Abstract--- This system proposes an automatic image registration using image segmentation and affine transformation for remote sensing application. The existing methods include SIFT along with OTSU thresholding. SIFT has limited performance when directly applied to remote sensing images. Even though SVD, OTSU and SIFT already used separately in registration. In proposed methodology first ever method that is the combination of SVD, OTSU Thresholding, SIFT and Affine transform. This combination provides more accurate registration compare to the existing methodologies. This combination allows for an accurate obtention of tie points for a pair of remote sensing images, being a powerful scheme for AIR.

Keyword--- Automatic Image Registration (AIR), Image Segmentation, Bi-Model Histogram, Scale Space Extrema Detection, Keypoint Descriptor, SVD, SIFT, OTSU, Affine Transform

I. INTRODUCTION

Image registration is one of the basic image processing operations in remote sensing. “Image registration is the process of matching two images so that corresponding coordinate points in the two images correspond to the same physical region of the scene being imaged” Both synthetic and real data have been considered in this work. Image registration 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.

With the increase in the number of images collected every day from different sensors, automated registration of multi sensor/multispectral images has become a very important issue. A wide range of registration techniques has been developed for many different types of applications and data.

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 neighborhoods, the feature spatial distribution, or the feature symbolic description.

Top row—feature detection (corners were used as the features in this case).

Middle row—feature matching by invariant descriptors (the corresponding pairs are marked by numbers).

Bottom left—transform model estimation exploiting the established correspondence. Bottom right—image resampling and transformation using appropriate interpolation technique.

Figure 1: Automatic Image Registration

System proposes an automatic image registration using image segmentation and Affine transform for remote sensing application. 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. The use of image segmentation as a step in image registration had been scarcely explored. Therefore, further improvements under the scope of methodologies for automatic image registration (AIR) may be achieved, particularly combining image segmentation with other methods.

The concept behind it is to detect image regions covariant to a class of transformations, which are then used as support regions to compute invariant descriptors, i.e., the detectors provide the regions which are used to compute the descriptors. A comprehensive review on the comparison of affine region detectors may be found in compared the performance of descriptors computed for local interest regions of gray-value images. There are three main classes of descriptors: distribution-based descriptors, spatial frequency techniques, and differential descriptors. They compared the descriptor performance for affine transformations, scale changes, rotation, blur, jpeg compression, and illumination changes.
Mikolajczyk and Schmid have found, based on their experiments, that the scale invariant feature transform (SIFT) which is a distribution-based descriptor—was among those which obtained the best results for most of the tests, with lower performance for textured scenes or when edges are not reliable.

The SIFT approach. The SIFT approach allows for the extraction of distinctive invariant features from images, which can be used to perform reliable matching between images presenting a substantial range of affine distortion, change in 3-D viewpoint, addition of noise, and change in illumination. Despite the several advantages of using the SIFT approach, it does not produce meaningful results when directly applied to remote sensing images.

Li et al. proposed an adaptation on the original method proposed by Lowe, where the feature descriptor is refined and the use of the Euclidean distance is replaced by a joint distance. The method proposed by Li et al. assumes that often remotely sensed images have no local distortions, and so, geometric distortions can be modeled by “shape-preserving mapping” model (translation, rotation, and scaling only). However, when an image with nadir looking is to be registered with an image with a considerable viewing angle (such as most of Satellite Pour l’Observation de la Terre (SPOT) images and other higher spatial resolution satellite images), this assumption fails. It is also not adequate for situations where the terrain relief has significant variations across the considered scene.

Mukherjee et al. proposed a method for detection of multiscale interest points for later registration of hyperspectral imagery. They proposed spectral descriptors for hyperspectral interest points that characterize each interest point based on the spectral information and its location and scale. Their method mainly differs from the Lowe’s keypoint detection algorithm in the sense that principal component analysis (PCA) is applied to the hyperspectral imagery, and nonlinear function for combining difference of Gaussian (DoG) responses along spectral dimension is applied prior to local extrema detection. They considered in their experiments four time-lapse images acquired by the Hyperion sensor, retaining a subset of the available spectral bands (such as uncalibrated or saturated data channels) and using random regions with around 200 scan lines. Nevertheless, this methodology is not appropriate for multi- or single-band images, which is still presently the main imagery source in remote sensing applications.

Sırmacı and Ünsalan [28] have recently used SIFT keypoints and graph theory applied to IKONOS images, under the scope of urban-area and building detection. They state that the standard SIFT implementation is not sufficient for urban-area and building detection from satellite images alone, since the presence of many similar and nearby buildings in the satellite images is a quite frequent problem [28]. Moreover, as mentioned by the authors, their building-detection method may not detect buildings if the contrast between their rooftop and the background is low. Although the work by Sırmacı and Ünsalan is about urban-area and building detection, and not AIR, this is another remote sensing application where the simple application of SIFT to remote sensing images is not sufficient.

II. PROPOSED SYSTEM

In this paper an efficient method for AIR is proposed, which combines image segmentation, SIFT, and affine transform. The reference and unregistered images may differ in translation, rotation, and scale and may present distortions associated to the terrain relief and significantly different spectral content. The methodology is described in Section II, the results of its application to different pair of images are illustrated in Section III, and the discussion is presented in Section IV.

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

\[
X_{REF} = f(X_{NEW}, Y_{NEW})
\]

\[
Y_{REF} = g(X_{NEW}, Y_{NEW})
\]

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. 1 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.

2.1 Single Band Conversion

A data reduction method will be applied to each image for converting multi band to single band image using SVD. Singular value decomposition (SVD) transform data take a form in which the first singular value has a great amount of the image information. With this, we can use only a few singular values to represent the image there by multi spectral image is reduced to single band image. A data reduction method will be applied to each image for converting multi band to single band image using SVD.
Singular Value Decomposition (SVD) is said to be a significant topic in linear algebra by many renowned mathematicians. SVD has many practical and theoretical values; Special feature of SVD is that it can be performed on any real (m, n) matrix. Let’s say we have a matrix A with m rows and n columns, with rank r and r \leq n \leq m. Then the A can be factorized into three matrices:

\[
A = U S V^T
\]

Figure 3: Illustration of Factoring A to USV^T

Where Matrix U is an m x m orthogonal matrix

\[
U = [u_1, u_2, ..., u_r, u_{r+1}, ..., u_m]
\]

Column vectors \(u_i\), for \(i = 1, 2, ..., m\), form an orthonormal set

\[
u_i^T u_j = \delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}
\]

And matrix V is an n x n orthogonal matrix \(V = [v_1, v_2, ..., v_r, v_{r+1}, ..., v_n]\) Column vectors \(v_i\), for \(i = 1, 2, ..., n\), form an orthonormal set

\[
v_i^T v_j = \delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}
\]

Here, S is an m x n diagonal matrix with singular values (SV) on the diagonal. The Matrix S can be showed in following

\[
S = \begin{bmatrix}
\sigma_1 & 0 & \cdots & 0 \\
0 & \sigma_2 & \cdots & 0 \\
\vdots & \ddots & \ddots & \vdots \\
0 & \cdots & \sigma_m & 0 \\
\end{bmatrix}
\]

For \(i = 1, 2, ..., n\), \(\sigma_i\) are called Singular Values(SV) of matrix A. It can be proved that \(\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{r}>0\), and \(\sigma_{r+1} = \sigma_{r+2} = \cdots = \sigma_n = 0\). The \(u_i\)’s and \(v_i\)’s are called right and left singular vectors of A.

2.2 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 \(P()\) be a homogeneity predicate defined on groups of connected pixels and \(J\) the first SVD 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.

The uniformity predicate \(P(Si) = true\) for all regions \(Si\) and \(P(Si \cup Sj) = false\) when \(Si\) is adjacent to \(Sj\). 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 [26]. The existing image segmentation methods include gray-level thresholding, iterative pixel classification, surface-based segmentation, edge detection, and methods based on fuzzy set theory [26]. Thresholding based methods can be classified according to global or local thresholding and also as either bilevel thresholding or multithresholding.

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 graylevel histogram.

According to this bilevel 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).

2.3 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: scalespace extrema detection, keypoint localization, orientation assignment, and keypoint descriptor.

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

\[
L(x, y, \sigma) = G(x, y, \sigma) \ast J(x, y)
\]

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

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

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 \(J(x, y)\)
\[ D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) \ast J(x, y) \\
= L(x, y, k\sigma) - L(x, y, \sigma). \]

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 keypoint 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 keypoint, having 36 bins covering the 360° 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 keypoint. Then, a thresholding-based procedure refined by a parabola fitting is used to accurately determine the keypoint orientation through the orientation histogram.

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

### 2.4 Affine Transform

Affine transformations are one of the least complicated operations that can be performed in image processing. However, it is still very useful, because it is one of the basic transformations that can be used on an image. Affine transformations do not really affect the values of the pixels. What is done during those transformations is simply a change in the pixels positions or order. But actually there will be no effect on the brightness or the contrast of the picture. It does not change the colours of the objects in the image, but rather their shapes. This kind of transformation is used on pictures which display some distortion. If one or several objects of a picture appear badly shaped on the image, then affine transformations can be used to make them look better.

In many imaging systems, detected images are subject to geometric distortion introduced by perspective irregularities wherein the position of the camera(s) with respect to the scene alters the apparent dimensions of the scene geometry. Applying an affine transformation to a uniformly distorted image can correct for a range of perspective distortions by transforming the measurements from the ideal coordinates to those actually used. (For example, this is useful in satellite imaging where geometrically correct ground maps are desired.)

### III. 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 keypoints, used for matching the keypoints derived from the image to be registered. In this paper, we have considered the nearest neighbor approach for keypoint matching. The nearest neighbor is defined as the keypoint 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.

### IV. Results

The image taken here is high spatial resolution data set, composed by the red, green, blue, and near-infrared bands of an orthophotograph and by an ALOS PRISM scene, covering a region from the North of Portugal. Image size of 385x385 segment is taken. And the reference image of size 251x200. Therefore, a first-order polynomial is sufficient to accurately register this piece of images.
The main scope of the project is to increase the Accuracy of Automatic Image Registration by considering the accurate set of tie points. In this project two stages are there first stage is converting multiband to single band image and then apply image segmentation and the second stage is get tie points by using SIFT transform and then apply Affine transform. Finally the image is registered. In the first stage, both base and reference images are considered. Initially the multiband images are converted to single band by applying SVD to the image. After converting to single band Image segmentation is applied to stretch the wanted region in both base and image to be registered. Otus thresholding is used in this system.

In second stage Sift transform applied to both images to get feature points. After getting the feature points affine transform is applied to image to be registered, so that it produce better matching with key points in reference image. Finally image registration is to be done to identify the object moment in the image by comparing the reference image.

**REFERENCES**


