

# MULTI-SPECTRAL REMOTE SENSING IMAGE REGISTRATION VIA SPATIAL RELATIONSHIP ANALYSIS ON SIFT KEYPOINTS

*Mahmudul Hasan<sup>1</sup>, Xiuping Jia<sup>1</sup>, Antonio Robles-Kelly<sup>2,3,1</sup>, Jun Zhou<sup>2,3,1</sup>, Mark R. Pickering<sup>1</sup>*

<sup>1</sup>School of Engineering and Information Technology, University College, The University of New South Wales, Australian Defence Force Academy, Canberra, ACT, Australia.

<sup>2</sup>Canberra Research Laboratory, NICTA, Canberra, ACT, Australia.

<sup>3</sup>College of Engineering and Computer Science, The Australian National University, Canberra, Australia.

## ABSTRACT

Multi-sensor image registration is a challenging task in remote sensing. Considering the fact that multi-sensor devices capture the images at different times, multi-spectral image registration is necessary for data fusion of the images. Several conventional methods for image registration suffer from poor performance due to their sensitivity to scale and intensity variation. The scale invariant feature transform (SIFT) is widely used for image registration and object recognition to address these problems. However, directly applying SIFT to remote sensing image registration often results in a very large number of feature points or keypoints but a small number of matching points with a high false alarm rate. We argue that this is due to the fact that spatial information is not considered during the SIFT-based matching process. This paper proposes a method to improve SIFT-based matching by taking advantage of neighborhood information. The proposed method generates more correct matching points as the relative structure in different remote sensing images are almost static.

*Index Terms*— Image registration, local weighted mean, SIFT.

## 1. INTRODUCTION

When image data is recorded by sensors on satellites and aircrafts it can contain errors in geometry which can be caused by the rotation of the earth during image acquisition, the finite scan rate of some sensors, the wide field of view of some sensors, the curvature of the earth, sensor non-idealities, variation in platform altitude, altitude and velocity, and panoramic effects related to imaging geometry, et al [1]. These geometrically distorted images can rarely be modeled by linear transformations such as shifting, rotation or affine. Registering two images taken at different times and with different sensors containing geometric errors often requires a non-linear transformation model. Sometimes due to local factors, even polynomials of

degree four [2] do not properly model the transformation between the images.

Rather than using one global mapping function to register the whole images, Goshtasby [3] used a number of local mapping functions, each tuned to map well in local neighborhoods. First the images are divided into triangular regions by triangulating the control points and then a linear mapping function is obtained by registering each pair of corresponding triangular regions in the images. Goshtasby [4] also proposed a local weighted mean (LWM) method when the distortion varies locally and a piecewise linear model is not sufficient. When spacing between the control points varies greatly, the piecewise linear method is found to produce a more accurate registration. However, when a very large set of control points is available and the control points contain positional inaccuracies, locally weighted mean method performs better [5]. Both these methods require a large number of control points to register the images accurately.

SIFT [6, 7] is a popular choice for automatic control point generation due to its scale and rotation invariance properties. Commonly the SIFT method is known to generate matching feature points distributed over a full range of positions and scales of the images. Multi-spectral image registration involves a great deal of illumination variation and might contain similar objects across the image which results in a very few number of correctly matched control points. Lowe [6] did not consider the neighborhood relationship of feature points in the spatial space of the images as the method was targeted at object retrieval where the locations, poses, and spatial relations of the objects to be retrieved can be quite different in two images. On the contrary, in the case of remote sensing image registration, we assume in most cases that the spatial relationship of the objects within an image does not experience a significant change within another image subject to local affine distortions. For example, Li [8], Yi [9], and Vural [10] suggested modifications to SIFT for better matching accuracy by imposing scale and orientation restrictions. What makes the spatial relationship more important is that

similar feature descriptors may be found in many locations, such as from buildings with similar shapes, which is common in remote sensing images. Thus, imposing a location restriction in feature point matching is the underlying principle of the technique we propose in this paper.

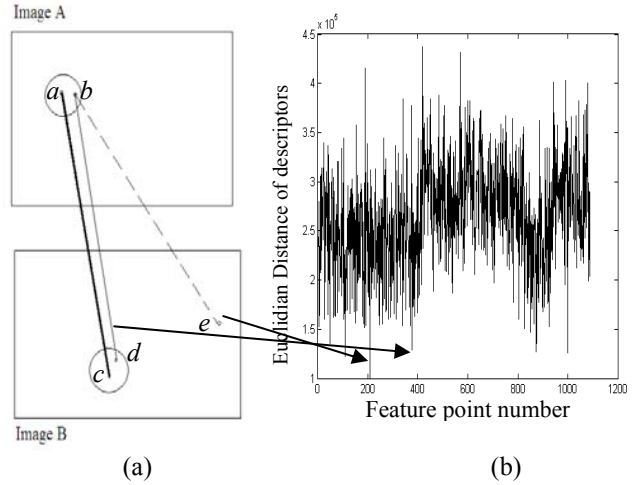
## 2. MATCHING FEATURE POINTS USING STRUCTURAL INFORMATION

The standard way of applying SIFT to image registration is as follows. The SIFT algorithm at first detects a set of feature points or keypoints in scale-space [11] by applying a difference-of-Gaussian (DoG) filter to a pyramid of Gaussian smoothed and resized images. Feature points with low contrast and located at edges are discarded. Then a 128-element feature descriptor is generated for each feature point using statistics of the gradient directions which are scale and rotation invariant. These descriptors are used to find matching points by calculating the ratio of the Euclidian distance between every feature point in the images to be registered. These matched feature points are used to determine the parameters of the transformation model between the images.

As we mentioned above, in multi-sensor remote sensing images, the spatial relationship between objects remains approximately the same. Thus, if we can find the matching position of a feature point, we can predict the matching position of the neighboring feature points. Note that each feature point is associated with a scale and an orientation via SIFT, so from a pair of matched feature points the scale difference for surrounding points can be predicted.

To illustrate the issue, we show an example in Figure 1. SIFT matching is applied to images *A* and *B*. The bold line shows a pair (*a*, *c*) of matched featured points in the two images. The dotted line shows the best match *e* of another feature point *b* in image *A*, while the correct match should be point *d*. In the proposed technique, *e* is not selected as a matched feature point for *b* because the spatial distance between points *c* and *e* is too large. Feature points *a* and *c* are matched while the counterpart for neighboring feature point *b* cannot be decided because the SIFT descriptors for points *d* and *e* are almost equally different from the SIFT descriptor of point *b*. This problem is made worse by the fact that a lot of similar descriptors can be found in typical remote sensing images. The idea to solve this problem is as follows. Still considering the example in Figure 1, assume points *a* and *c* are already matched with high confidence that the match is correct. We can predict that the feature points around *a* (shown in the circular window) can be found around *c*. So, for point *b* we only search the neighborhood of point *c* for a matching descriptor, which results in a correct match at point *d*. This process is iterated to recover more matching feature points and hence a more

accurate registration. RANSAC algorithm [12] is applied to eliminate the false matching points.



**Figure 1:** (a) Feature point matching by SIFT results *a* to *c* points as correct match but no match found for *b* point. (b) Euclidian distance of the descriptors of all the feature point on image B for the feature point *b* on image A.

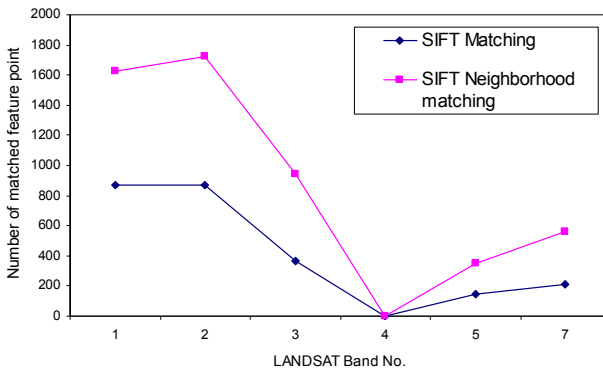
To determine the piecewise linear transformation function between the image pairs, first triangulation of control points are performed in one image which will automatically obtain corresponding triangles in the other image. Normally Dirichlet tessellation is used for triangulation. Then for each pair of triangles corresponding mapping functions are determined to register them. Boundary triangle planes are extended to determine mapping outside triangulated areas. The overall mapping function is then obtained by piecing together the linear mapping functions. Local weighted mean uses information about local control points only by forming polynomials to register local areas in the image. The algorithm calculates the radius of influence of the polynomial as the distance from the center control point to the farthest point used to infer the polynomial. The transformation of an arbitrary point is determined by the weighted mean of all polynomial passing over the point inside the radius of influence.

So, the automatic image registration algorithm can be summarized as follows:

1. Detect multi-scale feature points and their descriptors from both the target and reference images using the SIFT algorithm.
2. Find matched feature points in the target image for feature points in the reference image by rejecting all matches in which the Euclidean distance ratio of the descriptors is greater than 0.8 as described in [6].
3. To remove the false matches, apply the RANSAC method for the global affine transformation model on the matched feature points with a  $T_{RANSAC}$  pixel error

threshold. The resultant feature points are called primary matched feature points.

4. For each of the primary matched feature points make lists of their neighboring feature points on each image separately within a window of  $W$  pixels.
5. Instead of matching all the feature points in the target image with all the feature points in the reference image (similar to step 2), match between the two corresponding lists (obtained in step 4) of each primary matched feature points by rejecting all matches in which the Euclidean distance ratio of the descriptors is greater than 0.9.
6. As we assumed that the distortions will not cause a very big displacement of the relative positions of the feature points, eliminate the matching points that are too far away from their global affine position. All matched feature points having more than  $T_{affine}$  pixels positional difference from the position determined if global affine transformation is applied, are eliminated. The refined matched feature points are called secondary matched feature points.
7. Primary and secondary matched feature points are collectively selected as control points for registering the image pair. Calculate the piecewise linear (affine) or local weighted mean transformation function using the control points and apply the transformation function on the target image to construct the transformed image aligned with the reference image.

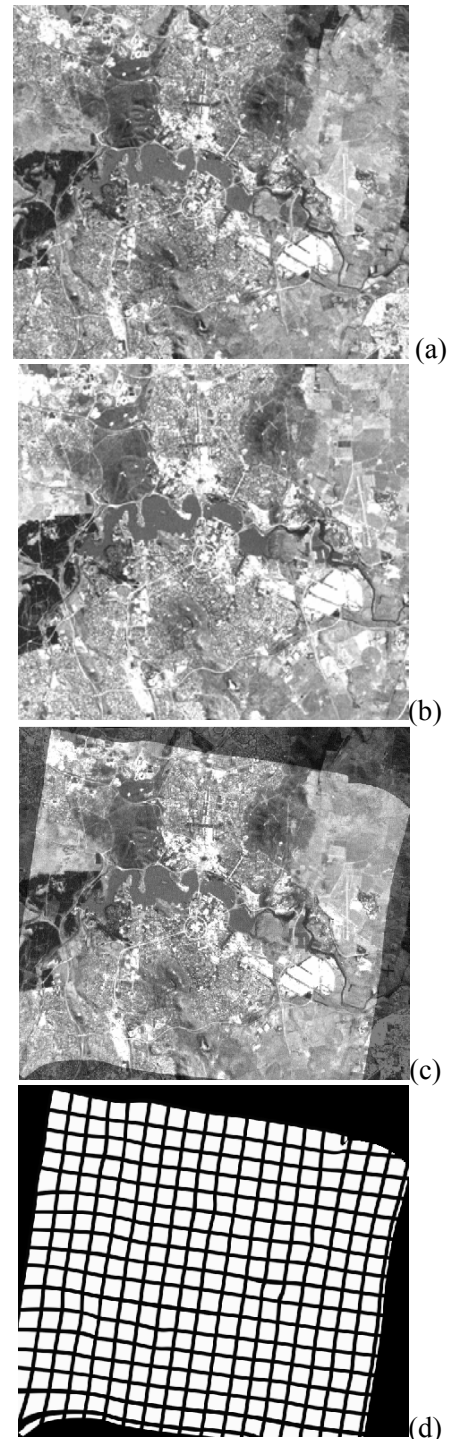


**Figure 2:** Number of feature points correctly matched in image registration between band 1 of LANDSAT image taken in 2001 and band 1, 2, 3, 4, 5 & 7 of LANDSAT image taken in 2000.

### 3. EXPERIMENTAL RESULTS

To evaluate the performance of the proposed technique we conducted an experiment using band 1 of a LANDSAT image taken in 2001 and bands 1, 2, 3, 4, 5 and 7 of a LANDSAT image taken in 2000 as shown in Figure 3(a) and 3(b) (band 1 only) respectively. Both SIFT and our

method (SIFT Neighborhood matching) were applied to find matching feature points.



**Figure 3:** (a) reference image, (b) target image, (c) resultant image after local weighted mean transformation registration using the proposed method, and (d) local distortions corrected in registration shown in grid representation.

The year 2001 band 1 image produced 9901 feature points and the year 2000 data produced 7383, 8240, 8087, 8337, 7870 and 7186 feature points for bands 1, 2, 3, 4, 5, and 7 respectively. According to the algorithm described in the previous section, the RANSAC method was then applied to remove the false matches and these matching control points are used to calculate the local weighted mean (LWM) transformation function. LWM transformation function is selected over piecewise affine transformation function because the matching control points are expected to have some positional inaccuracy which suits LWM transformation function. The resultant LWM transformation function is used to transform the target image to match the reference image. In this implementation we used TRANSAC=10 pixels error threshold for RANSAC method and Taffine=15 pixels in step 6 for feature points refinement to allow at most 15 pixels local distortion. The neighborhood window used in step 4 of the proposed algorithm was a radius of  $W=60$  pixels in this implementation.

Figure 2 shows the numbers of feature points correctly matched using the two algorithms. Both the methods were unable to register band 1 of the LANDSAT image taken in 2001 with band 4 of the LANDSAT image taken in 2000 because band 1 is captured at visible blue and band 4 is captured at near infra-red channel which caused a large variation in pixel intensities between the images. Although the number of matched feature point is large for band 1 and band 2 for SIFT matching, other bands produce a poor number of matches. SIFT produced less than 11% matching feature points of the total feature points, while the proposed neighborhood matching algorithm improved the matching performance to almost twice this amount in most cases.

Figure 3(a) shows the band 1 of LANDSAT image taken in 2001 around Canberra (image size 659×593), which is considered as reference image. Figure 3(b) shows the band 1 of LANDSAT image taken in 2000 at the same area with local distortion (image size 503×454), which is considered as target image. Figure 3(c) shows the results of registration using the proposed method. Figure 3(d) shows grid line images to illustrate the local distortions as well.

#### 4. CONCLUSION

We have proposed a new SIFT neighborhood matching method. This method increases the number of matched control points using SIFT based automatic image registration when combined with a local weighted mean (LWM) transformation model. The piecewise linear or local weighted mean transformation model can register images more accurately than global affine model when images contain local distortions. Experimental result shows example of SIFT neighborhood matching method along with local weighted mean transformation model in multi-spectral image registration.

#### 5. REFERENCES

- [1] J. A. Richards, and X. Jia, *Remote Sensing Digital Image Analysis*, 4th ed., Verlag: Springer 2006.
- [2] D. G. Leckie, "Use of polynomial transformations for registration of airborne digital line scan images," *Proceedings of the 6th Canadian Symposium on Remote Sensing, Halifax*, pp. 635-641, May 21-23, 1981.
- [3] A. Goshtasby, "Piecewise linear mapping functions for image registration," *Pattern Recognition*, vol. 19, no. 6, pp. 459-466, 1986.
- [4] A. Goshtasby, "Image registration by local approximation methods," *Image and Vision Computing*, vol. 6, no. 4, pp. 255-261, 1988.
- [5] L. Zagorchev, and A. Goshtasby, "A comparative study of transformation functions for nonrigid image registration," *IEEE Transactions on Image Processing*, vol. 15, pp. 529-538, 2006.
- [6] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91-110, 2004.
- [7] P. Schwind, S. Suri, P. Reinartz, and A. Siebert, "Applicability of the SIFT operator to geometric SAR image registration," *International Journal of Remote Sensing*, vol. 31, no. 6, 1959-1980, 2010.
- [8] Q. Li, G. Wang, J. Liu and S. Chen, "Robust Scale-Invariant Feature Matching for Remote Sensing Image Registration," *IEEE Geoscience and Remote Sensing Letters*, vol. 6, no. 2, pp. 287-291, 2009.
- [9] Z. Yi, C. Zhiguo, and X. Yang, "Multi-spectral remote image registration based on SIFT," *Electronics Letters*, vol. 44, no. 2, pp. 107-108, 2008.
- [10] M. F. Vural, Y. Yardimci, and A. Temizel, "Registration of Multispectral Satellite Images With Orientation-Restricted SIFT," *IEEE International Geoscience & Remote Sensing Symposium (IGARSS)*, 2009.
- [11] T. Lindeberg, *Scale-space theory in computer vision*, Kluwer Academic Publishers, 1994.
- [12] M. A. Fischler, and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381-395, 1981.