

## Image Registration Using Scale Invariant Feature Transform

N.NagaRaju<sup>1</sup>, T.Satya savitri<sup>2</sup>, Ch.A.Swamy<sup>3</sup>

<sup>1</sup>Department of Electronics and Communication Engineering, Institute of Aeronautical Engineering, Dundigal, Hyderabad, Andhra Pradesh, India.

<sup>2</sup>Department of Electronics and Communication Engineering, Jawaharlal Nehru Technological University, Hyderabad, Andhra Pradesh, India.

<sup>3</sup>Department of Electronics and Communication Engineering, Marri Laxman Reddy Institute of Technology and Management, Dundigal, Hyderabad, Andhra Pradesh, India.

[nagaraj.neela@gmail.com](mailto:nagaraj.neela@gmail.com), [tirumalasadya@rediff.com](mailto:tirumalasadya@rediff.com), [chaswamy@gmail.com](mailto:chaswamy@gmail.com)

*Abstract- Image registration is a key operation to spatially align two or more images for comparing the difference between them or exploiting complementary information from those images. The main objective regarding automatic registration of satellite images is to obtain an accurate set of tie points and then apply the transformation function which is most suitable to the pair of images to be registered. An efficient method for AIR (Automatic Image Registration) is proposed, which combines image segmentation and scale invariant feature transform (SIFT), complemented by an outlier removal stage. The reference and unregistered images may differ in translation, rotation, and scale and may present distortions. The performance of this method is evaluated through measures such as Nred, RMSall, RMSloo, Pquad, Bad Point Proportion: BPP(r), Scat, Cost Function ( $\phi$ ).*

**Keywords— Automatic Image Registration (AIR), Image Segmentation, Optical Images, Scale Invariant Feature Transform (SIFT), Principle Component Analysis (PCA)**

### I. INTRODUCTION

The main concept regarding automatic registration of satellite images is to obtain an accurate set of tie points and then apply the transformation function which is most suitable to the pair of images to be registered. A considerably large number of approaches may be found in the literature regarding the automatic obtention of tie points being mainly area- or feature-based methods by means of the image intensity values in their close neighbourhoods, the feature spatial distribution, or the feature symbolic description.

Image registration is still far from being a commonly automatized process, in particular regarding remote sensing applications. Although several methods have been proposed in the last few years geometric correction of satellite images is, in practice, mostly a manual work. The manual procedure is associated to inter and intra-operative subjectives beyond being a time consuming task.

Automatic image registration (AIR) is still a present challenge for the remote sensing community. Although a wide variety of AIR methods have been proposed in the last few years, there are several drawbacks which avoid their common use in practice. The recently proposed scale invariant feature transform (SIFT) approach has already revealed to be a powerful tool for the obtention of tie points in general image processing tasks, but it has a limited performance when directly applied to remote sensing images. In this project, a new AIR method is proposed based on the combination of image segmentation and SIFT, and Principle component Analyses (PCA) complemented by a robust procedure of outlier removal.

This combination allows for an accurate obtention of tie points for a pair of remote sensing images, being a powerful scheme for AIR. Both synthetic and real data have been considered in this work for the evaluation of the proposed methodology, comprising medium and high spatial resolution images, and single band, multispectral, and hyper spectral images. A set of measures which allow for an objective evaluation of the geometric correction process quality has been used. The proposed methodology allows for a fully automatic registration of pairs of remote sensing images, leading to a subpixel accuracy for the whole considered data set. Furthermore, it is able to account for differences in spectral content, rotation, scale, translation, different viewpoint, and change in illumination.

Image segmentation comprises a wide variety of methods, either for monochrome or color images (or to a single or multiple bands of satellite images). Most image segmentation methods can be classified according to their nature: histogram thresholding, feature space clustering, region-based approaches, edge detection approaches, fuzzy approaches, neural networks, physics-based approaches, and any combination of these. Any of these generally intends to transform any image to a binary image: objects and background.

In this paper, a robust and efficient method for AIR is proposed which combines image segmentation and SIFT,

complemented by an outlier removal stage. The reference and unregistered images may differ in translation, rotation, and scale and may present distortions associated to the terrain. The methodology is described in Section II, the results of its application to image are illustrated and discussion is presented in Section III, conclusion in Section IV.

## II. PROPOSED METHODOLOGY

Let us consider  $(X_{REF}, Y_{REF})$  as the coordinates of a point from the reference image and  $(X_{NEW}, Y_{NEW})$  as (pixel, line) of the corresponding point in the new image to be registered. The relation between  $(X_{REF}, Y_{REF})$  and  $(X_{NEW}, Y_{NEW})$  may be written as

$$\begin{aligned} X_{REF} &= f(X_{NEW}, Y_{NEW}) \\ Y_{REF} &= g(X_{NEW}, Y_{NEW}) \end{aligned}$$

where  $f$  and  $g$  are the functions which better describe the relation between the coordinates of the two images. The type of function may depend on several factors, such as the sensor acquisition model and terrain distortion, among others. In the presence of a set of  $N$  conjugate points, the previous equations may be solved for the function coefficients, through the most appropriate method in each case (usually the least square method). The main difficulty relies on an automatic and accurate identification of the  $N$  conjugate points, which is a particular challenge in several remote sensing applications. The main steps of the proposed methodology for AIR are shown in Fig. and include the following: conversion to single-band image, image segmentation, SIFT descriptors, obtention of a set of matching candidates, outlier removal, and final set of tie points and corresponding geometric correction measures. These steps will be separately described in the following.

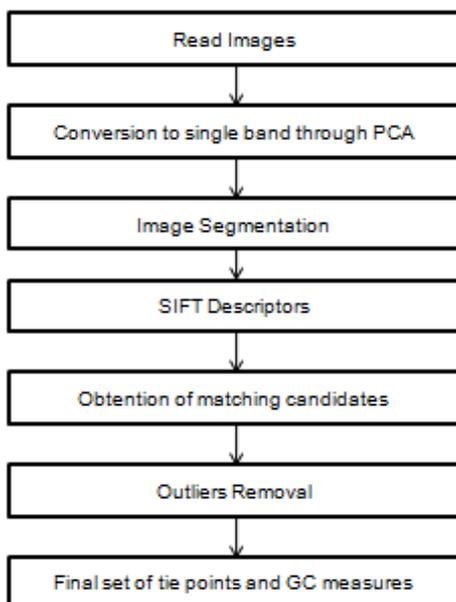


Fig 1 Main steps of the proposed methodology for AIR

### A. Conversion to Single Band through PCA

In the case that one or both images to be registered have more than one spectral band, a data reduction method should be individually applied to each image to facilitate the later segmentation stage. This step should account for minimizing the loss of important information for the later stages of the methodology, which is a consequence of applying data reduction methods. It should allow for considering one single band, which explains the majority of the image variability. The method of PCA was first derived and published by Hotelling. Principal components are still a basic tool for image description used in numerous applications, allowing, in general, for the reduction of the data dimension. Under the scope of image processing, PCA allows for the reduction of the number of bands of an image through a linear combination of them. Additionally, it allows for considering one single band, which explains the majority of the image variability. Let  $I$  be an image of size  $m \times n \times k$  pixels, where  $k$  is the number of spectral bands. This image stack  $I$  is rearranged as a new image  $I_{vector}$  of size  $mn \times k$  pixels, where each column of  $I_{vector}$  corresponds to the stacked pixels of the corresponding band of  $I$ .

Let us consider  $v = (v_1, v_2, \dots, v_k)^T$ , which is a  $k$ -element vector formed by the values of a particular pixel position across the  $k$  spectral bands, i.e., each line of  $I_{vector}$ . Defining  $m_v = E\{v\}$  as the expected value of  $v$  and  $C_v = E\{(v - m_v)(v - m_v)^T\}$  as the respective covariance matrix, let  $A$  be a matrix whose rows are formed from the eigenvectors of  $C_v$  in such an order that the first row of  $A$  is the eigenvector corresponding to the largest eigen value and so on. The Hotelling transform

$$y_v = A(v - m_v)$$

is what is also known as the principal component transform. Through this transform, it is possible to concentrate on a single band (the first principal component hereby assigned as  $J$ ) most of the variability explained by the original  $k$  bands. This is a widely known method of data reduction.

### B. Image Segmentation

The segmentation of an image allows for its “simplification,” since it significantly reduces the number of different pixel values. Although it is also associated with a loss of information on the image content, the decision of using an original or segmented image will depend on the context of the AIR method. Image segmentation is a process of partitioning an image into nonintersecting regions such that each region is homogeneous and the union of two adjacent regions is not homogeneous. Let  $P(\cdot)$  be a homogeneity predicate defined on groups of connected pixels and  $J$  the first principal component (of size  $m \times n$  pixels) of  $I$ , obtained as described in the previous section. Segmentation is a partitioning of image  $J$  into a set of  $l$  connected regions such that

$$\bigcup_{i=1}^7 S_i = J \quad \text{with } S_i \cap S_j = \phi, \quad i \neq j$$

and the uniformity predicate  $P(S_i) = \text{true}$  for all regions  $S_i$  and  $P(S_i \cup S_j) = \text{false}$  when  $S_i$  is adjacent to  $S_j$ .

A large number of segmentation methods can be found in the literature, but there is no single method which can be considered good for all images, nor are all methods equally good for a particular type of image. The existing image segmentation methods include gray-level thresholding, iterative pixel classification, surface-based segmentation, edge detection, and methods based on fuzzy set theory. Thresholding based methods can be classified according to global or local thresholding and also as either bi-level thresholding or multi-thresholding. For the aforementioned facts, we decided to consider the nonparametric and unsupervised Otsu's thresholding method.

The Otsu's thresholding method may be recommended as the simplest and standard method for automatic threshold selection, which can be applied to various practical problems. Although the Otsu's thresholding method is usually applied to images with a bimodal histogram, it may also provide a meaningful result for unimodal or multimodal histograms where a precise delineation of the objects present on the scene is not a requirement. Some examples are illustrated in, where the histogram shape is nearly unimodal and a meaningful segmentation is obtained. The key concept behind this method is to obtain an optimal threshold that maximizes a function of the threshold level. The optimal threshold is selected by a discriminant criterion, in order to maximize the separability of the resultant classes in gray levels. The procedure utilizes only the zeroth- and the first-order cumulative moments of the gray level histogram.

According to this bi-level thresholding, the image  $J$  pixels are assigned as 0 or 1. Then, the connected components in the binary image are identified and assigned a number, and objects with size less than 0.1% of the image size are removed in order to reduce the computation effort, without compromising the method performance. The labeled image with the small regions removed is then stretched to a 16-b unsigned precision in order to improve the detector obtention at the next step (SIFT).

### C. SIFT

One of the most powerful approaches for the obtention of local descriptors is the SIFT. The SIFT approach transforms image data into scale-invariant coordinates relative to local features and is based on four major stages: scale space extrema detection, key point localization, orientation assignment, and key point descriptor.

Let  $I(x, y)$  be an image and  $L(x, y, \sigma)$  the scale space of  $I$ , which is defined as

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y)$$

where  $*$  is the convolution operation in  $x$  and  $y$  and  $G(x, y, \sigma)$  is a variable-scale Gaussian defined as

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}$$

The scale-space extrema detection begins with the detection of local maxima and minima of  $D(x, y, \sigma)$ , defined as the convolution of a difference of Gaussian with the image  $I(x, y)$

$$\begin{aligned} D(x, y, \sigma) &= (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \\ &= L(x, y, k\sigma) - L(x, y, \sigma) \end{aligned}$$

The detection is performed by searching over all scales and image locations in order to identify potential interest points that are invariant to scale and orientation.

Once a set of keypoint candidates is obtained, the next step is to accurately localize them. This is performed by rejecting those keypoints, which have low contrast or are poorly localized along an edge, by a detailed fit to the nearby data for location, scale, and ratio of principal curvatures. Unstable extrema with low contrast are detected by considering a threshold over the extremum of the Taylor expansion (up to the quadratic terms) of  $D(x, y, \sigma)$ .

The third stage of the SIFT approach is the orientation assignment to each keypoint, based on local image gradient directions. This allows for the representation of each key point relative to this orientation, achieving invariance to image rotation. It is performed through an orientation histogram formed from the gradient orientations of sample points within a region around the key point, having 36 bins covering the 360-degree range of orientations. Each sample added to the histogram is weighted by its gradient magnitude and by a Gaussian-weighted circular window with a  $\sigma$  that is 1-5 times that of the scale of the key point. Then, a thresholding-based procedure refined by a parabola fitting is used to accurately determine the key point orientation through the orientation histogram.

The last stage of the SIFT approach is the key point descriptor. The previously described steps assigned the location, scale, and orientation of each key point. The motivation for the computation of a more complex descriptor is to obtain a highly distinctive key point and invariant as possible to variations such as change in illumination or 3-D viewpoint.

### D. Obtention of Matching Candidates

Under the scope of automatic registration of satellite images, since several distortion effects may be present in an acquired image (as already mentioned in the Introduction), it is desirable to have a reference image with as little distortions as possible (such as an orthoimage with no shadow effects and similar spectral content). Having that in mind, the SIFT descriptors of the reference image may be used as a reference database of key points, used for matching the key points derived from the image to be registered. In this paper, we have considered the nearest neighbor approach for key point matching. The nearest neighbor is defined as the key point with minimum Euclidean

distance for the invariant descriptor vector. An effective measure for a matching validation is the ratio between the distance of the closest neighbor and the distance to the second closest neighbor, hereafter assigned as  $d_{ratio}$ . An efficient nearest neighbor indexing is performed through the best-bin-first algorithm.

*E. Outlier Removal*

Even after the identification of matching candidates after removal of incorrect initial matches as described in the previous section, remote sensing images still produce unreliable tie points which lead to a further incorrect geometric correction. Each SIFT key point specifies four parameters: 2-D location ( $x$  and  $y$ ), scale, and orientation ( $\theta$ ). The proposed refinement of excluding unreliable tie points is an iterative procedure based on the principle that the “correct set” of matching key points corresponds to a denser region of the 2-D representation of the horizontal and vertical distances between the matching candidates.

An example of a set of matching candidates is shown in Fig. 2(a), corresponding to the segmentation of the images shown in Fig. 3. Each iteration of the outlier removal procedure consists in the analysis of the bivariate histogram associated to the previously described scatter plot, retaining only those bins with an absolute frequency greater than 10% of the maximum absolute frequency. The number of bins is a sensitive point and is selected according to the sturges rule. This rule provides an objective indication of the number of bins as being  $1 + 3.322 \log_{10} N$ , where  $N$  is the number of observations. The histogram in Fig. 2(b) corresponds to the bivariate histogram of the data in Fig.2(a). The procedure stops when the registration accuracy measure  $RMS_{all}$  (described in the next section) is below one or the maximum number of iterations is achieved. Based on our experiments, a maximum number of ten iterations are sufficient for the procedure to converge to a subset of valid matching keypoints. The brighter points in Fig. 2(a) correspond to the points considered valid after the first iteration of outlier removal.

*F. Final Set of Tie Points and Geometric Correction Measures*

The final set of tie points is composed by the initial matching candidates after the removal of the outliers detecting according to the procedure described in the previous section. The performance of the proposed methodology for AIR was evaluated through measures, which allow for an objective and automatic evaluation of the image registration process quality, based on theoretical acceptance levels. The measures include the following:  $N_{red}$  (number of redundant points),  $RMS_{all}$  (rmse considering all contr-

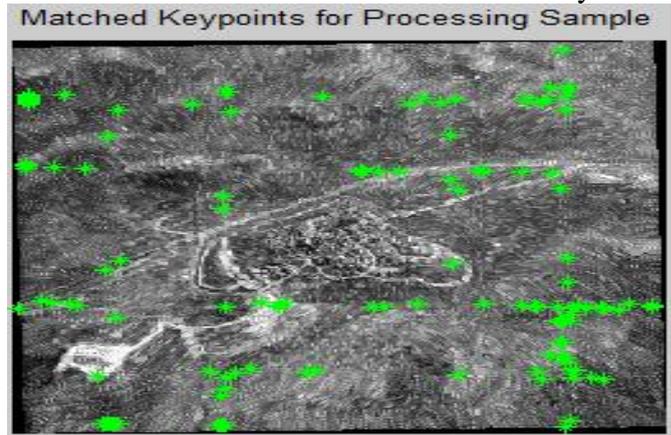


Fig 2(a) matched key points

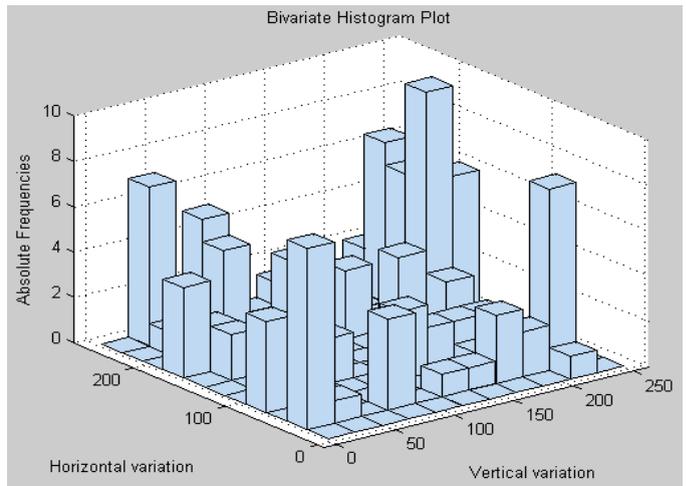


Fig 2(a) Bivariate histogram of the values in 2(a)

ol Points (CPs) together),  $RMS_{loo}$  (rmse computation of the CP residuals, based on the leave-one-out method),  $p_{quad}$  (statistical evaluation of residual distribution across the quadrants),  $BPP(1.0)$  (bad point proportion with norm higher than 1.0),  $S_{kew}$  (statistical evaluation regarding the presence of a preference axis on the residual scatter plot),  $S_{cat}$  (statistical evaluation of the goodness of CP distribution across the image), and  $\phi$  (a weighted combination of the seven previously referred measures). Both rms measures are normalized to the pixel size.

III. RESULTS AND DISCUSSION

The methodology for AIR was applied to remote sensing images. For each pair, the most proper transformation function was applied according to the involved distortion. The performance of the method was evaluated through measures recently proposed in the literature [x]. The images were processed on a computer with an Intel Core i5 2410 @2.30-GHz processor and 4.0 GB of physical memory, using MATLAB 7.0.1 version.

The segment of 280x280 pixel image was obtained from Landsat band 5. The obtained measures as a function of  $d_{ratio}$  are

Shown in Fig. 4, and the final set of keypoints (location, scale, and orientation representation) for  $d_{ratio} = 0.7$  (the value which led to the lowest value of  $\phi$ ) is shown in Fig. 4(c). The low rms values, which are achieved even for low values of  $d_{ratio}$ , are noticeable. For  $d_{ratio} = 0.7$  ( $RMS_{all} = 0.90$ ,  $RMS_{loo} = 0.92$ , and  $\phi = 0.511$ ), which led to the obtention of 23 tie points.



Fig 3(a) original image

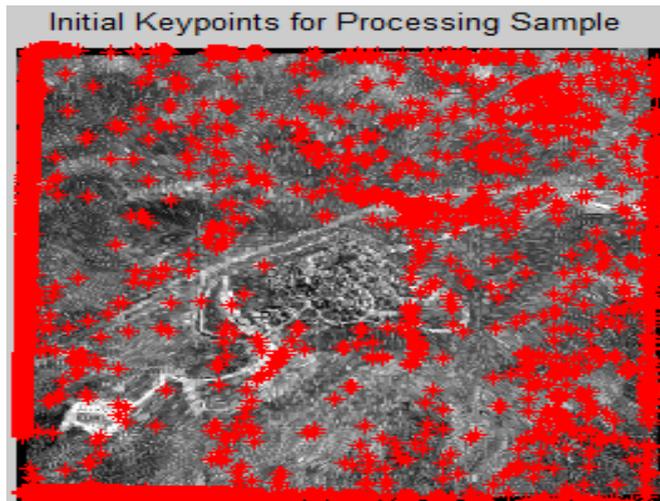


Fig 3(b) initial key points of 3(a) image

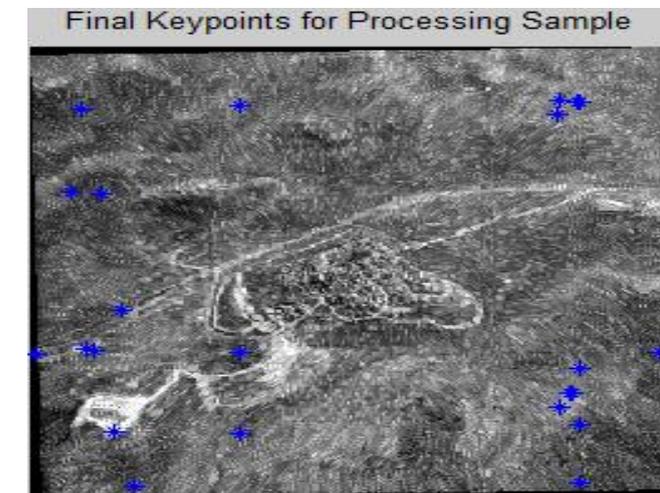
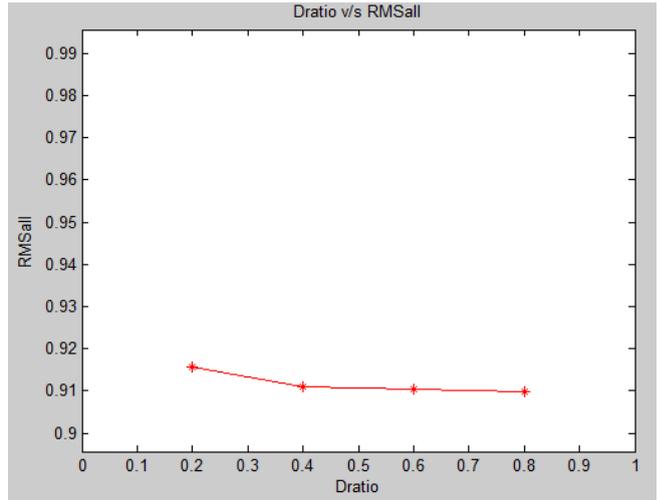
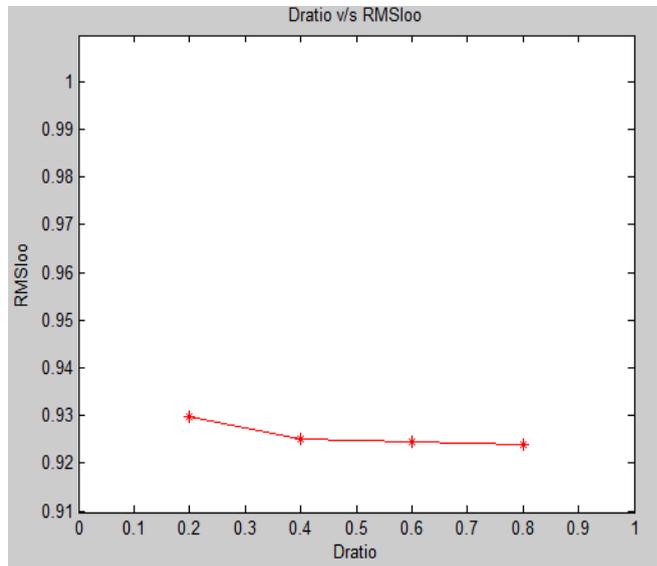


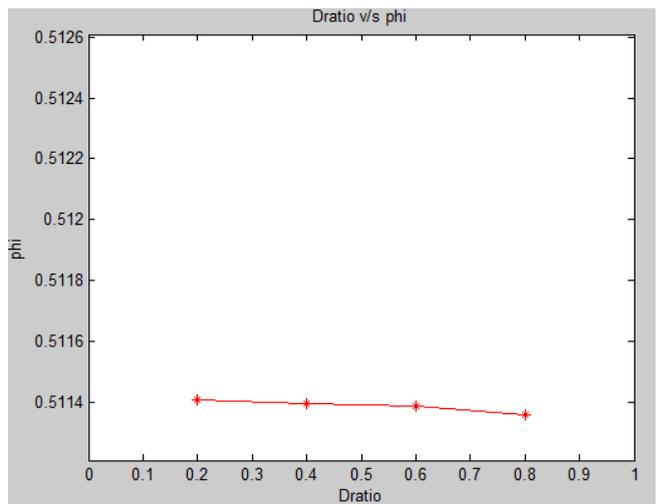
Fig 3(c) final key points after outlier removal



(a)



(b)



(c)

Fig 4. Measures (a) RMSall, (b) RMSloo, and (c) phi as a function of Dratio regarding the registration of images described.

The proposed methodology for AIR was compared with three methods for the four previously described pairs of images. The first was based on the correlation coefficient (CC), where each image was divided into tiles of size  $128 \times 128$  pixels allowing for the obtention of a set of tie points based on the center of the tile and the identified peak of the similarity surface. The similarity measure mutual information (MI) was similarly considered as the second method of comparison, whereas SIFT was the third considered method.

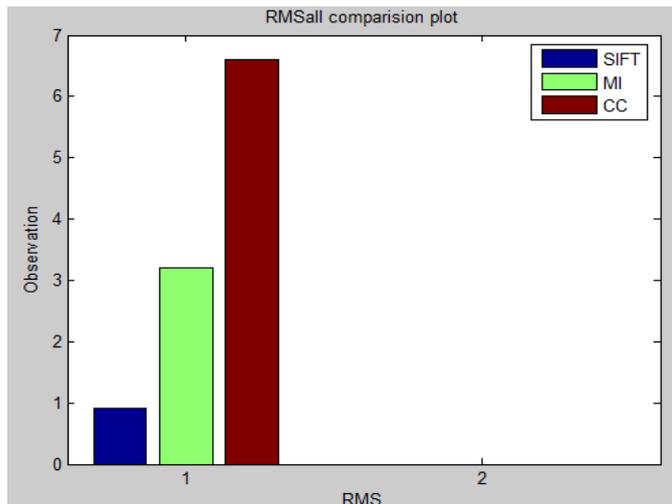


Fig 5(a) comparison of RMS<sub>all</sub>

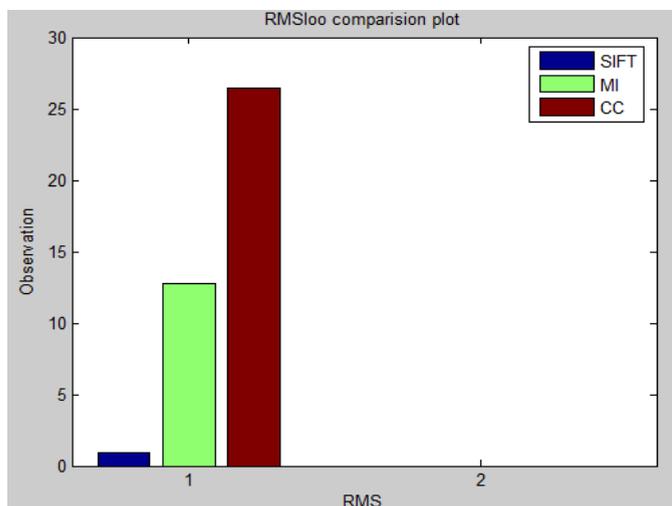


Fig 5(b) comparison of RMS<sub>loo</sub>

#### IV. CONCLUSION

To my knowledge, a sensitivity analysis on the  $d_{ratio}$  parameter has been performed in this work for the first time, regarding SIFT-based image registration methods. It allowed for a gain of sensitivity of its use with respect to different types of satellite images, as well as showing some misregistration results when using an inadequate value for parameter  $d_{ratio}$ . The proposed methodology for AIR revealed to be robust to a

variation of this parameter, not providing any registration when the choice for the parameter is not adequate. This is a highly desirable characteristic when dealing with AIR methods. Under the scope of an AIR methodology, it is strictly necessary that all stages are automatic, which includes the “image segmentation” phase. However, fully automatic image segmentation is still a present subject of research, in particular for natural (including remote sensing) images. Nevertheless, in order to provide some sensitivity analysis of the proposed AIR methodology on the segmentation method, the k-means clustering technique was also considered as an alternative to the Otsu’s thresholding method.

The performance of the proposed method is evaluated through measures such as  $N_{red} = 0.0480$ ,  $RMS_{all} = 0.9099$ ,  $RMS_{loo} = 0.9240$ ,  $P_{quad} = 0.0098$ , Bad Point Proportion:  $BPP(r) = 0.0435$ ,  $S_{cat} = 0.0055$ ,  $S_{kew} = 0.0037$ , Cost Function ( $\phi$ ) = 0.5113.

#### ACKNOWLEDGMENT

The authors thank the management of Institute of Aeronautical Engineering, Dundigal and Marri Laxman Reddy Institute of Technology and Management, Dundigal for providing the necessary facilities to undertake the above work.

#### REFERENCES

- i. D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.
- ii. L. G. Brown, “A survey of image registration techniques,” *Comput. Surv.*, vol. 24, no. 4, pp. 325–376, Dec. 1992.
- iii. N. R. Pal and S. K. Pal, “A review on image segmentation techniques,” *Pattern Recognit.*, vol. 26, no. 9, pp. 1277–1294, Sep. 1993.
- iv. L. Cheng, J. Gong, X. Yang, C. Fan, and P. Han, “Robust affine invariant feature extraction for image matching,” *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 2, pp. 246–250, Apr. 2008.
- v. P. Dare and I. Dowman, “An improved model for automatic feature-based registration of SAR and SPOT images,” *Proc. ISPRS J. Photogramm. Remote Sens.*, vol. 56, no. 1, pp. 13–28, Jun. 2001.
- vi. R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. New York: Wiley-Interscience, 2000.
- vii. B. Zitová and J. Flusser, “Image registration methods: A survey,” *Image Vis. Comput.*, vol. 21, no. 11, pp. 977–1000, Oct. 2003.
- viii. L. M. G. Fonseca and B. S. Manjunath, “Registration techniques for multisensor remotely sensed imagery,” *Photogramm. Eng. Remote Sens.*, vol. 62, no. 9, pp. 1049–1056, Sep. 1996.
- ix. H. Hotelling, “Analysis of a complex of statistical variables into principal components,” *J. Educ. Psychol.*, vol. 24, no. 7, pp. 498–520, Oct. 1933.
- x. J. Ma, J. C.-W. Chan, and F. Canters, “Fully automatic subpixel image registration of multiangle CHRIS/Proba data,” *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 7, pp. 2829–2839, Jul. 2010.