A Novel Method for Multi-image Matching Synthesizing Image and Object-space Information

YUAN Xiuxiao  MING Yang

Abstract  A novel method for multi-image matching by synthesizing image and object-space information is proposed. Firstly, four levels of image pyramids are generated according to the rule that the next pyramid level is generated from the previous level using the average gray values of the 3 by 3 pixels, and the first level of pyramid image is generated from the original image. The initial horizontal parallaxes between the reference image and each searching image are calculated at the highest level of the image pyramid. Secondly, corresponding image points are searched in each stereo image pair from the third level of image pyramid, and the matching results in all stereo pairs are integrated in the object space, by which the mismatched image points can be eliminated and more accurate spatial information can be obtained for the subsequent pyramid image matching. The matching method based on correlation coefficient with geometric constraints and global relaxation matching is introduced in the process of image matching. Finally, the feasibility of the method proposed in this paper is verified by the experiments using a set of digital frame aerial images with big overlap. Compared with the traditional image matching method with two images, the accuracy of the digital surface model (DSM) generated using the proposed method shows that the multi-image matching method can eliminate the mismatched points effectively and can improve the matching success rate significantly.

Keywords multi-image matching; digital surface model (DSM); cross correlation matching with geometric constraints; relaxation matching; matching success rate

CLC number  P231.5

Introduction

Digital surface model (DSM) has wide application in the domain of urban planning, virtual reality, transportation, and communication. It is also one of the key research contents in the field of photogrammetry. Automatically generating DSM from stereo image pairs entails mainly the two core techniques of image matching and surface reconstruction, and the key lies in how to obtain dense, reliable conjugate points. With the wide application of digital aerial cameras, the availability of large overlap digital frame aerial images is getting easier and easier, and will become a common operational pattern in the future of aerial surveying photography. As one object target may appear in multiple photography, this can provide large amount of redundant information for

Received on March 15, 2009.

Supported by the National Natural Science Foundation of China (Nos. 40771176, 40721001).

YUAN Xiuxiao, School of Remote Sensing and Information Engineering, Wuhan University, 129 Luoyu Road, Wuhan 430079, China; State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, 129 Luoyu Road, Wuhan 430079, China.

E-mail: yxxphiyw@public.wh.hb.cn
photogrammetric data processing. The emphasis in this paper is on how to make full use of the redundant information to improve the success rate and reliability of image matching and to provide dense 3D point clouds for automatic DSM generation.

It is well known that the multi-image matching approach leads to a reduction of problems caused by occlusions, repetitive texture, and so on. Gabet et al. presented a highly redundant correlation process for automatic generation of high-resolution urban DSM\[1\]. Firstly, the image acquisition specification as image sequences leads to stereo image pairs with various base-height ratios, then various stereovision methods are used and the disparity maps are merged, thus attributing to each pixel the most probable and accurate elevation. However, its computation is expensive. Okutomi and Kanade presented the concept of similarity measure for multiple images\[2\], which is defined with respect to the height to relate the reference image with all other search images simultaneously. According to the same idea, a similarity measure of average normalized cross correlation\[3\] and that of sum of normalized cross-correlation (SNCC)\[4\] for the matching to generate high-resolution DSM were adopted, respectively. For the above methods, the computation of such similarity measure is very complicated if the baseline between the images varies a lot. And this will result in a larger projective error if the image exterior orientation elements are inaccurate, even if image points in different images of the same object are not necessarily like the identified. To this end, Zinich and Webb raised an image-matching scheme that the candidates derived from image matching are projected directly to the object space and the correct candidates are determined by relaxation optimization \[5\]. To some extent, this method will make up for some deficiencies, but to find the right combination needs complex analysis.

In view of this, this paper studies a multi-image matching method comprehensively utilizing image and object-space information for automatic DSM generation. The basic philosophy is that the central image is taken as the reference image and the rest are taken as the search images among the overlapping

![Flow chart of DSM generation based on multiple-image matching](image-url)
image sequences which separately constitute stereo image pairs with the reference image. Firstly, a coarse matching is carried out on various stereo image pairs using a geometrically constrained cross-correlation matching algorithm. Secondly, the corresponding points between reference image and search images are determined through relaxation matching. Finally, the matching results of each stereo image pair are blended in the object space by a multi-ray forward intersection procedure with blunder detection function, and the accurate elevation information is obtained. The above various steps are integrated with coarse to fine strategy, and the obtained information from the upper pyramid level matching is used as guidance for the matching in the next level of pyramid image. Fig. 1 is the flow chart of matching algorithm of three images to generate DSM.

1 Key technologies

1.1 Selection of reference image

The reference image may be the central image, perhaps the one with the best quality, or may also be each one in turn. The central image is generally chosen. When the occlusion situation is quite serious, it should be better to select each image as the reference image; the information provided by each image may be used, but the computation amount is very big.

1.2 Generation of image pyramids

Image pyramid generation may use the methods of averaging pixels, wavelet transformation, Laplace law, and so on. The method of averaging pixels is widely used in practice due to the merits of usefulness and simplicity. This paper uses the 3 by 3 pixels average method to produce four levels of image pyramids. First of all, calculate the average grey value of each 3 by 3 cell of original image, and assign it to the corresponding pixel of the first level of the image pyramid, generating the first level of image pyramid, and so on, until the generation of the fourth level of image pyramid.

1.3 Determination of initial parallaxes

The determination of the initial parallaxes between the reference image and the search images is carried out at the fourth level of image pyramid, and the image exterior orientation elements and the average elevation $Z$ of the region of image coverage are used. Firstly, a feature point $(x_i, y_i)$ near image center is projected into the average elevation plane to obtain its object-space coordinate $(X, Y, Z)$, then the object point is back-projected into the search images by the collinearity equation to get the image coordinate $(x_2, y_2)$, with the coordinate difference $(\Delta x = x_2 - x_1, \Delta y = y_2 - y_1)$ calculated to obtain the approximate offset between images. When topographic relief is larger, it may be asked to execute the template matching using the maximum correlation coefficient method to obtain a more accurate initial parallax. From the experiments it is discovered that the initial parallax accuracy is about 100-200 pixels according to the method mentioned above.

1.4 Cross-correlation matching with geometric constraints

The interior and exterior orientation elements of each image are assumed to be known. Each pyramid level of reference image is divided into a regular grid mesh with cell size of 10 by 10 pixels. Firstly, one feature point in each cell is extracted by Förstner operator, and then a geometrically constrained cross-correlation matching algorithm is adopted to seek the conjugate candidate points in each search image. From the intersection of conjugate points at a higher pyramid level, the approximate object coordinate $(X, Y, Z_0)$ is obtained, the certain elevation error $\Delta Z$ is calculated from the elevation of neighbor points, then the correct position of point $p$ should lie between $p_{\text{min}}$ and $p_{\text{max}}$ relative to the height values of $(Z_0 - \Delta Z)$ and $(Z_0 + \Delta Z)$, respectively. If the points between $p_{\text{min}}$ and $p_{\text{max}}$ are back-projected onto the search images, the corresponding segment of the epipolar line is easily defined and the correct matched point should lie along it. Due to the errors of the image orientation parameters, image matching is usually searching along the epipolar segment with bandwidth of one to two pixels. Taking the normalized cross-correlation coefficient as the similarity measure (the threshold is 0.5), the match candidates are derived.
1.5 Relaxation matching

Relaxation matching is the method that uses the neighborhood context information, and considers the compatibility between objects, and enables the conjugate points to obtain more support from the neighborhood. During iterative procedure, the probability of correct candidate point increases, while that of wrong ones reduces unceasingly. When the iteration terminates, the correct match point is obtained. For the feature point \( I_i \) in the reference image, the match candidates are determined by cross-correlation matching with geometric constraints. Given \( m \) candidates \( I_j \) (\( j = 1,2,\cdots,m \)) in the search images, the corresponding correlation coefficients are \( \rho_j \) (\( j = 1,2,\cdots,m \)), then the match probability of \( I_i \leftarrow I_j \) is

\[
P(i,j) = \frac{\rho_j}{\sum_{k=1}^{m} \rho_k} \tag{1}
\]

After finding match candidates of each feature point \( I_i \), the relaxation optimization is performed in its eight neighborhoods. In order to link the matching results of the neighboring feature points to each other, a compatible coefficient \( C(i;j;k,l) \), which quantifies the compatibility between the match \( I_i \leftarrow I_j \) and a neighboring match \( I_k \leftarrow I_l \), is defined as:

\[
C(i;j;k,l) = \frac{T}{\exp((\Delta p_x^2 + \Delta p_y^2) / \beta)} \tag{2}
\]

where \( \Delta p_x = (x_i - x_j) - (x_k - x_l) \) expresses the difference of the \( x \)-parallaxes at the point \( I_i \) and its neighboring point \( I_k \), and \( \Delta p_y = (y_i - y_j) - (y_k - y_l) \) is the difference of the \( y \)-parallaxes. The bigger \( \Delta p \) is, the smaller the compatibility is. This corresponds to a smoothness constraint on the image matching results, and it provides an ability to bridge over the poor texture areas by assuming that parallax surface varies smoothly over the image. \( T \) and \( \beta \) are constant and set to 1 and 25 in the experiment, respectively.

In the relaxation scheme, the global consistency of image matching can be achieved by an iterative scheme where the probabilities \( P(i,j) \) are updated by the following rule:

\[
P^{(n+1)}(i,j) = \frac{P^{(n)}(i,j)Q^{(n)}(i,j)}{\sum_{s=1}^{m} P^{(n)}(i,s)Q^{(n)}(i,s)} \tag{3}
\]

where \( Q^{(n)}(i,j) = P^{(n)}(i,j)(c_0 + c_1 \sum_{l=1}^{m} P^{(n)}(k,l)C(i,j;k,l)) \). Here \( c_0, c_1 \) are the relaxation coefficients. \( \Omega(I_i) \) is the neighborhood of the point \( I_i \).

For the point \( I_i \), if one of the match candidate probabilities \( P(i,j) \) exceeds 0.9, or a pre-defined iterative number has been reached, the iteration will be terminated. The match point that gains the highest probability \( P(i,j) \) is selected as the conjugate point.

1.6 Fusion of matching results in object space

After performing the match on stereo image pairs constituted by reference image and each search image according to the above method, the feature point could successfully match on certain stereo image pairs, but fail on other stereo image pairs, or the match result would not satisfy the geometric constrains in the object space. Hence, the multi-ray forward intersection method is applied to realize the fusion of matching results of various stereo image pairs in object space.

When the interior and exterior orientation elements are obtained and the object coordinate of feature point is unknown, the error equation derived from the collinearity equation is written as:

\[
\begin{align*}
\nu_x &= -a_{11}\Delta X - a_{12}\Delta Y - a_{13}\Delta Z - (x - x^0) \\
\nu_y &= -a_{21}\Delta X - a_{22}\Delta Y - a_{23}\Delta Z - (y - y^0)
\end{align*} \tag{4}
\]

The meaning of each symbol in Eq. (4) can be referred to literature [9].

When all corresponding points in \( n \) search images are successfully matched, \( 2n+2 \) error equations as Eq. (4) can be formed. Because the number of unknowns is only three, it can be solved according to least squares adjustment. The varying weight iteration blunder detection method, which is based on posterior variance component estimation[10], is employed. If there are some false match points, they should be eliminated during the calculation of object space coordinates. When the iteration converges, if the root mean squares error of image coordinates is lower than the predefined threshold (one pixel size of the current level of image pyramid), it is retained or deleted. This strongly guarantees the accuracy and reliability of spatial information, which will be used to direct the match of various stereo image pairs at the lower level of image pyramid.
2 Experiments and analysis

2.1 Experimental test design

In order to validate the effect of the multi-image match strategy mentioned above, five pieces of overlapping images within one strip from a set of grey aerial digital frame images are selected to carry into image matching and to automatically generate a DSM. The pixel size is 9 µm, corresponding to the ground sampled distance of 0.1 m, and the forward overlap is about 80%. To obtain all image orientation parameters and the object space coordinates of 2,526 photogrammetric points, GPS-supported bundle block adjustment with four full ground control points on the corner of the adjusted block is carried out on the whole set of 176 images using the POS-supported bundle adjustment software WuCAPS[11]. From the statistical root mean square error of the coordinates of 45 ground check points, the planimetric and vertical accuracy of photogrammetric points are ±7.1 cm and ±4.8 cm on the ground, respectively.

Fig. 2 shows that the terrain covered by the test images is smooth, but includes a lot of man-made buildings, a large piece of area of repetitive texture (shown as block a), and poor texture (shown as block b). According to the match strategy in this paper, image matching is performed, the large number of discrete object points is obtained from TIN, and a regular DSM is generated by interpolation, as shown in Fig. 3.

![Fig.2 Overview of the experimental images](image)

![Fig.3 DSM generated by multiple-image matching](image)

From Fig. 3, it can easily be found that the resulting DSM can reproduce the general shape of the terrain relief. For the area of repetitive texture (block a), the disturbance caused multi-peak values is efficiently removed; for that of poor texture (block b), a satisfactory result is obtained; and the man-made buildings are depicted well.

| Image combinations | Matching success rate (%) | MAE/m | RMSE/m |
|--------------------|---------------------------|-------|--------|
| 33/34              | 68.9                      | 0.224 | 0.571  |
| 33/34/35           | 76.7                      | 0.183 | 0.366  |
| 33/34/35/36        | 77.3                      | 0.170 | 0.355  |
| 32/33/34/35/36     | 79.2                      | 0.153 | 0.326  |

Note: MAE is the mean absolute error, and RMSE is the root mean square error.

2.2 Comparison test of different image combinations

In order to quantitatively evaluate the accuracy of the resulting DSM, here the photogrammetric coordinates of 176 object points in the test area are used as the reference values, and the height differences of these points and that interpolated from DSM are statistically analyzed. For the five images in Fig. 2,
image 34 is selected as the reference image in the experiment, and different stereo image pairs are combined with other images. Table 1 compares DSM generated by matching with the results of different stereo image pairs. Table 2 shows the height differences of partial pass points.

| ID   | PE      | IE | error | IE   | error | IE   | error | IE   | error | IE   | error |
|------|---------|----|--------|------|--------|------|--------|------|--------|------|--------|
| 5021000 | 504.621 | 510.080 | -5.459 | 504.546 | 0.074 | 504.546 | 0.074 | 504.556 | 0.065 |
| 5010615 | 503.762 | 503.331 | 0.431 | 503.334 | 0.428 | 503.334 | 0.428 | 503.772 | 0.010 |
| 5041681 | 506.013 | 505.748 | 0.265 | 505.939 | 0.074 | 505.972 | 0.041 | 505.979 | 0.034 |
| 5051819 | 503.286 | 503.466 | 0.180 | 503.466 | 0.180 | 503.257 | 0.029 | 503.259 | 0.027 |
| 5051737 | 507.770 | 507.560 | 0.210 | 507.560 | 0.210 | 507.652 | 0.118 | 507.652 | 0.118 |
| 4061362 | 506.636 | 506.374 | 0.262 | 506.493 | 0.143 | 506.560 | 0.076 | 506.560 | 0.076 |
| 5031463 | 513.997 | 512.040 | 1.958 | 514.209 | 0.212 | 514.209 | 0.212 | 514.152 | 0.155 |
| 5031464 | 514.078 | 514.328 | -0.251 | 514.225 | 0.147 | 514.225 | 0.147 | 514.137 | 0.059 |
| 6031331 | 505.426 | 505.767 | -0.340 | 505.285 | 0.141 | 505.285 | 0.141 | 505.285 | 0.141 |
| 4030965 | 502.211 | 502.566 | -0.356 | 502.234 | 0.023 | 502.234 | 0.023 | 502.229 | 0.018 |
| 5061595 | 511.021 | 510.535 | 0.486 | 510.842 | 0.179 | 510.815 | 0.206 | 510.815 | 0.206 |
| 5041681 | 506.013 | 505.748 | 0.265 | 505.939 | 0.074 | 505.972 | 0.041 | 505.979 | 0.034 |
| 4041203 | 503.855 | 504.133 | -0.279 | 503.904 | 0.049 | 503.824 | 0.031 | 503.824 | 0.031 |

From the result of point 502100 in Table 2, it is found that the elevation obtained from the stereo pair of images 33 and 34 has an evident error of -5.459 m. Through multi-ray forward intersection with blunder detection function, it decreases to 0.065 m for the result of five images. Similarly, for point 503146, the elevation error 1.958 m for the result of two images decreases to -0.155 m for that of five images.

3) Compared with traditional two-image matching, the accuracy of DSM generated from the matching result of three-image matching shows a significant improvement. The reason is, as stated in 1), integrating information in image and object-space can improve the matching success rate. And multi-ray forward intersection can improve the elevation accuracy due to enlargement of intersection angle, which will increase the intensity of points and enhance the quality of single point, and finally reduce the interpolation error during the DSM generation and refine the DSM. However, when more images participate in image matching, the wrong match points are well detected and eliminated, the increase of matching success rate has a slower pace, and the role of the multi-ray intersection to improve elevation accuracy is limited, despite the refinement of the DSM accuracy. Therefore, it is not more obvious.

As only a limited number of ground check points...
are used to evaluate the resulting DSMs, the differences among them may be not completely expressed. Fig.4 shows three enlarged views for some local regions of automatically generated DSM by two-image matching and five-image matching.

Fig. 4 shows that DSM generated by two-images still has much granular noise in open and poor texture areas, and the distortion on building edge is very serious. This explains why there exists a large number of false match points. However, when using the method proposed in this paper, the DSM generated from five images looks much cleaner.

2.3 Comparison test of different matching methods

In order to confirm the superiority of the method proposed in this paper, two common matching methods were adopted to perform an experiment on test images. Method I is from literature [1], which constructs stereo pairs using the reference image with other search images to perform matching and merge them in object space to obtain the final DSM. Method II is from literature [4], which takes SNCC as the similarity measure and the optimal elevation is obtained by relaxation matching. Table 3 compares the DSM generated by three matching methods, and Fig. 5 hints at some partial DSM.

| Method | Matching success rate/% | MAE/m | RMSE/m | Time-consuming /h |
|--------|-------------------------|-------|--------|------------------|
| I      | 73.7                    | 0.157 | 0.340  | 1.91             |
| II     | 80.2                    | 0.177 | 0.327  | 1.88             |
| our    | 79.2                    | 0.153 | 0.326  | 1.10             |

By analyzing the results in Table 3 and Fig. 5, the following conclusions can be made:

1) Our method has the same matching success rate, about 80%, as method II, and the accuracy of the DSMs generated by the two kinds of methods is better than that of method I.

2) Contrary to method I, our method takes advantage of the multi-ray intersection technique to realize the fusion of various stereo pairs matching results in each level of image pyramid, which makes the stereo pairs no longer isolated and information from image and object-space comprehensively utilized. For method II, full use is made of photometric information simultaneously with multiple images, to improve the success rate and reliability of image matching, and finally to refine the generated DSM. From the results of Fig. 5 (depicted as block A), it is easily seen that the result of method I has quite obvious distortion in the building edge, but method II and our method can obtain the satisfactory results.

3) The superiority of method II lies in the ability to simultaneously use photometric and geometric information from multiple images, but the price is that it has to define a similarity measure in object space, such as SNCC. In order to calculate the SNCC, the window wrapping procedure is unavoidable. In literature [4], a corrective measure is adopted to alleviate the computation burden to a certain degree, but from Table 3 it is still considerable. For our method, the match is performed in the image space, but geometric restraint in object space of multiple images is also used, so a similar result as in method II is obtained while the time consumed by the CPU is reduced dramatically.
3 Conclusion

The matching strategy for multiple images is to first use the matching result in image to purify the information in object space, and in turn this is used to guide and constrain the matching for all images. In the process of image matching, the coarse to fine principle is obeyed, and cross-correlation matching with geometric constraints and relaxation matching is adopted, which can efficiently improve the success rate and reliability of image matching. Through a DSM generation experiment from different combinations of digital frame aerial images with large overlap, it is found that not only can our method pay regard to the independence of image matching, which allows the use of a variety of matured image matching algorithms, but also it considers the geometric restraint of multiple images, making the matching of each stereo pair not isolated. By the integrated use of information in image and object space, the wrong match points in individual stereo pairs can be efficiently eliminated, guaranteeing the high reliability of matching results. Compared with the traditional matching with just two images, that with three images can efficiently delete the bad points, and improve the matching success rate, and the generated DSM has an evident refinement. However, with more images participating in the matching, because the wrong match points are well detected and eliminated, the increase of matching success rate has a slower pace, and the role of multi-ray forward intersection to improve elevation accuracy is limited. Despite the refinement of the DSM accuracy, it is not more and more obvious.

References

[1] Gabet L, Giraudon G, Renouard L (1997) Automatic generation of high resolution urban zone digital elevation models [J]. International Journal of Photogrammetry and Remote Sensing, 52(1): 33-47
[2] Okutomi M, Kanade T (1993) A multiple-baseline stereo [J]. IEEE Transaction on Pattern Analysis and Machine Intelligence, 15(4):353-363
[3] Canu D, Ayache N, Sirat J A (1995) Accurate and robust stereovision with a large number of aerial images [C]. Proceedings of SPIE on Satellite Remote Sensing II, Paris
[4] Zhang Li (2005) Automatic digital surface model (DSM) generation from linear array images [D]. Zurich: ETH
[5] Zitnick C, Webb J A (1996) Multi-baseline stereo using surface extraction [R]. Technical Report CMU-CS-96-196, Carnegie Mellon University
[6] Zhang Zuxun, Zhang Jianqing (2002) Digital photogrammetry [M]. Wuhan: Wuhan University Press (in Chinese)
[7] Mallat S (1989) A theory for multiresolution signal decomposition: the wavelet representation [J]. IEEE Transaction on Pattern Analysis and Machine Intelligence, 11(7): 674-693
[8] Peter J, Adlison H (1983) The Laplcian pyramid as a compact image code [J]. IEEE Transaction on Communication, 31(4): 532-540
[9] Li Deren, Zheng Zhaobao (1992) Analytical photogrammetry [M]. Beijing: The Publishing House of Surveying and Mapping (in Chinese)
[10] Li Deren, Yuan Xiuxiao (2002) Error processing and reliability theory [M]. Wuhan: Wuhan University Press (in Chinese)
[11] Yuan Xiuxiao (2008) A novel method of systematic error compensation for a position and orientation system [J]. Progress in Natural Science, 18(8): 953-963