A Coarse-to-Fine Approach for Remote-Sensing Image Registration Based on a Local Method

Sang Rok Lee
School of Information and Mechatronics
Gwangju Institute of Science and Technology
261 Chemdan-gwagiro, Buk-gu, Gwangju, Republic of Korea
Emails: leesr@gist.ac.kr

Abstract- Multispectral satellite imagery registration is a fundamental step for remote sensing applications such as global change detection, feature classification, and image fusion. Since image registration via the manual selection of control points is a repetitive and time-intensive task, a more efficient automatic coarse-to-fine algorithm for multispectral remote sensing image registration is proposed in this paper. First, for the coarse registration, the Haar Wavelet Transform (HWT) is adopted to produce lower-resolution levels of reference and input images; then, the Speeded Up Robust Features (SURF) algorithm is utilized for quickly searching for matching points. After the coarse registration is completed, the Harris operator is used to extract feature points, and initial correspondences are established using the normalized cross-correlation to achieve the fine registration. Second, in order to remove mismatched points, the RANdom SAmple Consensus (RANSAC) algorithm is applied to putative correspondences. Due to large amount of satellite image data available, we used block processing in the refined registration to increase the efficiency in memory use. Finally, the final transformation function is obtained via the local weighted mean method in order to deal with local geometric differences between the reference and input images. Compared to global registration by manually selecting the control points, the proposed method is fully automatic and computationally efficient. Experimental results with well-known data sets (Worldview-1,2 and Quickbird remote-sensing images) again demonstrate the accuracy and effectiveness of the proposed algorithm for multispectral remote-sensing image registration.

Index terms*:* Multispectral satellite image, Image registration, SURF, RANSAC

I. INTRODUCTION

Image registration in multispectral satellite images is a crucial problem for remote sensing applications, and remains challenging because of the inherent nonlinearity in intensity changes. Image registration is the process of overlaying two or more images of the same scene taken at different times, from different viewpoints and by different sensors [1]. Specifically, its objective is to determine a transformation function that is similar to the actual distortion function. During the image registration process, traditional methods in satellite imagery often manually select control points that are then used to determine the parameters of a transformation function. The main drawback to these approaches is that they are very laborious and time-consuming, in addition to the fact that an expert is required to choose the control points in the remotely sensed images. Therefore, automatic registration of remote sensing images is highly desired.

In general, existing image registration techniques can be divided into two categories: intensity-based and feature-based methods [1]. In intensity-based methods, the similarities between pixel intensities are used to determine the transformation function between two images. Frequently, mutual information and maximum likelihood are used as a similarity measures in these methods. However, intensity-based methods are not suitable for multispectral satellite image registration due to presence of different electromagnetic reflectances, i.e., the intensity values do not change linearly [4]. In contrast, feature-based methods are more appropriate for images with large distortions or those obtained from different sensors.

In this paper, we limit our discussion to feature-based methods and use feature points to register satellite images. The major steps in feature-based image registration techniques include the following. First, control points from the reference and sensed images are detected and matched. Second, parameters of the transformation function are estimated using the previous control points. Finally, the estimated transformation model registers the sensed image to the reference image.

In literature, there have been a number of recent attempts to develop an automatic registration algorithm for remote sensing images. For instance, Yi et al. [4] proposed a Scale Invariant Feature Transform (SIFT) based on a scale restriction for satellite image registration. In addition, Song and Zhang [3] retrofitted the Speeded Up Robust Features (SURF) algorithm with an effective criterion for a higher correct matching rate and defined a new similarity measure based on trajectories generated from Lissajous figures. Previously, Kim and Im [2] employed the

normalized cross correlation to estimate the transform model for matching control points and used the RANdom SAmple Consensus (RANSAC) algorithm to reject outliers. Note that the RANSAC algorithm was first presented by Fischler and Bolles [8] to fit a model to experimental data; Goshtasby [5] then introduced piecewise linear mapping functions in which affine transformations could be applied separately to each triangular region of the image.

In this paper, we propose a coarse-to-fine registration algorithm for satellite images. First, in order to coarsely register satellite images, we obtain a set of control point pairs by applying SURF, and then compute the parameters of an affine transformation model. Once the coarse registration is completed, more control points are extracted by a Harris corner detector to refine the registration. Finally, the transformation function is obtained via the local weighted mean method. Since satellite images contain local geometric distortions caused by spatially variant terrain relief and by different satellite viewpoints, one transformation function alone cannot accurately register; multiple function are required.

The remainder of paper is organized as follows. Section II presents some preliminaries for the proposed algorithm, and Section III introduces the proposed algorithm in further detail. Next, in Section IV we present some experiments using high-resolution satellite images. Finally, conclusions are made in Section 5.

II. PRELIMINARIES

In this section, in order to make the paper self-contained, we provide a set of preliminary information. First, we explain how to extract feature points from the reference and the sensed images. Then, the local weighted means used for image registration in this paper are provided.

The Harris operator, also called the Plessey corner detector, is a well-known interest point detector and has been used for many years in computer vision [7]; however, it does not incorporate descriptors. This detector first computes a Harris matrix A at each pixel in an image.

$$\mathrm{A} = \begin{bmatrix} \langle I_x^2 \rangle & \langle I_x I_y \rangle \\ \langle I_x I_y \rangle & \langle I_y^2 \rangle \end{bmatrix}, \tag{1}$$

where I_x^2, I_y^2 and $I_x I_y$ are the second-order derivatives of the image intensities in the x, y, and xy directions. Here, the angle brackets denote averaging.

When the eigenvalues λ_1, λ_2 of Harris matrix A are rather large, an interest point is identified. To reduce the computational complexity, the eigenvalue computation can be replaced by the following response function [7].

$$R(\mathrm{A}) = \det(\mathrm{A}) - \kappa \mathrm{trace}^2(\mathrm{A}), \tag{2}$$

where κ is a tunable parameter whose values in the range 0.04 to 0.15 have been reported as feasible in literature.

If $R(\mathrm{A}) > 0$, then a corner is found. The Harris corner detector is invariant to rotation but not to larger scale changes.

The SURF algorithm not only detects interest points but also computes the descriptors. In addition, it is invariant to scale changes, 2D shift, and rotation. The SURF detector is based on a Hessian matrix H as function of point $X = (x, y)$ in an image and scale σ [10].

$$H(X,\sigma) = \begin{bmatrix} L_{xx}(X,\sigma) & L_{xy}(X,\sigma) \\ L_{xy}(X,\sigma) & L_{yy}(X,\sigma) \end{bmatrix}, \tag{3}$$

where $L_{xx}(X,\sigma)$ refers to the convolution of the second order Gaussian derivative $\frac{\partial^2 g(\sigma)}{\partial x^2}$ with the image at point X, and similarly for $L_{xy}(X,\sigma)$ and $L_{yy}(X,\sigma)$.

SURF approximates the Laplacian of Gaussian (LoG) by using box filter representations of the respective kernels, in contrast to SIFT, which approximates LoG using a Difference of Gaussian (DoG) [9]. By using box filters in conjunction with the integral image, SURF is several times faster and more robust against different image transformations than SIFT [11]. Specifically, the approximated determinant of the Hessian is given as

$$\det(H_{approx}) = D_{xx}D_{yy} - (0.9D_{xy})^2, \tag{4}$$

where D_{xx}, D_{yy}, and D_{xy} denote the weighted box filter approximations in the x, y, and xy directions, respectively.

The determinant of the Hessian matrix is then used in selecting the location and scale. For this task, non-maximal suppression in a $3\times3\times3$ neighbourhood is first used to find a set of candidate points.

Here, the SURF descriptor describes how the pixel intensities are distributed within a scale dependent neighbourhood of each interest point detected by the Fast-Hessian filter. To achieve invariance to rotation, an orientation is assigned to each detected point of interest using Haar

wavelet responses. Next, SURF descriptors are constructed by delineating a square window around the interest points in the direction of orientation. The descriptor window is then divided into 4×4 sub-regions. For each of these sub-regions, Haar wavelet responses are calculated for 25 uniformly distributed sample points. Finally, if the wavelet responses in the x and y directions, denoted by dx and dy, are summed, the feature vector for each sub-region is given by

$$v_{subregion} = [\sum dx, \sum dy, \sum |dx|, \sum |dy|]. \tag{5}$$

Therefore, the descriptor vectors from the 4×4 sub-regions result in length of $4\times4\times4 = 64$.

The local weighted mean is a local transformation function used to provide smooth transition across adjacent areas in a resampled image and only requires information about local control points to register local areas in the images [6].

Given n control point pairs such as (x_i, y_i) in the reference image and (u_i, v_i) in the sensed image, the weight function can be calculated as

$$W_i(R) = \begin{cases} 1-3R^2+2R^3, & 0 \le R \le 1 \\ 0, & \mathrm{R} > 1 \end{cases}, \tag{6}$$

$$R = \sqrt{[(x-x_i)^2 + (y-y_i)^2]} \ / \ R_n, \tag{7}$$

where R_n is the distance from (x_i, y_i) to the $(n-1)$th nearest point in the reference image.

By using the weighted mean method, we can obtain a transformation function based on the weighted mean of local polynomials.

$$x = \frac{\sum_{i=1}^{n} W_i(u,v) P_i(u,v)}{\sum_{i=1}^{n} W_i(u,v)}, \quad y = \frac{\sum_{i=1}^{n} W_i(u,v) Q_i(u,v)}{\sum_{i=1}^{n} W_i(u,v)}, \tag{8}$$

where $P_i(u,v)$ and $Q_i(u,v)$ are elements of the local transformation functions that project the i th point and $(n-1)$ of its closest control points in the input image to matching control points in the reference image.

III. COARSE-TO-FINE APPROACH

In this paper, the coarse-to-fine registration process consists of two main steps. As a pre-registration procedure, the initial coarse registration roughly registers the input image to the

reference image via a Haar wavelet transform in order to produce lower-resolution levels of images, where SURF is used to automatically extract control points. After this coarse registration, a fine registration process based on local weighted mean is performed, using the control points detected by a Harris corner detector. Further details of the overall procedure of the proposed algorithm are provided in the following sections.

The coarse registration process is performed as follows. It is carried out in the following three steps.

1) The first step in this process is to perform the Haar wavelet transform of the reference and sensed images. This transform is used to construct an image pyramid, due to its simplicity, speed, and accuracy. Specifically, the wavelet coefficients can be written as

$$a(m,n) = h_{\varphi}(-m) * [h_{\varphi}(n) * f\,|_{n=2i}]\,|_{m=2i}, \tag{9}$$

$$d^{j}(m,n) = h_{\varphi,\psi}(-m) * [h_{\varphi,\psi}(-n) * f\,|_{n=2i}]\,|_{m=2i}, \tag{10}$$

where $a(m,n)$ is an approximation and $d^{j}(m,n)\,|_{j=H,V,D}$ are the horizontal, vertical, and diagonal detail subbands of the input images. The input images are decomposed into the above wavelet coefficients up to third level using Haar wavelets.

2) In the second step, feature points in both images are extracted and matched using third-level approximation subimages, as input into the SURF algorithm.

3) Finally, given the two sets of corresponding control points from Step 2, the parameters of an affine transformation are estimated. The affine transformation is defined as

$$x = a\,u + b\,v + c, \tag{11}$$

$$y = d\,u + e\,v + f. \tag{12}$$

Based on the coarse transformation model, the sensed image is transformed. However, the aim of this process is not perfect registration.

The objective of fine registration is to estimate the final transformation model. To achieve this task, more control points are extracted by the Harris corner detector and matched using the normalized cross-correlation. After the matching control point pairs have been identified, RANSAC is used to eliminate the outlier control point pairs. The fine registration process includes the following steps and uses 10×10 block processing to reduce memory requirements.

1) The first step of the algorithm is to detect corner features in each image using the Harris corner detector (2).

2) A set of control point pairs is then computed using the normalized cross-correlation. The implementation closely follows the following formula.

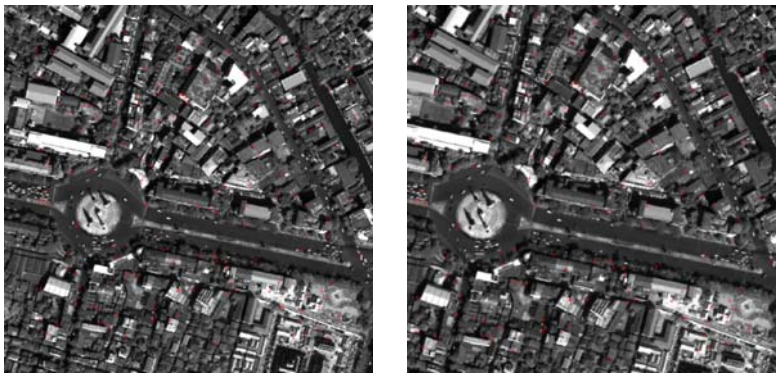
$$\gamma(u,v)=\frac{\sum_{x,y}[f(x,y)-\overline{f}_{u,v}][t(x-u,y-v)-\overline{t}]}{\left\{\sum_{x,y}[f(x,y)-\overline{f}_{u,v}]^2\sum_{x,y}[t(x-u,y-v)-\overline{t}]^2\right\}^{0.5}} \tag{13}$$

3) Next, the RANSAC algorithm is applied to the putative correspondences to estimate the homography and the inlier correspondences, which are consistent with this estimate because many of the putative correspondences obtained in the previous step are incorrect [8]. Note that though most incorrect correspondences have been removed, a few may still remain. Hence, in order to reclassify the inliers using an improved estimate of the homography, the maximum likelihood(ML) cost function minimization is carried out using the Levenberg-Marquardt algorithm, which provides a numerical solution to the non-linear optimization problem. The cost function is chosen as

$$\sum_i d(x_i,\hat{H}u_i)^2 \tag{14}$$

where x_i and u_i represent 2D point correspondences in the reference and sensed images, respectively. In addition, $\hat{H}$ is the estimated homography, and $d(x_i,\hat{H}u_i)$ denotes the Euclidean distance between x_i and $\hat{H}u_i$. Fig. 1 shows the results of outlier rejection using RANSAC.

4) Given the corresponding control point sets (x_i,y_i) and (u_i,v_i), the final registration of the sensed image with respect to the reference image can then be accomplished using the local weighted mean (8).



(a) (b)

Fig. 1. Control point detection on two Worldview-1 panchromatic image patches of a region of Bangkok, Thailand, obtained using the proposed algorithm. (Image courtesy of Digital Globe.) (a) The set of selected control points (red crosses), with the reference image superimposed. (b) The corresponding set of the selected control points (red crosses), with the sensed image superimposed.

IV. EXPERIMENTAL RESULTS

Three sets of remote sensing images derived from Digital Globe, Inc. were used to test our algorithm. The dataset can be divided into two categories: 1) panchromatic image pairs and 2) multispectral image pairs. Each test set consists of a reference and sensed image. Information about each dataset is provided below.

a. Test Data 1

Worldview-1 panchromatic images acquired on Jan. 3, 2009 over Bangkok, Thailand are shown in Fig. 2. The pair of Fig. 2(a) and (b) have 0.5 m resolution at nadir and an 11 bit per pixel dynamic range. In this test set, there is an urban area that was used to test the algorithm capability of processing a dense area. Both images are 4500×4499 pixels in size.

b. Test Data 2

Worldview-2 panchromatic and coastal blue spectral band images acquired on Dec. 10, 2009 over Rome, Italy are shown in Fig. 3. A coastal blue band is useful in bathymetric studies and the focus is on the 400–450 nm wavelength. The panchromatic and multispectral images have 0.46 m

and 1.84m spatial resolution at nadir respectively. In this case, the coastal blue band image of low spatial resolution was registered to the high resolution panchromatic image. The images in Fig. 3(a) and (b) are 4600 $\times$ 4604 and 1150 $\times$ 1151 pixels in size.

c. Test Data 3

QuickBird panchromatic and near-infrared band images acquired on Jul. 4, 2005 over Boulder, USA are shown in Fig. 4. The near-infrared band has been shown to be effective for estimating moisture content and plant biomass in the 760–900 nm wavelength range. In addition, the panchromatic and multispectral images have 0.6 m and 2.4 m spatial resolution at nadir, respectively. Here, the near-infrared band image of low spatial resolution was registered to the high resolution panchromatic image. The images in Fig. 4(a) and (b) are 3310 $\times$ 3260 and 937 $\times$ 915 pixels in size.

(a) (b) (c)

Fig 2. Registration of Worldview-1 panchromatic images of Bangkok, Thailand acquired at different angles. (Image courtesy of Digital Globe.) (a) The reference image. (b) The sensed image. (c) The registration result using the proposed method.

(a) (b) (c)

Fig. 3. Registration of Worldview-2 panchromatic and coastal blue band image of Rome, Italy. (Image courtesy of Digital Globe.) (a) The reference image. (b) The sensed image. (c) The registration result using the proposed method.

(a) (b) (c)

Fig 4. Registration of Quickbird panchromatic and near-infrared band images of Boulder, USA. (Image courtesy of Digital Globe.)(a) The reference image. (b) The sensed image. (c) The registration result using the proposed method.

d. Performance Evaluation

To evaluate the accuracy of the proposed registration approach, the Root Mean Square Error (RMSE) was calculated using the control points obtained by the proposed matching step,

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{n}\left\| X_i - \hat{X}_i \right\|^2}, \tag{15}$$

where is the total number of matched control points, X_i are the (x_i, y_i) coordinates in the reference image, and $\hat{X}_i$ are the estimated $(\hat{x}_i, \hat{y}_i)$ based on the final transformation model.

The RMSE results of the proposed approach are then compared to those obtained using manually selected control point pairs, as shown in Table 1. For an objective comparison, we manually selected 10 pairs of control points to register the two images. In the table, it can be observed that image registration using the proposed algorithm results in a lower RMSE than manual method. Also, the results show that all RMSE values are less than 2.8 pixels in the proposed automated algorithm. For a further assessment, the visual results of the three experiments (Worldview-1, Worldview-2, and Quickbird) are shown in Figs. 2(c), 3(c), and 4(c), overlapping the reference and registered images. Via a visual inspection, it can be seen that the registration results is accurate and valid for all test cases.

Experiment	RMSE	
	Proposed method	Manual method
Worldview-1	2.6548	3.4231
Worldview-2	2.7564	3.6789
Quickbird	2.6124	3.5434

Table 1. RMSE values for the proposed and manual methods.

V. CONCLUSION

In this paper, We proposed the automatic coarse to fine algorithm for a multispectral remote sensing image registration. To register the sensed image to the reference image coarsely, our algorithm incorporate the wavelet transform to exploit the low frequency image at lower-resolution level. Also by using the Harris corner combined with RANSAC, accurate feature points are detected and matched in fine registration process so that it eliminates incorrectly matched points in multispectral images which have intensity changes. We tested the proposed approach using three sets of satellite images derived from DigitalGlobe, Inc. Test results show that the proposed method based on local weighted mean is more accurate and effective for multispectral satellite imagery registration than a global registration by manual selection.

REFERENCES

[1] B. Zitova and J. Flusser, “Image registration methods: A survey”, Image Vis. Comput., Vol. 21, No, 11, pp. 977-1000, 2004.

[2] T. Kim and Y. Im, “Automatic satellite image registration by combination of matching and random sample consensus”, IEEE Trans. on Geosci. Remote Sens., Vol. 41, No. 5, pp. 1111-1117, 2003.

[3] Z. L. Song and J. Zhang, “Remote sensing image registration based on retrofitted SURF algorithm and trajectories generated from Lissajous figures”, IEEE Geosci. Remote Sens. Lette., Vol. 7, No.3, pp. 491-495, 2010.

[4] Z. Yi, C. Zhiguo and X. Yang, “Multi-spectral remote image registration based on SIFT”, IEEE Electronics Letter, Vol. 44, No. 2, pp. 107-108, 2008.

[5] A. Goshtasby, “Piecewise linear mapping functions for image registration”, Pattern Recognition, Vol. 19, No. 6, pp. 459-466, 1986.

[6] A. Goshtasby, “Image registration by local approximation methods”, Image and Vision Computing, Vol. 6, No. 4, pp. 255-261, 1988.

[7] C. Harris and M. Stephens, “A combined corner and edge detector”, in Proc. Alvey Vis. Conf., pp. 147-151, 1988.

[8] M. Fischler and R. Bolles, “Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography”, Commun. ACM, Vol. 24, No. 6, pp. 381-395. 1981.

[9] D. Lowe, “Distinctive image features from scale-invariant keypoints”, Int. J. Comput. Vis., Vol. 60, No. 2, pp. 91-110, 2004.

[10] H. Bay, A. Ess, T. Tuytelaars and L. Van Gool, “SURF: Speeded-up robust features”, Comput. Vis. Image Underst., Vol. 110, No. 3, pp. 346-359, 2008.

[11] J. Bauer, N. Sunderhauf and P. Protzel, “Comparing several implementations of two recently published feature detectors”, in Proc. Int. Conf. Intelli. Autonomous Systems, 2007.

[12] Y. Bentoutou, N. Taleb, K. Kpalma, and J. Ronsin, “An Automatic Image Registration for Applications in Remote Sensing”, IEEE Trans. on Geosci. Remote Sens., Vol. 43, No. 9, pp. 2127-2137, 2005.

[13] S. Leprince, P. Muse, and J. Avouac, “In-Flight CCD Distortion Calibration for Pushbroom Satellites Based on Subpixel Correlation”, IEEE Trans. on Geosci. Remote Sens., Vol. 43, No. 9, pp. 2675-2683, 2008.

[14] H. Rimminen, J. Lindstrom, and R. Sepponen, “Positioning Accuracy And Multi-Target Separation With A Human Tracking System Using Near Field Imaging”, International Journal On Smart Sensing and Intelligent, Vol. 2, No. 1, 2009.

[15] J. Lee, and K. Yao, “Multi-Target, Multi-Sensor Tracking Based on Quality-of-Information And Formal Bayesian Frameworks”, International Journal On Smart Sensing and Intelligent, Vol. 1, No. 4, 2008.