

# STEREO IMAGE DENSE MATCHING BY INTEGRATING SIFT AND SGM ALGORITHM

Yuanxiu Zhou<sup>1</sup>, Yan Song<sup>1\*</sup>, Jintao Lu<sup>2</sup>

<sup>1</sup> Faculty of Information Engineering, China University of Geosciences, Wuhan, China – zhouyuanxiu@cug.edu.cn,  
moonriver\_song@163.com

<sup>2</sup> China Railway Engineering Consulting Group Co., Ltd - cecljt@126.com

**KEY WORDS:** Stereo Matching, Semi-Global Matching, SIFT, Dense Matching, Disparity Estimation, Census

## ABSTRACT:

Semi-global matching(SGM) performs the dynamic programming by treating the different path directions equally. It does not consider the impact of different path directions on cost aggregation, and with the expansion of the disparity search range, the accuracy and efficiency of the algorithm drastically decrease. This paper presents a dense matching algorithm by integrating SIFT and SGM. It takes the successful matching pairs matched by SIFT as control points to direct the path in dynamic programming with truncating error propagation. Besides, matching accuracy can be improved by using the gradient direction of the detected feature points to modify the weights of the paths in different directions. The experimental results based on Middlebury stereo data sets and CE-3 lunar data sets demonstrate that the proposed algorithm can effectively cut off the error propagation, reduce disparity search range and improve matching accuracy.

## 1. INTRODUCTION

Generating dense accurate disparity maps is the essential step and crucial technology for many applications such as digital surface model generation and three-dimensional reconstruction. By extracting local correspondences between two or more reference images, the depth information of the images can be obtained. Due to the complicated scene of stereo matching in the real world, such problems as shadows and occlusions make stereo matching a hot and difficult issue in the field of digital photogrammetry and computer vision.

According to the matching strategies (Scharstein et al., 2002), stereo matching algorithm can be divided into local stereo matching algorithm, semi-global stereo matching algorithm and global stereo matching algorithm. The local matching algorithm constructs the matching cost function with pixel and its surrounding small areas as constraints, which takes advantage of high efficiency and real-time performance but does not consider the overall consistency. The global matching algorithm essentially minimizes the global energy function, including graph cuts (Boykov et al., 1999) and belief propagation algorithms (Sun et al., 2002), which can effectively overcome image occlusion and maintain non-continuity, has great advantages in stability, reliability, but with a costly calculation.

The semi-global matching algorithm (Hirschmuller, 2005) takes the optimal one-dimensional energy in multiple directions to approximate two-dimensional global optimum, is among the top-performing algorithms in dense matching. It takes advantage of high precision of global matching algorithm and low time complexity of local matching algorithm and is also the method used in this paper. However, mis-match is likely to occur where has large variations in parallax. With the search range becoming larger, calculation efficiency decreases and time-consuming increases.

Aiming at these problems, pyramid strategies (Hermann and Klette, 2013) were used in the SGM to reduce the search range by providing initial disparity maps, but this method did not improve the accuracy rate. Chen et al. (2017) proposed taking advantage of the region growing and SGM algorithms to correct the aggregation path while accelerating parallax search speed, but the results depended on the range and accuracy of the region growth. Most of the improved algorithms of SGM focus on improving the initial cost calculation or adaptively selecting the matching parameters with image information. Li et al. (2017) presented an SGM algorithm based on ADCensus, which made the SGM more robust by taking advantage of the AD similarity measure. Zhu et al. (2017) proposed a matching method that considered texture features to better preserve edge features. However, these above methods did not consider the effect of the aggregation path direction on the matching accuracy.

In this paper, a modified SGM algorithm integrated with SIFT (Lowe 2004) is proposed to enhance the quality of the estimated depth map while decreasing the computation. The main work as follows: To reduce parallax search range, improve matching precision and direct the mis-match in dynamic programming, initial matching results using SIFT with object-oriented segmentation are used. Besides, the correctness of the matching can be improved by finding the relationship between the main direction of the detected feature points and the path in the dynamic programming with modifying the weights of the paths in different directions.

The remainder of this paper is structured as follows. In Section2, the algorithm background on SIFT and semi-global matching are reviewed, and the methodology are described in details. Section3 shows the results of the experiments. Conclusion and future work are drawn in Section4.

\* Corresponding author.

E-mail address: moonriver\_song@163.com(Y Song)

## 2. STEREO MATCHING

Stereo matching is to find the correspondence between images of the same scene under different imaging angle. A matching search is performed on the selected matching primitive on the horizontal epipolar line of the corrected stereo image pair, calculate the corresponding parallax to restore the depth information of the object in the three-dimensional space.

### 2.1 Scale-Invariant Feature Transform

SIFT (Scale-invariant feature transform) is a point feature detection and description algorithm based on scale space, which maintains invariance to rotation, scaling, and brightness changes, and has strong robustness in stereo matching problems. In this paper, SIFT is used to generate massive feature points with high accuracy and robustness, and takes advantage of the implicit disparity and information of the detected feature points and successful matching pairs as constraints to guide the calculation process of SGM.

The main steps of the SIFT feature matching algorithm are as follows: firstly, construct scale space based on DOG pyramid and spatial extremum points are detected; secondly, accurately locate the stable key points and distribute the information of the stable key points to generate 128-dimensional feature description, then Euclidean distance is used as the similarity measure criterion to match the feature descriptors and on the basis of local distance check, RANSAC algorithm is used for further eliminate false matches. Finally, epipolar constraint is used to improve the accuracy of the match.

### 2.2 Semi-Global Matching

Taking multipath dynamic programming as a global energy minimization calculation strategy, SGM retains the characteristics of high dynamic planning and can generate dense disparity maps pixel by pixel. It can be divided into three steps: matching cost calculation, matching cost aggregation, disparity determination.

**2.2.1 Matching cost calculation :** In this paper, the SGM algorithm uses census transform as the matching cost, which is a local non-parametric transform. The encoding is performed in the order of the pixel gray levels in the window. Grey spatial information and local texture information can be preserved by census transform. Set the centre pixel  $p$  of the window as a reference, census transform compares the gray of other pixels in the window with  $p$  in sequence, where pixels with a smaller gray are 1, else 0, and compiles the result into a binary Hamming code. Defined as:

$$\delta(I(p), I(q)) = \begin{cases} 1 & I(p) < I(q) \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

$$C(p, d) = \bigotimes_{q \in W(p)} \delta(I(p), I(q)) \quad (2)$$

Where  $I(p)$ ,  $I(q)$  are the gray values of the pixels  $p$ ,  $q$ ;  $W(p)$  is a transform window centered on  $p$ , and  $C(p)$  represents a bit string transformed by Equation (2). Then calculate the Hamming distance between the left and right corresponding window bit strings  $C_l(p)$ ,  $C_r(p)$ , and get the initial matching cost as follows:

$$C(p, d) = \text{Hamming}[C_l(p), C_r(p + d)], d \in [d_{\min}, d_{\max}] \quad (3)$$

Where  $d_{\min}$  is the minimum parallax search range and  $d_{\max}$  is the maximum parallax search range.

**2.2.2 Disparity calculation :** Based on the matching cost of the stereo pair in the census transform calculation, the SGM algorithm realizes the optimal parallax acquisition by minimizing the energy function. As shown in the fig.1, a global optimal constraint is approximated by using the optimal matching path in eight directions.

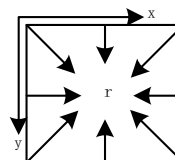


Figure 1. 8 paths

The matching cost of each point is the sum of the accumulation of the eight directional matching costs, which is calculated by dynamic programming. The calculation of the cost function  $L_r(p, d)$  of the pixel  $p$  in the disparity  $d$  along the direction of  $r$  is as in Equation 3.

$$L_r(p, d) = C(p, d) + \min \begin{pmatrix} L_r(p-r, d), \\ L_r(p-r, d-1) + P_1, \\ L_r(p-r, d+1) + P_1, \\ \min_i L_r(p-r, i) + P_2, \end{pmatrix} - \min_k L_r(p-r, k) \quad (4)$$

In the Equation: The first term is the initial matching costs; the second term is the minimum matching cost of previous point  $p-r$  including a penalty; the third term is added just to avoid  $L_r$  being too big. To ensure continuous smoothing of the overall parallax, the penalties for excessive disparity discrepancies will not be too great. Penalty  $P_1$  with disparity difference of 1 and penalty  $P_2$  with disparity greater than 1 are set to ensure an exact matching of discontinuous parallax, rather than being over smoothed.

**2.2.3 Disparity Determination :** After the matching cost  $S(p, d)$  of all pixels is calculated by Equation 4, the disparity of each pixel  $p$  can be obtained by Equation 5, where  $d_p$  is the value that minimizes the total matching cost.

$$d_p = \min_d S(p, d) \quad (5)$$

$$S(p, d) = \sum_r L_r(p, d) \quad (6)$$

### 2.3 SGM Integrated by SIFT

The integration of SIFT and SGM algorithm uses the detected feature points and matched points to constrain the SGM process, reduce the disparity search range and limit the spread of wrong disparity information. Each path cost aggregation in dynamic programming can be modified through extracting the main direction of each feature points.

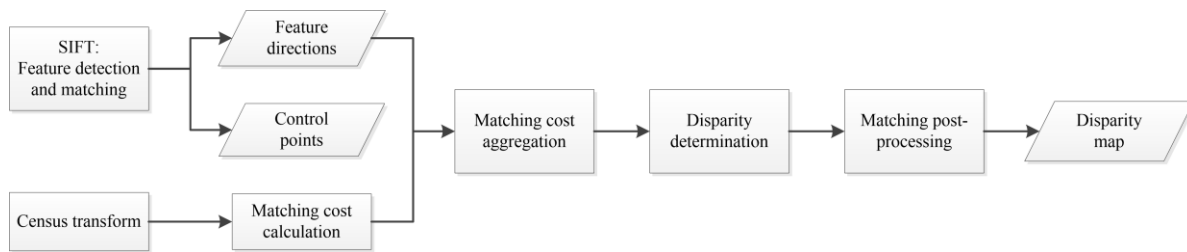


Figure 2. Processing steps for disparity estimation using the integrating SIFT and SGM algorithm

### 2.3.1 Parallax Correction and Search Range Reduction Based on Control points:

Through dynamic programming, the correct disparity can be calculated in most areas. But once there exist mismatching pixels, the wrong disparity will deliver to the following image pixels. Lin and Zhang (2006) proposed control points of parallax, which refers to absolutely reliable points in the disparity map. To solve the problem above, in this paper, the control points are used. SIFT can detect various local features that are insensitive to light and radiation changes. Through feature extraction, feature matching and excluding outliers, massive correspondence points can be obtained. From the paper (Wu et al., 2015), successful matching points generated by SIFT has a good corrective effect on the path aggregation of SGM algorithm. For pixels located in the control points, only the correct disparity accumulates cost.

Based on the assumption that the disparity is continuously changing in the interior region of the object (Marr and Poggio, 1979), this paper uses segmentation and control points to reduce the search range for pixels around the control points. Firstly, object-oriented segmentation is performed on the base image and the result is superimposed with the control points. The cost is calculated pixel by pixel. For pixel  $p$ , if its Manhattan distance from the nearest control point in the same partition is  $w$  which is less than the maximum disparity search range, the parallax search range can be reduced from the  $[d_{min}, d_{max}]$  to  $[d-w, d+w]$  based on the assumption that disparity in the same partition changes continuously. The  $d$  is the disparity of the nearest control points,  $d_{min}$  is the minimum parallax search range and  $d_{max}$  is the maximum parallax search range. In addition, when dynamic programming in the pixel where the control point is located, only the set disparity value can be given, thereby the propagation of the false matching cost can be truncated. By reducing the search range of dynamic programming on each pixel, the occupied memory can be reduced and the accuracy of the algorithm can be improved.

### 2.3.2 Aggregation Path Weight Correction Based on Feature Orientation:

Most SGM improved algorithm focus on improving the calculation of matching cost and setting the parameters of the penalty item to raise the accuracy, the path direction in cost aggregation is rarely considered. To study the impact of the aggregation path directions on the matching results, the SGM algorithm is divided into 8 aggregation paths, clockwise from  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ ,  $135^\circ$ ,  $180^\circ$ ,  $225^\circ$ ,  $270^\circ$  and  $315^\circ$ . Experiments were performed on Cones in the Middlebury dataset. The matching results of each direction are shown in Fig.3.

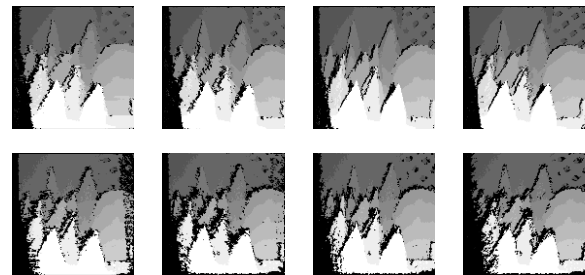


Figure 3. Matching results for different path directions

From the experiments, the matching effect of single path in different directions is various. Instead of the same weight to the paths in all directions, in this paper, feature direction is used to affect the weight of aggregation path. Based on the feature detection, the gradient direction and amplitude of the feature points are calculated, and the histogram is used for statistics. After the gradient direction histogram is constructed, the gradient information in different directions of each feature point is used to improve the path aggregation. And a relatively smaller weight is assigned to the path, which has a relatively larger gradient. An overview of the processing steps is given in Fig. 2.

## 3. DATA AND RESULTS

The established Middlebury stereo data sets (Aloe, Cones) and the stereo pair of CE-3 lunar data sets are used to verify the superiority of the modified SGM algorithm. Percentage of bad matching pixels is used as the evaluation criteria for dense matching and the matching quality of SIFT is measured by Root Mean Squared Error. The census transform is used as the matching cost and 8 path aggregations is adopted in the experiment.

### 3.1 Matching Quality Evaluation

In order to verify the effectiveness of this algorithm, the matching accuracy is assessed with the evaluation index which can be computed as follow,

(a) Percentage of Bad Matching Pixels (PBM)

$$P_{PBM} = \frac{1}{N} \sum_{(x,y)} (|d_e(x,y) - d_i(x,y)|) < \delta \quad (7)$$

where  $N$  is the total number of pixels,  $d_e(x,y)$  is the disparity value of the pixel in the experimental disparity map,  $d_i(x,y)$  is

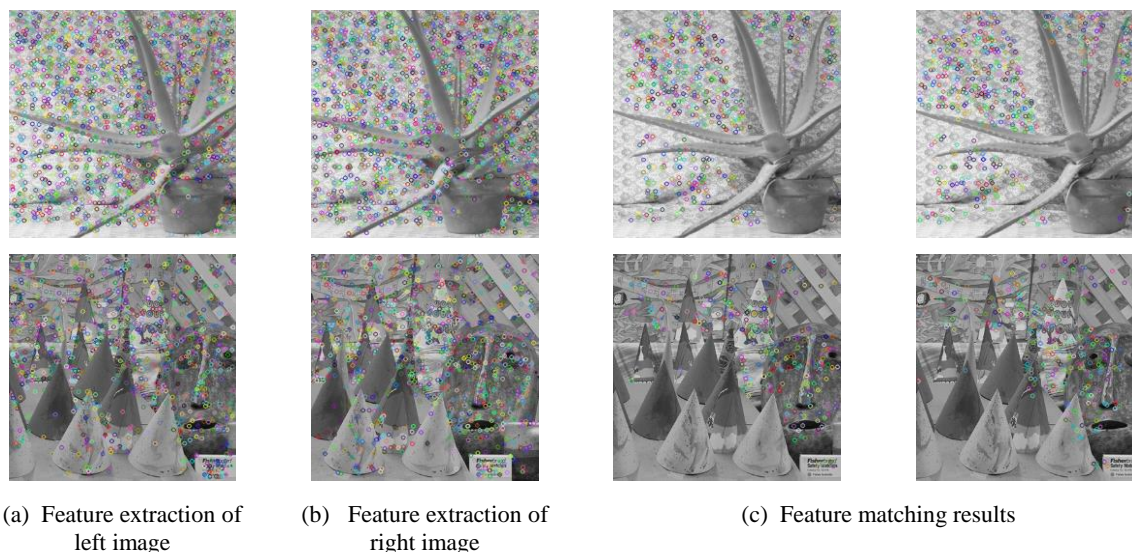


Figure 4. Feature extraction and matching results of Cones and Aloe

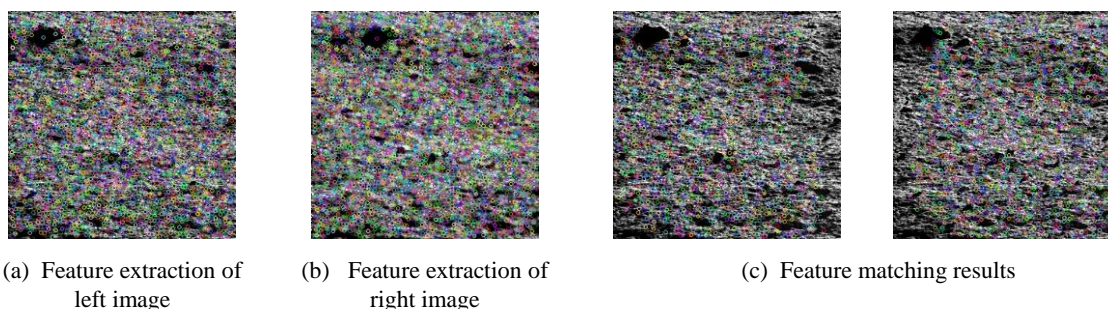


Figure 5. Feature extraction and matching results of ChangE-3 lunar stereo pair

the pixel value in the true disparity map. The threshold  $\delta$  is set to 1, if the matching result is within one pixel of the true disparity value, it is considered as the correct match.

(b) Root Mean Squared Error (RMSE)

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (m_1 x_{2i} + m_2 y_{2i} + t_x - x_{1i})^2 + (m_3 x_{2i} + m_4 y_{2i} + t_y - y_{1i})^2}{N}} \quad (8)$$

Where  $m_i$  ( $j=1, 2, 3, 4$ ) and  $t_x, t_y$  are the geometric transformation parameters obtained in the experiment;  $x_{1i}$  and  $y_{1i}$  are the feature points in the left image;  $x_{2i}$  and  $y_{2i}$  are the feature points corresponding to the left image in the right image;  $i=1, 2, \dots, N$ ,  $N$  is the number of control points.

### 3.2 Middlebury Images

The results of Middlebury stereo data set (Aloe, Cones) are displayed in Fig.6. Fig. 6(a) shows the left image of Aloe and Cones. The disparity ranges of Aloe and Cones are both 64 pixels. The results of feature extraction and feature matching using Middlebury Images are shown in Fig. 4(a) to (c). Massive evenly distributed feature points can be detected to provide gradient information and help to correct the weight of the aggregation path.

The matching accuracy and the number of correctly matched pairs are shown in Table 1. Fig. 6(b) shows the true parallax image of Aloe and Cones, and the parallax images obtained by the general SGM algorithm and the improved functions are shown in Fig. 6(c) to 6(f). According to the evaluation of control points extraction and dense matching, although there are a few false parallax control points, the accuracy of the final matching results will be effectively improved, which indicates that using matched points to reduce the search range will affect the accuracy of SGM while adding control points. Using the feature direction to affect the weight of the aggregation path can also improve the matching quality, and a higher accuracy will be brought by these two improved functions.

	Number of feature points	Number of matching pairs	RMSE (unit: pixel)
Aloe	1973/2123	692	1.043
Cones	1045/1026	324	1.316
CE-3 data	3450/3949	951	0.8189

Table 1. SIFT matching results

### 3.3 Real World Images

As the crucial link of China's Moon Exploration Program, the Chang'e 3 (CE-3) detector has successfully landed on the moon softly on 14th, Dec, 2013, which is China's first unmanned lunar detector. The stereo pair of CE-3 lunar data sets are provided by a panoramic camera carried on the rover (also



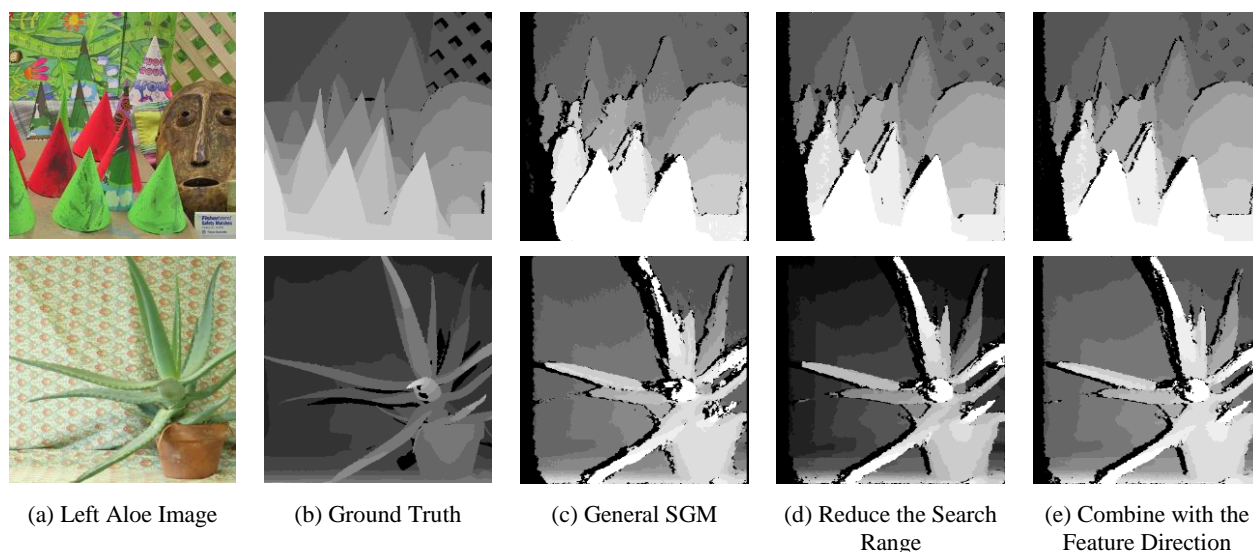


Figure 6. Disparity maps obtained with improved functions for Cones and Aloe.

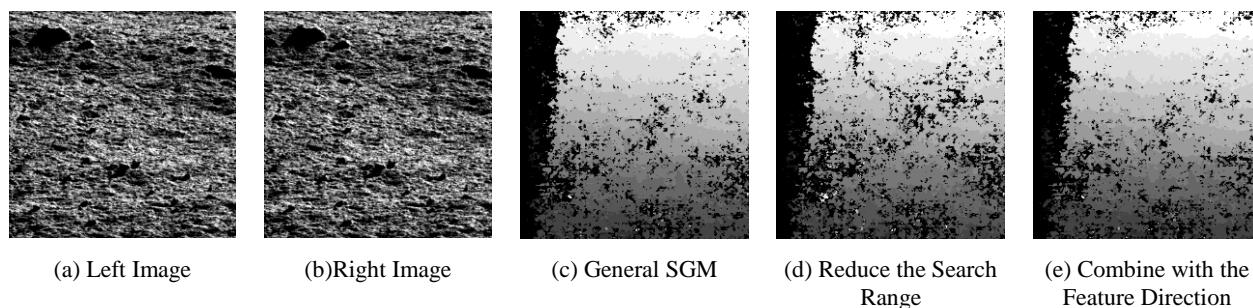


Figure 7. Disparity maps obtained with improved functions for stereo images of the lunar surface.

known as the “Yutu” lunar rover), which is a good data support for lunar terrain reconstruction.

The image of the lunar surface used in the experiment is shown in Fig. 7. The parallax range is 0~64 pixels. Affected by the lunar's undulating terrain and the solar approximately parallel irradiation, the difficulties of unbalanced illumination distribution are brought to the image matching. As can be seen from Fig. 4, SIFT extract massive high-precision feature points for lunar images with weak texture and repeated texture. The results of general SGM algorithm and improved SGM algorithm for dense matching of stereo image pair of CE-3 data as shown in Fig. 7(c), Fig. 7(d) and Fig. 7(e), and in lack of real parallax data, CE-3 is not evaluated in Table2.

	General SGM	Combine with the Feature Direction	Reduce the Search Range
Aloe	75.60%	76.55%	76.83 %
Cones	78.33%	81.35%	81.41 %

Table 2. Accuracy of improved functions for Cones and Aloe

#### 4. RESULT AND FUTURE WORK

This paper presents a dense matching algorithm by integrating SIFT and SGM, takes advantage of the implicit disparity and gradient information of the successful matching points generated by SIFT as constraints to guide the cost calculation process. Experimental results show that this algorithm can

improve the matching accuracy by cutting off the error propagation, reducing disparity search range and assigning different weights to each aggregation path, is a reasonable and effective stereo image dense matching algorithm. However, the efficiency and precision of the algorithm depend on the prior information, such as the matching quality of feature points. Therefore, our future is to enhance the SIFT algorithm to obtain more and higher precision control points and make full use of gradient information to correct the weight of aggregate path.

#### ACKNOWLEDGEMENTS

This work was supported by Natural Science Fund of Hubei Province (No.2016CFB690) and The Fund of Key Laboratory of Technology for Safeguarding of Maritime Rights and Interests and Application (No. SCS1610).

#### REFERENCES

- Boykov, Y., Veksler, O., Zabih, R., 2002. Fast approximate energy minimization via graph cuts, The Proceedings of the Seventh IEEE International Conference on Computer Vision, pp. 377-384 vol.371.
- Chen, Z., Hao, L.I., Zhou, Y., Wen, S., University, H., 2017. Semi-global dense matching method based on region growing. Science of Surveying & Mapping.

Hermann, S., Klette, R., 2012. Iterative Semi-Global Matching for Robust Driver Assistance Systems. *Lecture Notes in Computer Science* 7726, pp. 465-478.

Hirschmuller, H., 2005. Accurate and Efficient Stereo Processing by Semi-Global Matching and Mutual Information, *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, pp. 807-814.

Li, S., Li, H., Wang, S., Chen, Z., 2017. Improved Semi Global Matching Method Based on ADCensus. *Journal of Gansu Sciences*.

Lin, G.Y., Zhang, W.G., 2006. A Control-point-based Stereo Matching Algorithm for Layered and Bidirectional Dynamic Programming. *Information & Control* 35, pp. 411-416.

Lowe, D.G., 2004. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision* 60, pp. 91-110.

Marr, D., Poggio, T., 1979. A computational theory of human stereo vision. *Proc R Soc Lond B Biol Sci* 204, pp. 301-328.

Scharstein, D., Szeliski, R., 2001. A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms, *smbv*, p. 0131.

Sun, J., Zheng, N.N., Shum, H.Y., 2002. Stereo matching using belief propagation, *European Conference on Computer Vision*, pp. 510-524.

Wu, J., Yao, Z., Cheng, M., 2015. Airborne oblique stereo image dense matching by integrating SIFT and SGM algorithm. *Journal of Remote Sensing*.

Zhu, Q., Chen, C., Han, H.U., Ding, Y., 2017. An Adaptive Dense Matching Method for Airborne Images Using Texture Information. *Acta Geodaetica Et Cartographica Sinica*.