A Comparative Study of a Novel Coarse to Fine Automatic Image Registration using MS-SIFT & SIFT

Sneha Anne Jacob, Soumya Sara Koshy

Abstract—Automatic image registration is a challenging task, especially for remote sensing images. Image registration is a process for finding the precise match between two images of the same scene, taken at the same or different times, using same or different sensors, and from the same or different viewpoints. It is very important to have a registration approach which is fast, accurate, and robust in nature. For this purpose, a novel method for automatic image registration is required. This method consists of a coarse registration step and a fine-tuning step. To begin with, coarse registration step is implemented by the mode-seeking scale-invariant feature transform (MS-SIFT) .The method presented in this paper exploits the fact that each SIFT feature is associated with a scale, orientation, and position to perform mode seeking to remove outlier keypoints inorder to enhance the registration results, hence the name Mode Seeking SIFT (MS-SIFT) .

The comparative study includes the comparison of MS-SIFT and SIFT using mutual information(MI) registration results in terms of the average execution time and RMSE value.

Index Terms—Automatic Image Registration, Mode, MS-SIFT, Mutual Information(MI), SIFT.

---

1 Introduction

THE process of geometrical matching of two or more images of the same scene taken at different times, from different viewpoints, and/or by different sensors is called image registration. The two such images are called the reference and sensed images. Image registration is a vital step in all image analysis tasks and registration is required in remote sensing for multispectral classification, environmental monitoring, change detection, image mosaicing, weather forecasting, creating super-resolution images, integrating information into geographic information systems (GIS)), in medical image processing, in cartography, in computer vision etc. [1]

Automatic image registration is still a challenge due to the presence of difficulties within the remote sensing field. The difficulties such as both geometric deformations (translation effect, rotation and scale distortion, occlusion, and viewpoint difference) and radiometric discrepancies (illumination change and sensor and spectral content difference) are very common in remote sensing. So inorder to improve the performance of the existing registration methods,further research studies are required.

Image registration methods can be broadly classified into two categories: intensity- and feature-based methods [1], [5]. Feature-based methods first extract salient features and then match them using similarity measures to establish the geometric correspondence between two images. One of the main advantages of these approaches is that they are fast and robust to noises, complex geometric distortions, and significant radiometric differences. The commonly used features include point, edge, contour, and region, and the well-known feature matching methods include invariant descriptor, spatial relation,and relaxation methods [1]. The scale-invariant feature transform(SIFT) is capable of extracting distinctive invariant features from images, and it can be applied to perform reliable matching across a substantial range of affine distortion, change in 3-Dviewpoint, addition of noise, and change in illumination [2].But there exist some problems when it is directly applied to remote sensing images, such as ,the number of the detected feature matches may be small,and their distribution may be uneven due to the complex content nature of remote sensing images [6].

In this paper we study about an efficient automatic image registration method based on the scale-invariant feature transform (SIFT) [2] equipped with a mode seeking process [3] and the comparison of this method with the SIFT based method using MI [4].

MS-SIFT, as in [3] performs reliable filtering of outlying feature correspondences (keypoints) by mode seeking of scale ratios, orientation differences, horizontal and vertical differences. The inherent information of each SIFT key point i.e., scale, orientation, and position is used to compute a prospective transformation for each match (i.e., corresponding key points).In principle, we perform mode seeking in 4-D space, which is done in practice for each of the four components (scale, rotation, and vertical and horizontal translations)separately. This is followed by effective removal of outlying correspondences and a refined computation of the transformation.

Modes are used as in [3], because they are accurate and for a variety of multitemporal and multispectral images, the histogram modes are unique and evident (at least 40% higher than the next peak). Moreover, a mode of a distribution can be es-

ISSN 2229-5518

timated even when there exist a large number of outliers [3].

2 METHODOLOGY

Image registration is defined in [4] as follows. Given a pair of 2-D gray-level images between which there exist some geometric and radiometric differences. Let $f_R(x, y)$ and $f_S(x, y)$ represent the reference and sensed images, respectively, where coordinates $(x, y) \in \Delta \subset R^2$ and Δ is a region of interest. To register these two images is to find the optimal geometric transformation $T_\mu(\cdot)$ by which $f_S(T_\mu(x, y))$ best matches $f_R(x, y)$ for all (x, y), where $\boldsymbol{\mu}$ is a set of transformation parameters. Here, we select the affine transformation model, which is widely used in the registration of remote sensing images, and it can be written as:

$$T_{\boldsymbol{\mu}}(x, y) = \begin{bmatrix} a_{11} & a_{12} & \delta_x \\ a_{21} & a_{22} & \delta_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

where the transformation origin is considered to be the upper left corner of the reference image, $(a_{11}, a_{12}, a_{21}, a_{22})$ represent the rotation, scale, and shear differences, and $(\delta x, \delta y)$ are the shifts between the two images.

The methodology was implemented using MATLAB (2012).The algorithm uses the input images, reference image and sensed image. It consists of 2 steps—coarse registration(preregistration) and fine-tuning step. The preregistration consists of SIFT equipped with an outlier removal method and fine-tuning step includes the computation of required transformation/transformation parameters inorder to obtain a precise match between the reference and sensed images such that the images are geometrically aligned so as to obtain accurate registration results.

2.1 Automatic Image Registration using SIFT

***1)* SIFT Matching**

The preregistration process begins with the SIFT matching, which contains five steps: scale-space extrema detection, keypoint localization,orientation assignment, keypoint descriptor, and keypoint matching [2].

***2)* Outlier Removal**

First forma scale histogram like in [4].The denser cluster in the scale histogram corresponds to the true scale difference between the images. The keypoint pairs that contribute to the cluster are the correct matches, while the ones that are scattered and away from the cluster are considered as incorrect matches and they are eliminated. The outlier removal process is performed in an iterative fashion: discard the most likely mismatches first, and then compute the rmse based on the remaining matches; the iteration stops when the rmse is below a certain threshold or the maximum number of iterations is achieved [4].The coarse results thus obtained provide an excellent initial solution for the subsequent fine-tuning process.

3) **Maximization of MI**

From the definition of MI in [4] , it is shown that the geometric correction parameter μ is the optimal solution when the MI value is maximal. Thus the problem of image registration is mapped as an optimization problem, which can be expressed as in [4]:

$$\mu^* = \arg\left(opt\left(S(\mu)\right)\right)$$

where S is the MI defined previously and μ* is a set of the optimal transformation parameters corresponding to the maximum of MI. With the parameter μ* the transformed sensed image $f_S(T_{\mu^*}(x, y))$ is correctly aligned with the reference image $f_R(x, y)$.The multi-resolution framework works iteratively from the coarsest level of the image pyramid to the finest level of the image pyramid. For all cases, the MI between the whole overlap of subband images of the reference and sensed images is computed at each level and maximized successively, and the search is performed on an interval around the optimal transformation parameters found at the previous level and is refined at the next level as in [4].

2.2 Automatic Image Registration using MS-SIFT

The first step is same as the first step of previously illustrated method except for keypoint matching where it is done using mode seeking as follows[3]:

1) Find for each keypoint in the reference image its nearest neighbor in a Euclidean distance sense in the sensed image. Let us denote the set of the resulting correspondences by:

$$\{(x_n, y_n) \leftrightarrow (x_n{}', y_n{}')\}_{n=1,2,\ldots N}$$

where, $(x_n; y_n)$ and $(x_n'; y_n')$ are spatial locations of the SIFT keypoints in the reference and transformed images,respectively.

2) The next step is to form histograms of scale ratios and orientation differences between the correspondence pairs found in the previous step.

3) Find the maximum value of each histogram and compute the corresponding modes s_{mode} and $\Delta\Theta_{mode}$ by a weighted average of the maximum value and its two adjacent bins (i.e., the bins to its left and right).

4) These modes are used to rotate and scale the position differences, in both the Xand Y directions, between nearest neighbor pairs as follows [3]:

$$\Delta x = x - s_{mode}(x' cos(\Delta\Theta_{mode}) - y' sin(\Delta\Theta_{mode}))$$
$$\Delta y = y - s_{mode}(x' sin(\Delta\Theta_{mode}) + y' cos(\Delta\Theta_{mode}))$$

Modified Outlier Removal

The outlier removal is performed in [3] as follows:

1) Compute the histograms of the differences Δx,Δy and find their modes and denote it as by Δx_{mode} and Δy_{mode} , respectively.

2) Obtain the quadruple $(s_{mode}, \Delta\Theta_{mode}, \Delta x_{mode}, \Delta y_{mode})^T$ and filter outliers with respect to the initial correspondences.

3) For that define as in [3] the following two logical filters as:

ISSN 2229-5518

$$F_1 = |\Delta x - \Delta x_{mode}| \geq \Delta x_{thresh}$$
$$F_2 = |\Delta y - \Delta y_{mode}| \geq \Delta y_{thresh}$$

where Δx_{thresh} and Δy_{thresh} denote, respectively, predefined thresholds of horizontal and vertical differences, in terms of corresponding histogram bin widths (measured in pixels).

This outlier filter will reject all correspondences for which F1 or F2 holds. All remaining correspondences are considered inliers.

Similarity Transformation

The next step is to compute the similarity transformation resulting from the above said correspondences by a one step OLS as in [7].This is done by first computing the transformation that aligns the centroids of the (remaining) point sets, then computing the scale factor that aligns their spatial variances, and finally computing the rotation that minimizes the sum of squared distances [3].

3 RESULTS & DISCUSSION

The experimental study and comparison will be applied on a number of remote sensed images and graphs are plotted for a selected number of five images. The parameters that are under the comparative study are the average execution time and root mean square measure (RMSE).

After applying the automatic image registration process using SIFT and MS-SIFT individually on each case ultimately the root mean square between the correct matches and the matches removed after outlier removal method is estimated and also the average time needed to evaluate these results. It is found that the average execution time drastically reduces while using MS-SIFT, thus making it a faster method and the error measure(RMSE) also decreases making MS-SIFT more accurate method than SIFT. The sample images used are as follows:

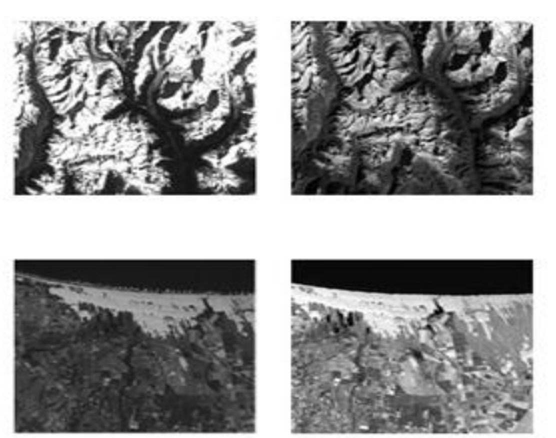

Fig.1. Sample Input Images-Reference image1 and sensed image1 (1st row); Reference image2 and sensed image2 (2nd row)

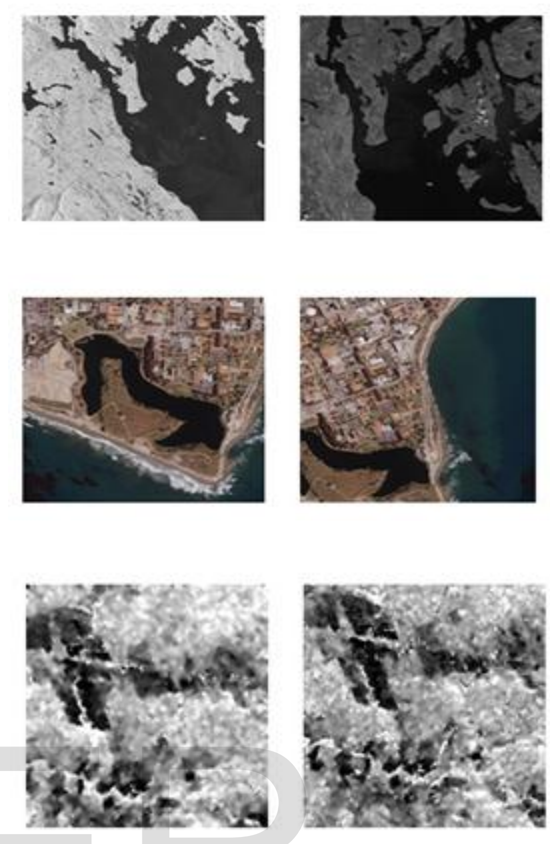

Fig.2. Sample Input Images-Reference image3 and sensed image3 (1st row); Reference image4 and sensed image4 (2nd row);Reference image5 and sensed image5 (3rd row)

The table below shows the average execution time:

Image	Time taken by SIFT	Time taken by MS-SIFT
1	41.78	18.87
2	29.41	16.11
3	50.48	23.86
4	28.04	15.46
5	24.73	15.25

Table 1: Average Execution time taken for SIFT and MS-SIFT

The table below shows the error measure:

Image	SIFT	MS-SIFT
1	3.28	1.78
2	3.06	1.98
3	1.52	1.27
4	6.26	3.94
5	0.39	0.33

Table 2: RMSE value for SIFT and MS-SIFT

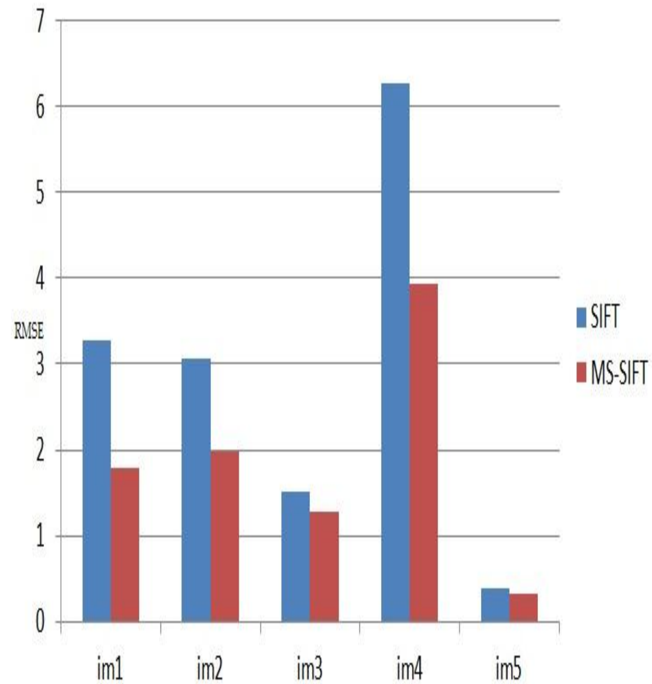

Fig.4. Error measure between SIFT and MS-SIFT

Fig.3 Average time ratio between SIFT and MS-SIFT

Conclusion

This paper aims at comparing result of registration methods using SIFT and MS-SIFT. The results has shown that by using MS_SIFT it is a very good registration method with acceptable accuracy compared to other methods and consumes less time for execution.Thus a simple, fast and accurate registration method is obtained. This can be further extended by multi-mode seeking method which is the futurework.

REFERENCES

[1] B. Zitová and J. Flusser, "Image registration methods: A survey," *Image Vis. Comput.*, vol. 21, no. 11, pp. 977–1000, Oct. 2003.

[2] D.Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis,*vol 60, no. 2 pp.91-110, Nov. 2004.

[3] Benny Kupfer, Nathan S. Netanyahu and IlanShimshoni, "An Efficient SIFT-Based Mode-Seeking Algorithm for Sub-Pixel Registration of Remotely Sensed Images," *IEEE Geoscience and Remote Sensing Letters,*vol.12, no.2, Feb 2015.

[4] Maoguo Gong, Shengmeng Zhao, Licheng Jiao, DayongTian, and Shuang Wang, "A Novel Coarse-to-Fine Scheme for Automatic Image Registration Based on SIFT and Mutual Information",*IEEE Transactions on Geoscience And Remote Sensing,* vol. 52, no. 7, July 2014

[5] L. G. Brown, "A survey of image registration techniques," *ACM Comput. Surv.*, vol. 24, no. 4, pp. 325–376, Dec. 1992.

[6] A. Sedaghat, M. Mokhtarzade, and H. Ebadi, "Uniform robust scale invariant feature matching for optical remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4516–4527, Nov. 2011.

[7] D. M. Mount, N. S. Netanyahu, and S. Ratanasanya, "New approaches to robust, point-based image registration," in *Image Registration for Remote Sensing*, J. LeMoigne, N. S. Netanyahu, and R. D. Eastman, ds. Cambridge, U.K.: Cambridge Univ. Press, Mar. 2011.