AN IMPROVED ENERGY SEGMENTATION BASED STEREO MATCHING ALGORITHM

Jinsheng Xiao[a], Liheng Yang[a], Jian Zhou[b, *], Huajing Li[a], Bijun Li[b], Ling Ding[c, †]

[a] School of Electronic Information, Wuhan University, Wuhan 430072
[b] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, 430079
[c] School of Computer Science, Hubei University of Education, Wuhan 430205, China

KEY WORDS: Stereo Matching, Energy Segmentation, Gradient Information.

ABSTRACT:

The stereo matching algorithm is commonly used in dealing with disparity discontinuous areas. However, the algorithm may have higher computational complexity and lack of well-aligned input images. To improve that, this paper presents an improved energy segmentation based stereo matching algorithm. The proposed algorithm incorporates gradient information of each segmented region into the energy function according to the facts that gradients inside the same segmentation should be closer than those from different segmentation. The invalid matching pixels are interpolated with vertical-horizontal nearest neighboring pixels to keep the consistency of stereo image pairs in the initial disparity map. Thus, the proposed method can reduce the deviations and the number of executions of the Random Sample Consensus (RANSAC) which arises from misalignment of input image pairs. Experiments results demonstrate that the proposed algorithm has a better disparity result while the running time is decreased by about 20%.

1. INTRODUCTION

The stereo vision techniques obtain the depth information from the 3D scene by simulating the human vision system, thus can be widely applied in the fields of robotic vision, intelligent driving, outer space exploration and military application.

Currently, there are two main directions in stereo vision research: One is the improvement to the traditional methods, which has been commonly applied for image processing (Poggi et al., 2019) (Sanchez-Rodriguez and Aceves-Lopez, 2018). The other is to obtain the depth information by training a convolutional neural network. Jure Žbontar (Žbontar and LeCun, 2016) firstly addressed the problem of the matching cost computation by learning a similarity measure on small image patches using a convolutional neural network (CNN). The output of the CNN is used to initialize the stereo matching cost. Although CNN has a higher matching precision (Kendall, 2018), it requires powerful equipment and longtime of training, thus is not fit for real-time applications, besides, the disparity results depend heavily on the training image set. Compared with the method based on deep learning (Yin et al., 2019), the stereo matching based on the segmentation still has extraordinary performance on the disparity results and has a large room for improvements (Liu et al., 2016). Consequently, we choose to improve the traditional methods based on energy segmentation.

The traditional stereo matching methods can be divided into three categories (Bebeselea-Sterp et al., 2017): local stereo matching, global stereo matching and semi-global stereo matching. The local matching method aggregates the matching cost by summing or averaging over a supported region centered at the main pixel. Combining the cost computation and aggregation approach with WTA (winner-take-all) strategy, Zhang et al. (Zhang et al., 2012) designed a new local stereo method called binary stereo matching. The algorithm mainly includes binary

and integer computations; thus, it is fast and fits for embedded or mobile devices. The local methods (Hernandez-Beltran et al., 2018) compute the similarity between pixels by comparing windows around the pixels of interest. Thus, the results are highly dependent on the cost functions and window sizes. Besides, the results are not so good in the regions with weak texture and occlusion.

The global algorithms (Yang et al., 2019), on the other hand, make explicit smoothness assumptions and then solve the global optimization problem. These algorithms usually do not perform aggregation, but seek a disparity assignment which can minimize a global cost function for all the pixels. A Bayesian approach for stereo matching was proposed by (Geiger et al., 2011), which is able to compute accurate disparity maps of high resolution images at a speed close to the real-time frame rate. However, almost all the global methods have a common drawback of lower speed and higher memory consumption, which often does not scale well with image size.

In 2008, a method named Semi-Global Matching (SGM) was proposed by Hirschmüller (Hirschmuller, 2008), which calculates the matching cost hierarchically applying the mutual information. Inspired by the SGM aggregation scheme, (Schonberger et al., 2018) proposed SGM-Forest, a new learning-based method that fuses disparity proposals estimated using scanline optimization.

To solve the problems of stripes and mismatching, Yamaguchi et al. (Yamaguchi et al., 2014) proposed a slanted plane smoothing (SPS) model for jointly recovering an image segmentation, a dense depth estimate and the boundary labels from a static scene form two frames of a stereo image pair. The approach first computes a semi-dense SGM depth map and then the SGM depth undergoes an effective slanted plane method. This can estimate dense stereo fields after considering the impact of segmentation, occlusion and outliers. The SPS method obtains a more accurate dense depth map while properly processing with

---

* Correspondence: jianzhou@whu.edu.cn
† Correspondence: dn0715dn@163.com

occlusions. Even so, the approach still has the problems of over-smoothing and too long running time.

To further improve the performance of the SGM method, a stereo matching algorithm based on an improved energy segmentation is proposed. The pipeline of the proposed approach is illustrated in Fig. 1.

Our main contributions are as following:

1) A new function of gradient information is incorporated into the energy functions in image segmentation to better retain the edges. The energy functions applied in image segmentation are modified according to the fact that gradient values inside the same segmentation should be closer than those from different segmentations, therefore, the edges are better retained.

2) The interpolation of invalid matching pixels is performed to remove the mismatch squares in the disparity map after SGM. The invalid matching pixels are interpolated with vertical-horizontal nearest neighboring pixels to eliminate the square errors in the disparity image.

3) In addition, the amount of calculations is reduced by simplifying the most time-consuming parts in the slanted plane smoothing algorithm. The algorithm is speeded up by simplifying and modifying the most time-consuming parts of disparity slanted plane fitting.

The experimental results show that the proposed method can obtain superior disparity images and better real-time performance.

2. SLANTED PLANE SMOOTHING ENERGY FOR SGM

Slanted Plane Smoothing (Yamaguchi et al., 2014) is an optimizing process for SGM . This method works quite well in the initial SGM disparity map with occlusion, weak texture and discontinuity areas.

The input image I is firstly evenly segmented into many regions. In each segment, a slanted disparity plane is constructed, which preserves an estimated disparity of every single pixel. Let $\theta_i = (A_i, B_i, C_i)$ be the disparity plane of segment i. At each pixel p with coordinates (p_x, p_y), the estimated disparity can be computed as Eq. (1):

$$\hat{d}(p, \theta_i) = A_i p_x + B_i p_y + C_i \qquad (1)$$

According to the difference between the initial disparity value and estimated value of the pixels in the segment, the pixels can be divided into outliers and inliers.

The energy of the system is defined to be the sum of energies encoding appearance, location, disparity, smoothness and boundary energies as Eq. (2) (Yamaguchi et al., 2014).

$$E(s, \theta, f, o, I, d) = \sum_p E_{col}(p, c_{s_q}) + \lambda_{pos} \sum_p E_{pos}(p, \mu_{s_q})$$
$$+ \lambda_{depth} \sum_p E_{depth}(p, \hat{d}, f_p)$$
$$+ \lambda_{smo} \sum_{\{i,j\} \in N_{seg}} E_{smo}(\theta_i, \theta_j, o_{i,j})$$

$$+ \lambda_{com} \sum_{\{i,j\} \in N_{seg}} E_{com}(o_{i,j})$$
$$+ \lambda_{bou} \sum E_{bou}(s_p, s_q) \qquad (2)$$

Where p and q indicate the pixels, s_p is a line segment containing p, θ_i is the disparity plane of segment i, f is an *outlier flag* to each pixel, o is the label of the neighboring segments i and j. λ corresponds to the weight of each energy term.The calculation method of s_p is shown in Eq. (3):

$$S_p = \arg\min_{S_p} \left\{ \begin{array}{l} \sum_p E_{col}(p, c_{s_q}) \\ + \lambda_{depth} \sum_p E_{depth}(p, \theta_{s_p}, f_p) \\ + \lambda_{pos} \sum_p E_{pos}(p, \mu_{s_q}) \\ + \lambda_{bou} \sum E_{bou}(s_p, s_q) \\ + \lambda_{gra} \sum_p E_{gra}(p, G_{s_p}) \end{array} \right\} \qquad (3)$$

The details of different energy components are as follows:

(1) Color Energy E_{col}: E_{col} is the SSD (Sum-of-Squared Differences) between the color of boundary pixels p and the average color of the segment s_q while q are the neighboring pixels of p. This term encourages pixels in the same segment to be close in color.

(2) Position Energy E_{pos}: It denotes the squares of the distance between coordinate of p and average coordinates of s_q, where q are the neighboring pixels of p. This energy term prefers well-shaped segments. The value of λ_{pos} is set to half of λ_{bou}.

(3) Depth Energy E_{depth}: Depth energy of all inliers is defined as the SSD between their initial disparities d and estimated disparities $\hat{d}$ while outliers are defined as constants to avoid abnormal disturbance. This term encourages the plane estimates to fit the SGM estimates. Usually, λ_{depth} is relatively large.

(4) Smoothness Energy E_{smo}: This term measures the smoothness between adjacent segments. $o_{i,j}$ assigns the line label between segment i, j. If two neighboring segments are co-planar, the two planes θ_i, θ_j should be merged into one segment, otherwise, if they form a hinge they should be separated by a boundary. And if they form an occlusion, the disparity of the two planes should be a constant value.

(5) Complexity Energy E_{com}: This term is used to measure the complexity between two neighboring segments. It encourages two planes to be coplanar where the cost of boundary type ranked as: Occlusion>Hinge>Coplanar= 0.

(6) Boundary Energy E_{bou}: This term encourages segments to be regular and prefers straight boundaries. Let p be a boundary pixel and q is the set of 8-neighboring pixels of p. If q does not belong to segment s_p , boundary energy will accumulate. The value of λ_{bou} is a little smaller than λ_{depth}.

Note that not all these terms are used in every step. So the value of λ sometimes can be set to zero. According to the definition of energy function for the slanted plane smoothing, we can describe the whole process as follows.

Firstly, we evenly segment the input image into several parts, and then we adjust the segmentation according to the idea of
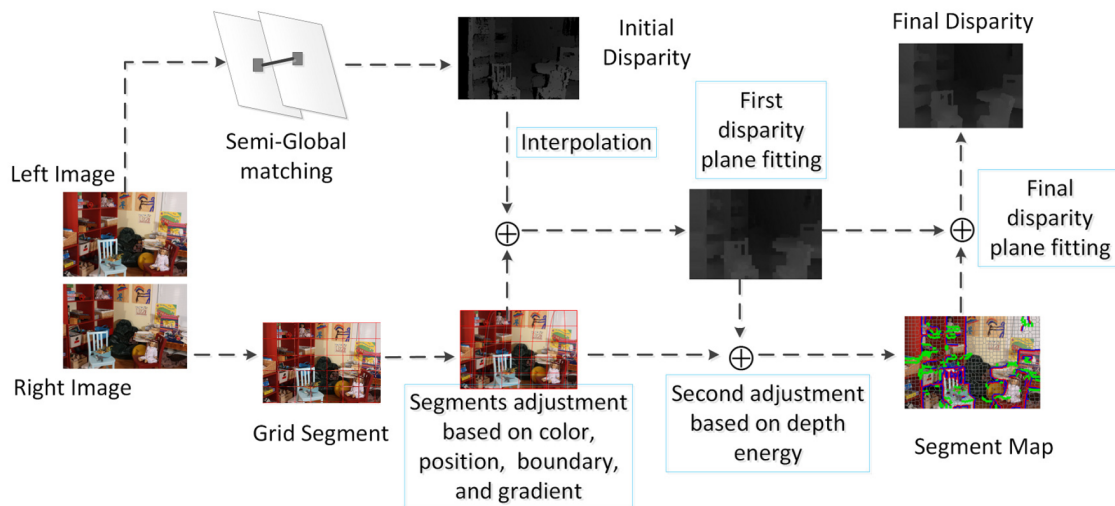
Figure 1. The flow chart of the proposed stereo matching algorithm.

minimizing the functions of color energy, position energy and boundary energy. After that, the initial disparity plane is fitted based on the SGM results taking the adjusted segmentations as a unit. The fitting is done by RANSAC algorithm combined with the least squares method. Secondly, we adjust the initial fitting plane according to the depth energy function, and then we will get a new set of boundary pixels. All the boundary pixels will be classified into three types: coplanar, hinge and occlusion by minimizing the complexity energy. Finally, we minimize the smoothness energy of the fitting plane according to the boundary labels and get the final dense and smooth disparity map. The details are introduced in the next section.

3. THE IMPROVED ENERGY SEGMENTATION ALGORITHM

The slanted plane smoothing energy will be used to perform disparity plane fitting separately for occlusion regions. Therefore, the whole process is computationally expensive and the image segmentation is not quite complete. Furthermore, if the input stereo image pairs are misaligned in rows, the initial disparity plane fitting applying SGM will result in some small mismatched squares. The left image in Fig. 2 is an outdoor scene captured by our own camera. Due to the effect of illumination and the camera, the rectified images could not be well aligned in rows. Thus, SGM differences can form a large number of blocks, indicating that the match is invalid.

To solve these problems, we propose a series of improvements. Firstly, in order to better retain the edge, a novel function of gradient information is incorporated into the energy functions in image segmentation processing. Then, the interpolation of invalid matching pixels is performed to eliminate the mismatch squares in the disparity map obtained by applying the SGM method. In addition, the computational complexity is reduced by simplifying the initial and final disparity plane fitting, which are the most time-consuming parts in the slanted plane smoothing algorithm. The flowchart of the proposed stereo matching algorithm based on the improved energy segmentation is shown in Fig. 1, where the improvements are marked with blue boxes.

The steps of our proposed method can be described as follows.

(1) Prepare the input stereo image pairs, in which the left one is a reference image and the right one is a matching image.

(2) Get an initial disparity map applying the SGM method.

(3) Segment the left image to a regular grid and achieve the first adjustment. Adjust the segment to get the smallest segmentation energy, including color energy, position energy, boundary energy and gradient energy to generate a more accurate segmentation.

(4) After interpolating the invalid pixels in the initial SGM map, the disparity plane of each segment is fitted applying RANSAC and the least squares method by minimizing the depth energy, where the calculation is optimized to speeding the processing.

(5) The second adjustment to the segmentation. Adjust the segmentation to get the minimal segmentation energy after incorporating the newly obtained depth energy. Then divide the boundary pixels into three types: occlusion, hinge and coplanar by minimizing the complexity energy.

(6) The final smoothing operation. The smoothness energy is calculated according to the boundary types so that the disparity plane is re-fitted with the least squares method by minimizing the smoothness energy. Then the final disparity map is obtained. By improving the iteration step in the final fitting process, the over-smoothing can be suppressed and less time consuming.

3.1 Energy Refinement with Gradient Information

In order to effectively segment the contours of different targets, we supplement the energy function defined in Eq. (2) with the gradient information, which will lead to a more reasonable segmentation.

To balance the accuracy and time consumption, Roberts operators (Roberts, 1963) are applied to extract gradient information of the input images. By computing the sum of the squares of the differences between diagonally adjacent pixels, the Roberts operator approximates the gradient of an image through discrete differentiation. The results of this operation will highlight the changes of intensity in a diagonal direction.
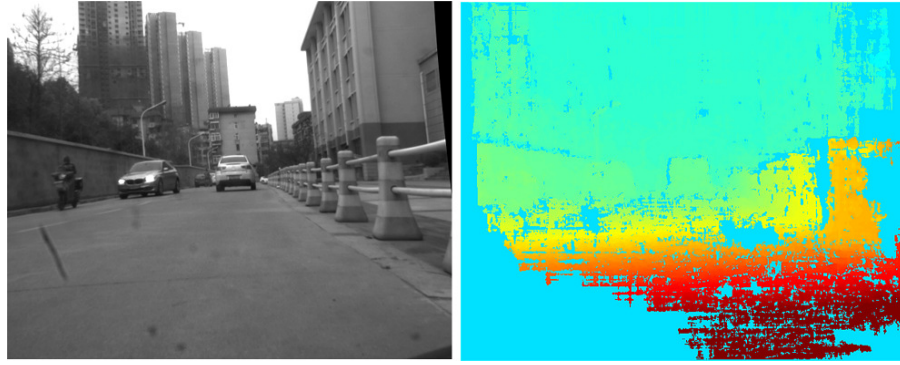
Figure 2. Left: The rectified left image. Right: The disparity map obtained by the SGM method.

In the image segmentation, the gradient information computed by Roberts operators is used to supplement the segmentation energy function. Where gradient energy is:

$$E_{gra}(p, G_{s_q}) = \|p - G_{s_q}\|^2 \tag{4}$$

Here, q represents the neighboring pixels of p, G_{s_q} indicates the mean gradient of segment s_q. Gradient energy is the SSD between the gradient of boundary pixels p and G_{s_q}.
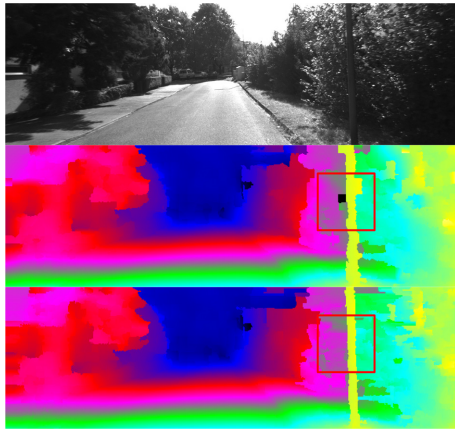

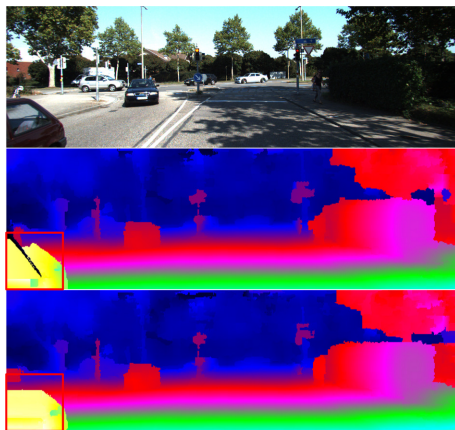Figure 3. The original disparity map and the improved disparity map (a).

Figure 4. The original disparity map and the improved disparity map (b).

The performance of the improved SPS is evaluated, as can be seen in Fig. 3 and Fig. 4. Among which, Fig. 3 shows the final disparity maps of the original and improved energy function

on the KITTI2012. It's apparently that after the refinement, the black hole besides the telegraph pole is removed. Similarly, Fig. 4 shows the difference on KITTI2015. On the original disparity map, the car is scratched by a thick line while in the improved disparity map, the car is not scratched.

The verification dataset for adding gradient energy in Table 1 comes from Middlebury (Scharstein, 2017). The most widely used objective indicator for Middlebury is the bad2.0 (that is, the percentage of pixels whose disparity error are more than 2 compared to ground truth) for a single image, and the ground truth is non-occluded mask. Therefore, p1, p2, p3, p4, p5, and p6 in Table 1 represent 6 pairs of stereo images randomly selected from Middlebury. As shown in Table 1, the pixel error ratio can be reduced approximately 0.5% with the refinement.

Table 1. bad2.0 comparisons before and after improved(%)

Inputs	p1	p2	p3	p4	p5	p6
Original SPS	7.64	6.07	9.40	13.43	12.14	10.24
Proposed	**7.51**	**5.80**	**9.40**	**13.10**	**11.65**	**9.42**

3.2 Speed optimizing of disparity plane fitting

By analyzing the computational complexity of the whole slanted plane smoothing algorithm, we find that the initial and final disparity plane fitting steps are the most time-consuming parts. To clarify this, we first introduce the whole process of the slanted plane smoothing algorithm and then we focus on the initial and final disparity plane fitting process.

In the initial fitting, the original algorithm is combined with RANSAC (Xiao et al., 2020) and used to fit the disparity plane of the initial disparity map. Specifically, the disparity plane is determined by randomly selecting three different points within the segmentation, and inliers and outliers are distinguished according to their disparity distance between the initial disparity and the fitting slanted plane. Then, the plane with the most inliers are obtained after several iterations. After all the inliers are determined, the disparity slanted plane is again fitted using the least squares method. *Least squares* means that the overall solution minimizes the sum of the squares of the residuals made in the results of each single equation, where the sum S is expressed as:

$$S = \sum_{i=0}^{n-1} (ax_i + by_i + c - d_i)^2 \tag{5}$$

So the following equation should be satisfied to get the smallest S:

$$\frac{\partial S}{\partial a} = 0, \quad \frac{\partial S}{\partial b} = 0, \quad \frac{\partial S}{\partial c} = 0 \qquad (6)$$

thus

$$\begin{cases} a \sum x_i{}^2 + b \sum x_i y_i + c \sum x_i = \sum x_i d_i \\ a \sum x_i y_i + b \sum y_i{}^2 + c \sum y_i = \sum y_i d_i \\ a \sum x_i + b \sum y_i + cn = \sum d_i \end{cases} \qquad (7)$$

Where n is the number of inliers. d_i are the initial disparities of inliers (x_i, y_i) in the segmentation. After the coefficients a^*, b^*, c^* are solved, the initial disparity plane can be denoted as

$$d^* = a^* x + b^* y + c^* \qquad (8)$$

Now the computational complexity of the initial disparity fitting will be analyzed here. Suppose the height and width of an input image are H and W respectively. The image is segmented into M segmentations, where every segmentation has N pixels, thus $H \cdot W = M \cdot N$. Assume that the computational complexity needed for removing outliers in the initial disparity plane fitting is T, needed for formula Eq. (7) is L, then the calculation amount needed for traversing all pixels within one segmentation will be $T \cdot N + L$. Therefore, the initial fitting calculation complexity of the entire image is $M(T \cdot N + L)$, which is quite large. Since those mismatching pixels will be corrected after SGM by interpolating disparities of their adjacent pixels, there should be few mismatched pixels left. Thus, it is unnecessary to iterate the RANSAC process due to much time cost. As a result, the RANSAC is executed once in the initial disparity fitting. Thus, the calculation complexity of the initial fitting will not exceed $M(N + L)$.

In the final disparity plane fitting step, the smoothing procedure is iterated to get a dense and smooth disparity map by minimizing the smoothness energy function. In the iterative process, the disparities of the previous fitting are required to adjust the segmentation to which the boundary pixels belong, thus the calculation complexity is quite large. Therefore, an iteration termination function is proposed to balance the quality and the speed of plane fitting as below:

$$\frac{E_i - E_{i+1}}{E_i} \leq \varepsilon \qquad (9)$$

where E_i and E_{i+1} denote the bad 2.0 (Scharstein, 2017) (percentage of pixels whose disparity error are more than 2 compared to ground truth) of the disparity after the i_{th} and $(i+1)_{th}$ iterations, respectively. ε is set to 0.01 in the experiment in this paper.

In the following Fig. 5, the bad 2.0 drops sharply during the first 4 iterations and the error drops very little from 4 iterations to 10 iterations.

We compared the disparity results of different iteration times in Fig. 6. As the figure shows, the visual quality is almost the same after 4 iterations. In that case, we can iterate fewer times to reduce the calculation.

On the other hand, to compensate the possible loss of performance caused by fewer iteration times, the interpolation for ini-
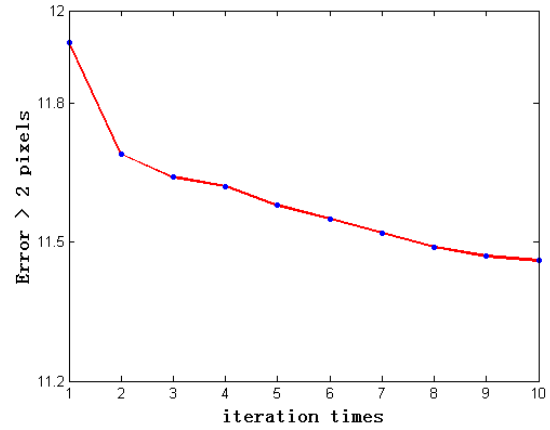

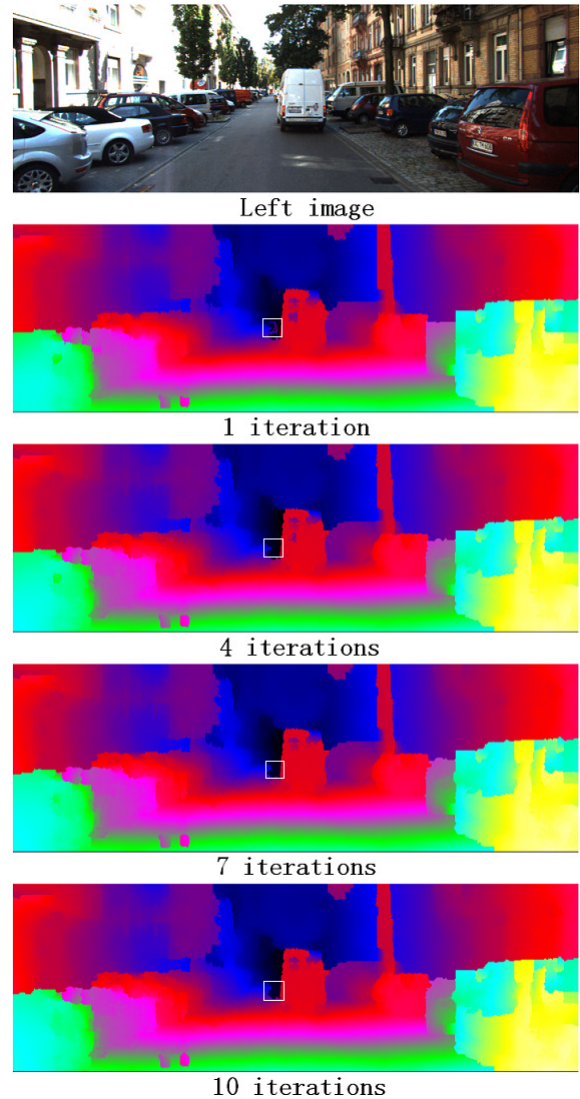Figure 5. Bad 2.0 of different iteration times.

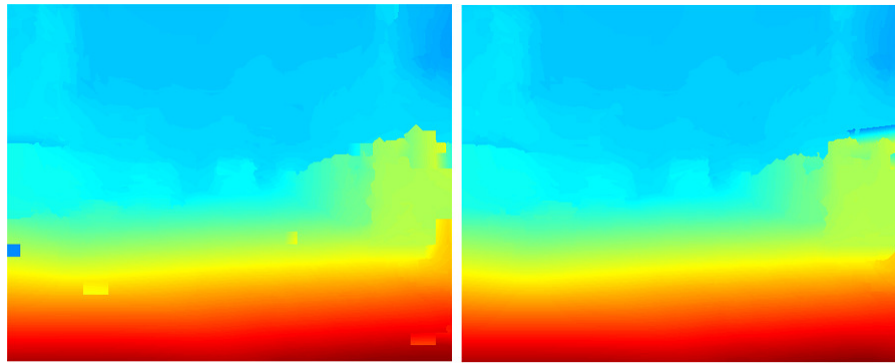Figure 6. The disparity maps in different iteration times.

Figure 7. Left: the original SPS disparity. Right: the disparity after interpolation.

tial disparity and the gradient energy are utilized. The experimental results demonstrate that the speed and performance are improved.

Table 2 shows an average time-consuming comparison of the improved and original SPS algorithms. Five datasets, selected from KITTI , named d1,d2,d3,d4,d5 respectively, are used to test. The algorithmic computation time is reduced approximately 8% in our improved method. All experiments are carried out on in Windows 7 operation system with Intel core i5 6500 processor.

Table 2. Time-consuming comparison(in seconds)

Datasets	d1	d2	d3	d4	d5
Original SPS	3.06	3.08	3.05	3.03	3.10
Proposed	**2.80**	**2.85**	**2.76**	**2.70**	**2.96**

3.3 Interpolation for initial disparity

In practical, it is difficult to obtain strictly row-aligned stereo image pairs, which will cause many mismatched points in the initial disparity map obtained by using the SGM method. After the left-right consistency checking, those mismatched points will be forcibly set to 0. (As can be seen from the right image in Fig. 2). Too much invalid disparity data will have a severe negative effect on the final fitted disparity plane (Fig. 7(Left)). Therefore, it is essential to use the disparities of the adjacent pixels to interpolate these mismatched regions.

The efficient nearest neighbor interpolation method is chosen to save the run time. The interpolation is implemented both in the horizontal and vertical direction to avoid the appearance of streaks in the disparity map. After the interpolation, we can gain a denser disparity map without blocks as shown in the right image in Fig. 7.

4. EXPERIMENTS AND RESULTS

The performance of the proposed stereo matching algorithm is evaluated on the challenging KITTI dataset (Geiger et al., 2013), which contains KITTI 2012 and KITTI 2015. In 2012, there are 200 pairs (i.e., 400 images), and in 2015 there are 194 pairs (i.e., 388 images). In order to validate the proposed algorithm, 40 pairs of images randomly selected from these two datasets are tested, we employ the same paramenters for all experiments, $\lambda_{gra} = 600$, other paramenters are set according to

reference (Yamaguchi et al., 2014). The ground truth is semi-dense covering approximately 30% of the pixels. The objective error comparisons of bad 2,3,4,5, are used here which fully demonstrates the improvements of our algorithm. The result is shown in Table 3 and Table 4.

Table 3. Comparison with the mainstream methods on the test set of KITTI2012

	>2 pixels	>3 pixels	>4 pixels	>5 pixels	Runtime
Binarystereo	40.93%	36.77%	34.26%	32.01%	137s
Elas	22.72%	21.07%	20.23%	19.66%	1.44s
SGM	6.28%	4.98%	4.14%	3.57%	1.38s
RWR	19.54%	17.22%	15.38%	14.07%	10.38s
MC-CNN	4.28%	2.63%	2.02%	1.72%	67s
SPS	4.86%	3.79%	3.17%	2.76%	4.17s
Proposed	4.56%	3.53%	2.92%	2.51%	2.85s

Table 4. Comparison with the mainstream methods on the test set of KITTI2015

	>2 pixels	>3 pixels	>4 pixels	>5 pixels	Runtime
Binarystereo	43.16%	39.44%	36.93%	35.01%	140s
Elas	24.09%	19.21%	17.59%	16.82%	2.12s
SGM	10.03%	6.93%	5.47%	4.48%	1.99s
RWR	28.01%	17.95%	13.26%	11.04%	12s
MC-CNN	6.38%	3.27%	2.37%	1.97%	67s
SPS	7.15%	4.58%	3.46%	2.93%	7s
Proposed	7.17%	4.51%	3.44%	2.81%	5.74s

In the experiments, the proposed approach are compared with several other methods, including the Binary stereo matching method (Binarystereo (Zhang et al., 2012)), the Efficient large-scale stereo matching method (Elas (Geiger et al., 2011)), the Semiglobal Matching method (SGM (Hirschmuller, 2008)), the Random walk with restart algorithm (RWR (Lee et al., 2015)), the Matching cost with convolutional neural network (MC-CNN (Žbontar and LeCun, 2016)), and the Slanted plane smoothing (SPS (Yamaguchi et al., 2014)).

Table 3 and Table 4 show the error which is measured as the percentage of pixels whose true and predicted disparities differ more than 2, 3, 4, or 5 pixels. For example, in terms of the > 2 pixels metric on KITTI2012, the proposed algorithm yields 4.56%, while the best performing method MC-CNN (Žbontar and LeCun, 2016) provides 4.28%. In general, the proposed
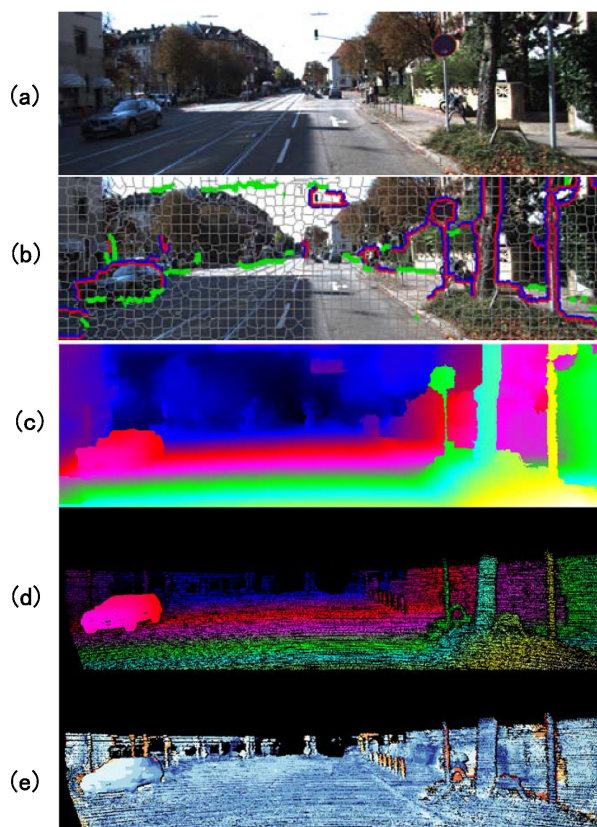
Figure 8. Qualitative results of our improved slanted plane model. (a) Left image; (b) Segments and boundaries; (c) Disparity map; (d) Ground truth; (e) Error graph
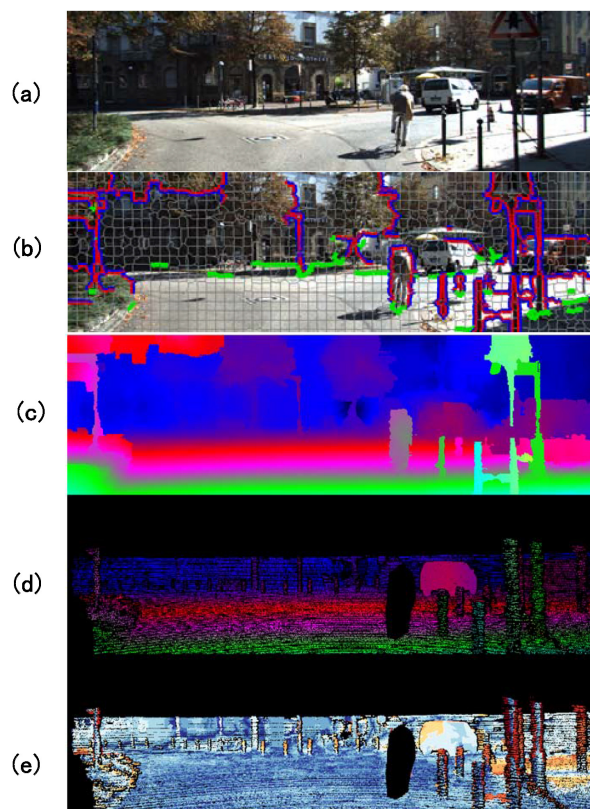

Figure 9. Qualitative results of our improved slanted plane model. (a) Left image; (b) Segments and boundaries; (c) Disparity map; (d) Ground truth; (e) Error graph

method performs better both on subjective and objective indicators in most cases. Although the deep learning method MC-CNN (Žbontar and LeCun, 2016) has the best results compared with all the methods in the table, it costs considerable amount of time.

Fig 8 and Fig 9 illustrate the qualitative results on dataset KITTI 2012 and KITTI2015. Among them, the error graph refers to the differences between the ground truth and the disparity map. As can be seen from the disparity map, the proposed algorithm can obtain smooth and dense disparity and retain the edge and texture mostly. And it is worth pointing out that the proposed algorithm is able to estimate the hinge or occlusion boundaries accurately. Like the cars, the trunks and the poles in Fig 8 and Fig 9. Totally speaking, the proposed matching algorithm is promising.

5. CONCLUSION

In this work, an improved energy segmentation algorithm is proposed for efficient stereo matching. In this paper, the SGM is applied to get an initial disparity map, and then those mismatched pixels are interpolated with vertical-horizontal nearest neighboring pixels. In the first step of disparity plane fitting, the number of iterations is reduced by executing the RANSAC just once and setting an iteration termination function to accelerate the process. Besides, it is essential to incorporate the gradient information into the energy definition to better retain the edges. Therefore, the disparity is refined by energy functions incorporating the gradient information. After the refinement of energy

function definition, the disparity results improve a lot. Experimental results demonstrated that the proposed algorithm outperforms the conventional stereo methods significantly. However, the algorithm in this paper still cannot get the best results in terms of accuracy, so the follow-up work may focus on improving the accuracy as much as possible while ensuring the operation speed.

REFERENCES

Bebeselea-Sterp, E., Brad, R., Brad, R., 2017. A Comparative Study of Stereovision Algorithms. *International Journal of Advanced Computer Science and Applications*, 8.

Geiger, A., Lenz, P., Stiller, C., Urtasun, R., 2013. Vision Meets Robotics: The KITTI Dataset. *Int. J. Rob. Res.*, 32(11), 1231C1237. https://doi.org/10.1177/0278364913491297.

Geiger, A., Roser, M., Urtasun, R., 2011. Efficient large-scale stereo matching. R. Kimmel, R. Klette, A. Sugimoto (eds), *Computer Vision – ACCV 2010*, Springer Berlin Heidelberg, Berlin, Heidelberg, 25–38.

Hernandez-Beltran, J., Diaz-Ramirez, V., Juarez-Salazar, R., 2018. Stereo matching using adaptive windows and correlation filtering. 2.

Hirschmuller, H., 2008. Stereo Processing by Semiglobal Matching and Mutual Information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2), 328-341.

Kendall, A., 2018. Geometry and Uncertainty in Deep Learning for Computer Vision. PhD thesis, University of Cambridge.

Lee, S., Lee, J. H., Lim, J., Suh, I. H., 2015. Robust stereo matching using adaptive random walk with restart algorithm. *Image Vis. Comput.*, 37, 1–11. https://doi.org/10.1016/j.imavis.2015.01.003.

Liu, D., Nicolescu, R., Klette, R., 2016. Stereo-Based Bokeh Effects for Photography. *Mach. Vision Appl.*, 27(8), 1325C1337. https://doi.org/10.1007/s00138-016-0775-5.

Poggi, M., Pallotti, D., Tosi, F., Mattoccia, S., 2019. Guided stereo matching. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Roberts, L. G., 1963. *Machine Perception of Three-Dimensional Solids*. Outstanding Dissertations in the Computer Sciences, Garland Publishing, New York.

Sanchez-Rodriguez, J.-P., Aceves-Lopez, A., 2018. A survey on stereo vision-based autonomous navigation for multi-rotor MUAVs. *Robotica*, 36, 1-19.

Scharstein, D., S.-R., 2017. http://vision.middlebury.edu/stereo/7.

Schonberger, J. L., Sinha, S. N., Pollefeys, M., 2018. Learning to fuse proposals from multiple scanline optimizations in semiglobal matching. *Proceedings of the European Conference on Computer Vision (ECCV)*.

Xiao, J., Xiong, W., Yao, Y., Li, L., Klette, R., 2020. Lane Detection Algorithm Based on Road Structure and Extended Kalman Filter. *International Journal of Digital Crime and Forensics*, 12, 1-20.

Yamaguchi, K., McAllester, D., Urtasun, R., 2014. Efficient joint segmentation, occlusion labeling, stereo and flow estimation. D. Fleet, T. Pajdla, B. Schiele, T. Tuytelaars (eds), *Computer Vision – ECCV 2014*, Springer International Publishing, Cham, 756–771.

Yang, W.-J., Tsai, Z.-S., Chung, P.-C., Cheng, Y.-T., 2019. An adaptive cost aggregation method based on bilateral filter and canny edge detector with segmented area for stereo matching. *International Workshop on Advanced Image Technology (IWAIT) 2019*, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 11049, 110491J.

Yin, Z., Darrell, T., Yu, F., 2019. Hierarchical discrete distribution decomposition for match density estimation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Žbontar, J., LeCun, Y., 2016. Stereo Matching by Training a Convolutional Neural Network to Compare Image Patches. *Journal of Machine Learning Research*, 17(65), 1-32. http://jmlr.org/papers/v17/15-535.html.

Zhang, K., Li, J., Li, Y., Hu, W., Sun, L., Yang, S., 2012. Binary stereo matching. *Proceedings of the 21st International Conference on Pattern Recognition*, IEEE Computer Society, 356–359.