Automatic Keypoints Extraction from UAV Image with Refine and Improved Scale Invariant Features Transform (RI-SIFT)

Dibs, H., Idrees, M. O.,* Saeidi, V. and Mansor, S.
Geospatial Information Science Research Center (GISRC), Faculty of Engineering, Universiti Putra Malaysia
43400 UPM, Serdang, Selangor, Malaysia, E-mail: dare.idrees@gmail.com*

Abstract
In this study, the performance of Refine and Improved Scale Invariant Features Transform (RI-SIFT) recently developed and patented to automatically extract key points from UAV images was examined. First the RI-SIFT algorithm was used to detect and extract CPs from two overlapping UAV images. To evaluate the performance of RI-SIFT, the original SIFT which employs nearest neighbour (NN) algorithms was used to extract keypoints from the same adjacent UAV images. Finally, the quality of the points extracted with RI-SIFT was evaluated by feeding them into polynomial, adjust, and spline transform mosaicing algorithms to stitch the images. The result indicates that RI-SIFT performed better than SIFT and NN with 271, 1415, and 1557 points extracted respectively. Also, spline transform gives the most accurate mosaicked image with sub-pixel RMSE value of 1.0925 pixels equivalent to 0.10051m, followed by adjust transform with root mean square error (RSME) value of 1.956821 pixel (0.17611m) while polynomial transform produced the least accuracy result.

1. Introduction

Image mosaicing has been in practice long before the age of digital computers. But advances in computer technology encourage the natural drive to develop automated computational techniques to extract keypoints from images for registration and stitching to produce a panoramic image. This development became much more important with the availability of remotely sensed imagery from different sensors and platforms, with varying data acquisition characteristics. Common keypoints extraction algorithms such as the Scale Invariant Feature Transform – SIFT (Brown, 2007 and Lowe, 1999, 2004) exploit universally stable features through an optimized process that consider all the images at the same time to identify and extract tie points. Stable points are recognized on the basis of spatial and radiometric relations among corresponding feature points within overlapping images. The disadvantage of these methods is their sensitivity to noise and background clutters that cause mismatch (Goncalves et al., 2011 and Hasan et al., 2010).

This consequently generates large numbers of false keypoint pairs (otherwise referred to as control points or CPs in this paper) that increases geometric distortion when used for image registration and mosaicing without removing outliers (Dibs et al., 2015a and Goncalves et al., 2011). SIFT is one of the most widely used method to extract feature points in the field of remote sensing because of its ability to handle image with varying orientation and scale (Dibs et al., 2015a, El rube et al., 2011, Goncalves et al., 2011, Lowe, 1999, Meng et al., 2006, Morel and Yu, 2009, Vural et al., 2009 and Zhang et al., 2009). However, using it directly for remote sensing imageries have been reported to produce inaccurate results and sometimes contain incorrect matches caused by false pair of points (El rube et al., 2011, Hasan et al., 2010 and Mukherjee, 2009). Several attempts have been made to improve the accuracy of SIFT extracted CPs (Dellinger et al., 2015, Huang et al., 2004, Mikolajczyk and Schmid, 2002 and Morel and Yu, 2009). Producing panoramic image from data acquired from new generation of sensors (near-equatorial satellite sensors and UAV) is indeed challenging.

The two major challenges are i) the problem of resolving the nonlinear relationship of remote sensing image intensity and ii) geometrically distorted image with SIFT extracted points when they are used as transformation parameters for registration (Dibs et al., 2015a and Sun et al., 2010). Recently, the RI-SIFT algorithm has been developed and patented (Intellectual Property 2015700713/ Objective-2) to automatically extract high-quality points with better accuracy and precision from satellite imagery. Detailed information about the RI-SIFT method can be found

in Dibs et al., (2015b). The method was first experimented with near-equatorial satellite data (Malaysian RazakSAT) to automatically extract CPs for band-to-band image registration. Interestingly, UAV, which is also a remote sensing system, operates at lower altitude (few hundreds of meters above the surface of earth) to generate large quantities of image. Each of these images covers a relatively small area that needs to be stitched together to produce mosaic image. Considering the fact that UAV and satellite system data share certain characteristics motivates examining RI-SIFT algorithm in this study to test its effectiveness, versatility, and reliability to automatically extract accurate and precise points from UAV imagery and to use those points for the purpose of mosaicing.

2. Materials and Methods

High-resolution imagery was acquired using a Multi-rotor UAV (Figure 1a). The vehicle was flown at 327.84m altitude above ground surface. The system used a Canon - PowerShot SX-230 HS (12.1 Megapixel) digital camera to collect sequence of overlapping RGB images. Each image has a ground coverage area of 1.55 km^2, with 60% forward overlap along the flight line. Two adjacent images were selected for this study (Figure 1b).

Figure 1: Multi-rotor UAV (a) assembled with a Canon - PowerShot SX-230 HS digital camera and all other accessories, fully prepared for data collection and ready for take-off. And (b) is the two adjacent UAV images used for testing the performance of RI-SIFT algorithm for keypoints extraction

Dibs et al., (2015b) developed and patented the RI-SIFT, as mentioned earlier, to extract control points from the most invariant features irrespective of change in rotation, scale, and illumination from near-equatorial satellite imagery, The researchers utilized RasakSAT data, covering the Kuala Lumpur–Pekan area in Malaysia. The method efficiently matched CPs with better accuracy than the original SIFT algorithm. Contrary to the prevailing conditions where images exhibit characteristics like similar noise types, stretching, rotation and skewness, UAV image scenes have single RGB image band and are more subjective to varying illumination, orientation and scale due to platform instability and changing flight directions. This makes investigating the capability of the RI-SIFT algorithm for extracting keypoints from images collected using UAV platform worthy.

2.1 Automatic Keypoint Extraction with RI-SIFT

Automatic CPs extraction was accomplished with the RI-SIFT algorithm (Dibs et al., 2015b). First, we developed the RI-SIFT algorithm into a front-end application package designed in MATLAB tool with easy to use graphic user interface (GUI) that allows users to communicate and visualize the result. RI-SIFT technique comprises of two main components: 1) the SIFT algorithm which automatically extract keypoints and 2) the sum of absolute difference (SAD) algorithm (Equation 1) to remove SIFT extracted points that have incorrect matching.

SAD algorithm has been widely used in applications such as computer vision, and object recognition (Dibs et al., 2015b), however in this study, we refined SIFT extracted CPs with this algorithm. RI-SIFT involve three successive processing stages. First, keypoints were extracted automatically with SIFT algorithm and stored as coordinate point database. Next, the generated image coordinates and the two referenced images are input into SAD algorithm. Finally, SAD measures similarity in intensity within the areas around the keypoints of the sensed image and the corresponding points in the reference image (Dibs et al., 2015b). During this final stage, false point pair matches are removed base on threshold enforced by distance and location criteria such that for each descriptor in the first image the corresponding descriptor in the second image is kept if the distance is below the threshold value and if it falls within acceptable neighbourhood location.

$$SAD = \sum\sum |A(x,y) - B(x,y)|$$

Equation 1

Where A and B are blocks, and x and y are the pixel indices of matrices A and B respectively. SAD samples local image intensities around the keypoints as a function of features in the keypoint database. The algorithm's initial task is to compute Euclidean distances between one descriptor and its corresponding point in the second image and the nearest neighbour is chosen. Next, distances to the second and first closest neighbour are compared applying a threshold on the ratio of their respective distances to filter false matches (Dellinger et al., 2015 and Wang et al., 2009). Keypoint with minimum Euclidean distance for the invariant descriptor vector are retained while outliers are rejected (Goncalves et al., 2011). Two feature points are matched when their difference is small. Figure 2 demonstrates how RI-SIFT GUI has been used to load the two overlapping imagery and to input values for both the empirical threshold and the number of keypoints to be extracted.

In this work, threshold value of 200 numbers of points was used to remove false pair of points. Also, we set 1000 as the maximum number of points that can be extracted. It should be noted that this does not imply that 1000 points must be extracted from the operation. Rather, it is a rule enforced to further eliminate the tendencies of including false point matches. Figure 3 shows the matching results of SIFT and RI-SIFT. To measure the reliability of RI-SIFT, the matched keypoint pairs were compared with the original SIFT and the nearest neighbour (NN) (Goncalves et al., 2011, Li et al., 2009 and Lowe, 2004) feature points extraction algorithms.

2.2 Image Mosaicing

In this study, we used three different transformation functions: polynomial (first, second, and third order), spline, and adjust transform to mosaic the two adjacent overlapping UAV images. Details of these methods can be found in (Brown, 2007, Dellinger et al., 2015, Dibs et al., 2015a, Goncalves et al., 2011, Song, 2009, Sun et al., 2010 and Wang et al., 2009). These three methods are traditional mosaicing algorithms that focus on globally consistent alignment through a large scale optimized procedure that takes into account all images at the same time. The decision to use these three transformation algorithms for image mosaicing (and registration) is to test their behaviour with different geometric properties of UAV image. The first order polynomial is a linear transform that can overcome linear distortion in image (Goncalves et al., 2011and Hasan et al., 2010), while second order can resolve non-linear distortion (Dibs et al., 2015b).

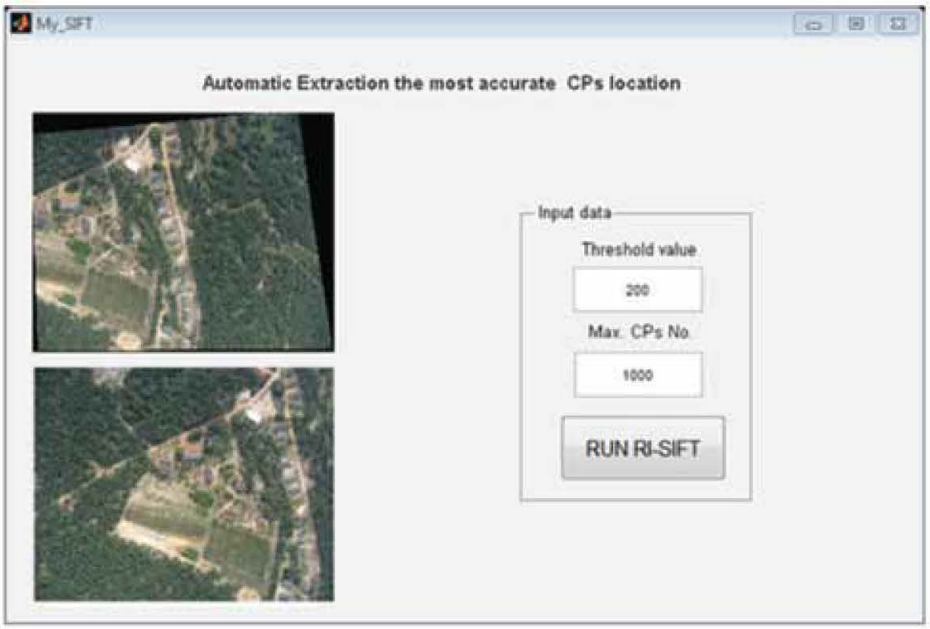

Figure 2: RI-SIFT graphic user interface developed in MATLAB – the interface is interactive, allow user to input the overlapping images, assign threshold and maximum numbers of keypoints and at the same time visualizing the result

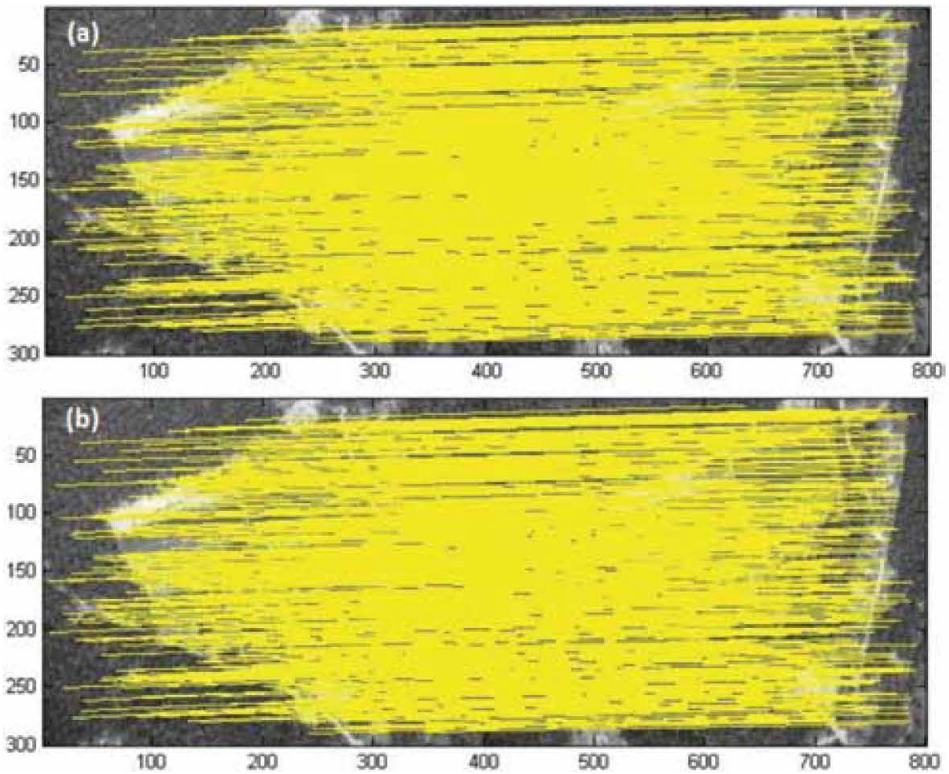

Figure 3: Visualizing matched corresponding points in the adjacent images: (a) is the matched points using the original SIFT algorithm while (b) are those generated using the RI-SIFT algorithm. Note that unlike the statistical report, the size of the viewing window does not visually reveal the differences in the number of matched points

Polynomials of up to third order are commonly used but orders higher than three are seldom used. However, care must be taken when using higher order polynomial because they have unpredictable behaviour, although may perform well with low-frequency images (Richard and Jia, 2006). The choice of the order of the polynomial transformation depends on the trade-off between accuracy and cost of optimisation. Spline, like the polynomial transform, caters for incorrect registration. Although it is widely used for solving non-linear distortion in remotely sensed images, it requires several ground control points (Zagorchev and Goshtasby, 2006). Also, the adjust transform is a spatial adjustment transformation algorithm commonly used in GIS and image processing packages to reposition source image to correspond with a correctly positioned target layer. The transformation process requires translation of origin and scale change to produce planimetrically accurate image alignment. Adjust transforms can be affine, conformal or projective.

3. Results and Discussion

3.1 Keypoints Extraction

Prior to image mosaicing, the first step was to extract and match candidate points in the two images using the RI-SIFT algorithm. Matching of extracted points was implemented using sum of absolute difference (SAD) which employs the intensities of nearest neighbour pixels. At the end RI-SIFT extracted keypoints were input into the transformation function to solve the unknown transformation coefficients for image mosaic and alignment. Finally, we compared RMSEs as performance indicator for all the three transformation methods.

Figure 3 shows the outcome of SIFT and RI-SIFT's corresponding points matching from the two overlapping images as displayed on the designed interface. Although the window did not show a clear distinction in terms of the number of matched points due to technical limitations with the interface design, the matching report produced shows that the original SIFT generates larger number of point matches than the modified RI-SIFT (see Table 1).

Table 1: Comparison of the number of CPs extracted using different methods

CP extraction method	Number of CPs extracted
Nearest neighbor	1557
SIFT	1415
RI-SIFT	271

To refine the keypoints with RI-SIFT, SAD measures similarity in intensity between the reference and sensed images by calculating the absolute differences between each point in the reference image window and the corresponding point in the search window. Then the computed differences are integrated to identify their similarities. To eliminate false pair of points, SAD imposes distance threshold on the locations of SIFT extracted points and points outside the threshold are rejected. Detailed information on how to extract CPs in RI-SIFT can be found in Dibs et al., (2015b). Table 1 shows the number of keypoints obtained from the automatic extraction of keypoints using three different algorithms. It can be seen that NN and SIFT generates large number of points with 1557 and 1415 points respectively. Conversely, the number of extracted keypoints with RI-SIFT were significantly reduced to less than 20% of NN and SIFT. This demonstrates the efficiency and robustness of RI-SIFT for automatic extraction of keypoints. NN performed poorly because rejection of points is based on low contrast by comparing the distance of the closest neighbour of the image patches. Contrast comparison is usually sensitive to noise and to features that arise from background clutter.

This generates poorly localized points that result in many false matches in addition to the correctly matched points leading to high number of extracted points (Dellinger et al., 2015 and Dibs et al., 2015b). SIFT algorithm exploits invariance as a function of dimensionality of the feature space (location, scale, orientation, image intensity). This will, in combination with correctly matched points, multiply the number of other false matches within similar distances. With respect to the results obtained by (Dibs et al. 2015b) where he used RI-SIFT to extract keypoints from multi-spectra image acquired with near-equatorial satellites, it is observed that the algorithm also performs well with UAV. This is in spite of the differences in platform altitude, solar and sensor zenith and azimuth angles, capturing time, illumination, spatial and spectral resolutions.

3.2 Image Mosaicing

Image mosaic fuses several views of remotely sensed images to produce a panoramic image of the scene. The process requires accurate determination of the inter-frame spatial relationships from where transformation parameters will be defined. To test the RI-SIFT method, the high-quality points obtained were utilised for mosaicing using polynomial, adjust, and spline transform in ENVI and ArcMap software. Table 2 shows the RMSE of each mosaic method with varying accuracies across the methods.

Table 2: RMSE values of different mosaic algorithms

Method	RMSE (pixel)	RMSE (meter)
Polynomial – 1st order	28.344	2.607
Polynomial – 2nd order	12.784	1.176
Polynomial – 3rd order	11.316	1.041
Adjust transform	1.957	0.176
Spline transform	1.093	0.101

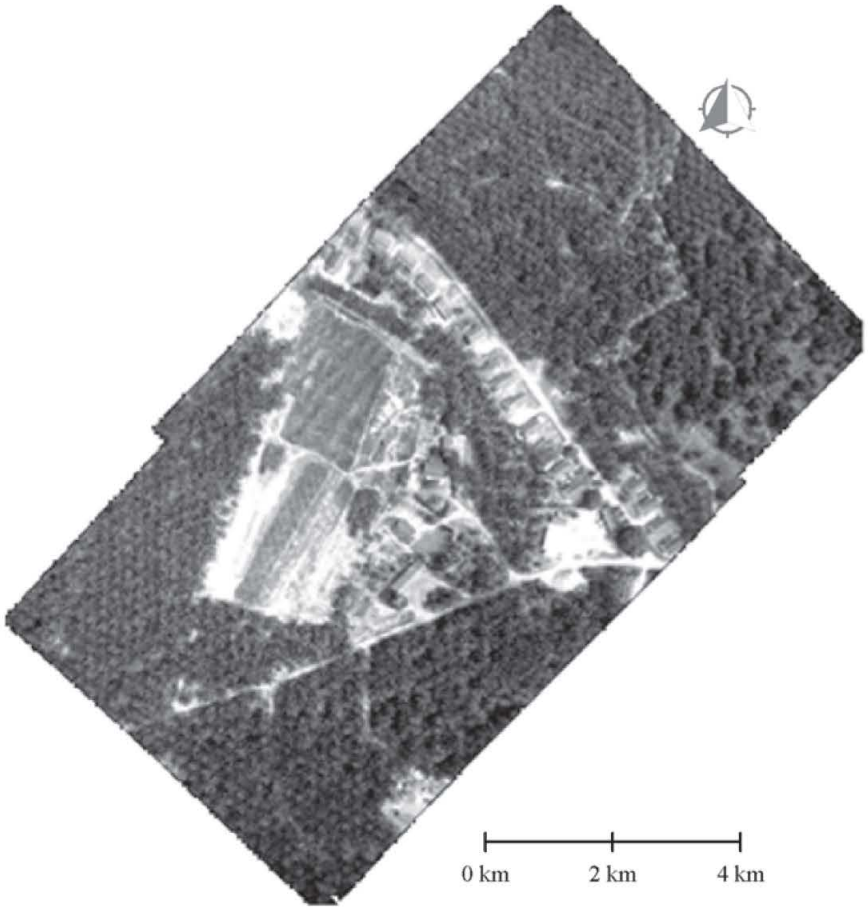

Figure 4: Final mosaic image using spline transform algorithm which performed best among all the tested image stitching approaches

It can be observed that polynomial transform has an accuracy range between 11.32 and 28.34 pixels. Considering the ground resolution of the UAV image with 0.0915098m/pixel, it means that the mosaicing is accurate to within 1.04m and 2.61m. The interpretation of which indicates that the algorithm performed poorly for stitching overlapping UAV image sequence. The other methods, Adjust and Spline transforms, perform much better than polynomial transform. With RMSE value of ~1.96 pixels, adjust transform is accurate to about 0.176m. This result is acceptable (Sun et al., 2010) for land use or land cover map but for large scale map applications such as urban road and building updating, spline transform is more suitable. Among the three methods, Spline performs excellently with sub-pixel RMSE value of 0.10051m. This reflects the highest accuracy achieved for UAV image mosaic. From this study, we discovered that Spline transform gives the best accuracy and were therefore selected to mosaic the two adjacent images and the result is shown in Figure 4. The use of RI-SIFT extracted CPs yields better accuracy of the mosaicked UAV images in contrast to the work of Sun et al., (2010) and Li et al. (2012) that employed SIFT extracted point to mosaic UAV images using Adjacency transfer algorithm and frame-to-frame respectively. The former requires all image frames to be transformed to the same coordinate system which, in the end,

affects the precision and shape of a mosaicked image. Also, the latter needs a careful step-by-step operation that is time consuming and not efficient where large numbers of image frames are involved. RI-SIFT extracts image coordinates directly as CPs and thereby eliminates the need to transform image frames before mosaicing. Figure 4 is the Spline mosaicked output image. It can be observed that the two images correctly aligned, but photometric misalignments induced by changing lighting conditions and exposures can be noticed at the boundary of the overlapping images that indicates lack of proper blend. Future investigation with RI-SIFT will look into (1) extracting keypoints from multiple image sequence and, (2) toner blending of the output mosaicked image.

4. Conclusions

Considering the findings of this study by the application of RI-SIFT to extract accurate and precise keypoints from UAV, it can be concluded that the method performs well. RI-SIFT shows superiority over the original SIFT by streamlining extracted points to less than 20 per cent of the number of points extracted with SIFT algorithm. The implication of this is that RI-SIFT is an efficient and reliable method that can extract keypoints that meets the requirements of mosaicing or registration of UAV images to form a panoramic image. Furthermore, the method has the capability to extract image coordinates as CPs that automatically georeference the mosaicked image.

References

Brown, M., 2007, Automatic Panoramic Image Stitching using Invariant Features. *International Journal of Computer Vision*, 74(1), 59–73. doi:10.1007/s11263-006-002-3.

Dellinger, F., Delon, J., Gousseau, Y., Michel, J. and Tupin, F., 2015, SAR-SIFT : A SIFT- Like Algorithm for SAR Images. *IEEE Transactions on Geoscience and Remote Sensing*, 53(1), 453–466.

Dibs, H., Mansor, S. and Ahmad, N., 2015b, A Novel, Fast and Robust Automatic Method for Refining SIFT Keypoints from Remote Sensing and Computer Science Imagery. *In 2015 IEEE International Conference on Aerospace Electronics and Remote Sensing Technology*. Bali, Indonesia; 3-5 December 2015. 1–7.

Dibs, H., Mansor, S., Ahmad, N. and Pradhan, B., 2015a, Band-to-Band Registration Model for Near-Equatorial Earth Observation Satellite Images with the use of Automatic Control Point Extraction. *International Journal of Remote Sensing*, 36(8), 2184–2200. doi:10.1080/01431161.2015.1034891

El rube, I., Sharkas, M., Salman, A. and Salem, A., 2011, Automatic Selection of Control Points for Remote Sensing Image Registration Based on Multi-Scale SIFT. *2011 International Conference on Signal, Image Processing and Applications with workshop of ICEEA*, 21, 46–50.

Goncalves, H., Corte-Real, L. and Goncalves, J. A., 2011, Automatic Image Registration through Image Segmentation and SIFT. *IEEE Transactions on Geoscience and Remote Sensing*, 49(7), 2589–2600. doi:10.1109/TGRS.2011.2109389

Hasan, M., Jia, X. and Robles-Kelly, A., 2010, Multi-Spectral Remote Sensing Image Registration via Spatial Relationship Analysis on Sift Keypoints. *In Geoscience and Remote Sensing Symposium (IGARSS), 2010 IEEE International*. 1011–1014. Honolulu, 25-30 July 2010: IEEE Explorer. doi:10.1109/IGARSS.2010.5653482

Huang, X. H. X., Sun, Y. S. Y., Metaxas, D., Sauer, F. and Xu, C. X. C., 2004, Hybrid Image Registration based on Configural Matching of Scale-Invariant Salient Region Features. *2004 Conference on Computer Vision and Pattern Recognition Workshop*. doi:10.1109/CVPR.2004.88

Li, Q., Wang, G., Liu, J. and Chen, S., 2009, Robust Scale-Invariant Feature Matching for Remote Sensing Image Registration. *IEEE Geoscience and Remote Sensing Letters*, 6(2), 287–291. doi:10.1109/LGRS.2008.2011751.

Li, M., Li, D., & Fan, D. (2012), A Study on Automatic UAV Image Mosaic Method for Paroxysmal Disaster. ISPRS - International Archives of the Photogrammetry, *Remote Sensing and Spatial Information Sciences*, XXXIX-B6 (September), 123–128. doi:10.5194/isprsarchives-XXXIX-B6-123-2012

Lowe, D. G., 1999, Object Recognition from Local Scale-Invariant Features. *IEEE International Conference on Computer Vision*, (Sept. 1999), 1–8. doi:10.1109/ICCV.1999.790410.

Lowe, D. G., 2004, Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 1–28.

Meng, Y., Tiddeman, B. and Et Al., 2006, Implementing the Scale Invariant Feature Transform (SIFT) Method. Citeseer. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.102.180&rep=rep1&type=pdf

Mikolajczyk, K. and Schmid, C., 2002, An Affine Invariant Interest Point Detector. Computer Vision - ECCV 2002, 2350, 128–142. doi:10.1007/3-540-47969-4_9.

Morel, J. M. and Yu, G., 2009, ASIFT: A New Framework for Fully Affine Invariant Image Comparison. *SIAM Journal on Imaging Sciences*, 2(2), 438–469. doi:10.1137/080732730.

Mukherjee, A., 2009, Interest Points for Hyperspectral Image Data. *IEEE Transaction on Geoscience and Remote Sensing*, 47(3), 748–760. doi:10.1109/TGRS.2008.2011280.

Richards, J. and Jia, X., 2006, *Remote Sensing Digital Image Analysis: An Introduction*. *Springer* (4th ed.). Berlin, Germany: Springer. http://doi.org/10.1007/978-3-642-30062-2

Song, Z., 2009, A New Remote Sensing Image Registration Approach Based on Retrofitted SIFT Algorithm And a Novel Similarity Measure. *IEEE Explore*, 1–5. doi:10.1109/CISE.2009.5366676.

Sun, J., Guo, L., Gong, J., Li, Y. and Li, R., 2010, Study on Automatic Method for UAVRS Images Registration and Mosaic Based on Adjacency Transfer Algorithm. *2010 International Conference on Multimedia Technology*, ICMT 2010, 1–4. doi:10.1109/ICMULT.2010.5631199.

Vural, M. F., Yardimci, Y. and Temizel, A., 2009, Registration of Multispectral Satellite Images with Orientation-Restricted SIFT. *International Geoscience and Remote Sensing Symposium (IGARSS)*, 3, 243–246. doi:10.1109/IGARSS.2009.5417801.

Wang, G. H., Zhang, S. B., Wang, H. Bin, Li, C. H., Tang, X. M., Tian, J. J. and Tian, J., 2009, An Algorithm of Parameters Adaptive Scale-Invariant Feature for High Precision Matching Of Multi-Source Remote Sensing Image. *2009 Joint Urban Remote Sensing Event*. doi:10.1109/URS.2009.5137515.

Zagorchev, L. and Goshtasby, A., 2006, A Comparative Study of Transformation Functions for Nonrigid Image Registration. *IEEE Transactions on Image Processing : A Publication of the IEEE Signal Processing Society*, 15(3), 529–538. doi:10.1109/TIP.2005.863114.

Zhang, Y., Guo, Y., Gu, Y. and Street, W. D., 2009, Robust Feature Matching and Selection Methods for Multisensor Image Registration. In Geoscience and Remote Sensing Symposium, *2009 IEEE International, IGARSS 2009*. III–255 – III–258. Cape Town, 12-17 July 2009: IEEE Explorer. doi:10.1109/IGARSS.2009.5417786