A least square matching optimization method of low altitude remote sensing images based on self-adaptive patch

Nan Yang1,2,3, Yaping Zhang1,* & Jialin Li1
1Harbin Institute of Technology, School of Transportation Science and Engineering, 73 Huanghe Road, Harbin, China, 150090;
2Shenzhen Research and Development Center of State Key Laboratory for Information Engineering in Surveying, Mapping and Remote Sensing, Keyuan Road, Shenzhen China, 518057.
3Key laboratory of Urban Land Resource Monitoring and Simulation Ministry of Land and Resource of China, 8007 Hongli Road, Futian District Shenzhen China, 518034;
Email: candyyanghot@126.com

Abstract: This paper presents a novel matching optimization method based on self-adaptive template. The proposed method is designed to be effective for enhancing the accuracy of stereo matching. In order to improve the similarity of the initial matching windows and fully exploit the pixels around the corresponding image points, a self-adaptive patch is introduced instead of a constant patch. Then, an error equation is built to compute the optimal point according to space geometry relationship and epipolar line constraint. At last, a least square adjustment method is used to calculate the coordinate of the corresponding 3D point in the Object Space Coordinate System. Comparison studies and experimental results prove the high accuracy of the proposed algorithm in low-altitude remote sensing image point cloud optimization.

1. Introduction
It is one of the simplest and fastest ways to reconstruct a fine-detailed three-dimensional (3D) object or scene through a large number of images shot from low altitude remote sensing platform[1-3], as well as it can save the information about color and texture in the images, which is superior to the 3D laser scanning[4]. In the process of 3D reconstruction based on images, one of the most significant procedure is image matching[5]. Since the accuracy of image matching can directly ordain the quality of the reconstructed 3D model, a huge number of researchers have spent several decades to improve the accuracy and efficiency of the image matching, like point matching based on the intensity of the pixel[6], point or line matching based on the distinct feature in the images[7-8], as well as dense matching[9] and multi-view stereo matching[10]. However, although a great deal of image matching refinement algorithms had greatly enhanced the precision of image matching, some kinds of problems like noise, deep discontinuities, occlusions and so forth are still encountered in most matching refinement approaches[11].

Nowadays, a patch-based matching optimization method is popularly used to enhance the accuracy of single pair corresponding pixels matching[12]. The main idea of this method is: 1) calculating object point through an initial matching (feature matching); 2) utilizing this object point as center to build a patch in the object space coordinate system; 3) projecting the object points on the patch to reference image and search image to get matching windows in the two images; 4) constructing error equation
based on projective geometry relationship and using least square adjustment method to calculate the corresponding pixels and the final object point. It is can be demonstrated that the patch in this method which is built in the object space actually is local tangent patch in geometry relationship. However, as generally common sense, the local tangent patch only exists on the continuous area of the surface, so that the patch-based optimization methods would result in error matching or blurring of the edge lines on the discontinuities of surface like buildings in the urban images. Moreover, the initial elements which is introduced in the least square adjustment can directly determine the final results. Only if the initial elements are close to the final results the process of image matching can be robust and the number of iteration can be controlled in an acceptable range.

To improve the matching accuracy of points, a self-adaptive patch is employed in this study. The self-adaptive patch is built according to the Normalized Cross Correlation (NCC) of the matching window around the initial corresponding pixels. By employing the self-adaptive patch, more effective feature around the initial corresponding pixels can be taken part in the process of matching to enhance the matching accuracy. Low-altitude remote sensing image data at Northwestern University Campus and the YangJiang area are used to verify the feasibility of this method. The experiments show that the proposed algorithm can effectively improve the robust of the image matching, and achieve more accurate 3D point cloud data.

The remainder of the paper is structured as follows: the proposed method is introduced in Section 2; in Section 3 experiments are conducted to verify the feasibility of the algorithm in terms of reliability and matching accuracy; finally, the conclusions are stated in Section 4.

2. Method
In this section, the point cloud optimization method based on self-adaptive patch is developed. The proposed method aims at raising stability of image matching and improving the matching accuracy. The proposed algorithm utilizes four steps to optimize the object points in point cloud data: 1) Calculation of image points in two images; 2) Setting self-adaptive patch in object space; 3) Construction of error equation; 4) Calculation of optimized object point by Least Square Adjustment. By introducing self-adaptive patch in the matching refinement process, the robust and accuracy of image matching is improved in the proposed method.

2.1. Calculation of image points in two images
According to the geometry relationship between object points and the image points, the optical center of the camera, the object point on the ground and its corresponding image point must locate on a straight ray in the object space. In accordance with the collinear relationship and the epipolar line constraint, the initial corresponding image points \((x_0, y_0)\) and \((x_i, y_i)\) in the reference image and search image respectively can be calculated as:

\[
\begin{align*}
    x_{(0)} &= -\frac{a_{(0)}}{a_{(0)}}(X - X_{(0)}) + b_{(0)}(Y - Y_{(0)}) + c_{(0)}(Z - Z_{(0)}) \\
    y_{(0)} &= -\frac{b_{(0)}}{a_{(0)}}(X - X_{(0)}) + c_{(0)}(Y - Y_{(0)}) + b_{(0)}(Z - Z_{(0)})
\end{align*}
\]

In the collinear equation above, \((X, Y, Z)\) is the coordinate of object point in the object space coordinate system, \((Xs0, Ys0, Zs0)\) and \((Xsi, Ysi, Zsi)\) are the coordinates of the two images’ optical center in the object space coordinate system, \((Xsi, Ysi, Zsi)\) are rotation matrix of reference image and search image \(i\).
2.2. Setting patch in object space

After the pair of image points are calculated, a self-adaptive patch is determined by the Normalization Cross Correlation (NCC) between the two image windows in this part. The proposed method utilize two sub steps to determine the patch.

At the first sub step, the initial corresponding image points which were calculated in 2.1 are used as main points. Then 9 pairs 7×7 image windows are built on the reference and search images surround main points. As illustrated in the Figure 1, traditional intensity-based image matching method employed initial image points in the reference image as center and construct a fixed image window as template, and then used NCC between the template and the image window to find the corresponding image point in the search image. The NCC is closer to 1, the template and the image window in the search image are more similar, and the center of the template and the image window are much more likely to be corresponding points. However, due to the intricate texture of the urban images, using initial image point as center to build template cannot achieve best initial matching. So that, in the first sub step, the proposed method construct 9 pairs of image window which main points are located at the center, northwest, north, northeast, west, east, southwest, south and southeast respectively. And then 9 values about the NCC of the image window pairs are calculated. The proposed method employs the image window pair which is maximum as the initial matching window. As example in Figure 2, the image window pair in sub figure (e) gets the largest NCC in the test, so the initial matching window is determined as the main point located at the southwest of the template.

![Figure 1. Initial corresponding image points and traditional matching window](image1)

![Figure 2. Calculation the NCC according to the location of the main points](image2)

As the initial location of the main image point is determined, the size of the patch have to be fixed at the second sub step. At this step, the initial matching window is expanded by a certain direction and a certain step (one pixel in this paper) based on the main point. As illustrated in the Figure 3, the location of the main point is on the southwest of the image window, so the direction of the expansion is north.
and east. In sub figure (a), a pair of 8×8 image windows are built and the NCC of them is 0.906 which is larger than 0.896, so the expansion process continues, so as 9×9 and 10×10. When the size of image window is 11×11 in sub figure (d), the NCC between the image window pair is 0.862 which is smaller than 0.911, the expansion ends and the size of the patch is determined as 10×10 points.

2.3. Construction of error equation

The main idea of matching refinement algorithm is to find a pair of image windows in the reference image and search image respectively. When the differences between the each corresponding pair of pixels in two image windows reach the minimum, the two main pixels in the corresponding image windows could be the best matching pixels. According to the LSM method, the error equation is:

\[ v = dh_{0i} + g_i(x_0, y_0) \cdot dh_{0i} + h_{0i} \left( \frac{\partial g_i}{\partial x_i} \cdot dx_i + \frac{\partial g_i}{\partial y_i} \cdot dy_i \right) - \left( g_0(x_0, y_0) - h_{0i} \cdot g_i(x_0, y_0) \right) \]  

(2)

In the error equation above, \( v \) is the projection error, \( h_{0i} \) and \( h_{1i} \) are the radiation distortion coefficients between the reference image and search image \( i \). \( dh_{0i} \) and \( dh_{1i} \) are corrections of parameter \( h_{0i} \) and \( h_{1i} \). \( g_0(x_0, y_0) \) and \( g_i(x_i, y_i) \) are intensity value of corresponding pixels \( (x_0, y_0) \) and \( (x_i, y_i) \) which are located in the image window of the reference image and search image \( i \) respectively. \( \frac{\partial g_i}{\partial x_i}, \frac{\partial g_i}{\partial y_i} \) is the derivative value of pixel intensity in the x and y direction. \( (dx_{01}, dy_{01}) \) is the correction values of the image points \( (x_0, y_0) \).

Because the interval of the adjacent pixels is extremely small, the difference between neighboring pixels is employed to calculate \( \frac{\partial g_i(x_i, y_i)}{\partial x_i} \) and \( \frac{\partial g_i(x_i, y_i)}{\partial y_i} \) in the error equation (1) instead of differential. The derivative value of pixel intensity in the x and y direction can be approximately expressed as:

\[ \frac{\partial g_i}{\partial x_i} = \frac{1}{2} \left( g_i(x_i + 1, y_i) - g_i(x_i - 1, y_i) \right) \]  

(3)

\[ \frac{\partial g_i}{\partial y_i} = \frac{1}{2} \left( g_i(x_i, y_i + 1) - g_i(x_i, y_i - 1) \right) \]

By introducing formula (3) to error equation (2), the error equation can be simplified a linear equation which use corrections as unknown numbers. Then a Least Square Adjustment can be used to solve the final optimized matching result.

2.4. Calculation of optimized object point by Least Square Adjustment

Since a pair of corresponding pixels can build only one error equation and it is impossible to seek solution by only one error equation, a pair of image matching windows which contain several pixels surrounding the pair of corresponding pixels is essential. Moreover, in 2.2, the proposed method construct a most similar initial matching window around the main initial corresponding image points, so the pixels which are located at the same position of the reference image and search image windows respectively are theoretical corresponding pixels. By introducing these corresponding pixels in equation
(2), more than one error equation therefore can be built, and if the number of error equations is larger than the number of unknown quantities, least square adjustment can be employed to get more accuracy result through these redundant pixels.

3. Experiments and results

In this section, two sets of point cloud data are used in the experiment. The data is generated from a pair of Unmanned Aerial Vehicle (UAV) images captured in Northwestern University Campus (Xi’an, Shanxi province) and a pair of Aerial Image which are captured at nadir in Yangjiang (Guangdong province). Detailed information of the images and initial point cloud are illustrated in Table.1, Table.2, Figure.4 and Figure.5.

Table.1. Parameters of the photography in Northwest University (46337 points)

| Camera Name | CCD Size (mm) | Image Size (pixel) | Pixel size (μm) | Focal Length (mm) | Flying Height (m) | Ground Resolution (m) |
|-------------|---------------|--------------------|-----------------|-------------------|-------------------|-----------------------|
| Canon EOS 400D | 22.16×14.77 | 3888×2592 | 5.7 | 24 | 600 | 0.118 |

Table.2. Parameters of the photography in Yangjiang (165352 points)

| Camera Name | CCD Size (mm) | Image Size (pixel) | Pixel size (μm) | Focal Length (mm) | Flying Height (m) | Ground Resolution (m) |
|-------------|---------------|--------------------|-----------------|-------------------|-------------------|-----------------------|
| SWDC-5 | 49.24×36.47 | 8206×6078 | 6 | 82 | 800 | 0.058 |

Figure 4. Initial point cloud of Northwestern University Campus

Figure 5. Initial point cloud of Yangjiang

3.1. Parameters setting

In the proposed method, the camera parameters and the pose of the images are known. To solve error equations, an iteration is used that is set to stop when the parameter corrections dh0i and dh1i are less than 10-5 other parameter corrections are all less than 10-3 or the iteration is more than 300 times. The matching results with patch size of 7×7 and self-adaptive size are compared. If the parameter corrections can be convergent and the iteration times is acceptable, the corresponding image pixels are matched successfully.
In the process of individual point matching, the Normalized Correlation Coefficient can reflect the degree of approximation between image windows. Theoretically, the NCC is closer to 1, the correlation between image windows is higher, and the main pixels of the two image windows are more probable to be a pair of corresponding points. Therefore, when the NCC value between the image matching windows is larger than a fixed threshold, the main pixels can be considered as valid matching, and their corresponding object point on the ground can be added in the point cloud. In the experiments, an NCC value 0.6 is set as threshold.

Moreover, for stereo matching with low altitude images by photography, the error of altitude is the largest, hence the altitude difference is always considered as the criteria to evaluate the accuracy. The experiments employed altitude difference between the difference between the Z-coordinates calculated after the end of the iteration and the Z-coordinates calculated by the forward intersection.

3.2. Result and analysis

The comparison performances between patch-based LSM and the proposed method are illustrated in Fig.6 and Fig.7. The red bar charts (sub figure (a)) in the two figures compared the percentage of successful matching in the two data sets. The bar chart indicate that the proposed self-adaptive patch performs higher rate of successful matching (61.89% in Northwestern University images and 65.76% in Yangjiang images) than traditional constant patch (55.93% and 59.30% respectively), since more pixels information are involved in the matching process, and this can decrease the weight of the noise and enhance the robust of image matching. It can also explain the reason for which the proposed method has higher rate in valued matching as comparison performance illustrated in green bar chart (sub figure (b)) in the two figures (37.31% in Northwestern University images and 50.43% in Yangjiang images compares with 35.23% and 46.62% in two image sets respectively). And the NCC of initial matching windows in self-adaptive patch is higher than constant patch is another reason lead to the superior result in the proposed method.

By comparing the average elevation difference in matching tens thousands of points between constant patch and self-adaptive patch in each blue bar chart (sub figure (c)) of the two figures, the results can be illustrated that the error of the proposed method (0.917 in Northwestern University images and 0.461 in Yangjiang images) is lower than traditional constant method (0.977 and 0.531 respectively), so that the proposed method which builds initial matching window according to the feature and the intensity of the pixels surrounding the initial corresponding image points can improve the accuracy of stereo matching. In comparison between the two data sets, the matching accuracy of Yangjiang images which is less than 0.5 is superior to Northwestern University images due to the higher resolution of the camera. Moreover, the Yangjiang images produced more effect in accuracy improvement, it probably because the texture in this image sets are more complex than the Northwestern University images.

Figure 6. Comparison performance of Northwestern University Campus
4. Conclusion
In this paper, we presented a novel method for improving the accuracy of point cloud reconstructed by low altitude remote sensing images. The proposed method introduced a self-adaptive patch instead of a constant patch to build error equation, which can improve the initial NCC of the matching windows and introduce the better initial elements value to the Least Square Adjustment in order to raise the stability and the accuracy of stereo matching. Meanwhile, the matching results and the accuracy of the algorithm are validated experimentally.

The comparison of successful matching and altitude difference has demonstrated that the proposed approach is superior to the patch-based LSM method especially in images which have complex texture. A possible future work is matching pixels according to deep learning since deep learning can detect more effective information in the images which can be utilized to improve the accuracy of image matching.

Acknowledgments
The project was supported by the Open Fund of Key Laboratory of Urban Land Resources Monitoring and Simulation, Ministry of Land and Resources (KF-2018-03-026).

References
[1] Kelly, T., Femiani, J., Wonka, P., & Mitra, N. J. (2017). Bigsur: large-scale structured urban reconstruction. ACM Transactions on Graphics, 36(6), 1-16.
[2] Jiang, S., & Jiang, W. (2017). Efficient structure from motion for oblique uav images based on maximal spanning tree expansions. Isprs Journal of Photogrammetry & Remote Sensing.
[3] Chen F, Yu H, Ha Y. Quality Estimation and Optimization of Adaptive Stereo Matching Algorithms for Smart Vehicles[J]. ACM Transactions on Embedded Computing Systems (TECS), 2020.
[4] Zhang J, Ramanagopalg M S, Vasudevan R, et al. LiStereo: Generate Dense Depth Maps from LIDAR and Stereo Imagery[J]. 2019.
[5] Knappitsch, A., Park, J., Zhou, Q.Y., & Koltun,V.(2017). Tanks and temples: benchmarking large-scale scene reconstruction. ACM Transactions on Graphics, 36(4).
[6] Ackermann, F. (1984). Digital image correlation: performance and potential application in photogrammetry. Photogrammetric Record, 11.
[7] Lowe D G. 2004. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 60(2): 91-110.
[8] Von Gioi R G, Jakubowicz J, Morel J M, et al. 2010. LSD: A Fast Line Segment Detector with a False Detection Control[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 32(4): 722-732.
[9] Hirschmuller H. Stereo processing by semiglobal matching and mutual information [J]. IEEE Transactions on pattern analysis and machine intelligence, 2008, 30(2): 328-341.
[10] Rothermel M, Wenzel K, Fritsch D, et al. SURE: Photogrammetric surface reconstruction from imagery [C] //Proceedings LC3D Workshop, Berlin. 2012, 8.
[11] Furukawa, Y.. (2010). Accurate, dense, and robust multiview stereopsis. IEEE Transactions on Pattern Analysis and Machine Intelligence, 32.
[12] Ozyesil, O., Voroninski, V., Basri, R., & Singer, A.. (2017). A survey of structure from motion. Acta Numerica, 26.