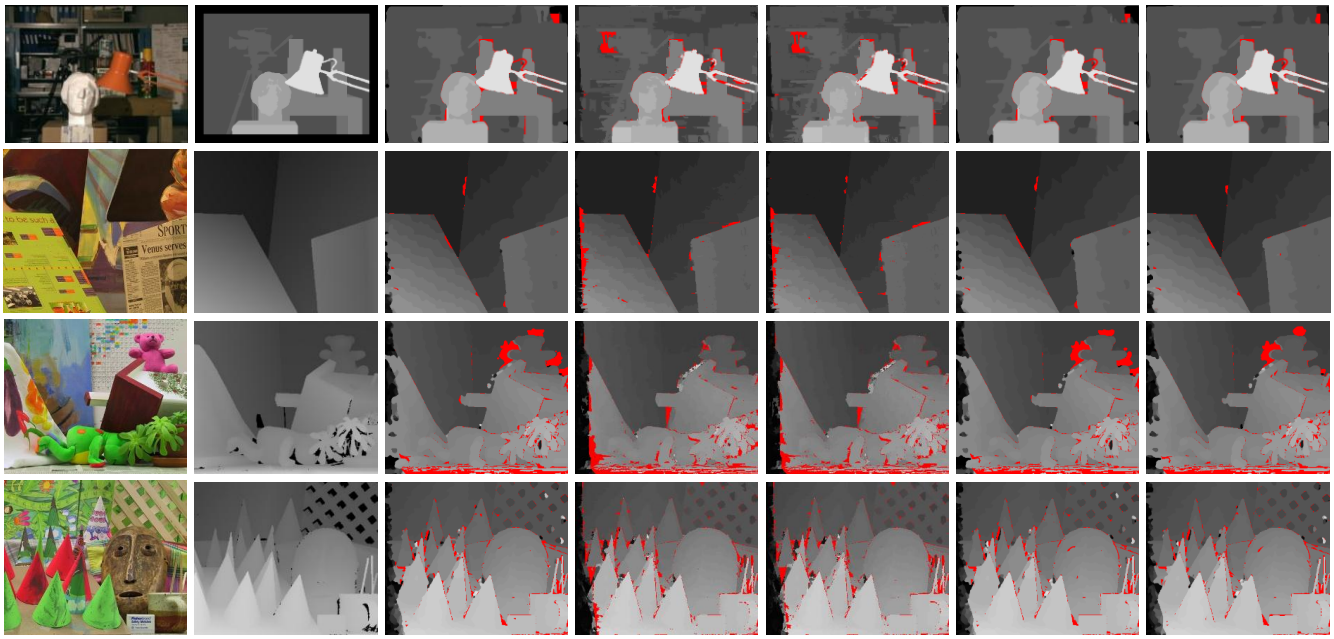
To evaluate the effectiveness of the proposed algorithm, experiments were performed using standard test image pairs obtained from the Middlebury Data Platform website [18]. The test images used in the study were corrected for camera distortion using polar line correction. The algorithm was as follows:

$$R = \frac{1}{N}\sum_{(x,y)}\left(\left|d_C(x,y) - d_T(x,y)\right| > \delta_d\right). \quad (13)$$

where N is the total number of pixel points in the disparity map, $d_C(x,y)$ is the disparity value at point (x,y) on the disparity map, $d_T(x,y)$ is the standard disparity value corresponding to point (x,y), and δ is the error threshold. When the error threshold is equal to 1, the current matching point in the disparity map differs from the real disparity map by more than one pixel, and the point is set as the mismatched pixel.

Algorithm comparison tests were performed on the Tsukuba, Venus, Teddy, and Cones images using four sets of standard stereo images from the Middlebury Dataset Platform (http://vision.middlebury.edu). The parallax search ranges of the four groups of images are [0, 15], [0, 19], [0, 59], [0, 59], with corresponding parallax scaling scales of 16, 8, 4, and 4, respectively. Fig. 5 compares the disparity maps obtained by the proposed algorithm with algorithms from the literature (ST-2, CS-NL, CS-ST and IST-2 algorithms) on four sets of test images. The mismatched pixels in the non-occlusion regions are marked in red. It can be observed that the proposed algorithm exhibits fewer mismatches amongst all algorithms. The CS-NL and CS-ST algorithms exhibit more error points in the left camera area of the Tsukuba image. However, in the bear region of the Teddy image, the two aforementioned algorithms, along with our proposed algorithm, achieve improved results compared with the ST-2 and IST-2 algorithms. In addition, the processing of the object edges is more refined. Due to the effective information provided by the proposed algorithm cost function, the matching results obtained from the weak texture region are more accurate compared with the literature algorithms. Such regions include the table lamp area of the Tsukuba image and the wooden frame area of the Cones image.

Table I compares the data of the proposed and literature algorithms for an error threshold of 1. All optimal values are highlighted in bold. The percentage of the erroneous pixels in the Nonocc/All/Disc. regions are used to evaluate the performance of the algorithm. The proposed algorithm performs best for all regions within the Venus, Teddy, and Cones images. This demonstrates the high accuracy of the proposed grouping strategy, as well as the effectiveness of the processing of edge regions and the resultant smoother image. In addition, based on the overall performance of the algorithms in the non-occlusion region, the proposed algorithm is closer to the disparity value of the corresponding region in the real disparity map. Moreover, the disparity map obtained in the weak texture region generally exhibits fewer mismatch points.



(a) (b) (c) (d) (e) (f) (g)

Fig. 5. Comparison of four groups of image experiments. Pixels with erroneous disparities are marked in red. (a) Test image, (b) actual disparity map, (c) ST-2 algorithm disparity map, (d) CS-NL algorithm disparity map, (e) CS-ST algorithm disparity map, (f) IST-2 algorithm disparity map, (g) proposed algorithm disparity map.

TABLE I
MISMATCH RATE OF LITERATURE AND PROPOSED ALGORITHMS.

Algorithm	Tsukuba			Venus			Teddy			Cones			Avg.Nonocc
	Nonocc	All	Disc	Nonocc	All	Disc	Nonocc	All	Disc	Nonocc	All	Disc	
ST-2	2.04	2.71	8.96	0.43	1.01	5.35	6.96	13.91	16.53	3.28	11.28	9.34	3.18
CS-NL	2.96	3.61	10.44	1.62	2.64	12.68	7.47	16.09	19.47	4.72	14.36	13.18	4.19
CS-ST	3.10	3.84	11.01	1.46	2.56	12.49	7.36	16.15	19.76	4.74	14.50	13.43	4.16
IST-2	**1.68**	**2.34**	**7.92**	0.41	1.01	5.05	6.84	13.84	16.25	3.22	11.25	9.08	3.04
Proposed	1.72	2.37	7.99	**0.34**	**0.69**	**3.98**	**6.25**	**13.06**	**14.83**	**3.04**	**10.72**	**8.61**	**2.84**

TABLE II
COMPARISON OF ALGORITHM RUNNING TIMES.

Algorithm	Tsukuba	Venus	Teddy	Cones	Avg. Time(s)
ST-2	0.32	0.52	0.99	1.01	0.71
CS-NL	0.45	0.71	1.42	1.45	1.01
CS-ST	0.43	0.70	1.29	1.29	0.93
IST-2	0.31	0.52	0.98	0.98	**0.70**
Proposed	0.37	0.61	1.17	1.18	**0.83**

In order to verify the time complexity of the proposed algorithm, all algorithms are run on the same configured machine. Runtime comparisons for ST-2, CS-NL, CS-ST, IST-2 and the proposed algorithms on four sets of standard test images are reported in Table II . All optimal values are highlighted in bold. It can be seen that the ST-2 and IST-2 algorithms exhibit the fastest execution time, while the CS-NL algorithm has the longest execution time. The proposed algorithm achieves an equilibrium in terms of time efficiency and matching accuracy. The results in Table II demonstrate that although the running time of the proposed algorithm is not optimal, a higher matching accuracy is observed. Moreover, unlike the ST-2 and IST-2 algorithms, the algorithm efficiency is not excessively reduced.

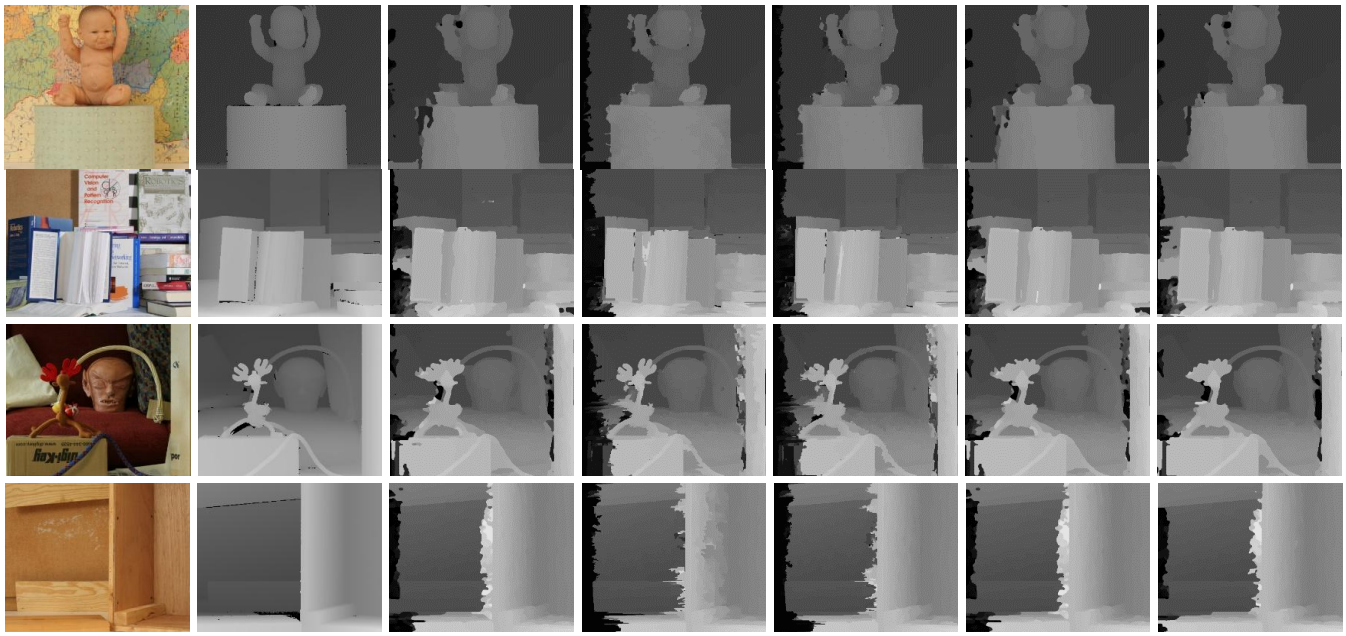
Further tests were performed on 15 stereo image pairs from the Middlebury dataset. Fig. 6 compares the disparity map of a portion selected from 15 sets of stereo image pairs.

There exists a large number of weakly textured regions in the Baby1 and Wood1 images. Due to the high similarity levels between pixels in this part of the region, ambiguous matching is likely to occur during the calculation process. The proposed new cost function can effectively provide reliable information support for matching and fixing incorrect matching points. In addition, thanks to the reasonable grouping of the segment tree structure, the proposed algorithm can segment similar regions in an image at a greater accuracy compared with other regions. It can be observed from the disparity maps of the Books and Reindeer images that the proposed algorithm exhibits fewer false matching points, with smoother and more accurate images compared with other algorithms.

The experimental results are shown in Table III , whereby only the mismatch rate in the non-occlusion region is recorded. All optimal values are highlighted in bold. The average mismatch rate of the proposed algorithm is slightly lower than that of the CS-BF algorithm. However, the proposed algorithm performs better than the CS-BF algorithm in several images (e.g. the Aloe and Cloth3 images). Moreover, the average mismatch rate of the proposed algorithm is lower than those of the other algorithms. The average mismatch rates of the ST-2 and IST-2 algorithms are approximately equal, while the CS-NL algorithm performs the worst. Although the proposed algorithm performs poorly for several stereo image pairs, it outperforms the other algorithms in most stereo image pairs. In terms of time efficiency, the ST-2 and IST-2 algorithms exhibit faster runtimes compared to the remaining algorithms. However, the CS-NL and CS-ST algorithms perform a greater amount of cross-scale space calculations, and thus complexity levels increase, leading to longer running times. According to the experimental results, the proposed algorithm is able to produce high-precision disparity map results by limiting excessive running times.

IV. CONCLUSION

In the current study, we present a novel matching costs method for stereo matching. First, in order to determine a reasonable weight distribution in the matching process, multi-dimensional color and gradient features are introduced to enhance the appearance modeling. In addition, a new grouping strategy is proposed to divide the pixels that conform to the parallax consistency into the same spanning tree. Preliminary results demonstrate the proposed method achieves promising results in terms of aggregation accuracy and efficiency on the Middlebury dataset. First, we would like to test our method with more challenging outdoor stereo data sets, such as the KITTI Vision Benchmark [19]. Second, we would like to test with various segmentation methods, since the performance of the algorithm is closely related to the segmentation results. Finally, we would like to extend this method to perform more general edge-preserving image processing tasks.



(a) (b) (c) (d) (e) (f) (g)

Fig. 6. Comparison of four groups of image experiments. (a) Test image, (b) real disparity map, (c) ST-2 algorithm disparity map, (d) CS-NL algorithm disparity map, (e) CS-ST algorithm disparity map, (f) IST-2 algorithm disparity map, and (g) proposed method.

TABLE III
MISMATCH RATE OF 15 STEREO IMAGE PAIRS.

Stereo Pairs	ST-2	CS-NL	CS-ST	IST-2	Proposed
Aloe	4.74	6.57	6.59	**4.48**	4.76
Art	10.71	13.87	13.21	10.79	**10.37**
Baby1	4.24	7.88	4.79	4.34	**3.65**
Books	9.34	12.47	10.98	9.06	**8.67**
Cloth1	0.52	1.15	1.16	**0.51**	0.56
Cloth2	3.64	5.36	5.34	**3.60**	3.87
Cloth3	2.16	3.16	3.06	2.17	**2.09**
Cloth4	1.46	2.40	2.18	1.38	**1.32**
Dolls	6.14	7.62	7.11	6.16	**5.90**
Laundry	12.05	17.85	18.46	12.47	**11.12**
Moebius	**8.27**	11.66	10.11	8.66	8.33
Reindeer	7.01	11.93	10.35	6.95	**5.50**
Rocks1	**2.77**	3.96	3.92	2.71	2.79
Rocks2	2.17	3.38	3.14	**2.05**	2.19
Wood1	4.42	10.86	6.33	4.70	**4.17**
Avg. Error	5.31	8.01	7.12	5.34	**5.02**
Avg. Time(s)	**1.02**	1.54	1.33	1.01	**1.25**

REFERENCES

[1] Yang Ding, QinXiao Li, Liu Ming, Tang GuoJian, "Analysis of the autopilot design," *2016 11th International Conference on Computer Science & Education (ICCSE)*. Nagoya, 06 October 2016.

[2] Oh SeungSub, Hahm Jehun, Jang Hyunjung, Lee Soyeon, Suh Jinho, "A study on the disaster response scenarios using robot technology," 2017 14th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI). South Korea, 27 July 2017.

[3] Yu. Zhao ,Zh. Yang ,Ch. Song , D. Xiong, "Vehicle dynamic model-based integrated navigation system for land vehicles," *2018 25th Saint Petersburg International Conference on Integrated Navigation Systems (ICINS)*. Russia, 09 July 2018.

[4] Jérôme Thevenot, Miguel Bordallo López, Abdenour Hadid,"A Survey on Computer Vision for Assistive Medical Diagnosis From Faces," *IEEE Journal of Biomedical and Health Informatics*.22: 1497 – 1511, Sept. 2018.

[5] Ye Chen, RuHui Huang, and Yi Zhu, "A Cumulative Error Suppression Method for UAV Visual Positioning System based on Historical Visiting Information," Engineering Letters, vol. 25, no.4, pp424-430, 2017.

[6] D. Scharstein ,R. Szeliski, R. Zabih, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Proceedings IEEE Workshop on Stereo and Multi-Baseline Vision(SMBV)*, Kauai, 07 August 2002.

[7] Yang Q," A non-local cost aggregation method for stereo matching.,"2012 *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, Providence, 26 July 2012.

[8] Yang Q, "Stereo matching using tree filtering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*. HongKong,37:834 - 846 ,April 2015

[9] Xing Mei, Xun Sun, Weiming Dong, Haitao Wang, Xiaopeng Zhang," Segment-tree based cost aggregation for stereo matching,"*2013 IEEE Conference on Computer Vision and Pattern Recognition*. Portland, 03 October 2013.

[10] Kang Zhang, Yuqiang Fang, Dongbo Min, Lifeng Sun ,Shiqiang Yang , Shuicheng Yan. Qi Tian,"Cross-scale cost aggregation for stereo matching," *2014 IEEE Conference on Computer Vision and Pattern Recognition*. Columbus, 25 September 2014.

[11] Ziyang Ma, Kaiming He ,Yichen Wei ,Jian Sun ,Enhua Wu," Constant time weighted median filtering for stereo matching and beyond," *2013 IEEE International Conference on Computer Vision*. Sydney, 03 March 2014.

[12] Li Linchen, Yu Xin, Zhang Shunli,"3D cost aggre-gation with multiple minimum spanning trees for stereo matching," *Applied Optics*. 56(12):3411-3420,2017 .

[13] Bleyer M," Patch Match stereo-stereo matching with slanted support windows", *British Machine Vision Conference*.14:1-14,2011 .

[14] Simon Perreault, Patrick Hebert," Median filtering in constant time," *IEEE Transactions on Image Processing*, 2389 – 2394, 13 August 2007.

[15] Yao P, Zhang H, Xue Y, "Segment-tree based cost aggregation for stereo matching with enhanced segment advantage, "*2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. New Orleans, 2027-2031, 19 June 2017.

[16] Felzenszwalb P F, Huttenlocher D P, "Efficient graph-based image segment, "*International journal of computer vision*, 59(2): 167-181,2004.

[17] Mutto C D, Zanuttigh P, Cortelazzo G M," Fusion of geometry and color information for scene segment," *IEEE Journal of Selected Topics in Signal Processing*, ,505-521, 2012.

[18] Westling P , Heiko Hirschmüller. High-Resolution Stereo Datasets with Subpixel-Accurate Ground Truth[J]. 2014.

[19] A. Geiger, P. Lenz, and R. Urtasun. Are we ready for autonomous driving the kitti vision benchmark suite. CVPR, 3354–3361, 2012 ,

Cheng Han was born in Jilin, Jilin Province, China in 1978. He received the B.S. degree in industrial automation from Jilin Institute of Chemical Technology, Jilin, in 2002, M.S. degree in computer application technology from Changchun University of Science and Technology, Changchun, in 2006, and the Ph.D. degree in Measuring & Testing Technologies and Instruments from Changchun University of Science and Technology, Changchun, in 2010.

From 2006 to 2013, he was a Lecturer with School of Computer Science and Technology from Changchun University of Science and Technology. Since 2013, he has been an Assistant Professor with the National and Local Combined Engineering Research Center of Special Film Technology and Equipment. He is the author of three books, more than 40 articles, and more than 20 patents. His research interests include virtual reality and digital media, computer vision, and computer simulation. He is a Reviewer of the journal Infrared and Laser Engineering, ACTA PHOTONICA SINICA.

Dr. Han is a Member of China Computer Federation and a Director of Jilin Society of Image and Graphics. He was a recipient of the first and third prize of Technological Invention Award in Jilin Province in 2012 and 2013 respectively, and the first prize of Science and Technology Progress Award in Jilin Province in 2017.