Research Article

Binocular Images Dense Matching considering Image Adaptive Color Weights and Feature Points

Zhenghui Xu and Jingxue Wang

School of Geomatics, Liaoning Technical University, Fuxin 123000, China
Collaborative Innovation Institute of Geospatial Information Service, Liaoning Technical University, Fuxin 123000, China

Correspondence should be addressed to Jingxue Wang: wangjingxue@lntu.edu.cn

Received 20 January 2022; Revised 27 February 2022; Accepted 16 May 2022; Published 30 May 2022

Copyright © 2022 Zhenghui Xu and Jingxue Wang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

When the matching cost function in Semiglobal Matching is unstable, the inaccurate matching cost values will be propagated in the cost aggregation process. It will lead to a serious mismatching phenomenon. To address the problem, a binocular images dense matching method considering image adaptive color weights and feature points was proposed. Firstly, the Color Birchfield Tomasi (CBT) matching cost calculation method was proposed to obtain a stable initial cost volume, which combined image adaptive color weights and gradient information. Secondly, the Scale-invariant Feature Transform matching algorithm was used to extract the a priori feature points from binocular images. Then, the feature points were filtrated. The cost volume was optimized by using their coordinate information and disparity information. Finally, an aggregation path segmentation rectification method was adopted to optimize the SGM aggregation paths and reduce the propagation of incorrect paths. Experimental results demonstrate that the proposed method can effectively improve the stability and accuracy of dense matching, reduce the mismatching phenomenon, and finally produce high-quality disparity maps.

1. Introduction

In recent years, stereo matching has played an essential role in the field of photogrammetry and computer vision [1]. Stereo matching mainly refers to finding pixels in stereo images corresponding to the same scene point. The corresponding points obtained from the binocular imaging model can be used to recover the depth of the scene or its 3D coordinate information. Therefore, stereo matching has a wide application space in 3D reconstruction [2], medicine [3], face recognition [4], automatic driving [5], and many other fields. As for stereo matching, it can be divided into two main categories. One is the feature matching method based on image feature point information, and the other is the dense matching method based on image pixel information.

Image feature matching is actually a process of feature extraction, description, and matching [6]. Firstly, the interest features and attributes of multiple images are extracted. Then, the parametric description is performed. Finally, similarity matching is performed on the extracted features. At present, there are many commonly used feature point matching algorithms, such as Harris [7], Forster [8], SIFT (Scale-invariant Feature Transform) [9], and deep learning methods [10]. Due to the features of rotation invariance, scale invariance and affine invariance [11], SIFT has become a commonly used algorithm in feature matching. Although the extracted coordinates of the corresponding points are accurate, the number of corresponding points obtained by feature matching is few, which will result in insufficient detailed information in 3D reconstruction. Hence, the methods that make sparse feature points become dense have been widely concerned. Aurenhammer constructed Voronoi polygons based on sparse feature points to divide the image and used the SSD (Sum of Squared Difference) method to process each pixel in the polygon area to obtain dense points [12]. A matching method of adaptive propagation was proposed by [13]. On
the basis of feature matching, corresponding triangulations were built for the two images. And then feature dense matching was carried out inside the triangulations. Delaunay triangulations were established after SIFT feature point extraction, and the dense point set was obtained by iterative processing based on the triangle center of gravity [14]. Although these algorithms can get a dense set of corresponding points, they depend on high precision initial feature points. When the features in the region are not obvious, it is easy to cause incorrect point matching. And the methods above are also complicated.

Dense matching is a simple and direct method to obtain corresponding points in stereo vision. It is mainly dependent on the gray level of image pixels. The disparity per pixel is calculated by establishing the corresponding matching relationship between pixels in two images. Owing to the advantages of low cost and high density of matching points [15], dense matching is widely used in change detection [16], mapping [17, 18], smart city [19], and other fields. Compared with feature matching, dense matching can obtain more matching points. But on the other hand, the accuracy of the matching points is inadequate, and the mismatching phenomenon is serious. At present, dense matching methods can be divided into the global method [20], SGM (Semiglobal Matching) [21], local method [22], and deep learning method [23]. There are many advantages of SGM, such as good universality, satisfactory efficiency, and high matching accuracy. SGM does not rely on data sets [24]. So SGM is the most commonly utilized mainstream method.

SGM dense matching method mainly involves four steps: matching cost calculation, cost aggregation, disparity calculation, and disparity refinement. The stability of matching cost calculation methods and the accuracy of cost aggregation paths directly affect the accuracy of dense matching. In the matching cost calculation methods, local windows were often used [25–28]. The matching cost values were computed by setting regular local windows on binocular images, respectively, and evaluating the correlation of pixels in the windows. Although these methods are simple, the matching accuracy depends heavily on the size of the window. MI (Mutual Information) [20] method is not sensitive to illumination information, but it is complicated and needs iteration. So MI method is not commonly used. Combining the AD (Absolute Difference) method with the Census method, Mci proposed the AD-Census joint cost calculation method [29], which can not only preserve image edges but improve the accuracy of matching cost results as well. Although the joint cost is beneficial to make the cost volume with multiple image features [30], it will also weaken some original features and make the calculation more complicated. Although the BT (Birchfield Tomasi) method [31] can maintain the continuity of disparities, it ignores the color information of the image itself and is not conducive to the edge preservation of the disparity map. Therefore, the BT method still has a stability issue. In cost aggregation, Gehrig improved the SGM and proposed a real-time dense matching method [32]. Rothermel proposed the T-SGM to accelerate SGM [33]. In T-SGM, the original image is stratified and downsampled before the cost aggregation process. However, the image downampling process will affect the quality of the final result. Meanwhile, the above methods ignore the path propagation effect in cost aggregation, which will cause the incorrect matching cost values to be continuously propagated and affect the accuracy of the final disparity map.

In summary, in order to improve the accuracy of dense matching, we analyzed the unique advantages of the two matching methods and proposed a binocular images dense matching method combining image adaptive color weights and feature points. The main contributions of the method are as follows: (1) A more stable CBT matching cost calculation method was proposed. In the calculation process, the color information of each pixel can be adaptively weighted, which can better reflect the color information of the image itself. And it is beneficial to preserve the edge of the disparity map by adding the constraint of image gradient information. (2) A segmentation correction cost aggregation method was adopted to reduce the mismatching phenomenon. The initial cost volume was optimized according to the image prior to feature point information. During the SGM cost aggregation, the aggregation paths were corrected in segments to reduce the error path, avoid the propagation of the incorrect matching cost values, and improve the accuracy of the whole dense matching.

2. BT Cost Calculation Method Combined with Image Adaptive Color Weight

In the real world, the depth of the scene is continuous. Discretization errors will be generated during image sampling. As a result, the image depth is discontinuous when the camera is used to obtain the real scene image. For binocular dense matching, the image depth discontinuity is actually the image disparity discontinuity. The discontinuity of image depth and disparity will directly lead to the mismatching phenomenon in the process of stereo matching, which will have a negative impact on the matching accuracy and the 3D reconstruction effect. The linear interpolation calculation method was adopted in BT method [31], and gray images were used to calculate directly. The subpixel disparity values were obtained by calculating the subpixel matching costs of the image. The linear interpolation method is simpler than the subpixel matching method, and it can effectively avoid the image depth discontinuity and reduce the image sampling errors. So compared with other matching cost calculation methods, the advantage of the BT method [30] is that it can maintain the continuity of disparities. A sketch map of the linear interpolation method is shown in Figure 1.

In Figure 1, $I_L$ and $I_R$ are the left image and the right image, respectively. $I_L$ is the reference image. $x_i$ and $x'_i$ are two points to be matched in $I_L$ and $I_R$, respectively. $I_{L'}$ is the linear interpolation result of $x_i$ with its left neighbor pixel $x_i - 1$. $I_{R'}$ is the linear interpolation result of $x'_i$ with its right neighbor pixel $x'_i + 1$. The calculation methods of $I_{L'}$ and $I_{R'}$ are shown as follows:
In the BT method, the image is sampled by linear interpolation. Then the similarity of two points is assessed by the pixels dissimilarity measurement method. The measurement result of the BT method is the BT value. Disparity level is denoted by \( d \). In the range of \( d \), if the BT value of the two matching points in binocular images is larger, that means the similarity between the two points is smaller, and the two points are dissimilar. When the BT value is the smallest, it indicates that the similarity between the two points is the greatest. At this time, the corresponding matching points are similar points. When the BT method is used to calculate the similarity of two points in binocular images, the left image is taken as the reference image. At the disparity level \( d \), \( x'_i = x_i + d \). The differences of pixel gray values for \( I_L \) with \( I'_G \), \( I'_R \) and \( I'_B \) computed, respectively. The minimum value of computation results is taken and regarded as \( d_L \). Then the right image is taken as the reference image. The differences of pixel gray values for \( I_R \) with \( I'_G \), \( I'_L \) and \( I'_I \) are computed, respectively. The minimum value of computation results is taken and regarded as \( d_R \). The final result \( d(x_i, x'_i) \) is the minimum value between \( d_L \) and \( d_R \), which is the result of matching cost calculation. The calculation methods of BT matching cost are shown as follows:

\[
\begin{align*}
I^*_R &= \frac{I_R(x'_i - 1) + I_R(x'_i)}{2}, \\
I^*_L &= \frac{I_L(x'_i) + I_L(x'_i + 1)}{2}.
\end{align*}
\]

In the BT method, \( I^*_R \) and \( I^*_L \) are selected for calculation, the two points are corresponding points. When channels \( R \) or \( G \) are selected for calculation, the two points are not corresponding points. (2) The gradient changes of pixels in the image are not considered in the BT method. In the calculation process, with the change of disparity \( d \), the two images maintain relative motion (as shown in Figure 2). The relative motion will lead to the gradient changes in the image. The lack of gradient constraint will result in the loss of the image edge information. So, the stability and edge constraint ability of the BT method should be improved. In Figure 2, \( p \) is a point in the cost volume after matching cost calculation.

\[
\begin{align*}
d_L &= \min \{|I_L(x_i) - I'_R(x'_i)|, |I'_L(x_i) - I_R(x'_i)|, |I'_L(x_i) - I'_R(x'_i)|\}, \\
d_R &= \min \{|I_R(x'_i) - I'_L(x_i)|, |I'_R(x'_i) - I_L(x_i)|, |I'_R(x'_i) - I'_L(x_i)|\}, \\
d(x_i, x'_i) &= \min(d_L, d_R).
\end{align*}
\]

To address the problems in the BT method, the CBT method combining image adaptive weight and gradient information was proposed. Firstly, the matching results for each color channel of the images were calculated by the BT method. The adaptive weight for each color component per pixel was calculated according to the image color information. The matching result \( d_c \) of the adaptive color weights for two images was obtained by weighted sum. Then the horizontal and vertical gradient information of the two images was then obtained at each disparity \( d \). BT values of horizontal gradient and vertical gradient for the two images were calculated separately. Finally, the weighted results of color and gradient information were computed. Equations (6)–(9) describe the computation process of the CBT method.

\[
\begin{align*}
d_c &= \sum_{c=R,G,B} \lambda_c \min(d_{c,x_i}, d_{c,x'_i}), \\
d_{c,x_i} &= \min(d_{c,x_i}, d_{c,x'_i}).
\end{align*}
\]
\[
\begin{align*}
\n d_{V_x} & = \min(d_{V_x}, d_{V_y}), \quad (8) \\
D_{\text{cost}} & = \omega_1 \min(d_c, \tau) + \omega_2 \min(d_{V_x}, \tau) \\
& + \omega_3 \min(d_{V_y}, \tau), \quad (9)
\end{align*}
\]

where, \(d_c\), \(d_{V_x}\) and \(d_{V_y}\) are the results of color information, horizontal gradient information, and vertical gradient information, respectively. \(c\) indicates the color channel \(R, G, B\). \(\lambda_c\) represents the adaptive color weight for each pixel, \(\lambda_c = \frac{x_c}{\sum c \in \{R, G, B\} x_c}\), where \(x_c\) represents the gray value of pixel \(x\) on channel \(c\). \((\nabla x_r, \nabla y_r)\) and \((\nabla x_L, \nabla y_L)\) are the gradient information of binocular images. \(\omega_1\) is the color weight. \(\omega_2\) and \(\omega_3\) represent the gradient weights. The sum of \(\omega_1\), \(\omega_2\) and \(\omega_3\) is 1. \(D_{\text{cost}}\) is the final result of cost volume. In this paper, \(\omega_1\), \(\omega_2\) and \(\omega_3\) are 0.6, 0.2, 0.2, respectively. \(\tau\) is the truncation error.

\[
E(D) = \sum_p \left( C(p, d_p) + \sum_{q \in N_p} P_1 T \left[ |d_p - d_q| = 1 \right] + \sum_{q \in N_p} P_2 T \left[ |d_p - d_q| > 1 \right] \right), \quad (10)
\]

where \(p\) denotes a pixel. \(d_p\) refers to the disparity value of \(p\). \(C(p, d_p)\) represents the matching cost value of \(p\) at the disparity \(d_p\). \(P_1\) and \(P_2\) are the two penalty parameters of the external input, where \(P_2 > P_1\). \(q\) is the neighbor pixel of \(p\). \(T[\cdot]\) is a discriminant function, which is used to judge the relationship between \(d_p\) and \(d_q\). \(P_1\) penalty is imposed when \(|d_p - d_q| = 1\). \(P_2\) penalty is imposed when \(|d_p - d_q| > 1\). The last two terms in (10) are smoothing terms. The energy function is used to calculate the whole image from a global perspective [35]. Since the energy function \(E(D)\) is not differentiable, it actually becomes an NP problem throughout solving. In order to simplify the calculation of \(E(D)\), SGM considers optimizing point \(p\) from multiple directions. Then the results of optimization in each direction are added to get the final result (as shown in equations (11) and (12)), which is the cost aggregation of the SGM method.

\[
L_r(p, d) = C(p, d) + \min \left\{ \begin{array}{l}
L_r(p - r, d) \\
L_r(p - r, d - 1) + P_1 \\
L_r(p - r, d + 1) + P_1 \\
\min L_r(p - r, i) + P_2
\end{array} \right. - \min L_r(p - r, k), \quad (11)
\]

\[
S(p, d) = \sum_r L_r(p, d), \quad (12)
\]
where $r$ refers to direction. $L_r(p,d)$ is the cost value of $p$ in the direction $r$. The aggregation cost along a direction can be regarded as the information transmitted by each pixel at every disparity $d$ in the direction. $L_r(p-r,d)$ denotes the matching cost value of the previous point of $p$ in the current path. $\min L_r(p-r,k)$ is a constraint term to prevent the calculation result from being too large. $S(p,d)$ represents the final cost aggregation result.

SGM implements 2D constraints using a large number of 1-D constraints, which can reduce the amount of computation, improve the efficiency of operation, and achieve an approximate semiglobal effect. However, every disparity $d$ at every pixel is actually considered in SGM. In fact, the cost aggregation process is a direct correlation process of neighbor pixels and optimizes the current pixel cost value based on the neighbor pixel cost value. If the matching cost calculation yields an incorrect cost value, the incorrect cost value will be continuously transmitted along the current aggregation path in cost aggregation. The aggregation results of the adjacent pixels will be affected. The transmission of the incorrect matching cost will directly result in the mismatching phenomenon of the image. Therefore, we proposed using image feature points to rectify the aggregation path and improve the matching accuracy.

SIFT method was used to extract the image feature points in the paper. Since the coordinates of feature points obtained by SIFT are subpixel level, they should be converted to integer-pixel level. At the same time, mismatching points and repeated points should be eliminated to get accurate feature points. According to the coordinates of the feature points, accurate disparity information between the corresponding two points can be acquired. Information of a priori feature points was used for accurate positioning in 3D cost volume. The positions were taken as the guide points. Matching cost values corresponding to other candidate disparities of the point were set as invalid values to optimize the cost volume. In the next cost aggregation process, the path through the point is accurate and unique. This aggregation path constraint (as shown in Figure 3) is beneficial to improving the accuracy of dense matching and reducing the propagation of incorrect cost values. So the dynamic programming path of SGM can be rewritten into the form of the following equations:

$$
L'_r(p,d) = \begin{cases} 
L_r(p,d), & p \notin P, \\
C(p,d), & p - q' = d, \ p \in P, \\
\text{Invalid}, & p - q' \neq d, \ p \in P,
\end{cases}
$$

$$
S(p,d) = \sum_r L'_r(p,d),
$$

where $P$ is the set of corresponding points. $p$ refers to the point in the left image. $q'$ represents the corresponding point of $p$ in the right image. If the number of corresponding points in $P$ is larger, it will make the path constraint effect more obvious and make the accuracy of SGM higher.

The experimental process of the proposed method is shown in Figure 4. The specific experimental steps are as follows:

1. Binocular rectified images are input;
2. The proposed CBT method is used to obtain the initial matching cost volume;
3. SIFT feature matching is performed on binocular images to acquire feature points;
4. The feature points are optimized to obtain feature point coordinates information and accurate disparity information;
5. According to the information of a priori feature points, the cost volume is optimized. Except for the accurate disparity of the feature points, the corresponding cost values of the other candidate disparities are set to be invalid;
6. Cost aggregation of path constraints is performed;
7. WTA method is used to obtain the disparity map based on the left image;
8. Left and right consistency detection is carried out;
9. The final disparity map is output;
10. The quality of the disparity map is evaluated.

4. Dense Matching Experiment

4.1. Experimental Data. Cone and Teddy image pairs (as shown in Figures 5 and 6) in Middlebury public data set were selected for experiments in the paper. To guarantee that disparity values in the experimental results are accurate and reliable, the left and right consistency detection was conducted. To analyze the characteristics of the proposed CBT method, the CBT method was compared with AD, BT, Census [36], and AD-Census methods. Meanwhile, the proposed dense matching method was compared with the traditional SGM method. Since the relative gray values in the window are compared in the Census method, it is easy to produce two neighborhood windows with completely different gray levels in low texture and repeated texture regions, resulting in the occurrence of a mismatching phenomenon. On the other hand, Census relies too much on the stability of the center point in the window, and the correlation between pixels in the window is weak. So as to increase the stability of the Census, the window size for Census calculation was set to $3 \times 3$. The center pixel in the window was replaced with the mean value of the window. In this way, the correlation between pixels in the window can be increased, and the distortion of the center pixel can be avoided. Experimental results of Cone images are shown in Figures 7–10. Figure 9 is the local detail of the corresponding results. Figure 10 is the error map of the corresponding results. Experimental results of Teddy images are shown in Figures 11–14. Figure 13 is the local detail of the corresponding results. Figure 14 is the error map of the corresponding results.

As for the quality evaluation of disparity maps, after consistency detection, the places in the disparity map with a value larger than 0 are effective disparities. For binocular images without ground truth, the number of effective
**Figure 3:** Sketch map of SGM aggregation path correction. The red line represents the incorrect matching path and the green line indicates the correct matching path. According to the coordinates of feature points obtained by SIFT method, the accurate disparity information of the corresponding two points can be obtained. The green checkmarks are the exact matching point positions, which are obtained by the coordinates and the disparity information of feature points. The yellow crosses are invalid cost values. The matching paths cannot pass. When the incorrect matching path passes through a matching point, the matching path will be forcibly corrected to the matching point, which plays the role of path rectification.

**Figure 4:** The flow chart of the experiment.

**Figure 5:** Cone experimental images. (a) Left image, (b) right image, (c) ground truth.
disparities is an important indicator to measure the quality of the disparity maps [1]. So the percentage of effective disparities was recorded for every result, and the histograms were generated for comparative analysis (as shown in Figures 15 and 16) to illustrate the reliability of the proposed method. The elapsed times of different matching cost calculation methods are shown in Table 1. According to the ground truth, RMSE (Root Mean Square Error) and PBM (Percentage of Bad Matching Pixels) were used to assess the quality of experimental results. The quality evaluation results are shown in Tables 2–5. In Tables 3 and 5, “noncc” stands for the PBM of the nonoccluded area. If the RMSE value and the PBM value are smaller, it demonstrates the quality of the result is better.

4.2. Experimental Results and Analysis. According to Figures 7–11, the number of mismatching pixels in AD results is the most. Due to the interference of many factors in the calculation process, the stability of the AD method is not good enough. The BT method can improve the stability of the cost volume by calculating the subpixel disparities. However, there are still obvious mismatching regions in BT results, and the results have insufficient edge information and poor quality. As for Census results, the Census method has a good detection effect on the edge and corner features of images, making image edges more obvious. But the matching accuracy in the Census results is still not ideal. The advantages of AD and Census methods are combined by the AD-Census method. The AD-Census method can not only preserve the color information but also have a certain ability of image edge protection. As shown in Figures 7(e) and 7(f) and Figures 11(e) and 11(f), there is a superior result produced by the CBT method. Considering the adaptive color weight information of the image can make the results reflect the real color information of the image, which can obviously suppress the occurrence of a mismatching
phenomenon. Combining gradient information is beneficial to improve the matching stability and image edge constraint ability. Compared with the Census method, the CBT method can effectively avoid poor performance in low texture and repeated texture regions. Therefore, the edge preservation effect of the CBT method is better than those of the Census. According to Table 1, the computational efficiency of the CBT method is better than that of joint cost methods, which also indicates CBT method is simpler than the joint cost methods.
Compared to Figures 7 to 16, the results of the proposed dense matching method are better than that of the traditional SGM method. It can rectify the paths of cost aggregation, reduce the propagation of incorrect cost, and increase the accuracy of dense matching by using image prior feature points to optimize the cost volume. The proposed method can increase the detailed information of disparity maps and make the information of disparity maps become rich and complete. As for the error maps, the error map quality of the proposed method is the best, which indicates the corresponding disparity result has a better matching accuracy. By observing the percentage of
effective disparities in the histogram, the proposed method can increase the number of effective disparities, which indicates the integrity of the disparity maps is improved and the mismatching phenomenon has been suppressed. Therefore, the accuracy of dense matching can be effectively improved by using feature points to rectify the aggregation paths. By observing the experimental results, the results of the proposed method are better, which demonstrates the effectiveness of the method.

Figure 12: Teddy image results with the proposed method. (a–f) The same as in Figure 7.

Figure 13: Detail comparison of Teddy image results. (a) Figure 11(d), (b) Figure 11(e), (c) Figure 11(f), (d) Figure 12(d), (e) Figure 12(e), and (f) Figure 12(f).
From the evaluation results of RMSE and PBM in Tables 2–5, the quality of AD disparity results is the lowest and the mismatching rate is the highest. The quality of CBT results is the best, which indicates combining color information and gradient information of the image can effectively improve the stability of the method. The CBT method can make the disparity results close to the ground truth. Compared with the SGM method, the indicators of the proposed method are better, and the mismatching rate of the corresponding disparity results is lower. It proves that using feature point constraints is also helpful in enhancing the quality of disparity maps and improving the accuracy of dense matching. The objective quality evaluation results are consistent with the subjective quality evaluation results.

Therefore, from the experimental results and indicator evaluation results, the proposed dense matching method has certain superiority in vision and indicator, which demonstrates the proposed method is feasible.
The proposed method

Figure 16: The percentage of effective disparities in Teddy results.

Table 1: The elapsed times of different matching cost calculation methods.

|       | AD (s) | BT (s) | Census (s) | AD-census (s) | CBT-census (s) | CBT (s) |
|-------|--------|--------|------------|----------------|----------------|---------|
| Cone  | 0.43   | 0.33   | 2.14       | 2.54           | 3.93           | 1.67    |
| Teddy | 0.49   | 0.32   | 2.13       | 2.53           | 3.72           | 1.59    |

Table 2: RMSE evaluation results of Cone image.

|       | AD    | BT    | Census | AD-census | CBT-census | CBT   |
|-------|-------|-------|--------|-----------|------------|-------|
| SGM   | 17.74 | 14.63 | 15.92  | 15.42     | 13.83      | 13.51 |
| SIFT-SGM | 16.22 | 14.40 | 15.66  | 14.76     | 13.73      | 13.46 |

Table 3: PBM evaluation results of Cone image.

|       | AD    | BT    | Census | AD-census | CBT-census | CBT   |
|-------|-------|-------|--------|-----------|------------|-------|
| SGM   | 15.07 | 16.28 | 7.29   | 8.89      | 10.22      | 12.03 |
| SIFT-SGM | 9.89  | 11.25 | 6.96   | 8.51      | 9.47       | 11.24 |

Table 4: RMSE evaluation results of Teddy image.

|       | AD    | BT    | Census | AD-census | CBT-census | CBT   |
|-------|-------|-------|--------|-----------|------------|-------|
| SGM   | 13.95 | 12.44 | 14.07  | 13.69     | 12.20      | 11.85 |
| SIFT-SGM | 13.89 | 12.18 | 13.73  | 13.19     | 12.15      | 11.64 |

Table 5: PBM evaluation results of Teddy image.

|       | AD    | BT    | Census | AD-census | CBT-census | CBT   |
|-------|-------|-------|--------|-----------|------------|-------|
| SGM   | 11.08 | 13.29 | 7.57   | 9.93      | 11.31      | 13.63 |
| SIFT-SGM | 10.86 | 13.08 | 7.05   | 9.46      | 11.74      | 10.28 |
5. Conclusions
The mismatching phenomenon in traditional SGM was researched. The reasons for mismatching in the SGM method were analyzed. Combined with the properties of feature matching and dense matching, a binocular images dense matching method considering image adaptive color weights and feature points was proposed. The experimental results indicate that the proposed dense matching method has good stability. The method can reflect the real color information of images, and be beneficial to improving the whole quality and detailed information of the disparity map. In the meantime, the method can significantly reduce the mismatching phenomenon and enhance the accuracy of dense matching. Compared with other methods in this paper, the error matching rates of Cone and Teddy images in the proposed method are the lowest, which are 3.05 and 4.17, respectively. Hence, the proposed method has certain advantages in improving the accuracy of 3D image reconstruction.

Data Availability
The dataset used to support the findings of this study is included in the article, which is cited at relevant places within the text as [37].

Conflicts of Interest
The authors declare no conflicts of interest.

Acknowledgments
This research was funded by the National Natural Science Foundation of China, grant nos. 41871379 and 42071343, Liaoning Revitalization Talents Program, grant no. XLYC2007026, and Discipline Innovation Team of Liaoning L`he research was funded by the National Natural Science Foundation of China, grant nos. 41871379 and 42071343, Liaoning Revitalization Talents Program, grant no. XLYC2007026, and Discipline Innovation Team of Liaoning L’he research was funded by the National Natural Science Foundation of China, grant nos. 41871379 and 42071343, Liaoning Revitalization Talents Program, grant no. XLYC2007026, and Discipline Innovation Team of Liaoning

References
[1] Y. S. Li, Research on Key Techniques of Fast Stereo Matching for Photogrammetric Images, Wuhan University, Wuhan, Hubei, 2018.
[2] L. Ye and B. Wu, "Integrated image matching and segmentation for 3D surface reconstruction in urban areas," Photogrammetric Engineering & Remote Sensing, vol. 84, no. 3, pp. 135–148, 2018.
[3] X. Lai, X. Xu, J. Zhang, Y. Fang, and Z. Huang, "An efficient implementation of a census-based stereo matching and its applications in medical imaging," Journal of Medical Imaging and Health Informatics, vol. 9, no. 6, pp. 1152–1159, 2019.
[4] Z. Li, J. Yuan, B. Jia, Y. He, and L. Xie, "An effective face anti-spoofing method via stereo matching," IEEE Signal Processing Letters, vol. 28, pp. 847–851, 2021.
[5] P. X. Han, M. Zhao, and S. Y. Chen, "Fusion of texture, color and gradient information for stereo matching cost computation," in Proceedings of the 2017 2nd International Conference on Image Vision and Computing, vol. 22-29, pp. 118–121, Chengdu, China, June 2017.
[6] Z. X. Geng, B. M. Zhang, and D. Z. Fan, Digital Photogrammetry, Surveying and Mapping Press, 2010.
[7] C. Harris and M. Stephens, "A combined corner and edge detector," in Proceedings of the 4th Alvey Vision Conference, vol. 3, pp. 147–151, 1988.
[8] W. Förstner and E. Gilch, "A fast operator for detection and precise location of distinct points, corners and centres of circular features," in Proceedings of the ISPRS Intercommission Conference on Fast Processing of Photogrammetric Data, vol. 2, Switzerland, June, 1987.
[9] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," International Journal of Computer Vision, vol. 60, no. 2, pp. 91–110, 2004.
[10] G. Yao, A. Yilmaz, F. Meng, and L. Zhang, "Review of wide-baseline stereo image matching based on deep learning," Remote Sensing, vol. 13, no. 16, p. 3247, 2021.
[11] J. Wu, Z. X. Yao, and M. M. Cheng, "Airborne oblique stereo image dense matching by integrating SIFT and SGM algorithm," Journal of Remote Sensing, vol. 19, no. 03, pp. 431–442, 2015.
[12] F. Aurenhammer and D. Voronoi, "Voronoi diagrams—a survey of a fundamental geometric data structure," ACM Computing Surveys, vol. 23, no. 3, pp. 345–405, 1991.
[13] Q. Zhu, B. Wu, and J. Zhao, "A reliable image matching method based on self-adaptive triangle constraint," Chinese Journal of Computers, vol. 28, no. 10, pp. 1734–1739, 2005.
[14] X. Zhang, "A dense matching algorithm for remote sensing images based on reliable matched points constraint," Journal of Geoinformation Science, vol. 23, no. 8, pp. 1508–1523, 2021.
[15] D. C. Gong, Y. L. Han, and X. Huang, "Global refinement of building boundary with line feature constraints for stereo dense image matching," Acta Geodaetica et Cartographica Sinica, vol. 50, no. 6, pp. 833–846, 2021.
[16] X. M. Tang, H. Y. Wang, and X. Y. Zhu, "Technology and applications of surveying and mapping for ZY-3 satellites," Acta Geodaetica et Cartographica Sinica, vol. 46, no. 10, pp. 1482–1491, 2017.
[17] Z. X. Zhang, J. F. Zhu, and X. Y. Hu, "Measurable orthoimage elevation synchronous model and its applications in mapping," Acta Geodaetica et Cartographica Sinica, vol. 43, no. 1, pp. 5–12, 2014.
[18] X. B. Yang, "Digital surface model generation for high resolution satellite stereo image based on modified semi-global matching," Acta Geodaetica et Cartographica Sinica, vol. 47, no. 10, pp. 1372–1384, 2018.
[19] Y. J. Zhang, "A fast aerial image matching method using airborne LiDAR point cloud and POS data," Journal of Geodesy and Geoinformation Science, vol. 2, no. 1, pp. 26–36, 2019.
[20] L. Jian Sun, H. Y. Nan-Ning Zheng, and N. N. Heung-Yeung Shum, "Stereo matching using belief propagation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, no. 7, pp. 787–800, Copenhagen, May, 2003.
[21] H. Hirschmüller, "Stereo processing by semiglobal matching and mutual information," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, no. 2, pp. 328–41, 2008.
[22] K. J. Suk-Jin Yoon and I. S. In So Kweon, "Adaptive support-weight approach for correspondence search," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 4, pp. 650–656, 2006.
[23] Z. B. Yang, "Digital surface model generation for high resolution satellite stereo image based on modified semi-global matching," Acta Geodaetica et Cartographica Sinica, vol. 47, no. 10, pp. 1372–1384, 2018.
[24] Y. J. Zhang, "A fast aerial image matching method using airborne LiDAR point cloud and POS data," Journal of Geodesy and Geoinformation Science, vol. 2, no. 1, pp. 26–36, 2019.
[25] J. Jian Sun, H. Y. Nan-Ning Zheng, and N. N. Heung-Yeung Shum, "Stereo matching using belief propagation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, no. 7, pp. 787–800, Copenhagen, May, 2003.
[26] H. Hirschmüller, "Stereo processing by semiglobal matching and mutual information," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, no. 2, pp. 328–41, 2008.
[27] K. J. Suk-Jin Yoon and I. S. In So Kweon, "Adaptive support-weight approach for correspondence search," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 4, pp. 650–656, 2006.
[24] J. C. Liu, "Dense stereo matching strategy for oblique images that considers the plane directions in urban areas," IEEE Transactions on Geoscience and Remote Sensing, vol. 58, no. 7, pp. 1–8, 2020.
[25] A. G. Abiel, A. E. Miguel, and F. Berry, “Depth from motion algorithm and hardware architecture for smart cameras,” Sensors, vol. 19, no. 1, p. 53, 2019.
[26] Y. Lin, Y. X. Gao, and Y. Wang, “An improved sum of squared difference algorithm for automated distance measurement,” Frontiers in Physics, vol. 9, 2021.
[27] B. Cui and J.-C. Créput, "NCC based correspondence problem for first- and second-order graph matching," Sensors, vol. 20, no. 18, p. 5117, 2020.
[28] R. Zabih and J. Woodfill, “Non-parametric local transforms for computing visual correspondence,” in Proceedings of the European Conference on Computer Vision, pp. 151–158, Sweden, May 1994.
[29] X. Mei, S. Xun, Z. Mingcai, J. Shaohui, W. Haitao, and Z. Xiaopeng, "On building an accurate stereo matching system on graphics hardware," in Proceedings of the IEEE International Conference on Computer Vision Workshops, pp. 6–13, Barcelona, November 2012.
[30] D. B. Qiu, Research on Key Technology of Multi-View Oblique Aerial Image Matching, PLA Information Engineering University, Zhengzhou, 2017.
[31] S. Birchfield and C. Tomasi, “Depth discontinuities by pixel-to-pixel stereo,” International Journal of Computer Vision, vol. 35, no. 3, pp. 269–293, 1999.
[32] S. K. Gehrig, F. Eberli, and T. Meyer, "A real-time low-power stereo vision engine using semi-global matching,” Lecture Notes in Computer Science, in Proceedings of the 7th International Conference on Computer Vision Systems, vol. 13, pp. 134–143, Berlin, October 2009.
[33] M. Rothermel, W. Konrad, F. Dieter, and H. Norbert, SURE: Photogrammetric Surface Reconstruction from Imagery, LC3D Workshop, Berlin, 2012.
[34] B. Cyganek and J. P. Siebert, An Introduction to 3D Computer Vision Techniques and Algorithms, John Wiley & Sons, NJ, USA, 2009.
[35] T. Hu, B. Qi, T. Wu, X. Xu, and H. He, "Stereo matching using weighted dynamic programming on a single-direction four-connected tree,” Computer Vision and Image Understanding, vol. 116, no. 8, pp. 908–921, 2012.
[36] B. Cyganek, “Comparison of nonparametric transformations and bit vector matching for stereo correlation,” Lecture Notes in Computer Science, vol. 3322, pp. 534–547, 2004.
[37] D. Scharstein and R. Szeliski, "Middlebury stereo vision,” 2003, https://vision.middlebury.edu/stereo.