ADAPTIVE HIERARCHICAL DENSE MATCHING OF MULTI-VIEW AIRBORNE OBLIQUE IMAGERY

Z. C. Zhang *, C. G. Dai , S. Ji, M. Y. Zhao

Dept. of Photogrammetry and Remote Sensing, Zhengzhou Institute of Surveying and Mapping, Zhengzhou, China – paperrs2014@gmail.com, (paperrs2012, jisong_chxy, zhao_ming_yan)@163.com

Commission VI, WG III/1

KEY WORDS: Multi-view Oblique Imagery, Dense Matching, Image Pyramid, Matching Constraint, Delaunay Triangulation

ABSTRACT:

Traditional single-lens vertical photogrammetry can obtain object images from the air with rare lateral information of tall buildings. Multi-view airborne photogrammetry can get rich lateral texture of buildings, while the common area-based matching for oblique images may lose efficacy because of serious geometric distortion. A hierarchical dense matching algorithm is put forward here to match two oblique airborne images of different perspectives. Based on image hierarchical strategy and matching constraints, this algorithm delivers matching results from the upper layer of the pyramid to the below and implements per-pixel dense matching in the local Delaunay triangles between the original images. Experimental results show that the algorithm can effectively overcome the geometric distortion between different perspectives and achieve pixel-level dense matching entirely based on the image space.

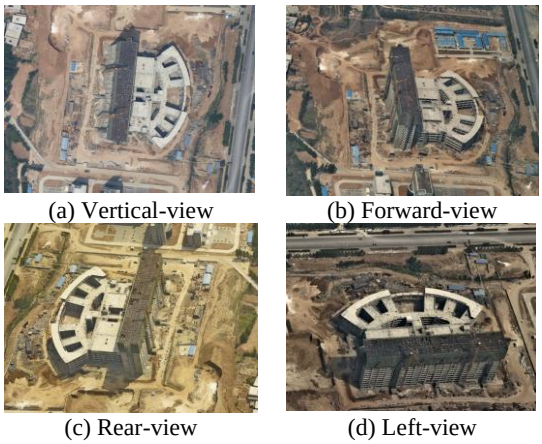
1. INTRODUCTION

The traditional single-lens photogrammetry cannot effectively obtain the lateral texture of tall buildings, while multi-view airborne photogrammetry can compensate for this shortcoming (Zhu et al., 2013). Combination of vertical images and oblique images into 3D modeling with abundant texture is an important trend in the development of photogrammetry (Gui et al. 2012; lo et al. 2013). This platform is equipped with multiple-perspective sensors to ensure more comprehensive access to objective information. The structure is usually mounted as one vertical lens with 4 (or 2) oblique lenses (Zhu et al., 2013).

Image matching is a critical part of digital photogrammetry, whose quality directly affects the accuracy of the DSM (Digital Surface Model). The popular area-based matching methodology for multi-view oblique images almost completely loses efficacy because of the large geometric distortion between multi-view images. Matching algorithms for images with large geometric distortion can be invariant feature-based matching methods, e.g. SIFT (Scale Invariant Feature Transform) (Lowe, 2004), SURF (Speeded-Up Robust Features) (Bay et al., 2009), ASIFT (Affine - Scale Invariant Feature Transform) (Moreal and Yu, 2009). Dense matching methodology is usually narrowing searching area based on comprehensive utilization of image-space and object-space information, e.g. GC^3 (Geometrically Constrained Cross-Correlation), MVLL (Modified Vertical Line Locus). But these algorithms are mainly used for images of the same perspective and experiments on multi-view oblique images are rare. In this paper, a dense matching algorithm specifically for multi-view oblique images is proposed based on invariant feature-based matching and geometric correction of area-based matching.

2. MULTI-VIEW AIRBORNE OBLIQUE IAMGES

Take five-view camera as an example, in multi-view airborne photogrammetry, the five lenses expose at the same time at a single shot, and images and their exterior orientation elements are obtained. The operation principals of the vertical lens are similar to those in the traditional photogrammetry. The tilt directions and placement characteristics determine the difference between oblique images and vertical images. The four oblique lenses are usually mounted as Malta shape, e.g. Chinese TOPDC-5, SWDC-5 and AMC580, of which the angles between vertical and oblique lenses generally range from 45 ° to 60 °. Figure 1 shows the vertical-, forward-, rear- and left-view images of the same scene taken by five-view AMC580 camera over Dengfeng, Henan province.



(a) Vertical-view (b) Forward-view

(c) Rear-view (d) Left-view

Figure 1. The multi-view images of AMC580

Figure 1 shows there is not only rotation of optic axis but also pitching differences between images of different views. Moreal and Yu (2009) put that longitude difference and latitude difference from geography could be used to quantitatively measure the two kinds of tilt degrees in oblique images. In figure 1, the tile angles of oblique lenses are all 45 °, so the longitude differences between (a) and (b), (c), (d) are 0 °, 180 °, 90 ° respectively and the latitude differences are all 45 ° (Zhang

* Corresponding author.

et al., 2014). The image contrast and brightness of different perspectives are of a significance difference and serious overlapping and redundancy exist in multi-view images. In addition, scales on one single oblique image are not constant, which is the most important difference from vertical images (Grenzdorffer et al., 2008; Gruber and Walcher, 2013; Petrie, 2008; Hohle, 2008). These characteristics of oblique images make the effects of feature-based matching and area-based matching decline sharply, and even fail completely.

3. THE OVERALL PIPELINE OF HIERARCHICAL MATCHING

Image matching is the process of searching the corresponding entity between the reference images and the matching images. Often the corresponding point cannot be found on the matching image and then this problem seems to be insoluble. Geometric constraints are often used to initially narrow the searching area thus to improve the matching efficiency and reliability. In computer vision, given certain pairs of corresponding points, the relative camera geometry and the entire geometry of two images can be obtained. The widely used constraint methods are homography and epipolar constraint, the relevant matrix being the homography H and the fundamental matrix F (Hartley and Zisserman, 2003; Yang et al., 2012).

Image frame of the frame camera imagery is considerably wide now, so three-layer hierarchical matching strategy is used to deliver and refine the matching results. The corresponding pairs and geometric relationship are passed down to the original layer, and the per-pixel matching is finally realized on the original images. Figure 2 shows the overall pipeline of matching process between vertical image and oblique image.

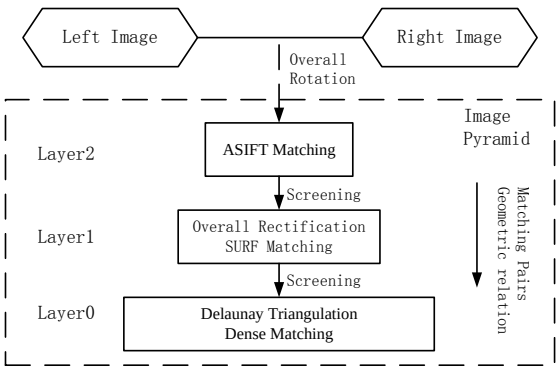


Figure 2. The overall pipeline of hierarchical dense matching

Although SIFT descriptor is invariant and stable towards image rotation, our experiments show that SIFT and ASIFT for images rotated to the same direction in advance can achieve better matching effect than for images without rotation. So overlapping images are rotated to the same perspective first so that the images will point in the same direction in general. Assume that the original image size is $w \times h$ (width × height) and the point coordinate on the rotated image is (x, y). Then there are three conditions of calculating the screen coordinates on the original images:

1. Before $180°$ rotation, the screen coordinate is $(w-x-1, h-y-1)$.
2. Before clockwise $90°$ rotation, the screen coordinate is $(y, h-x-1)$.
3. Before anticlockwise $90°$ rotation, the screen coordinate is $(w-y-1, x)$.

Then both the original images after rotation are resampled twice to obtain three-layer image pyramids, from bottom to top: layer 0 (the original layer), layer 1 (the middle layer), layer 2 (the highest layer). On the second layer, after low-pass filtering twice, the image size tends to be smaller and detailed information has been largely filtered out. The remains are mainly overall texture so scale- and rotation- invariant ASIFT is appropriate to implement here. With those delivered corresponding pairs, the distortion geometry is estimated between the two images on the layer 1. The matching image is rectified to the reference image with the geometric relationship. Assume that the imaged areas are both plane and the systematic error impact of lens distortion and atmospheric refraction is overlooked here. Then the perspective transformation can fully simulate this "image-to-image" coordinate correspondence between the two images (Yang et al., 2012). The efficient SURF matching algorithm is used between the reference image and the rectified matching image to achieve sub-pixel matching. Certainly the pixel coordinates on the rectified matching image should be projected reversely to the original Layer 1. To ensure the accuracy of rectification and inverse calculation, the bilinear interpolation or cubic convolution interpolation is adopted here.

When the corresponding pairs are passed to the original images on layer 0, Delaunay triangulation is constructed among the evenly-distributed matching pairs after screening. The empty circle and the maximum minimum angle characteristics of Delaunay triangulation can ensure the nearest points are used to restrain the dense matching process (Zhang et al., 2013; Wu et al., 2011). To ensure the Delaunay networks on both images are corresponding and consistent, the Delaunay network is constructed first on the reference image with the reliable matching features. Following this is construction of network on the matching image with the chain code sequence of the left network. Finally, per-pixel dense matching is implemented in corresponding triangles.

4. KEY TECHNOLOGY

4.1 Delaunay Triangulation on Layer 0

Delaunay triangular network is established with upper matching pairs on layer 0, which is followed by per-pixel dense matching in each Delaunay triangle. According to the theory of continuous disparity (Wu et al., 2011), the corresponding pixel of the reference image pixel must be located in or around the corresponding triangle. For a pair of corresponding triangles, our method calculates the circumscribed rectangles of both the triangles first and take the new rectangle with maximum length and width as the local rectification cell. The rectangular calculation cell should be expanded outward a certain number of pixels (half of the window size for correlation coefficient calculation) to ensure adequate surrounding pixels take part in the calculation of correlation coefficients. At that time, each pair of corresponding Delaunay triangles corresponds to one pair of "Rectangular Patches". In each pair of rectangular patches, dense matching is implemented for every pixel in the Delaunay triangle as following:

1. Search for the feature-based matching pairs in the rectangular range twice as wide as the rectangular patch to

calculate the perspective transformation geometry. The feature-based pairs are delivered from SURF on Layer 1.

2. Rectify the matching patch to the reference patch based on the perspective transformation matrix. Ideally, the pixels on the rectified patch exactly corresponds to the ones on the matching patch, i.e. the pixels with the same coordinates on both patches are ideally corresponding pixels. But owing to terrain deformation and undulation, there is a slight offset between the corresponding pixels.

3. For every point p on the reference triangle, open the same $n \times n$ searching window around the pixel of the same coordinate on the right patch. Epipolar geometry is used to further constrain the searching space. Normalized Cross Correlation (NCC) (Wu et al., 2011) is taken as the correlation measurement here which is calculated between the reference pixel and the matching pixels in the searching window. The pixel on the right patch with the maximum NCC which is larger than the given threshold is the final corresponding pixel.

Here we take rectangular patch as the rectification cell and the pixels in the triangles are only rectified once, thus the process of calculating NCC only involves "reading" gray level of pixels, and avoiding pixel projection and interpolation, which effectively improved operational efficiency. Besides, the NCC threshold is adaptively determined with the three pairs of triangular vertices. If the total number of corresponding pairs for calculating perspective transformation matrix is only 4 or 5, then the perspective matrix is not reliable enough and the threshold needs to be further increased by 0.1. Tests show that threshold of NCC generally ranges from 0.65 to 0.85.

4.2 Homography and Fundamental Matrix

In the imaging process, the perspective transformation can be represented with the homography *H*. Assume that the overlapping area or the imaged area is absolutely plane and only rotation, translation and scaling change exist, then all the corresponding points between the two images can be characterized with only one transformation matrix. The reference pixel coordinate is $(x, y, 1)^T$,the matching pixel coordinate is $(x', y', 1)^T$, the homography between them is in (Yang et al., 2012):

$$\mathrm{H} = \begin{pmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ a_7 & a_8 & a_9 \end{pmatrix} \tag{1}$$

Then there is the following relationship between the two coordinates (Yang et al., 2012):

$$\begin{cases} x' = \dfrac{a_1 x + a_2 y + a_3}{a_7 x + a_8 y + a_9} \\ y' = \dfrac{a_4 x + a_5 y + a_6}{a_7 x + a_8 y + a_9} \end{cases} \tag{2}$$

Epipolar geometry is widely used as image matching constraint: given the fundamental matrix F between two images and the reference pixel coordinate, the corresponding epipolar line can be determined on the matching image. Theoretically, the corresponding point on the matching image must be on the corresponding line, so epipolar geometry can transfer the two-dimensional searching into one-dimentioanal. *F* is determined by imaging camera position and orientation and can also be approximately obtained with at least 8 pairs of corresponding pixels (Hartley and Zisserman, 2003).

F is calculated with the delivered SURF matching pixels from layer 1. Take these pairs as check points, experiental results show that 98% of these points can achieve the accuracy of 5 pixels and 100% can achieve the accuracy of 6 pixels. The constraint effect is shown in Figure 3. The horizon-axis is the corresponding point sequence and the vertical-axis is the constrait offset. Figure 4 shows the constrait effect of 5-pixel accuracy of features and the corresponding epipolar lines. the circles are the matching points. By further comparison, epipolar constraint can achieve higher accuracy than homography constraint, i.e., *F* can better express the overall geometric transformation of two images than *H*. So on layer 0, homography constraint is used in the local triangles, while epipolar constraint is globally used.

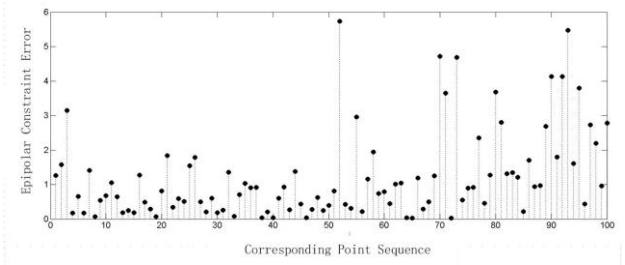


Figure 3. Constraint accuracy of epipolar geometry

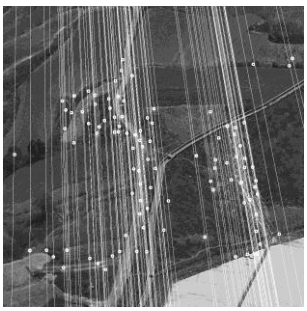


Figure 4. Epipolar geometry

When calculating *H* or *F*, more than necessary pairs of corresponding points are used to obtain optimal solutions under the RANSAC (Random Sample Consensus) (Fisehler and Bolles, 1981) criteria.

4.3 Feature Screening

In the upper two layers of the image pyramid, two types of corresponding features need to be screened. One is mismatching points derived from ASIFT and SURF, the other is dense matching points. For the first type, RANSAC can take the role to screen them. The over-dense matching pairs are screened in case of non-significant Delaunay constraint and low efficiency.

The main idea of feature screening is judging the distance of two pairs of corresponding points on the reference and the matching images. If one of the two distances is less than the threshold, one of the two pairs should be removed. Provided that *n* pairs of corresponding points are given, the number of judgement is C_n^2. For the i th and the j th corresponding points which are independent, assume that $\Delta DLij$ is the distance of

the two features on the reference image, $\Delta DRij$ is the distance on the matching image:

1. $\Delta DLij < T_1$ or $\Delta DRij < T_1$: The two pairs are too closely distributed, so remove one pair of them. T_1 is the repetitive and redundant threshold, which takes 5 pixels here.
2. $\Delta DLij \le T_2$ or $\Delta DRij \le T_2$: the two pairs are so close that the triangulation constraint is not significant, so remove one pair. T_2 is the minimum distance threshold, often determined by the terrain condition and matching accuracy and should increase gradually down along the image pyramid.
3. $\Delta DLij > T_2$ and $\Delta DRij > T_2$, both of the two pairs should be kept.

The premise of the screening process is the acquisition of reliable and high-precision matching features. After screening, the matching features can be used to establish Delaunay triangles.

5. EXPERIMENTS

The experimental data are both derived from the airborne images over Dengfeng taken by AMC580. They are cut from the original 10328 ×7760 oblique images. The data information is shown in Table 1.

NO	Image Size	Lens Perspective	Remarks
1	800×800	Vertical & Rear	Plain and smooth terrain
2	1000×1000	Vertical & Forward	Significant brightness difference

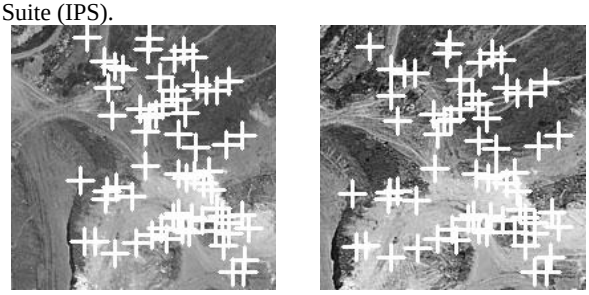
Table 1. The dataset information

Under the 3-layer image pyramid matching strategy, the reliable matching results and Delaunay triangle numbers are listed in Table 2. Group 1 is matching images of plain area, and the terrain is plane. Because of the overall direction is not the same, the rear-view image should be rotated by 180° first. Only 64 pairs of ASIFT matching pairs are remained after screening. 100 pairs of matching pairs are obtained on layer 1, which are delivered to Layer 0 to establish 186 Delaunay triangles. Per-pixel dense matching is conducted on layer 0 finally and 214,028 pairs of corresponding points are obtained. Group 2 is matching the vertical- and forward-view images. 340 Delaunay triangles are established on Layer 0 and 317,551 pairs of points are matched. Every pixel in the Delaunay triangles is meant to search for the corresponding pixel on the matching image. The matching success rate is the percentage of successful matching numbers to all the points on the reference Delaunay triangles. The matching success rates of the two experiments are both over 75%.

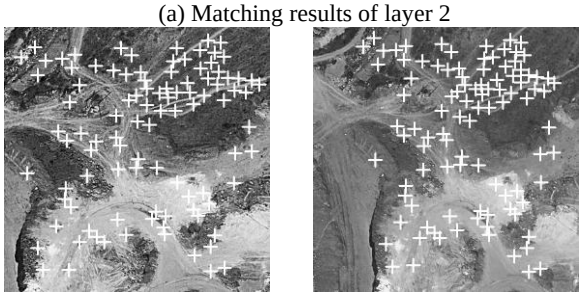
NO	Layer 2	Layer 1	Layer 0	Delaunay Triangle Number	Matching Success Rate
1	64	100	214,028	186	76.34%
2	68	178	317,551	340	81.68%

Table 2. Matching results

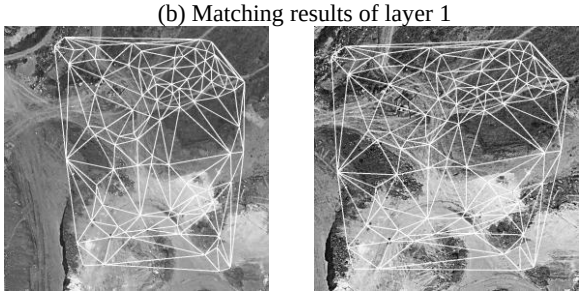
The final matching results of the two experiments are showed in Figure 5 and Figure 6. In these two figures, (a) and (b) are matching results of layer 2 and layer 1 respectively; (c) are the Delaunay triangulation on layer 0; (d) are point clouds generated from matching pairs assisted with the exterior elements derived from the commercial Icaros Photogrammetric Suite (IPS).



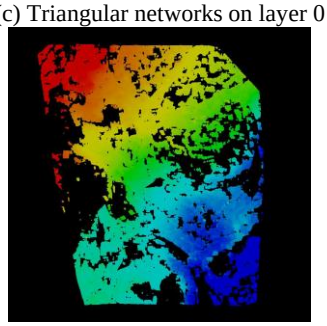
(a) Matching results of layer 2



(b) Matching results of layer 1

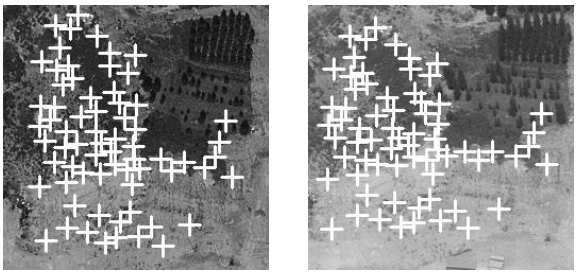


(c) Triangular networks on layer 0

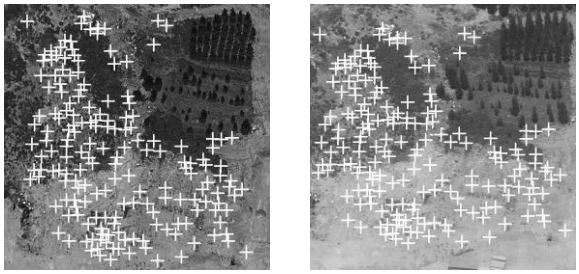


(d) Dense point cloud

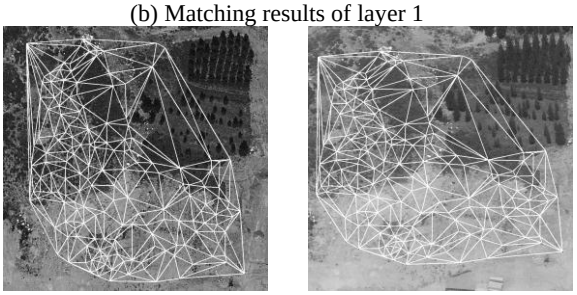
Figure 5. Matching results of Group 1



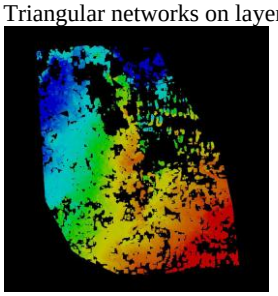
(a) Matching results of layer 2



(b) Matching results of layer 1



(c) Triangular networks on layer 0

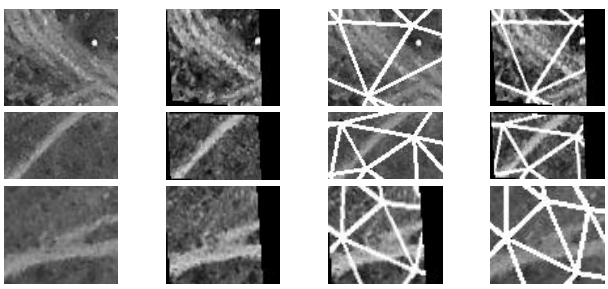


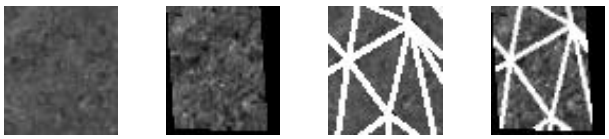
(d) Dense point cloud

Figure 6. Matching results of Group 2

By comparison of Figure 5(a) and (b), the feature-base matching of ASIFT on Layer 2 and the SURF on Layer 1 can achieve the similar distribution of matching features. In the textureless area, the ASIFT and SURF are difficult to achieve successful matching. In Figure 6(a) and (b), there is no successful matching in the upper right area (the vegetarian area) and these areas are usually constrained by large triangles and thus difficult to achieve a successful match in the following area-based matching (see the sparse point cloud on the upper right area in Figure 6(d)).

Certain pairs of the 186 Delaunay triangles and rectangular patches in Group 1 are shown in Figure 7. On each row, the first two patches are the corresponding rectangular patches and the following two are rectangular patches with the corresponding Delaunay triangles. The corresponding rectangular patches are of the same size. The left images shown in Figure 7 are the reference images while the right are the rectified matching images. So the corresponding patches on every row look similar to each other and the corresponding pixels are close with only several pixels' offset.





(a) Rectangular patches (b) Delaunay Triangles

Figure 7. Part rectangular patches of Group 1

This paper adopts the manual method to testify the matching accuracy. On the original images, 50 easily distinguishable corresponding pairs are selected and we find that the numbers of correct matching for the two groups are 49 and 46 respectively and the matching accuracy goes to pixel-level. There are two main sources of mismatches: one source is matching errors on the textureless or texture-repetitive area; the other is projection and interpolation error on layer 1 and layer 0.

6. CONCLUSIONS AND OUTLOOK

Without POS data, this paper implements the dense per-pixel matching which can effectively compensate for the longitude and latitude deformation of multi-view oblique images. The image hierarchical strategy and image space-based geometric constraints are used to restrain the searching area and refine the matching results. The Delaunay triangulation is used to compensate for the serious geometric distortion to achieve the successful pixel-level matching for at least 75% pixels on the restrained area of the reference multi-view images. Since the Delaunay triangles on the original images are rectified only once, the searching process can be implemented around the pixels of the same coordinates and thus avoid the repetitive projection and interpolation. For the challenges of matching textureless or texture-repetitive area, multi-view matching and object space information is to be further studied and combined into our algorithm.

ACKNOWLEDGEMENTS

We want to thank Shanghai Hangyao Information Technology Co., Ltd. for providing multi-view airborne oblique images.

REFERENCES

Bay H., Tuytelaars T. and Gool L. V., 2009. SURF: Speeded up robust features. European Conference on Computer Vision, 1, pp. 404-417.

Fisehler M. and Bolles R., 1981. Random Sample Consensus：A paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM, 24(6), pp. 381-395.

Grenzdorffer G. J., Guretzki M. and Friedlander I., 2008. Photogrammetric image acquisition and image analysis of oblique imagery. The Photogrammetric Record, 23, pp. 372-386.

Gruber M. and Walcher W., 2013. Oblique image collection-challenges and solutions. http://www.ifp.uni-stuttgart.de/publications.

Gui D., Lin Z., Zhang C., 2012. Research on construction of 3D building based on oblique images from UAV. Science of Surveying and Mapping, 37(4), pp.140-142.

Hartley R. and Zisserman A., 2003, Multiple View Geometry in Computer Vision. Cambridge University Press.

Hohle J., 2008. Photogrammetric measurements in oblique aerial images. Photogrammetrie Fernerkundung Geoinformation ,1, pp. 7-14.

Lo B., Geng Z., Wei X., 2013. Texture mapping of 3D city model based on Pictometry oblique image. Engineering of Surveying and Mapping, 22(1), pp. 70-74.

Lowe, D. G., 2004. Distinctive image features from scale-invariant key points. International Journal of Computer Vision, 60(2), pp. 91-110.

Moreal J. M. and Yu G., 2009. ASIFT: A new framework for fully affine invariant image comparison. SIAM Journal on Imaging Sciences, 2(2), pp. 438-469.

Petrie G., 2008. Systematic oblique aerial photography using multiple digital cameras. VIII International Scientific & Technical Conference, pp. 15-18.

Wu B., Zhang Y. and Zhu Q., 2011. A triangulation-based hierarchical image matching method for wide-baseline images. Photogrammetric Engineering & Remote Sensing, 77(7), pp.695-708.

Yang H, Zhang L., Yao G., 2012. An automatic image registration method with high accuracy based on local homography constraint. Acta Geodaetica et Cartographica Sinica, 41(3), pp. 401-408.

Zhang Y., Zhu Q., Wu B., 2013. A hierarchical stereo line matching method based on triangle constraint. Geomatics and Information Science of Wuhan University, 38(5), pp. 522-527.

Zhang Z., Dai C., Mo D., 2014. Effect of image tilt measure on ASIFT matching validity. Journal of Geomatics Science and Technology, 31(5), pp. 514-518.

Zhu Q., Xu G., Du Z., 2013. An technology overview of oblique photogrammetry, http://www.paper.edu.cn/releasepaper /content/201205-355.