EFFECTIVE LOCAL STEREO MATCHING BY EXTENDED TRIANGULAR INTERPOLATION

Chunrong Xia, Yang Yang, Ran Ju and Gangshan Wu

State Key Laboratory for Novel Software Technology, Nanjing University, China

chunrong.xia@gmail.com, charlie.yang.nju@gmail.com, dg1233013@smail.nju.edu.cn and gswu@nju.edu.cn

ABSTRACT

In this paper, we propose an effective local stereo matching method based on extended triangular interpolation. The whole image is covered with a triangular mesh by performing triangulation on a set of initial support points. Since the disparity interpolation in some areas is ineffective, especially in the triangle which appears across the boundaries of objects or is formed by initial matching outliers, we formulate a new matching model based on the Bayesian rule to address this challenge. In the model, we introduce the concept of the triangle's reliability, and utilize the disparity planes determined by the neighboring triangles as the Bayesian prior. With the model, the disparity interpolation can be effectively performed. Experiments on the standard stereo data sets show that the method is effective, especially in dealing with the boundaries of objects.

Index Terms— Stereo matching, support points, triangulation, disparity interpolation

1. INTRODUCTION

Stereo vision enables users to access the depth information of the captured object or scene, and is regarded as a foundation for further applications such as 3D reconstruction and content analysis. Stereo matching which is one of the typical stereo vision techniques, reads the depth information from pixel-wise correspondence of two or more images.

Matching accuracy and performing efficiency are two key concerns in stereo matching. However, according to previous studies on stereo matching methods, these two concerns seem conflicting. Generally, by strategy, stereo matching approaches can be divided into two categories: Global and local methods. Global methods can produce high precision depth maps, however, their time cost is very high. On the contrary, although the produced depth map is not as accurate as produced by the global, the local methods are efficient enough to be adopted in real-time or near-real-time applications. Considering the global and local methods, some researchers proposed the semi-global ones, they are the compromise of the both, though more efficient than the global, the semi-global is still a challenge to be parallelized for practical use. So according to the needs of practical applications, it is valuable for us to do further study on local methods.

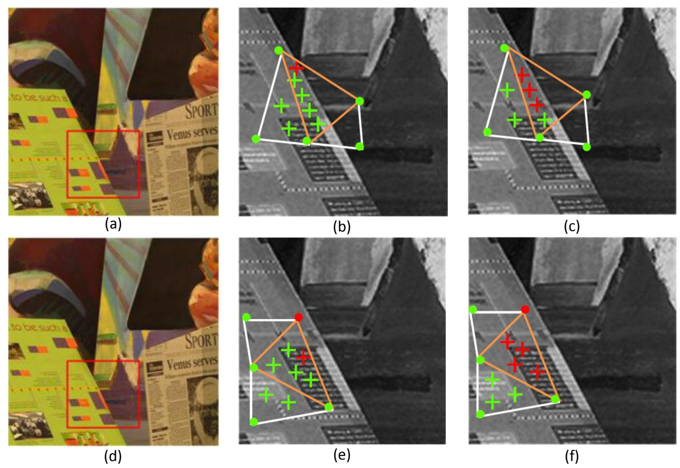


Fig. 1. (a) (d) are the original images. (b) (c) (e) (f) are the enlarged areas of the red boxes in (a)(d). (b) (c) contain the triangles across the borders of objects (with yellow edges) . (e) (f) contain the triangles formed by initial outliers (the red points).

As the depth values often keep continuous variation in local smooth scenes, disparity interpolation is widely adopted in local methods. Interpolation in triangular regions which are formed by triangulation on a set of support points is also studied by previous works. Assuming the disparities in the triangle are continuously varying, using the disparity plane defined by the triangle to do disparity interpolation is applicable. However, if the triangle appears crossing the borders of objects or one of its vertexes is a matching outlier, it will be invalid to interpolate. For example, in figure 1, simple disparity interpolation in the yellow edge triangles in image (c)(f) causes more outliers, as the red crosses.

For the points inside these invalid triangles, they are most likely on the neighboring reliable disparity planes, so their disparities can be effectively estimated by the neighboring triangles. For example, in figure 1, the neighbor reliable triangles can offer more valid disparities for the points in the yellow edge triangles in image (b)(e), as the green crosses.

This work is supported by the NSFC (No. 61021062) and the National Program of China (No. 2011AA01A202, No. 2011ZX05035-004-004HZ.).

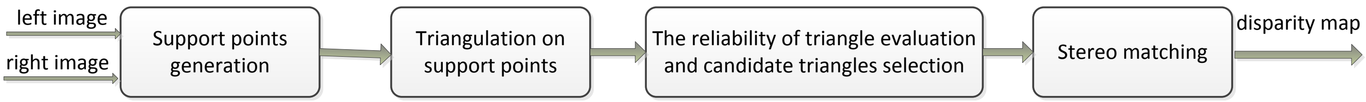


Fig. 2. An overview of the proposed algorithm.

To address this problem, we propose an effective matching method by extended triangular interpolation, a new matching model based on Bayesian rule is formulated in the paper. In the model, we take the reliability of the triangle into consideration. In accordance with most real-life scenario, we assume that the reliable triangle should be on the plane where its valid neighbors lie. To quantify the reliability of the triangle, we try to figure out the number of the neighboring triangles which are coplanar with it. The more neighbors meet such condition, the more reliable the triangle is. Experiment results show that our proposed method can effectively deal with the objects' boundaries, and the processing time is also acceptable.

The rest of the paper is organized as follows: In Section 2, we introduce the related work to this paper. Section 3 describes the proposed method in detail. And the experiments on two standard stereo data sets are given in Section 4, along with their results and analysis. Finally, we conclude our work and give some remarks on following work in Section 5.

2. RELATED WORK

Global methods are generally recognized as high precision matching algorithms. Most top-performing algorithms in the Middlebury web site [1] are global-based, like [2] [3], they incorporate segmentation-based prior constraints into the correlation based stereo model, and use segmented image regions as matching units, the regions are segmented by mean shift which is time-consuming, about 8 seconds to process an image of resolution 434×383 pixels, and then the convergence optimization process for the global energy minimization is also very slow. Their low efficiency keeps them away from practical applications.

On the contrary, local methods are more applicable than the global as their processing speed is much faster. Local methods measure correlation between pixels information inside a local support window centered on the reference pixel. [4] focuses on reducing the computational redundancy which exists among the disparity search range and the cost aggregation window, a repeated filtering for all disparity hypotheses and an sampling scheme inside the matching window are done to reduce the complexity. [5] proposes an adaptive weight approach which leverages the color and spatial similarity measures for the color images. It can provide more accurate results, but the computational complexity is considerable. For more details about local methods, we refer readers to [6] [7].

Nowadays, in order to make the matching speed reach real-time, many researchers use parallel techniques to realize the local methods, such as GPU-based methods [8] [9] [10]. Compared to the CPU-based local methods, their efficiency is an advantage. Though the parallel techniques can make the matching methods more efficient, they need expensive hardware investment as the cost. To the best of our knowledge, integrating the parallel processors into mobile devices is very inconvenient, and it is a great burden for the application system [11].

Previous study works show that CPU-based approaches can also run very fast, ELAS [12] is one of the typical representatives, it is efficient for high resolution images, and the authors also use ELAS for the practical applications in their later works: 3D reconstruction [13] and objects detection [14]. ELAS has shown its feasibility in those applications. By applying Delaunay triangulation on a set of initial matched correspondences, it efficiently exploits the disparity search space for the points in the corresponding triangle, it could achieve significant speedups but only needs a single CPU. Assuming the initial points are robust matched and without outliers, ELAS can perform effectively in the areas where the disparities are continuous. But along the boundaries of objects, or if the initial points contain outliers, the method will lose its effectiveness. Considering this problem, [15] has given a solution. After the triangulation, they classify all the triangles into matched and unmatched ones, they use the triangle size (the edge length and triangle area) to do triangles comparison. For the matched triangles, they use the simple disparity interpolation, and for the unmatched ones, they again classify them into semi-occluded and occluded ones, for the semi-occluded, they do sub-triangulation in each triangle, then simple disparity interpolation is done in each subtriangle. And for the occluded ones, they use the neighboring disparities to do estimate. One big disadvantage of their method is that the time cost is considerable because of two classification processes, and only using the semi-occluded triangle itself to calculate disparities may be ineffective in some cases. [16] also used triangulation in their proposed method. They assume all the pixels in each triangle have the same disparity value, and let the triangles as matching units. And in the refinement stage, they utilize the adjacent regions to do some adjustment. Their method can lead to a dense disparity map, but the superpixel strategy may cause over smoothness in some triangular regions.

3. APPROACH

The proposed approach contains the following steps: Firstly, we adopt the existing adaptive support weight method [5] to obtain initial support points. Secondly, Delaunay Triangulation is applied to organize these points into triangles, thus the whole image will be covered with a triangular mesh. Then, to make the disparity interpolation in each triangle more effective, we introduce the reliability of the triangle, and evaluate each triangle's reliability with its neighbors. Finally, we formulate a new matching model based on Bayesian rule to produce an accurate disparity map. An overview of the proposed approach is shown in figure 2.

3.1. Support points generation

In this stage, we apply uniform sampling on the left image to get a set of initial points. Then, we adopt the state of the art adaptive support weight method to search their best correspondences in the right image. The method aggregates matching cost in a support window based on both pixel similarity and geometric proximity. Given a pixel p, for any pixel q in its support region, the matching cost of q is weighted by the color difference c_{pq} and euclidian distance g_{pq} between p and q. The weight $\omega(p,q)$ is formulated as following:

$$\omega(p,q) = \exp(-(\frac{c_{pq}}{\sigma_c} + \frac{g_{pq}}{\sigma_g})) \tag{1}$$

σ_c and σ_g are used to control the color difference and euclidian distance for the weight function. Then the cost aggregation for each pixel is done:

$$C'(p,d) = \sum_{q\in\Omega_p} \omega(p,q)C(q,d) \tag{2}$$

Ω_p is the set of pixels in p's local window. And $C(q,d)$ is the pixel difference between point q in the left image and its correspondence $q_d = (x-d, y)$ in the right image. The pixel difference can be pixel-wise color dissimilarity, intensity dissimilarity, texture dissimilarity or combination of them. It's hard to say which dissimilarity function is better than others, it depends on the images themselves. In this paper, we use sobel filter [12] as the matching cost. Then for the initial sampling pixels with obvious texture features, we adopt a winner-take-all strategy to select the disparity which leads to the minimum matching cost.

$$D_L(p) = \arg\min_d C'(p,d) \tag{3}$$

Then the left sparse disparity map is generated. In a similar way, the right disparity map can also be computed. In order to make the support points more reliable, some post-processing is applied, such as left-right check [17], after that the remained matched ones are the final support points.

3.2. Triangulation

After obtaining the support points from the previous stage, Delaunay Triangulation is applied on the 2D support points to organize them as triangles, (see figure 3). The whole image is covered with a triangular mesh, all pixels are in this mesh.

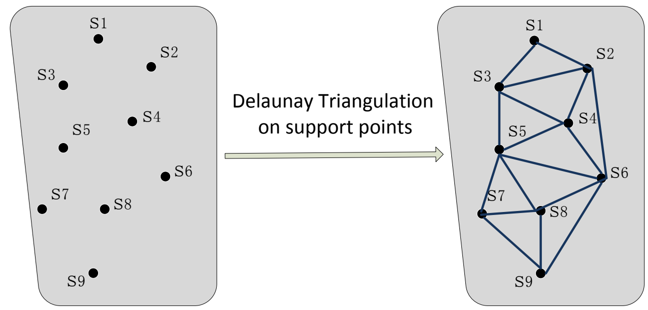


Fig. 3. Triangulation on support points.

3.3. The reliability of the triangle evaluation

According to most of the practical scenarios, the disparities in one object are most likely on the same plane, so the triangles inside one object also should share one plane. As we used the texture feature to calculate vertexes, the size and number of the triangles are determined by the image itself, and one object must be covered with a proper number of triangles. Generally, the disparity plane determined by a triangle can be denoted as (4).

$$d_p = a_{f_p}p_x + b_{f_p}p_y + c_{f_p} \tag{4}$$

p_x and p_y are the position coordinates of point p, and a_{f_p},b_{f_p} and c_{f_p} are the parameters of plane f_p.

However, for the invalid triangle, its plane can not truly reflect the disparity variation of the points in it. These invalid triangles may be across objects' boundaries or with incorrect vertexes. They are on dissimilar planes with their neighbors. For example, in figure 4, image (a) contains an invalid triangle appearing across the border of objects, image (b) has an invalid one whose vertex is an incorrect matching.

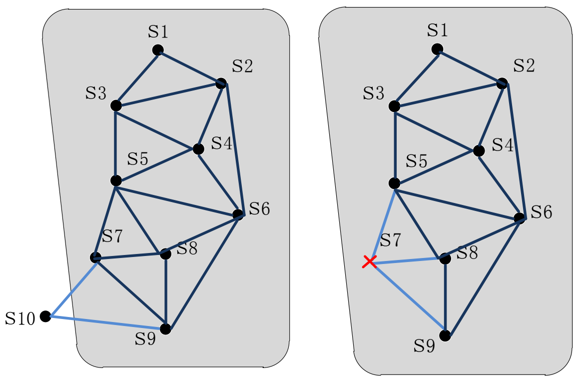
According to our assumption, the valid triangle should be on the plane where its neighboring valid triangles lie, so we can assess the triangle's reliability with its neighbors. To judge whether two triangles are on the same plane, we can calculate the angle between them. The angle between two planes equals the angle between their plane normal vectors.

For the disparity plane (4), its plane normal vectors are easy to calculate, $\overrightarrow{n'} = (a_{f_p}, b_{f_p}, -1)$ is one of them.

$$\theta = \arccos\frac{\overrightarrow{n}_{f_1}\cdot\overrightarrow{n_{f_2}}}{\|\overrightarrow{n_{f_1}}\|\cdot\|\overrightarrow{n_{f_2}}\|} \tag{5}$$

θ is the angle between f_1 and f_2. If $\theta < \tau_f$, f_1 and f_2 are thought to be on the same plane, τ_f is the pre-set threshold.

To measure the reliability of one triangle, only qualitatively checking whether all of its neighbors are coplanar with it is not proper for some triangles, for example, if the triangle is near the objects' borders but it is inside the objects and indeed



(a) triangle across the border (b) triangle containing the outlier

Fig. 4. Two cases of unreliable triangles.

valid, like the triangle formed by s7,s8,s9 in image (a) of figure 4, one of its neighbors (the triangle formed by s7,s9,s10) is across the borders and on a different plane, it may be classified into the invalid ones. To avoid this, we propose to find the number of its coplanar neighboring triangles to quantitatively measure the degree of reliability.

$$\gamma(f) = \frac{t_r + \tau_r}{t_n + \tau_r} \quad t_r \ < \ t_n \tag{6}$$

The number of adjacent triangles which meet the condition is t_r, t_n is the number of all the neighboring triangles and is set to 3 as discussed in section 3.4. τ_r is used to adjust the reliability.

3.4. Candidate disparity planes selection

For the invalid triangle, the disparities in it may be on the neighboring valid disparity planes. The best size for the neighborhood is determined by the sampling density and details of the image. Besides, the time cost is also a concern since the computation complexity is proportional to the size of neighborhood. In this paper, for the points in each triangle, we let the triangle itself and its three common edge neighbors as their candidate disparity planes.

$$\Theta_{f_p} = \{f \ , \ f_{n1}, \ f_{n2}, \ f_{n3}\} \tag{7}$$

Θ_{f_p} is the set of candidate disparity planes for point p, f is the triangle in which p lies. f_{n1}, f_{n2} and f_{n3} are the three neighboring triangles of f.

3.5. The model for stereo matching

We formulate the matching model using the Bayesian rule, the posterior probability over D given image I and disparity plane F can be written as $P(D|I,F) \propto$

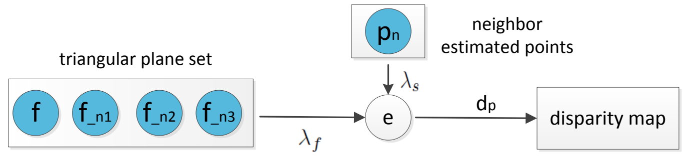


Fig. 5. Stereo matching model.

$P(I|D)P(F|D)P(D)$, maximizing the posterior is equivalent to minimize its negative log likelihood, then our objective is to find a disparity plane f_p that minimizes a local energy function for point p.

$$E(p, f_p) = E_{data}(p, f_p) + \lambda_s E_{smooth}(p) + \lambda_f E_f(f_p) \tag{8}$$

$E_{data}(p, f_p)$ is the data energy of p assuming p is on the plane f_p. It is the matching cost of p.

$$E_{smooth}(p) = \sum_{q \in N_4} \Delta d_{pq} \tag{9}$$

$$\Delta d_{pq} = \begin{cases} 1, & abs(d_p - d_q) > 1 \\ 0, & otherwise \end{cases} \tag{10}$$

$E_{smooth}(p)$ is a parametric smoothness prior that correlates neighboring estimated pixels [18], N_4 is the 4×4 neighborhood of p. Considering the efficiency, only one pass is processed, maybe a few estimated neighbors are available in the beginning, but this won't give a large bias since it's a soft constraint.

$$E_f(f_p) = \frac{1}{\gamma(f_p)} \tag{11}$$

The disparity plane f_p is defined by the corresponding triangle. According to the reliability of the triangle, a plane penalty term is added into the energy function, $\gamma(f_p)$ (6) is the reliability of f_p.

Then local optimization is done for (8), the winner-take-all strategy is also adopted here.

$$f(p) = \arg \min_{f_p \in \Theta_{f_p}} E(p, f_p) \tag{12}$$

With the winner disparity plane $f(p)$, according to the equation(4), the valid disparity interpolation d can be obtained for point p .

4. EXPERIMENTS

To verify the effectiveness of the proposed approach, we implement it on a PC with single CPU of 2.27GHz and 2G memory, the program language is C++. The parameters are constant across all images, considering the related works, they are set as table 1:λ_s and λ_f are respectively the smoothness and the unreliable disparity plane penalty parameters in (8). σ_c and σ_g are used to balance the color difference and euclidian distance in the adaptive support weight formulation (1). τ_f is the angle threshold to judge whether two triangles are on the same plane. τ_r is used to adjust the reliability.

Table 1. The prameters for experiments

λ_s	λ_f	σ_c	σ_g	τ_f	τ_r
5	2.5	4	2.5	0.1	1

The calculated disparity maps are evaluated by measuring the percent of bad matching pixels. If the disparity map is less than 100% density it will be interpolated using simple background interpolation as explained in the corresponding header file in the development kit [19]. The norms for evaluation are as follows:

- Out-N: Percentage of erroneous pixels in non-occluded areas
- Out-A: Percentage of erroneous pixels in total
- Out-D: Percentage of erroneous pixels in discontinuous areas.
- Avg-N: Average disparity / end-point error in non-occluded areas
- Avg-A: Average disparity / end-point error in total
- Err-th: Disparity error threshold

We first use the stereo image pairs issued by KITTI vision benchmark suite [19] to make experiments and evaluations. The data set is captured from the outdoor environment and there are 194 gray image pairs in it, the size of each image is 1241×376 pixels, the vertical resolution is much lower as restricted to the laser scanner which provides the ground truth of disparities. We have made a comparison with ELAS [12] which is a typical representative of local CPU-based methods, the results are shown in table 2.

The results in table 2 show that our proposed method can get lower error rate than ELAS on the Kitti data set. About the speed, though slower than ELAS, the proposed can still reach seconds speed, about 2.1s to process one image. [15] is also based on triangular interpolation and has considered the discontinuity of disparities, but it is much slower and needs about several minutes, only a little faster than the global graph-cut method.

Table 2. The comparison between our method and the ELAS on the Kitti stereo data set

Err-th	Method	Out-N	Out-A	Avg-N	Avg-A
2 pixels	**Our method**	**10.40**	**11.82**	**1.3 px**	**1.5 px**
2 pixels	ELAS	11.71	13.76	1.5 px	1.7 px
3 pixels	**Our method**	**7.20**	**8.43**	**1.3 px**	**1.5 px**
3 pixels	ELAS	8.34	9.58	1.5 px	1.7 px
4 pixels	**Our method**	**5.54**	**6.58**	**1.3 px**	**1.5 px**
4 pixels	ELAS	6.56	7.62	1.5 px	1.7 px
5 pixels	**Our method**	**4.51**	**5.40**	**1.3 px**	**1.5 px**
5 pixels	ELAS	5.40	6.32	1.5 px	1.7 px

We also have experimented on the standard data set from Middlebury web site [1]. From the results in table 3, we can find that, for the low resolution images: Cones, Teddy, Venus and Tsukuba, the proposed method could also get higher accuracy than ELAS. It performs better in all areas, especially in the discontinuous areas. Since we considered the initial outliers, our method also does better in the continuous places. The invalid triangles caused by the initial outliers are less than the one in the discontinuous areas, so the improvement in the discontinuous regions is more than the continuous.

Then, we have testified on higher-resolution images from Middlebury web site, taking the length of the paper into account, we only present two images here: cones, aloe with the resolution 868×766 pixels, the blue areas are occluded regions. From the results in figure 6, we find that our method indeed performs better and causes less matching outliers near the objects' borders.

Table 3. The comparison with ELAS on the benchmark Middlebury stereo data set

	Method	Cones	Teddy	Venus	Tsukuba
Out-N	ELAS	13.8	18.8	5.95	8.12
Out-N	**Our method**	**11.0**	**16.5**	**5.16**	**7.87**
Out-A	ELAS	23.5	27.1	7.59	10.3
Out-A	**Our method**	**21.0**	**25.1**	**6.82**	**10.0**
Out-D	ELAS	27.4	37.8	39.9	30.2
Out-D	**Our method**	**23.4**	**32.5**	**38.8**	**26.1**

5. CONCLUSION AND FUTURE WORK

In this paper, we described a novel local stereo matching method which is based on extended triangular interpolation. For the triangle which appears across the object's boundaries or formed by incorrect matches, simple interpolation in it is ineffective. To handle this, we propose to use the neighboring triangles to evaluate the triangle's reliability. With the candidate disparity planes formed by the triangle and its three neighbors, we proposed a new matching model which is based on the Bayesian rule. In the model, the candidate planes are adopted as the prior, and we also add a new parametric smoothness prior that correlates neighboring pixels in the model. We implemented the proposed method and testified it on two standard stereo data sets with a single CPU core, the results show that our method can generate more accurate disparity maps than ELAS, especially outperforms ELAS on the boundaries of objects, the proposed method is very useful for the following work, such as stereo image segmentation and objects detection. The processing time is also acceptable.

In the future, we will improve the efficiency of the proposed method and try to make it meet real-time requirements. And its effectiveness can also be improved with parameter optimization.

6. REFERENCES

[1] D. Scharstein and R. Szeliski, "The middlebury web site," `http://vision.middlebury.edu/stereo/`.

[2] Z. F. Wang and Z. G. Zheng, "A region based stereo matching algorithm using cooperative optimization," in *Computer Vision and Pattern Recognition–CVPR 2008.* IEEE, pp. 1–8.

[3] A. Klaus, M. Sormann, and K. Karner, "Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure," in *Pattern Recognition–ICPR 2006.* IEEE, vol. 3, pp. 15–18.

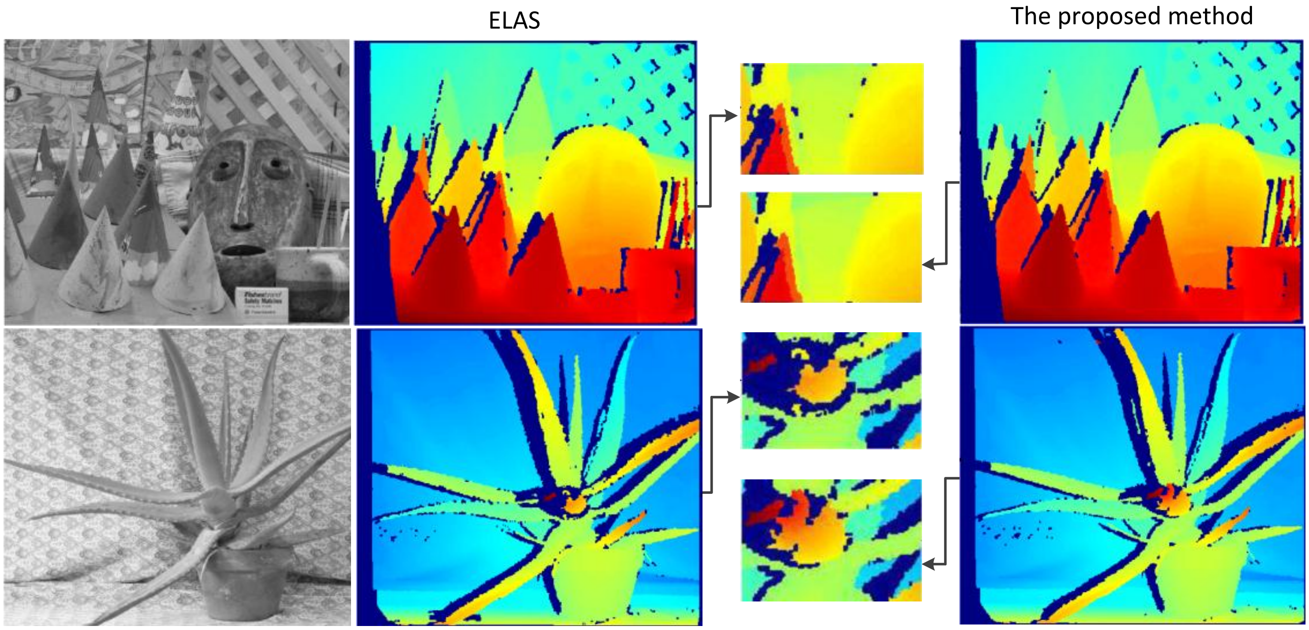


Fig. 6. The comparison results of cones and aloe between ELAS and the proposed method.

[4] D. B. Min, J. B. Lu, and M. Do, "A revisit to cost aggregation in stereo matching: How far can we reduce its computational redundancy?," in *Computer Vision–ICCV 2011*. IEEE, pp. 1567–1574.

[5] K. J. Yoon and I. S. Kweon, "Adaptive support-weight approach for correspondence search," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 4, pp. 650–656, 2006.

[6] K. Zhang, J. B. Lu, and G. Lafruit, "Cross-based local stereo matching using orthogonal integral images," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19, no. 7, pp. 1073–1079, 2009.

[7] F. Tombari, S. Mattoccia, L. Di Stefano, and E. Addimanda, "Classification and evaluation of cost aggregation methods for stereo correspondence," in *Computer Vision and Pattern Recognition–CVPR 2008*. IEEE, pp. 1–8.

[8] X. Mei, X. Sun, M. C. Zhou, S. H. Jiao, H. T. Wang, and X. P. Zhang, "On building an accurate stereo matching system on graphics hardware," in *Computer Vision Workshops–ICCV Workshops 2011*. IEEE, pp. 467–474.

[9] C. Richardt, D. Orr, I. Davies, A. Criminisi, and N. A. Dodgson, "Real-time spatiotemporal stereo matching using the dual-cross-bilateral grid," in *Computer Vision–ECCV 2010*, pp. 510–523. Springer.

[10] A. Hosni, M. Bleyer, C. Rhemann, M. Gelautz, and C. Rother, "Real-time local stereo matching using guided image filtering," in *Multimedia and Expo–ICME 2011*. IEEE, pp. 1–6.

[11] C. O. Park, J. H. Heo, D. H. Lee, and J. D. Cho, "Dynamic search range using sparse disparity map for fast stereo matching," in *Broadband Multimedia Systems and Broadcasting–BMSB 2012*. IEEE, pp. 1–4.

[12] A. Geiger, M. Roser, and R. Urtasun, "Efficient large-scale stereo matching," in *Computer Vision–ACCV 2010*, pp. 25–38. Springer.

[13] A. Geiger, J.Ziegler, and C. Stiller, "Stereoscan: Dense 3d reconstruction in real-time," in *Intelligent Vehicles Symposium–IV 2011*. IEEE, pp. 963–968.

[14] P. Lenz, J. Ziegler, A. Geiger, and M. Roser, "Sparse scene flow segmentation for moving object detection in urban environments," in *Intelligent Vehicles Symposium–IV 2011*. IEEE, pp. 926–932.

[15] S. H. Zhang, X. W. Qu, S. J. Ma, Z. R. Yang, and L. R. Kong, "A dense stereo matching algorithm based on triangulation," *J. Comput. Inform. Syst*, vol. 8, no. 1, pp. 283–292, 2012.

[16] G. P. Fickel, J. R. Claudio, R. Samadani, and T. Malzbender, "Stereo matching based on image triangulation for view synthesis," in *Image Processing–ICIP 2012*. IEEE, pp. 2733–2736.

[17] J. Sun, Y. Li, S. B. Kang, and H. Y. Shum, "Symmetric stereo matching for occlusion handling," in *Computer Vision and Pattern Recognition–CVPR 2005*. IEEE, vol. 2, pp. 399–406.

[18] L. Wang and R. G. Yang, "Global stereo matching leveraged by sparse ground control points," in *Computer Vision and Pattern Recognition–CVPR 2011*. IEEE, pp. 3033–3040.

[19] C. Stiller A. Geiger, P. Lenz and R. Urtasun, "The kitti vision benchmark suite," `http://www.cvlibs.net/datasets/kitti/eval_stereo_flow.php?benchmark=stereo`.