International Journal of Computer Science and Software Engineering (IJCSSE), Volume 4, Issue 11, November 2015
ISSN (Online): 2409-4285 www.IJCSSE.org Page: 308-314

Feature Based Automatic Multiview Image Registration

Sruthi Krishna[1] and Dr. Abraham Varghese[2]

[1, 2] Computer Science and Engineering, Adi Shankara Institute of Engineering and Technology, Kalady, India

[1]*sruthi4krishna@gmail.com*

ABSTRACT

Automatic image registration is a vital yet challenging task, particularly for remote sensing images. A fully automatic registration approach which is accurate, robust, and fast is required. The registration process is divided into six main steps: feature detection, feature extraction, feature matching, outlier detection and removal, transform model estimation and resampling. In the feature detection, keypoints (PCA-SIFT), blob (SURF), region (MSER) features are detected. The features are then matched to produce point correspondences. The alignment process uses the RANSAC algorithm to estimate an outlier free transformation using point correspondences. The target image is transformed by using transformation function, which results in automatic image registration. Mutual information is used for fine tuning registration. In this thesis, feature descriptors used are SURF (Speeded Up Robust Features), MSER (Maximally Stable Extremal Regions) and PCA-SIFT. They are invariant to zoom, noise, scale, rotation and illumination, hence very useful. The analysis is conducted for various image transformations such as image rotation, scaling and change in illumination. For all these transformations, various quality checking parameters such as recall and RMSE error are evaluated to analyse the performance of registration. In the registration of multiview images, the affine transformation model applied is not suitable. So a more appropriate transformation models such as thin-plate spline is subtitled the affine model in the proposed work.

Keywords: *Image registration, SURF, PCA-SIFT, MSER, RMSE, Thin plate spline, Affine Transformation.*

1. INTRODUCTION

Image registration is the procedure of searching spatial transform relations between two images or multiple images at different time, different perspective, or captured by the different sensors of the same scene, matching and overlaying one or more images. In remote sensing, a set of GCPs (Ground Control Points) are used to determine the parameters of a transformation function, which is then used to apply a deformation with an interpolation function on the new image without affecting the reference image. There are two methods: the first one is the manual registration, which is not feasible in the case where a large number of images must be registered, and with times cited from a few hours to 4-5 days for registering a single image. And the second one is the automatic registration techniques that require a little or no operator supervision. These techniques have the advantage of multi-temporal and / or multi-sensor and / or multi-resolution information.

Automatic image registration is still a challenge due to the presence of particular difficulties within the remote sensing field. The difficulties involved mainly include both geometric deformations (translation effect, rotation and scale distortion, occlusion, and viewpoint difference) and radiometric discrepancies (illumination change and sensor and spectral content difference). SIFT and its variants is capable of extracting distinctive invariant features from images, and it can be applied to preform reliable matching across a substantial range of affine distortion, change in 3D viewpoint, addition of noise, and change in illumination. Despite the attractive advantages of SIFT, there exist some problems when it is directly applied to remote sensing images, i.e., 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 . In addition, many outliers exist in feature matches on account of significant differences on the image intensity between the overlay regions of remote sensing images. Therefore, using SIFT alone cannot produce optimal results. The mutual information (MI) criterion is particularly suitable for multisensor image registration. MI, which represents a measure of statistical dependence between two images, is one of the most commonly used similarity measures in intensity-based methods. This process geometrically aligns the two images: (i) Reference image (ii) Sensed image

Image registration involves locating and matching similar regions in the two images to be registered. In manual registration, a human carries out these tasks visually using interactive software. In automatic registration, on the other hand, autonomous algorithms perform these tasks. In remote sensing, automated

procedures do not always over the needed reliability and accuracy, so manual registration is frequently used. The user extracts from both images distinctive locations, which are typically called control points (CPs), tie points, or reference points. First, the CPs in both images (and datasets) are interactively matched pairwise to achieve correspondence. Then, corresponding CPs are used to compute the parameters of a geometric transformation in question.
The flow diagram for process of Image Registration is as shown in Figure 1.

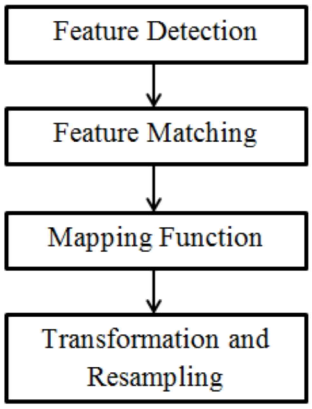

Fig. 1. steps of image registration

Functions of each block are as follows:

- Feature detection: Extraction of distinct regions, or features, to be matched.
- Feature matching: In this stride, the correspondence between the elements recognized in the detected picture and those identified in the reference picture is built up.
- Transform model estimation: The parameters of the mapping capacities are processed
- Image resampling and change: The detected picture is changed by method for the mapping function.

2. RELATED WORK

A. Feature Descriptors

PCA-SIFT
The PCA-SIFT descriptor (Principal Components Analysis SIFT) is a variety of SIFT with two main differences: (1) the descriptor is calculated for a region of size 39 x 39 sub-regions instead of 4x4 used in SIFT and (2) instead of 8 bins for the orientation, PCA-SIFT calculates the orientation in the x and y directions. The result is a vector of dimension 3042 (39 x 39 x 2), which is then reduced to 36 with the principal component analysis. [9]

GLOH
The GLOH descriptor (Gradient Location and Orientation Histogram) is an extension of the SIFT descriptor designed to increase the robustness and distinctiveness of the descriptors. Instead of using the Cartesian coordinate, GLOH makes use of the log-polar coordinate system and calculates descriptors in 17 sub regions: 8 in angular direction and 3 in radial direction at 3 different radii (6, 11 and 15). For each sub region, 16 gradients are computed, giving a vector of dimension 272 (16 x (8 + 3 x 3)), this is then reduced to 64 through principal components analysis. [9]

B. Transformation Approaches

1. Nearest neighbour, where the output pixel is given the value of the input pixel whose location is closest to the reverse sampled position (x, y). The advantage of nearest neighbour resampling is that the output image only contains intensity values present in the original image. However, it can produce aliasing “jaggies," particularly with rotation.
2. Bilinear, where the output pixel value is a linear interpolation of the local neighbourhood, usually the four surrounding input pixels. The advantage of bilinear interpolation is that it is fast. Also, its results are visually similar to those obtained by more complex interpolators, although it is not as accurate as the bicubic interpolator or the other higher-order methods.
3. Bicubic, where the output pixel value is obtained by a cubic polynomial interpolation of the values in a local neighbourhood.
4. Spline, where the output pixel value is computed by a polynomial spline interpolation (e.g., B-spline) of the local neighbourhood.
5. Sinc function, where the output pixel value is obtained by an interpolation based on the sinc function, sin(x, y)/r (where r = x2 + y2) over a local neighbourhood

3. PROPOSED METHODOLOGY AND DISCUSSION

A procedure of searching spatial transform relations between two images or multiple images taken at different time, different perspective, or captured by the different sensors of the same scene, matching and overlaying one or more images. There will be a reference image and sensed image. Feature descriptors are obtained from these two images using algorithms

SIFT, SURF, PCA-SIFT and MSER. Find the matching points between the two images. An algorithm called RANSAC is used to eliminate the outliers and then transform the sensed image with respect to reference image using affine and TPS transformation. The result is optimized using mutual information. Result of each of the algorithm is collected and analysed.

A. Feature Extraction and Matching

SIFT

Steps in SIFT algorithm are:

1. Developing a scale space. This is the initial preparation. An internal representation of the first picture is made to guarantee scale invariance. This is finished by creating a "scale space".
2. LoG Approximation. The Laplacian of Gaussian is extraordinary for discovering intriguing focuses (or key focuses) in a picture. Be that as it may, it's computationally extravagant. So cheat and rough it utilizing the representation made before.
3. Finding keypoints. With the super quick rough guess, attempt to discover key points. These are maxima and minima in the Difference of Gaussian picture.
4. Get rid of bad key points. Edges and low complexity districts are awful keypoints. Taking out these makes the calculation productive and powerful. A method like the Harris Corner Detector is utilized here.
5. Assigning an orientation to the keypoints. An orientation is computed for every key point. Any further estimation is done in respect to this orientation. This viably counteracts the impact of introduction, making it rotation invariant.
6. Generate SIFT features. Finally, with scale and rotation invariance in place, one more representation is produced. This assists extraordinarily with recognizing components. In the event that 50,000 elements arrive, with this representation, highlights that are searching for can be effortlessly recognized.

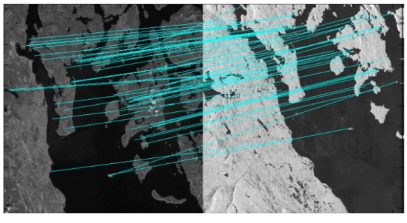

Fig. 2. Matching points

SURF

SURF stands for Speeded Up Robust Features. It is an algorithm which extracts some unique keypoints and descriptors from an image. A set of SURF keypoints and descriptors can be extracted from an image and then used later to detect the same image.

SURF [6] is comprised of a feature detector based on a Gaussian second derivative mask, and a feature descriptor that relies on local Haar wavelet responses. The purpose of SURF is to develop an algorithm that is faster and with the same performance than SIFT. SURF uses integral images and approximations for accomplishing higher speed than SIFT. These integral images are used for convolution. Like SIFT, SURF lives up to expectations in three stages: extraction, description and matching. The contrast in the middle of SIFT and SURF is that SURF extracts the features from an image using integral images and box filters SURF detector is mainly based on the approximated Hessian Matrix. On the other hand, the descriptor gives a distribution of Haar wavelet responses within the interest point neighbourhood. Both the detector and descriptor are used to reduce the computation time because descriptor has low dimensionality. So that SURF is better than previously used schemes with respect to repeatability, distinctiveness, robustness and speed.

Since Hessian matrix has good performance and accuracy. In image I, x = (x, y) is the given point, the Hessian matrix H(x,σ) in x at scale σ, it can be define as

$$\mathbf{H(x,\sigma)} = \begin{bmatrix} L_{xx}(x,\sigma) & L_{xy}(x,\sigma) \\ L_{yx}(x,\sigma) & L_{yy}(x,\sigma) \end{bmatrix}$$

Where $Lx\mathrm{x}(x, \sigma)$ is the convolution result of the second order derivative of Gaussian filter $(\partial^2 / \partial x^2) g(\sigma)$ with the image I in point x, and similarly for $Lxy(x, \sigma)$ and F$Lyy(x, \sigma)$.

SURF creates a “stack” without 2:1 down sampling for higher levels in the pyramid resulting in images of the same resolution.

Due to the use of integral images, SURF filters the stack using a box filter approximation of second-order Gaussian partial derivatives. Since integral images allow the computation of rectangular box filters in near constant time. The Gaussian second order partial derivative box filters of Dyy and Dxy are show in Figure 3.

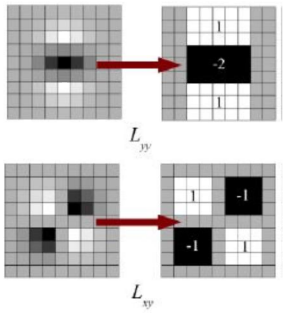

Fig. 3. The Gaussian Second order partial derivative box filters in y-(Dyy) and xy-direction (Dxy) [6].

PCA-SIFT

PCA-SIFT is a modification if SIFT, which changes how the keypoint descriptors are constructed. Main steps are:

1. Select a representative set of pictures and detect all keypoints in these pictures
2. For each keypoint: a. Extract an image patch around it with size 41x41 pixels b. Calculate horizontal and vertical gradients, resulting in a vector size 39 x 39 x 2 = 3042
3. Put all these vectors into a k x 3042 matrix A where k is the number of keypoints detected
4. Calculate the covariance matrix of A
5. Compute the eigenvectors and eigenvalues of covA
6. Select the first n eigenvectors: the projection matrix is a n x 3042 matrix composed of these eigenvectors
7. n can either be a fixed value determined empirically or set dynamically based on the eigenvalues
8. The projection matrix is only computed once and saved.

MSER

The MSER algorithm extracts stable regions that are invariant to translation and rotation. However, since the images in this thesis are of the same scale, the regions do not need to be invariant to scale.

1. Sweep threshold of intensity from black to white, performing a simple luminance thresholding of the image. If (intensity < threshold) ! white otherwise black. Figure 4 shows a sequence of binary image using different threshold values. This is the main idea behind the MSER procedure.
2. Extract connected components(extremal regions")
3. Find a threshold when an external region is \maximally stable"
4. Keep those regions descriptors as features.

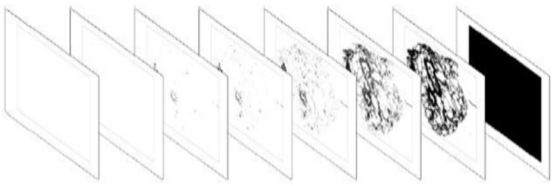

Fig. 4. Sequence of binary image [7]

B. *RANSAC Outlier Removal*

The algorithm of random sample and consensus (RANSAC) is a robust method of estimating parameter. The basic idea of the method is to eliminate deviation points using internal data constraints of data sets. When estimate the parameter, first design the objective function and then get the value of the parameters using the iterative estimation. All the data are divided into inliers and outliers, where the inliers meet to the estimation model and the outliers do not meet to the model. And the parameters of estimation model are obtained by inliers through continuous iteration.

A RANSAC algorithm provides a general technique for model fitting in the presence of outliers and consists of the following steps:

1. Pick a model.
2. Determine the minimal number of points needed to indicate the model.
3. Define a threshold on the inlier count.
4. Fit the model to a randomly selected minimal subset
5. Apply the transformation to the complete set of points and count inliers.
6. If the number of inliers exceeds the threshold, flag the fit as good and stop.
7. Otherwise rehash steps 4 to 6.

As an example take linear fitting given in the figure 4.5.

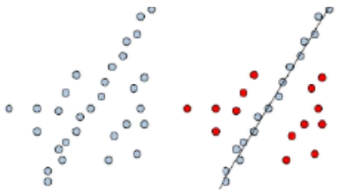

Fig. 5. Linear fitting using RANSAC [5]

The result of image registration is given in figure 6 and figure 7. And the values of scaling, rotating, shifting, MI and RMSE are given in table 1. Since SIFT and affine transformation is not suitable for multiview image registration a new transformation is used which is TPS. This is explained in the next section and the result of the same is in the table 2.

C. Multiview image registration using SIFT and TPS

For image registration SIFT is used along with affine transformation. And it successfully registers the images. Affine transformation is widely used in remote sensing image registration. But when it comes to multiview images with the difference in the acquisition angle and the terrain elevation, affine is not suitable. For that a new approach is proposed. For feature extraction, description and matching SIFT is used. And for resampling and transformation thin plate spline (TPS) method is used.
Sift is used for Feature extraction and matching. RANSAC is used for outlier removal. And TPS is using for transforming sensed image with respect to reference image.

Thin plate spline
TPS is a conventional mathematical tool for interpolating surfaces among scattered points in a 2D plane. It can handle the effects introduced by the acquisition angle and terrain elevation differences. Given two images, the objective is to deform an image so it coordinates the second one. Thin Plate Splines is one system that provides a smooth interpolation between the arrangements of control points. It interpolates a surface a surface that goes through every control point. An arrangement of 3 points will thus create a flat plane. It is anything but di cult to think about the control focuses as constraints on a bending surface. The ideal surface is one that bends the least. Fig demonstrates a sample of such a surface with 7 control points. The surface is compelled to go through all these 7 control focuses. An example is shown in figure 8

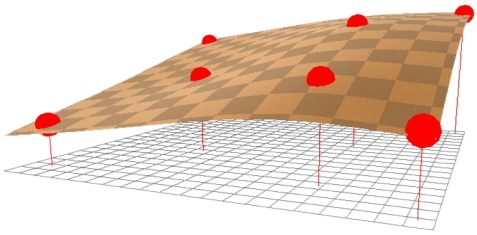

Fig. 8. A Thin Plate Splines that goes through a set of control points [8].

This least bent surface is given by the following equation:

$$f(x,y) = a1 + a2x + a3y + \Sigma_{i=1}^{n} w_i U(|P_i - (x,y)|)$$

A flat plane that best matches all control points (this can be seen as a least square fitting) is defined by the initial three terms relate to the linear part. n control points provides the bending forces and the last term corresponds to it. Each control point has a coefficientw_i. Also, the distance between the control point Pi and a position (x,y) is defined by $|P_i - (x,y)|$. This distance is used as a part of the function U defined by $U(r) = r^2 \log r^2$ So far, the coefficients a1,a2,a3, and w_i for every control point are unknown. All w_i forms the vector W. These unknowns are defined by

$$L^{-1}V = (W|a1a2a3)^T$$

3. EXPERIMENTAL RESULTS

The results of registration are given below.

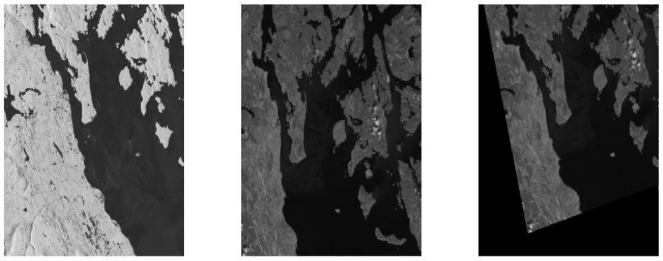

Fig. 6. Result of image registration using SURF

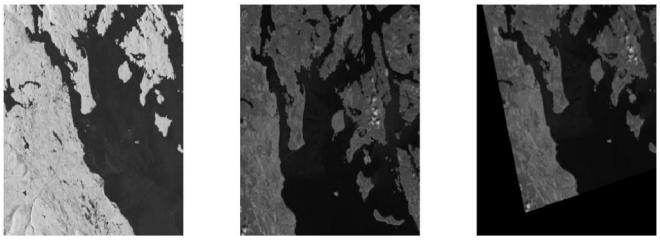

Fig. 7. Result of image registration using PCA-SIFT

The result of automatic image registration for images of Campbell River in British Columbia taken by ALOS-PALSAR in 2010 and by Landsat ETM+ in 1996 is given in the table 5.2. Registration parameters (a11, a12, a21, a22, x, and y), corresponding values of MI and RMSE are given in table. a11, a12, a21, a22 represent the rotation, scale, and shear differences, and x, y are the shifts between the two images. Sensed image is rotated, scaled and shifted to create a dataset of images using which RMSE is calculated by taking the average.
From the table it is clear that RMSE is lesser for registration using SIFT and the same obtain highest MI value. So it can be concluded that SIFT is the most suitable and robust for this application.

Table1: A Comparison of existing systems.

Methods	a11	a12	a21	a22	x	y	RMSE	MI
SIFT	0.884	-0.24	0.449	0.88	-18.88	-15.29	2.496	0.152
TPS-SIFT	0.971	-0.26	0.263	0.963	-30.18	6.979	2.992	0.105
SURF	0.225	0.147	-0.06	0.191	121.1	207.65	3.977	0.144
PCA-SIFT	-0.07	0.035	0.011	-0.09	150.45	110.76	4.36	0.016
MSER	0.537	1.019	0.765	-0.84	-37.48	73.52	3.766	0.013

The result of multiview image registration is shown below in the figure 9.

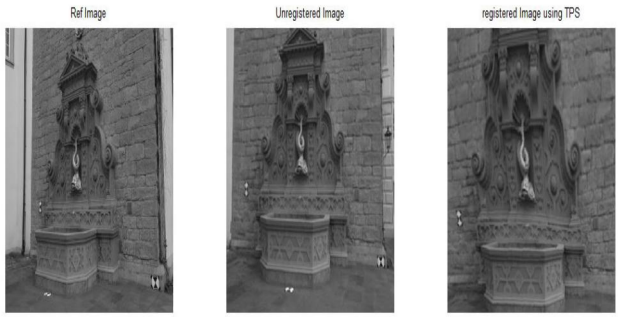

Fig. 9. Using SIFT and TPS

The registration process is repeated for multiview images using the same algorithms along with SIFT and TPS transformation. The result of the experiment is given in the table 6.1. There is an increase in RMSE for all algorithms except SIFT-TPS and MI is maximum for SIFT-TPS. So it is clear from the table that SIFT-TPS is well suitable for multiview image registration.

Table 2: Values for multiview image registration

Methods	a11	a12	a21	a22	x	y	RMSE	MI
SIFT	1.047	-0.11744	0.0017	0.972	56.68	-13.087	24.342	0.157
TPS-SIFT	1.2517	-0.478	0.049	1.006	166.930	-33.64	2.788	0.159
SURF	0.287	-0.076	0.061	0.516	216.43	125.20	142.89	0.155
PCA-SIFT	0.007	-0.0098	0.0065	0.0020	133.29	125.65	101.85	0.0105
MSER	0.2118	0.1247	0.294	-0.092	48.33	122.58	74.6491	0.0103

3. RESULT AND CONCLUSION

The work of different feature descriptor algorithms in image registration is analysed and evaluate the performance of it. It is compared with the proposed systems. By analysing each of the results it is clear that each of them have its own importance, advantages and limitations. Accordingly, an accuracy is higher for the approach using SIFT in terms of RMSE has been achieved, which can also be demonstrated by the largest value of MI (0.152). Moreover, visual inspection of the registered image validates that the transformed sensed image fits well with the reference image across the whole image overlap. Algorithms must be chosen based on the application. However when it comes to multiview images affine transformation is suitable hence a new approach of using TPS transformation with SIFT is proposed.

In this section defines various steps involved in TPS transformation and also the result along with it. With TPS Transformation it gives a larger 2D view. The table shows that SIFT algorithm with TPS transformation is the most suitable and robust for multiview image registration. Accordingly, accuracy in terms of rmse has been achieved, which can also be demonstrated by the largest value of MI (0.159). The result of this experiment is not constant for all cases. SIFT`s matching success attributes to that its feature representation has been carefully designed to be robust to localization error. PCA is known to be delicate to registration error. Using a small number of dimensions provides significant benefits in storage space and matching speed. SURF demonstrates its strength and quick speed in the analyses. It is realized that 'Quick Hessian' locator that utilized as a part of SURF is more than 3 times quicker that DOG (which was utilized as a part of SIFT) and 5 times speedier than Hessian-Laplace. SURF looks fast and good in most situations, but when the rotation is large, it also needs to improve this performance. SIFT demonstrates its dependability in every one of the analyses with the exception of time and it can recognize a large number of keypoints and discovers such a large number of matches.

SIFT has been shown to outperform the MSER approach due to the robustness and distinctiveness of the SIFT descriptors. Notwithstanding this, MSER, which uses a low dimensional descriptor, has been shown to successfully match certain series. However, it did not generally perform very well and was not sufficiently robust for registration in comparison with SIFT.

In the previous work the registration of multiview images with the difference in the acquisition angle and the terrain elevation, the affine transformation model applied is not suitable. So a more appropriate transformation models such as thin-plate spline is substituted the affine model in the proposed coarse-to-

fine scheme to handle the effects introduced by the acquisition angle and terrain elevation differences.
An excellent outlier removal method should be able to eliminate most false matches and preserve most correct matches as well.

REFERENCES

[1] M. Gong, S. Zhao, L. Jiao, D. Tian, and S. Wang, \A novel coarse-to-_ne scheme for automatic image registration based on sift and mutual information," Geoscience and Remote Sensing, IEEE Transactions on, vol. 52, no. 7, pp. 4328{4338, 2014.

[2] 2. H.-M. Chen, M. K. Arora, and P. K. Varshney, \Mutual information-based image registration for remote sensing data," International Journal of Remote Sensing, vol. 24, no. 18, pp. 3701{3706, 2003.603. \AI Shack," 2013. http://aishack.in/tutorials/siftscaleinvariant-featuretransform-log-approximation/.

[3] B. Zitova and J. Flusser, \Image registration methods: a survey," Image and vision computing, vol. 21, no. 11, pp. 977{1000, 2003.

[4] AI Shack," 2013. http://aishack.in/tutorials/sift-scale-invariant-featuretransform-features/.

[5] T. Gao, Y. Xu, T.-x. Xu, and L. Shuai, \Multi-scale image registration algorithm based on improved sift," Journal of Multimedia, vol. 8, no. 6, pp. 755{761, 2013.

[6] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, Speeded-up robust features (surf)," Computer vision and image understanding, vol. 110, no. 3, pp. 346{359, 2008.

[7] M. O'Malley and H. Al Hadad, Automatic registration of large images from light microscopy. Lund University.

[8] 2006. http://step.polymtl.ca/ rv101/thinplates/.

[9] K. Besbes and R. Bouchiha, \Automatic remote-sensing image registration using surf," 2013. 57

[10] J. Ma, J. C.-W. Chan, and F. Canters, \Fully automatic subpixel image registration of multiangle chris/proba data," Geoscience and Remote Sensing, IEEE Transactions on, vol. 48, no. 7, pp. 2829{2839, 2010.

[11] L. Cheng, J. Gong, X. Yang, C. Fan, and P. Han, Robust a_ne invariant feature extraction for image matching," Geoscience and Remote Sensing Letters, IEEE, vol. 5, no. 2, pp. 246{250, 2008.

[12] A. Cole-Rhodes, K. L. Johnson, J. LeMoigne, I. Zavorin, et al., \Multiresolution registration of remote sensing imagery by optimization of mutual information using a stochastic gradient," Image Processing, IEEE Transactions on, vol. 12, no. 12, pp. 1495{1511, 2003.

[13] S. Suri, P. Schwind, P. Reinartz, and J. Uhl, Combining mutual information and scale invariant feature transform for fast and robust multisensory sar image registration," in 75th Annual ASPRS Conference, 2009.

[14] Y. S. Heo, K. M. Lee, and S. U. Lee, \Mutual information-based stereo matching combined with sift descriptor in log-chromaticity color space," in Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, pp. 445{452, IEEE, 2009.

[15] A. Goshtasby, G. C. Stockman, and C. V. Page, \A region-based approach to digital image registration with subpixel accuracy," Geoscience and Remote Sensing, IEEE Transactions on, no. 3, pp. 390{399, 1986.

[16] J. Canny, \A computational approach to edge detection," Pattern Analysis and Machine Intelligence, IEEE Transactions on, no. 6, pp. 679{698, 1986.

[17] Z. Zheng, H. Wang, and E. K. Teoh, \Analysis of gray level corner detection," Pattern Recognition Letters, vol. 20, no. 2, pp. 149{162, 1999.