A ROBUST MATCHING METHOD FOR UNMMANED AERIAL VEHICLE IMAGES WITH DIFFERENT VIEWPOINT ANGLES BASED ON REGIONAL COHERENCY

Zhenfeng Shao [a, b, *], Congmin Li [a, b], Nan Yang [a, b]

[a] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, No. 129 Luoyu Road, Wuhan, Hubei, China
[b] Collaborative Innovation Center for Geospatial Technology，129 Luoyu Road, Wuhan 430079, China
shaozhenfeng@whu.edu.cn

KEY WORDS: Unmanned Aerial Vehicle Images, Image Matching, Regional Coherency, Affine Invariant, Feature Detection, Feature Description

ABSTRACT:

One of the main challenges confronting high-resolution remote sensing image matching is how to address the issue of geometric deformation between images, especially when the images are obtained from different viewpoints. In this paper, a robust matching method for Unmanned Aerial Vehicle images of different viewpoint angles based on regional coherency is proposed. The literature on the geometric transform analysis reveals that if transformations between different pixel pairs are different, they can't be expressed by a uniform affine transform. While for the same real scene, if the instantaneous field of view or the target depth changes is small, transformation between pixels in the whole image can be approximated by an affine transform. On the basis of this analysis, a region coherency matching method for Unmanned Aerial Vehicle images is proposed. In the proposed method, the simplified mapping from image view change to scale change and rotation change has been derived. Through this processing, the matching between view change images can be converted into the matching between rotation and scale changed images. In the method, firstly local image regions are detected and view changes between these local regions are mapped to rotation and scale change by performing local region simulation. And then, point feature detection and matching are implemented in the simulated image regions. Finally, a group of Unmanned Aerial Vehicle images are adopted to verify the performance of proposed matching method respectively, and a comparative analysis with other methods demonstrates the effectiveness of the proposed method.

1. INTRODUCTION

Image matching is an important issue in the photogrammetry field and an academic hotspot for research. It is one of the key technologies in target recognition, three-dimensional reconstruction, image retrieval (Gruen A and Zhang L, 2002; Richard et al., 2010). In the field of remote sensing, the performance of image matching determines the subsequent quality of digital surface models generated automatically, semi-automatic three-dimensional measuring of surface features, and automatic aerial triangulation results. Image matching can promote the applications on the update of large-scale maps, urban three-dimensional geological modelling (Gruen A and Akca D, 2005), urban planning, agricultural monitoring, etc. In order to find homologous point candidates across the images, many approaches for matching have evolved over the years. Ida J and Clive S F address image scanning for feature point detection, and specifically an evaluation of three different interest-point operators to support feature-based matching in convergent, multi-image network configurations (Ida J and Clive S F, 2010). Chunsun Z and Clive S F have put forward an automated registration of high-resolution satellite images based on a hierarchical image matching strategy and the similarity of grey levels (Chunsun Z and Clive S F, 2007). Wang R R proposed an automatic region registration method based on spatially assistant plane (Wang R R et al., 2010), but the accuracy will decrease when the reference image and the image to be matched have big difference in angle and resolution. Although automated approaches have quite a number of advantages, the quality of the results is still not satisfactory and, in some cases, far from acceptable. Even with the most advanced techniques, it is not yet possible to achieve the quality of results that a human operator can produce (Gruen A, 2012).

As to the different viewpoint images, the similarity of the same object on images will become smaller with the increasing of view change. Furthermore, due to the instability of the remote sensors caused by their flight platform at different spatial positions, there exist translation, rotation, scale and perspective changes in the external parameters between Unmanned Aerial Vehicle images. This results in image matching failure. Therefore, how to match the Unmanned Aerial Vehicle images is still not fully resolved and has drawn great attention from the international research community. At present, the point-matching algorithms are widely used, which are based on the assumption that similar neighboring pixels are disparity, and thus are difficult to qualify the matching on Unmanned Aerial Vehicle images. For the matching of Unmanned Aerial Vehicle images which have different viewpoint angles, there are two major difficulties:

(1) The changes of remote platform altitude, roll angle and pitch angle will result in scale difference between the images, which reduces the success rate and reliability of image matching algorithms;

(2) Multi-angle imaging will produce radiation distortion. Thus, there exits serious chromatic aberration between homogeneous points, which will result in the failure of image matching. Therefore, in order to achieve the robust matching of

* Corresponding author

Unmanned Aerial Vehicle images with multi-angles, it is necessary to analyze the imaging model to determine the geometric transformation model between Unmanned Aerial Vehicle images, and chromatic aberration between homogeneous points need to be adjusted.

For the above two issues, this paper proposes a new method based on remote sensing image matching perspective transformation model and invariant features to find a feasible solution for robust matching of Unmanned Aerial Vehicle images with different viewpoint angles.

2. RELATED WORK

Image matching is probably the most important function in digital photogrammetry and also in automated modelling and mapping (Gruen A, 2012). For the matching of different viewpoint images, the traditional methods are to improve the affine invariance of feature detector or descriptor, such as SIFT, Harris-Affine, Hessian-Affine (Mikolajczyk et al., 2004), MSER and so on. In (Lowe, 2004), a Gaussian weighting function is used to assign a weight to the gradient magnitude of each sample point when computing SIFT descriptor. It gives less emphasis on gradients that are far from the center of the descriptor. As a result, the problem caused by SIFT without affine invariance can be offset partially. In (Mikolajczyk et al., 2004), a set of initial points extracted at their characteristic scales based on the Harris-Laplace detection scheme are input to Harris-Affine detector, and an iterative algorithm is applied to adjust the location, scale and local area of every point so as to get an affine invariant point. Through the imitation of Harris-Affine, another affine invariant detector Hessian-Affine is proposed. The difference is that it starts from the Hessian rather than the Harris corners. In (Matas et al., 2004), the concept of terrain watershed is introduced to extract MSERs. The MSERs are the parts of the image where local binarization is stable over a large range of thresholds. The definition of MSER stability based on relative area change is invariant to affine transformations.

In recent years, another feasible solution to cope with the change of view in image matching is simulating the original image to every possible view, extracting features and matching respectively. In (Yu et al., 2009; Morel et al., 2009), (Morel et al., 2009) have proposed a full affine invariant framework ASIFT for different viewpoint images matching (shown in Fig.1). ASIFT simulates the reference image and the test image to cover the whole affine space. Then, SIFT is used to extract and compare the features from these simulations. After feature matching, the correspondent pairs are converted to the original images. ASIFT can find matches from the images even if they are from different viewpoints. In (Yu et al., 2012), another matching framework Iterative SIFT (ISIFT) for different viewpoint images is proposed (shown in Fig.2). In Fig.1 and Fig.2, the square images A and B represent the compared images, and the square image B′represents the estimated image about the transformed pose. Through the iterative algorithm, the geometric transformation between the image pair is estimated. According the estimated model, the test image (or the reference image) is simulated. Then the reference image (or the test image) is matched with the simulated image. And the matching results are converted to the original images.

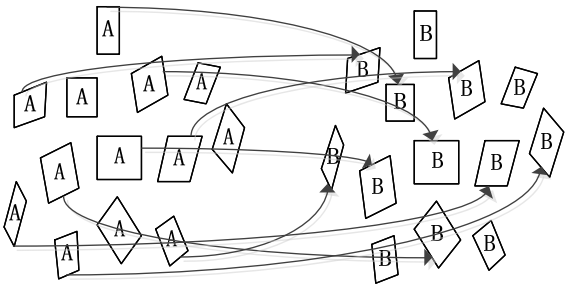


Fig.1 Schematic diagram of ASIFT method (Morel et al., 2009)

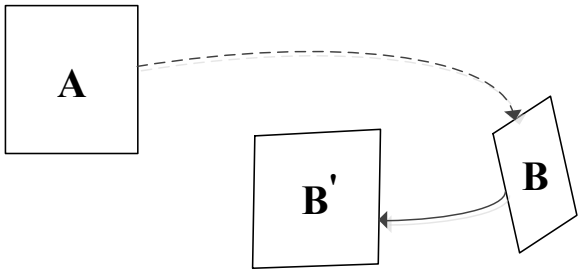


Fig.2 Schematic diagram of ISIFT method (Yu et al., 2012)

3. PROPOSED VIEW INVARIANT IMAGE MATCHING METHOD

Based on the geometric transform analysis, it can be drawn that for the same real scene, if the instantaneous field of view or the target depth changes are small, transformation between pixels in the whole image can be approximated by an affine transform. Otherwise, transformation between different pixel pairs is different. And they cannot be expressed by a uniform affine transform. On the basis of this analysis, a region congruency matching method for Unmanned Aerial Vehicle images with different viewpoint angles is proposed. In the proposed method, the simplified mapping from image view change to scale change and rotation has been derived. Through this processing, the matching problem between view change images can be converted into the matching between rotation and scale changed images.

3.1 Analysis of transformation between different viewpoint images

The transformation between a point on the ground and its corresponding appearance on the image can be described by the Holes perspective projection imaging model. The relationship between the images coordinates of a pixel and its corresponding world point's coordinates can be expressed as

$$s\left[u,v,1\right]^{\mathrm{T}} = \mathbf{M}_1\mathbf{M}_2\left[X,Y,Z,1\right]^{\mathrm{T}} \qquad (1)$$

Where s denotes the depth coordinate in projection direction of the camera coordinate system, the matrix $\mathbf{M_1}$ is the internal parameters of the camera including focal length, resolution and the image plane offset, and the matrix $\mathbf{M_2}$ represents the translation and rotation between the camera coordinate system and the world coordinate system called extrinsic parameters.

When a point on the ground is viewed from different angles, the relationship between the pixels coordinates of the point in the

images can be obtained by the camera imaging model as shown in the following Formula.

$$\begin{cases} u_2 = a_{11}u_1 + a_{12}v_1 + b_1 \\ v_2 = a_{21}u_1 + a_{22}v_1 + b_2 \end{cases} \quad (2)$$

The camera intrinsic parameters and extrinsic parameters are equal for every point, which can be seen as constant. But the coefficients a_{ij} and b_i in equation (2) are still related to the depth coordinates s_1 and s_2 . Obviously, when the ratio s_1/s_2 of the two corresponding pixels in the image pair is constant, the coefficients in Formula (2) are constant. Then the geometric relationship between all pixels in the image pair can be expressed by an affine transform. In practice, if the fields of view of the cameras are small or the depth variation in local object is negligible, the projection transformation of the corresponding regions can be approximated by an affine transform because s_1/s_2 is proximate to constant.

According to the analysis above, the geometric transformation between the corresponding local regions shown in Fig.1 can be approximated by an affine transform that is expressed by matrix $\mathbf{A}$ (Yu et al., 2009),

$$\mathbf{A} = \mathbf{H}_{\lambda}\mathbf{R}_1(\psi)\mathbf{T}(\theta)\mathbf{R}_2(\phi) \quad (3)$$

Where $\mathbf{H}_{\lambda}$ is the scale change, $\mathbf{R}_1(\psi)$ denotes the rotation, and $\mathbf{T}(\theta)\mathbf{R}_2(\phi)$ represents the view change.

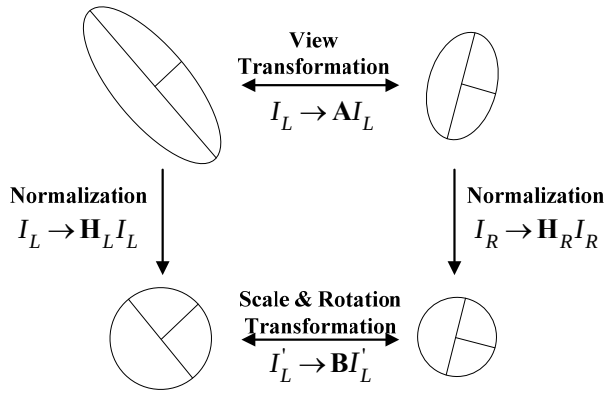


Fig.3 Transformation between a pair of correspondent local areas

In this paper multiple local areas from the two images are extracted using the method. In these local areas, the view field and target depth variation are small, so each local area can be approximated with an affine transformation model accordingly. This paper defines the characteristics of the local areas that can be approximated with an affine transformation model region congruency.

In Fig.3 the two elliptical areas are the two corresponding local areas with region congruency which are extracted and simulated from images of different viewpoint angles. And the two local elliptical regions are transformed to circular areas respectively, the matrix $\mathbf{B}$ is used to describe the transformation between them. According to (Baumberg, 2000),

$$\mu_L' = \mathbf{B}^{\mathrm{T}}\mu_R'\mathbf{B} \quad (4)$$

where μ_L' and μ_R' denote the second-order moments of the two circular areas. For the circular fields,

$$\mu_L' = \lambda_L \mathrm{E},\ \mu_R' = \lambda_R \mathrm{E} \quad (5)$$

It can be deduced from the Formula (4) and (5) that

$$(\lambda_L/\lambda_R)\mathrm{E} = \mathbf{B}^{\mathrm{T}}\mathbf{B} \quad (6)$$

Therefore, there are only scale change and rotation between the two areas, which could be mathematically expressed by

$$\mathbf{B} = \mathbf{H}_{\lambda'}\mathbf{R}_1(\psi') \quad (7)$$

Besides, this conclusion can be illustrated by conducting experiments on a pair of images taken from two different viewpoints (see Fig. 4).

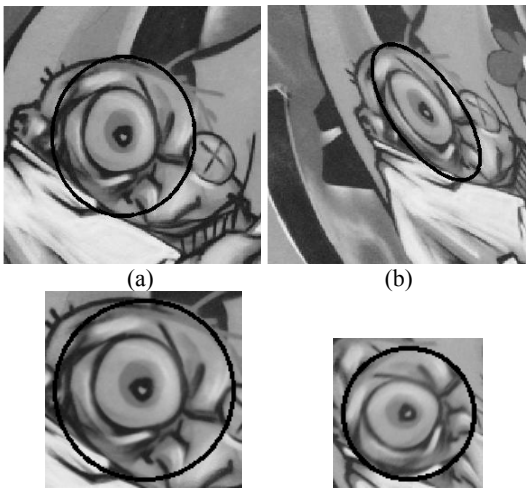


Fig.4 (a) & (b) are ellipse fitting results of the detected regions from two viewpoints. (c)& (d) are normalization regions by transforming elliptical regions to circular areas.

3.2 Three steps of the proposed matching method

Fig.5 shows the proposed method composed of three steps: (1) region detection and extraction; (2) region transformation; (3) feature extraction and image matching.

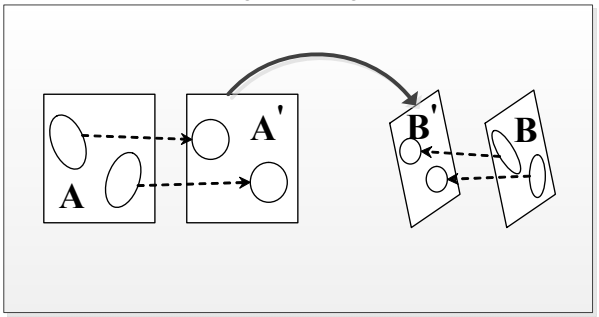


Fig.5 Schematic diagram of the proposed method

3.2.1 Region detection and extraction: In order to obtain the circular areas, local regions need to be extracted firstly from the images. Currently, there are many well-known regional feature extraction algorithms such as SR (Kadir et al., 2004), IBR (Tuytelaars et al., 2004), EBR (Tuytelaars et al., 2004), MSER (Matas et al., 2004), etc. In this paper the purpose of detecting the local area is to match images with large viewpoint angle change, so the local area feature detection methods used need to have a strong robustness to the change of viewpoint angle. (Mikolajczyk et al., 2005) shows that MSER has strong robustness to the change of viewpoint angle, so in this paper MSER operator is selected to extract image features of local area.

In the original MSER algorithm, although the extracted MSERs come in many different sizes, they are all detected at a single image resolution. When a scene is blurred or viewed from increasing distances, many details in the image disappear and different region boundaries are formed. In this case, the repetition rate of the local area extracted from different images obtained will decrease, which will affect the subsequent image matching. In order to improve the scale invariance of MSER, this paper uses the multi-resolution strategy to detect MSERs from different resolutions instead of detecting MSERs only in the original input image (Forssen et al., 2007). Specifically, the procedure adopted in our method is described as follows:
(1)Firstly, a scale pyramid is constructed by blurring and sub-sampling with a Gaussian kernel.
(2)Then, MSERs are detected separately at each resolution image according the method proposed in (Matas et al., 2004).
(3)Finally, duplicate MSERs are removed by eliminating fine scale MSERs with similar locations and sizes as MSERs detected at the next coarser scale.

3.2.2 Region transformation: The detected elliptical multi-resolution MSERs are transformed to circular areas (called CA_p) according to the values of macro axis and minor axis by the method as follows:

Assuming the macro axis of Elliptical Area (called EA_p) is l and its minor axis is w , the radius of the transformed circular area is calculated with $r = \sqrt{l \cdot w}$. Matrix $\mathbf{H}$ is used to express the geometric transformation between the elliptical region and the circular area, so $\mathbf{H}$ satisfy the formula

$$\left[\mathbf{H}(\mathbf{X}-\mathbf{X}_g)\right]^{\mathrm{T}}\left[\mathbf{H}(\mathbf{X}-\mathbf{X}_g)\right] = r^2 \qquad (8)$$

where $\mathbf{X}$ is a point on the ellipse, and $\mathbf{X_g}$ is the center of the ellipse. Since $\mathbf{X}$ is on the ellipse, thus

$$(\mathbf{X}-\mathbf{X_g})^{\mathrm{T}}\boldsymbol{\mu}^{-1}(\mathbf{X}-\mathbf{X_g}) = 1 \qquad (9)$$

where $\boldsymbol{\mu} = [\mu_{20}, \mu_{11}; \mu_{11}, \mu_{02}]$ is the second-order moment of the elliptical region. Calculate the equations (8) and (9),

$$\mathbf{H} = \frac{r}{[\mu_{20}(\mu_{20}\mu_{02} - \mu_{11}^{\ 2})]^{1/2}} \begin{bmatrix} (\mu_{20}\mu_{02} - \mu_{11}^{\ 2})^{1/2} & 0 \\ -\mu_{11} & \mu_{20} \end{bmatrix} \qquad (10)$$

Then, the elliptical region can be mapped in a circular area with the centre $\mathbf{X_g}$ and the radius r by the transformation matrix $\mathbf{H}$. Thus the detected elliptical multi-resolution MSERs are transformed to circular areas according to the following equations.

$$CA_p = \mathbf{H} \cdot EA_p \qquad (11)$$

3.2.3 Feature extraction and image matching: In this paper the scale invariant DoG detector is selected to extract the features and the SIFT feature descriptor is chosen to describe and obtain the initial matching results.

It is Inevitable that there are some incorrect correspondences in the initial results. The traditional method is to estimate the homography between the two images by using RANSAC algorithm. Those initial matches that do not conform to the estimated model should be eliminated as false matches. Generally the geometric transformation between all pixels of the two images cannot be approximated by an affine transform. In this condition, if the homography is used to identify wrong matches, many correct matches will be eliminated as a result. In this paper, in order to avoid the incurrence of this problem, epipolar constraint based on the fundamental matrix is used to eliminate wrong corresponding pairs with RANSAC. Fig.6 is the flowchart of the proposed matching method.

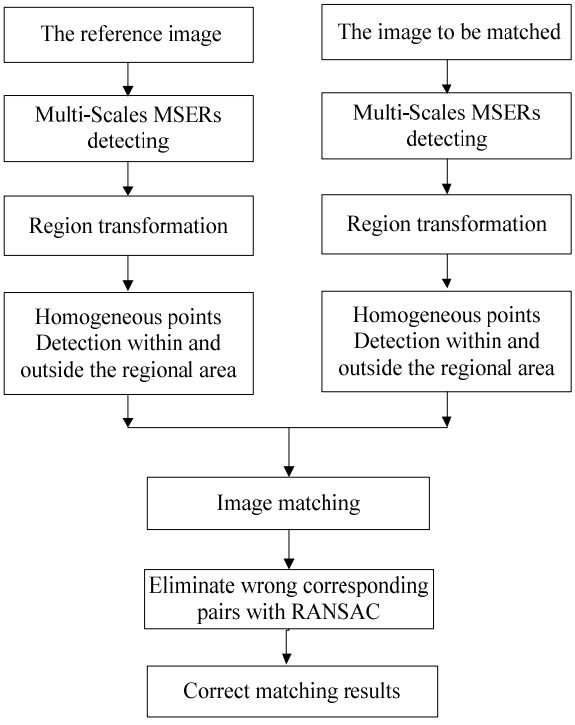


Fig.6 Flowchart of the proposed matching method

4. EXPERIMENTAL RESULTS AND ANALYSIS

4.1 Data sets

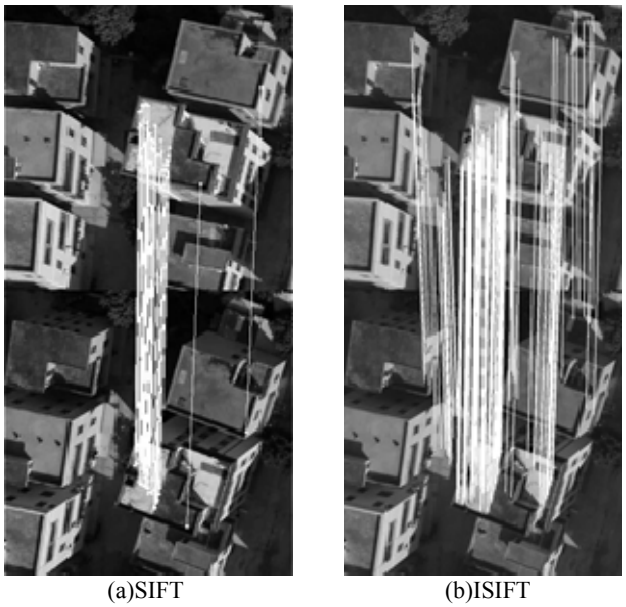
In order to evaluate the effectiveness of the proposed feature matching method, a pair of Unmanned Aerial Vehicle (UAV) images is selected as shown in Fig. 6.

(a) (b)

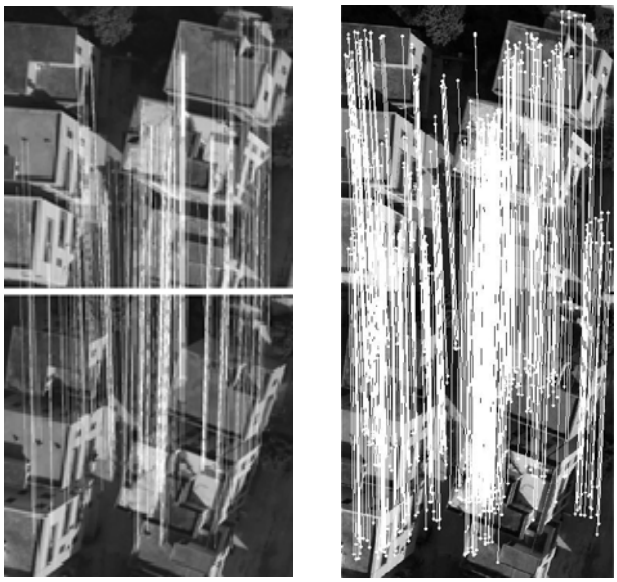
Fig.7 UAV image pair, the image size are 1000×1000 pixels

4.2 Experimental results and analysis

Among all the feature-based matching methods, SIFT and its two improved methods ISIFT and ASFT perform well in dealing with images taken from different viewpoints. Therefore, comparative experiments are conducted using these methods based on the above images. For the four approaches, matching results can be visual displayed in Fig. 8.



(a)SIFT (b)ISIFT



(c)ASIFT (d) Proposed method

Fig.8 Matching results of Fig. 7

Besides, the correct number of matches can be counted and shown in Table 1.

Matching method	Data of Fig.7
SIFT	48
ISIFT	144
ASIFT	274
Proposed method	750

Table 1. Statistical results of SIFT, ISIFT, ASIFT and Proposed method

As can be seen from the results in Table 1:

(1) To all the image pairs shown in Fig.8, the proposed method obtained the largest number of matching features compared to SIFT, ISIFT and ASIFT, while the proposed method got a minimum of correct matching features.

(2) Based on SIFT matching method, the ISIFT method is established by estimating the transformation matrix to generate simulated images, the feature matching is implemented on simulated images, which can obtain more correct features than those of the SIFT matching method.

(3) In Fig.7, the reference image and the image to be matched are real images captured from different perspectives. As the change of scene depth of the image features becomes very large, the entire two images do not obey the same affine transformation model, so transformation model of the ISIFT method obtained by iterative procedure only covers part of the image area, and cannot get a consistent simulated image with the real scene model based on the affine transformation model. Especially in Fig.3, the ISIFT method does not improve the results of the SIFT matching method, and similar to the SIFT method, the correct matching features are very limited.

(4) Compared to the SIFT and the ISIFT methods, the ASIFT and the proposed methods have gained better matching results, which indicate that both the ASIFT and the proposed methods

are more robust than the SIFT and the ISIFT methods for different viewpoint UAV images.

(5)For UAV images, the number of correct matching features acquired by ASIFT algorithm is far less than that of the proposed method.

The reason for this result is as follows: in the UAV images with complex scenes, there is a sampling interval when ASIFT method simulates the affine space. This causes the entire space to be discontinuous, and some of the planar region is not covered completely by the simulated affine space, so ASIFT method obtained fewer features than those of proposed method. The proposed method simulates each regional planar separately, so it can cover planar features better within the image area. These simulated regional planar regions guarantee that more points features will be matched.

(6)Experimental results of UAV images showed that the proposed method can obtain better matching results for different perspective images with undulating terrain.

5. CONCLUSIONS

In this paper a novel matching method on multi-view UAV images is proposed. The proposed method composed of three steps: region detection and extraction; region transformation; feature extraction and image matching. The robustness of local area detection determines the success of the whole image matching process. In the proposed method, the simplified mapping from image view change to scale and rotation change has been derived. Through this processing, the matching problem between view change images can be converted into the matching between rotation and scale changed images. A view change image matching method based on local image region simulation has been put forward. Then, point feature detection and matching are implemented in the simulated image regions. The proposed method simulates each regional planar separately, so it can cover planar features better within the image area. These simulated regional planar regions guarantee that more points features will be matched. Meanwhile, in order to detect the homogeneous points covering the whole image in order to realize the better matching, we also detect the homogeneous points outside the regional area and matching procedure is executed in these areas. Experimental results of UAV images showed that the proposed method can obtain better matching results than those of SIFT, ASIFT and ISIFT.

ACKNOWLEDGEMENTS (OPTIONAL)

This work was supported by National Science & Technology Specific Projects under Grant 2012YQ16018505, 2013BAH42F03, Program for New Century Excellent Talents in University under Grant NCET-12-0426 and the Basic Research Program of Hubei Province (2013CFA024).

REFERENCES

Baumberg A., 2000. Reliable feature matching across widely separated views. IEEE Conference on Computer Vision and Pattern Recognition 2000, 1, pp. 774-781.

Chunsun Z., Clive S. F. ,2007. Automated Registration of High-resolution Satellite Images. The Photogrammetric Record 22(117) , pp.75-87.

Forssen, P. E., and Lowe, D G, 2007. Shape descriptors for maximally stable extremal regions. IEEE 11th International Conference on Computer Vision 2007 (ICCV 2007), pp. 1-8.

Gruen A., 2012. Development and Status of Image Matching in Photogrammetry. The Photogrammetric Record, 27(137) , pp.36-57.

Gruen A., Zhang L. 2002. Automatic DTM generation from three-line scanner (TLS) images. International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences, 34(3/A) , pp.131-137.

Gruen A., Akca D., 2005. Least squares 3D surface and curve matching. ISPRS Journal of Photogrammetry and Remote Sensing, 59(3) , pp.151-174.

Ida J., Clive S. F., 2010.Interest Operators for Feature-based Matching in Close Range Photogrammetry, The Photogrammetric Record 25(129) , pp. 24-41.

Lowe, D. G., 2004. Distinctive image features from scale-invariant keypoints. International journal of computer vision, 60(2), pp. 91-110.

Matas, J., Chum, O., Urban, M., and Pajdla, T, 2004. Robust wide-baseline stereo from maximally stable extremal regions. Image and vision computing, 22(10), pp. 761-767.

Mikolajczyk, K., Schmid, C, 2004. Scale and affine invariant interest point detectors. International journal of computer vision, 60(1), pp. 63-86.

Mikolajczyk, K., Tuytelaars, T., et al., 2005. A comparison of affine region detectors. International journal of computer vision, 65(1-2), pp. 43-72.

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

Tuytelaars, T., and Van Gool, L, 2004. Matching widely separated views based on affine invariant regions. International journal of computer vision, 59(1), pp. 61-85.

Wang R. R., Wang J. J., You H. J. and Ma J. W., 2010. Automatic region registration method based on spatially assistant plane. Journal of Remote Sensing. 14 (3) , pp.1-6.

Yu, G., and Morel, J. M, 2009. A fully affine invariant image comparison method. IEEE International Conference on Acoustics, Speech and Signal Processing 2009 (ICASSP 2009), pp. 1597-1600.

Yu, Y., Huang, K., Chen, W., and Tan, T, 2012. A novel algorithm for view and illumination invariant image matching. IEEE Transactions on Image Processing, 21(1), pp. 229-240.