AUTOMATIC REGISTRATION OF HIGH-RESOLUTION IMAGES IN URBAN AREAS USING LOCAL PROPERTIES OF FEATURES

Youkyung Han*, PhD Student
Yongmin Kim*, PhD Student
Younggi Byun*, PhD
Jaewan Choi*, PhD
Dongyeob Han,** Assistant Professor
Yongil Kim*, Professor
*Department of Civil and Environmental Engineering
Seoul National University
Seoul 151-742, Korea
han602@snu.ac.kr, kym0210@snu.ac.kr, kko071@snu.ac.kr, choijw11@snu.ac.kr, yik@snu.ac.kr
**Department of Civil and Environmental Engineering
Chonnam National University
Yeosu 550-749, Korea
hozilla@chonnam.ac.kr

ABSTRACT

We propose an automatic image-to-image registration of high-resolution satellite images using local properties and geometrical locations of matching points to improve the registration accuracy. First, coefficients of global affine transformation between images are extracted using a scale-invariant feature transform (SIFT)-based method, and features of the sensed image are transformed to the reference coordinate system using these coefficients. Then, a spatial distance between a feature of the reference and features of the sensed images that have been transformed to the reference coordinates within a predefined buffer is additionally used to extract precise matching points. Finally, the spatial distance integrated with Euclidean distances of invariant vectors is employed for local matching. The optimal ranges of the proposed distance and the radius of the buffer for local matching are determined using a registration consistency measure. The average orientation differences between matching points of the two images are used for outlier elimination. A mapping function model consisting of an affine transformation and piecewise linear functions is applied to the matching points for automatic registration of high-resolution images. The proposed method can extract precise matching points and gives better registration results than the SIFT-based method alone.

KEYWORDS: automatic registration, high-resolution satellite image, registration consistency, scale-invariant feature transform (SIFT)

INTRODUCTION

Image registration is the process of geometrically overlaying two or more images of the same scene (Zitová and Flusser, 2003). It is an essential stage in many fields such as remote sensing, medical imaging, and computer vision. In remote sensing, it is one of the fundamental preprocessing steps for various applications such as image fusion, change detection, map updating, and so on. The tie-points, however, are usually extracted manually, consuming much time and labor (Kennedy and Cohen, 2003; Schowengerdt, 1997). Automatic image-to-image registration has therefore received much attention.

The majority of image registrations can be categorized by the following four steps; feature extraction, feature matching, transformation model estimation, and image registration (Zitová and Flusser, 2003). Automatic image registration is generally achieved by automating steps feature extraction and matching. A feature is an object that is distinctive and can be extracted from the points in the image. Features can be extracted by their representative points, such as the center of gravity of a homogeneous region, line endings, or corners. When features are extracted from images, they should be matched across the images. Each feature is described by vectors that sustain its essential properties such as invariance, uniqueness, stability, and independence, and the extractor finds corresponding features that have similar properties. When the corresponding pairs are matched, a transformation model is estimated using these pairs.

Recently, the necessity for higher resolution of satellite image and its precise registration has increased. High spatial resolution images, however, have complicated geometrical characteristics. They have relief displacement and local distortion because of objects with different heights, so have many errors if a rigid transformation is applied as the registration model (Hong and Zhang, 2008). Therefore, nonrigid transformations such as thin-plate splines or piecewise linear functions can be used for registration between high-resolution images.

There have been many studies of automatic registration between high-resolution images. The scale-invariant feature transform (SIFT) method, which was introduced by Lowe (2004) for extracting and matching features, is representative (Li *et al*., 2009; Song *et al*., 2010; Yang *et al*., 2007; Yu *et al*., 2008). Edge-based selection of the most distinctive control points (Bentoutou *et al*., 2005), and an approach combining phase correlation and normalized cross-correlation (Liu and Yan, 2008) have also been proposed. Combined area-based and feature-based registration methods have been proposed (Hong and Zhang, 2008; Zhang and Fraser, 2007). Area-based matching algorithms have been found useful for image scenes where no distinctive features are available (Du *et al*., 2008; Liu *et al*., 2006). A new algorithm for interest-point matching of high-resolution satellite images has also been proposed (Xiong and Zhang, 2009). These methods, however, commonly have a limitation when the algorithms are applied to high-resolution images of urban areas that have many high buildings and their shadows. The cast shadow region and exposed sides of buildings can change from image to image because of the characteristics of images such as acquisition time, observation pose, and off-nadir observation angle. The matching points extracted from these objects, therefore, should be removed when searching for matching-point pairs because they can cause registration errors.

The objective of this paper is to extract precise matching points to increase the accuracy of automatic image-to-image registration of high-resolution data. First, we extract features using the SIFT method, and the locations of features in the sensed image are transformed to the reference coordinate system obtained by affine transformation. Then, the spatial distances between features of the reference and sensed images are additionally used for a proposed similarity measure. The range of parameters suitable for local matching is found from registration consistency. The matched points extracted from shadow regions or objects that have height variations such as buildings or trees are removed using orientation differences between images. Finally, the matching points are subjected to combined piecewise linear functions and affine transformation for precise registration.

METHODOLOGY

The proposed algorithm uses only the features extracted using the local maximum of the magnitude of the Laplacian-of-Gaussian operator in both spatial and scale dimensions. The original feature descriptor and similarity measure used by Lowe (2004) tend to miss matching points that actually are matching pairs. The proposed method can extract these additional matching points.

The flowchart of the proposed method is shown in Figure 1. First, feature points of the sensed image and reference image are respectively extracted using the local maximum of the magnitude of the Laplacian-of-Gaussian operator in both the spatial and scale dimensions. The extracted features are identified by the SIFT descriptor. Feature matching between images occurs by taking the features from one image and using their descriptors to index into the data structure for the other image. The descriptor distance, measured as the Euclidean distance between vectors, is computed for each candidate match. The two closest matches for each descriptor are found and the ratio of the distances to the closest and second closest is calculated; if it is less than the predefined threshold, then these features become candidate matching points. Matched pairs of features that have large root mean squared error (RMSE) are removed from the matching-point sets, and the affine coefficients between the two images are estimated by using the remaining matching points in the least-squares method.

Next, the location of all features in the sensed image is transformed to the reference coordinate system. A spatial distance, which means the Euclidean distance between features of the reference image and features of the sensed image that are transformed to the reference image's coordinates by the affine transformation, is calculated. A proposed distance that combines the descriptor distance with the spatial distance is used for local matching. A specific spatial distance from each feature in the reference image that defines a circular buffer zone is applied during matching to extract the precise matching points and reduce the computation time. When one feature in the reference image is i and one of the features in the sensed image that is within the buffer zone of i is j, we calculate the distance for local matching as:

$$D(i,j) = ED(i,j)(1 + ND(i,j)) \tag{1}$$

where $ED(i,j)$ is the Euclidean distance between the two 128-dimensional descriptor, and $ND(i,j)$ is the spatial distance between two features in the sensed and reference image respectively that is normalized into [0, 1]. When the locations of features i and j are exactly equal, then $ND(i,j)$ becomes 0 and only the Euclidean distance between the description vectors is used for local matching. As $ND(i,j)$ increases, the possibility of matching as a matching point decreases. For one feature pair i and j that has less than the specific ratio of the closest to the second-closest distance, when the calculated $D(i,j)$ is smallest compared with other distances that were calculated for the feature i and the features of the sensed image within the buffer zone and smaller than a specific threshold, then this pair can be a matching point.

Appropriate ranges of the proposed distance, $D(i,j)$, and the radius of the buffer must be determined for precise local matching. A large number of matching points for the precise automatic image-to-image registration of high-resolution data is inevitable; however, matching points with inaccurate locations reduce the reliability of the transformation model. Therefore, we analyzed this trade-off relationship between the number of matching points and the reliability of the model by using registration consistency (Holden *et al*., 2000; Chen *et al*., 2003). In this paper, registration consistency is used as a measure to evaluate the performance of the implemented registration algorithm as the values of buffer and distance were being changed to find the optimal range. Defining $\mathrm{T}_{\mathrm{A,B}}$ as the transformation found by using image A as the sensed image and image B as the reference image, and $\mathrm{T}_{\mathrm{B,A}}$ for the reverse transformation, the registration consistency of $\mathrm{T}_{\mathrm{A,B}}$ and $\mathrm{T}_{\mathrm{B,A}}$ over the images A and B can be formulated as:

$$Registration\ Consistency = \frac{1}{N_A} \sum_{(x,y) \in (I_A \cap I)} \left\| (x,y) \quad T_{B,A} \circ T_{A,B}(x,y) \right\| \tag{2}$$

where (x,y) is the coordinate of a pixel in an image and the composition $\mathrm{T}_{\mathrm{B,A}} \circ \mathrm{T}_{\mathrm{A,B}}$ represents the transformation that applies $\mathrm{T}_{\mathrm{A,B}}$ first and then $\mathrm{T}_{\mathrm{B,A}}$. I is the overlap region of images A and B. I_{A} is the discrete domain of image A and N_{A} is the number of pixels of image A within the overlap region. If the value of the registration consistency is small, the constructed model between two images is relatively reliable. The difference between the numbers of matching points extracted between $\mathrm{T}_{\mathrm{A,B}}$ and $\mathrm{T}_{\mathrm{B,A}}$ is also used as an index for evaluating model reliability, because robust models tend to extract similar numbers of matching points regardless of the applied order. In remote sensing images, the cast shadow regions or objects such as buildings and trees that have height variation cause local distortion, so these objects should be excluded while extracting matching points. The difference in orientation between each matching-point pair is used for outlier elimination. The matching points extracted within shadows or from objects having height variation have orientation differences that vary according to the image's characteristics, compared with the general orientation difference of other matching points. Using the matching points extracted from the matching-point set, the average and standard deviation of orientation differences between the matching points of the two images are calculated. Then, a simple z-score test is used to detect the outliers. When the absolute z-value is greater than the threshold of one sigma, the matching point is judged as an outlier and eliminated from the matching-point sets.

Finally, the extracted matching points are used to construct the optimal triangulation that covers the convex hull of the points. Areas in the images inside the convex hulls are transformed by piecewise linear functions, and areas outside the convex hulls are transformed by a global affine transformation to increase the accuracy of registration. The piecewise linear functions deal with the registration process by dividing the images into triangular elements by Delaunay's triangulation method; these are then individually mapped through a linear transformation (Goshtasby, 1986). If matching points in the reference image are triangulated, corresponding triangles in the sensed image can be determined. Then, the affine transformation is used to map a region in the sensed image to the corresponding region in the reference image. With the piecewise linear functions, however, precise registration accuracy is only achieved within the convex hull of the points from which the triangles are obtained. For areas outside this convex hull, extrapolation must be performed, which often results in large errors (Arévalo and González, 2008). We therefore applied a global affine transformation to the areas outside the convex hull of the points; it was estimated from the points defining the convex hull. With this method, we tried to increase the registration accuracy not only inside the convex hull but also outside.

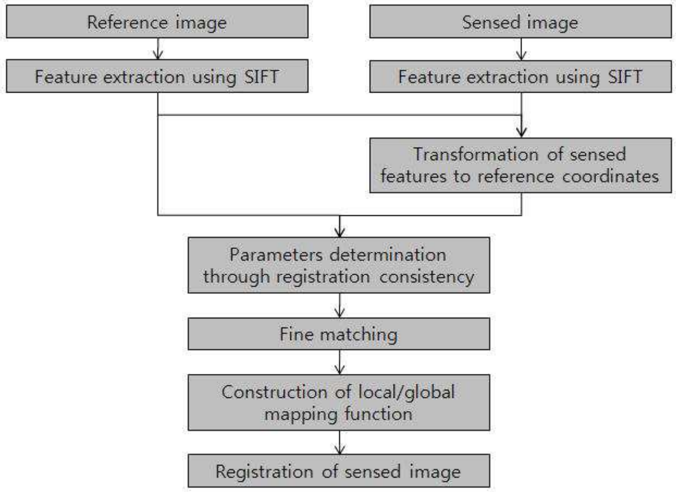


Figure 1. Flowchart of proposed method.

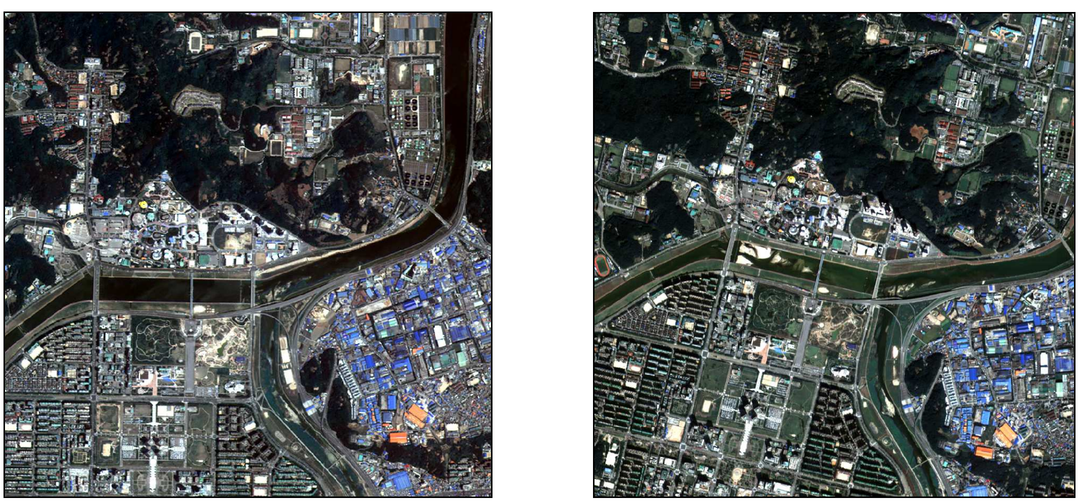
EXPERIMENTS AND DISCUSSION

Study Site

The study site for evaluating the proposed method includes an urban area, taken from KOMPSAT-2, located in Daejeon, South Korea (Figure 2). The reference and sensed images are temporally different multispectral images that have 4 m spatial resolution. The reference image was acquired on May 6, 2008, and the sensed image on October 5, 2007. All the images were normalized by histogram equalization in preprocessing to help extract the feature points. The specification of this site is summarized in Table 1.

Table 1. Sensor characteristics of study site

Sensor	Resolution		Date
KOMPSAT-2 (Multispectral)	4 m	Reference	2008/05/06
		Sensed	2007/10/05



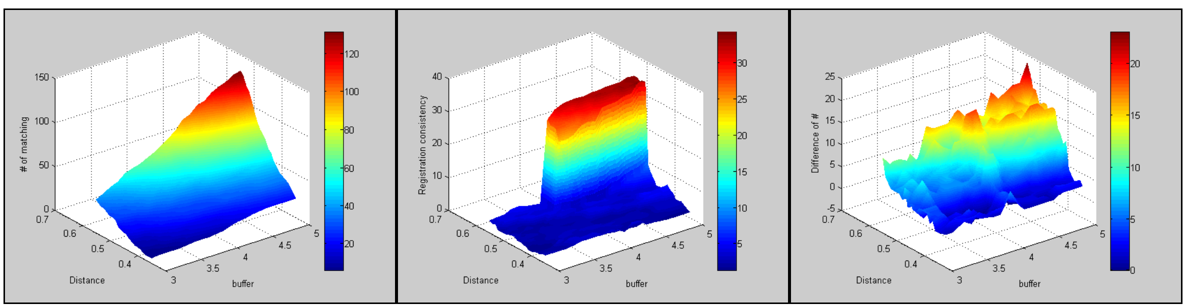
(a) (b)

Figure 2. Study site: (a) reference image and (b) sensed image.

Parameter Determination

As mentioned above, a competent registration result can be acquired when: i) the number of matching points is increased; ii) registration consistency is decreased; and iii) the difference between the numbers of matching points extracted from $T_{A,B}$ and $T_{B,A}$ is decreased. The number of matching points, registration consistency, and the difference between numbers of matching points were calculated from the study site as the values of buffer and distance were changed (Figure 3). Figure 3 (a) shows the number of matching points as the buffer and distance parameters were changed. When the distance and buffer parameters were increased, the number of matching points also increased. If the parameters are small, a reliable registration result cannot be obtained because few matching points are extracted, and the nonrigid transformation requires many matching points to ensure its successful application. In contrast, good registration consistency was not achieved when these two parameters were large (Figure 3 (b)). When the value of the distance exceeds 0.5, the registration consistency value becomes higher and is dramatically increased until it becomes meaningless, especially when the buffer value exceeds 3.8. While not strongly correlated between the number of points and the two parameters, the difference between the numbers of matching points and these parameters also show a similar relation. In particular, the small differences between the numbers of matching points in the range of interest are affected by the distance more than by the buffer size. Figure 3 (c) shows that reliable differences are presented when the distance value is less than 0.5, regardless of the buffer value.

These three results were combined to obtain a significant tendency. If the buffer and distance values are increased, a larger number of matching points is extracted while the registration consistency is high and the difference between the numbers of matching points is large, so that the registration accuracy and the model robustness decreased. Conversely, the reverse case can increase the registration's accuracy and model's robustness but the number of extracted matching points was too small to construct the nonrigid transformation model required for registering between high-resolution satellite images. From the experiments, we concluded that a reliable result is generally obtained when the buffer range is from 3.5 to 5.0, and the distance range is from 0.4 to 0.5.

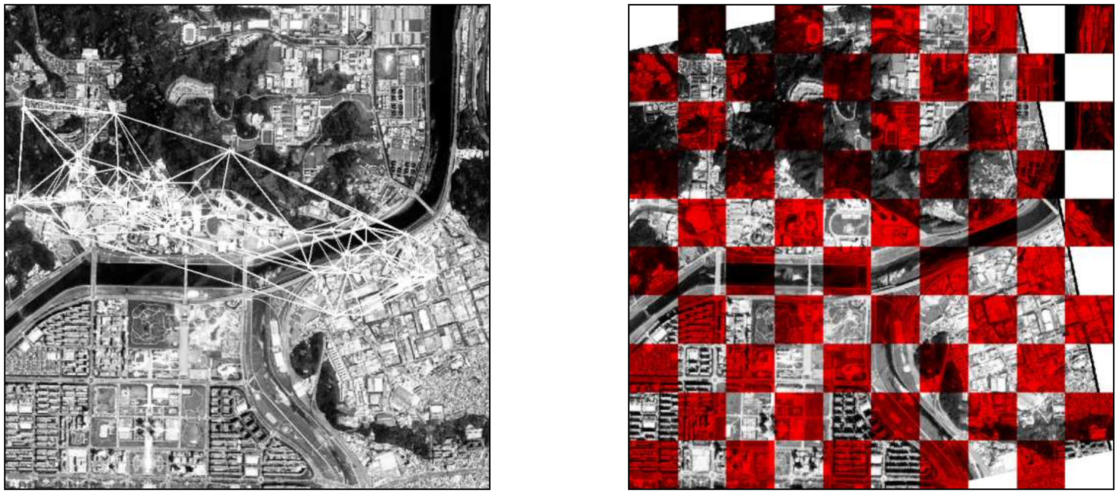


(a) (b) (c)

Figure 3. Evaluation of registration indexes for parameter determination: (a) the number of matching points; (b) registration consistency; and (c) the difference between numbers of matching points.

Registration Result

In the experiment, to extract enough matching points to register the sensed image reliably, the buffer and distance parameters were set as 5.0 and 0.5, respectively, which are the highest values within the ranges determined above. A total of 93 matching points were extracted, and triangular regions were constructed using these points (Figure 4 (a)). From Figure 4 (a), we can see that the outliers were effectively eliminated by using the orientation difference, because the matching points were not extracted from the regions such as buildings or hills that were high or had height variation. Figure 4 (b) shows the mosaic image generated by the proposed model that combines piecewise linear functions and a global affine transformation. Figure 5 is the magnified mosaic results for the visual assessment of the method. Figure 5 (a) is the magnified mosaic result from the SIFT method, and (b) is that from the proposed method. The result of the SIFT method is from the affine transformation estimated from the matching points that were removed as outliers having large RMSEs. The upper two rows of Figure 5 are magnified results from the area outside the triangular region, and the next two are those from inside that area. The proposed method shows more precise registration results than the SIFT method, regardless of whether the area is within the constructed triangulation.



(a) (b)

Figure 4. Registration result: (a) triangular construction on the reference image; and (b) mosaic result of proposed method.

CONCLUSION

We have proposed an automatic image-to-image registration algorithm for high-resolution satellite images using local properties and geometrical locations of features extracted by the SIFT method. The proposed method considers the spatial distance and orientation difference of matching point pairs between the reference and sensed images to extract precise matching points. We then apply a nonrigid transformation combined with piecewise linear functions and a global affine transformation. Suitable ranges of buffer and distance parameters can be determined using registration consistency. The proposed algorithm can extract precise matching points, and also shows a better registration result than the SIFT method alone. The two parameters can be adjusted within the predefined range according to the image's resolution or the sensor's properties to achieve reliable registration results. In future work, we will extend our experiments to other sensors having different resolutions or radiometric properties.

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 20100027762).

REFERENCES

Arévalo, V., and J. González, 2008. An experimental evaluation of non-rigid registration techniques on QuickBird satellite imagery. *International Journal of Remote Sensing,* 29(2):513–527.

Bentoutou, Y., N. Taleb, K. Kpalma, and J. Ronsin, 2005. An automatic image registration for applications in remote sensing. *IEEE Transactions on Geoscience and Remote Sensing,* 43(9):2127–2137.

Chen, H. -M, M. K. Arora, and P. K. Varshney, 2003. Mutual information-based registration for remote sensing data. *International Journal of Remote Sensing,* 24(18):3701–3706.

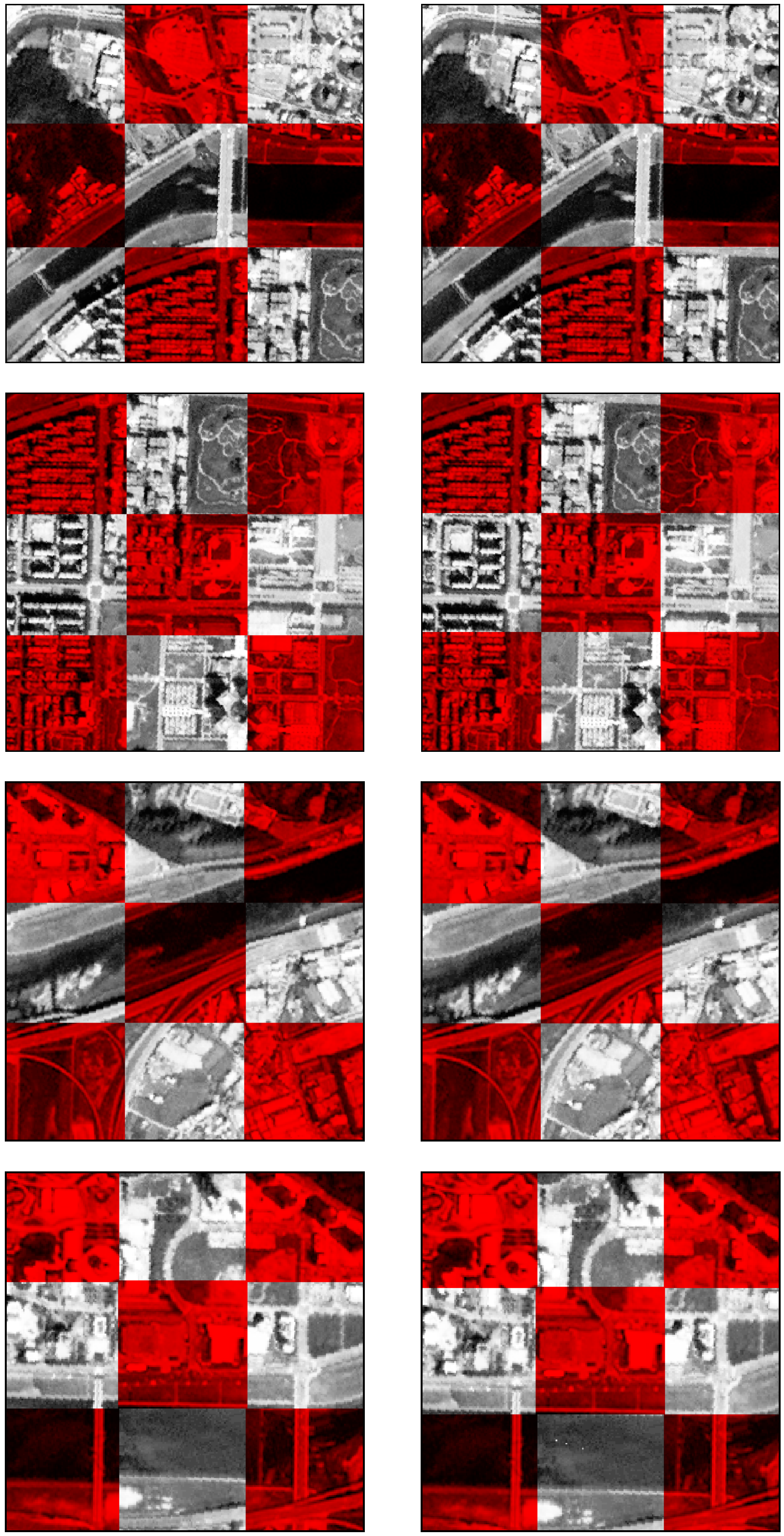
Du, Q., N. Raksuntorn, A. Orduyilmaz, and L. M. Bruce, 2008. Automatic registration and mosaicking for airborne multispectral image sequences. *Photogrammetric Engineering and Remote Sensing,* 74(2):169–181.

Goshtasby, A., 1986. Piecewise linear mapping functions for image registration. *Pattern Recognition,* 19(6):459–466.

Holden, M., D. Hill, E. Denton, J. Jarosz, T. Cox, T. Rohlfing, J. Goodey, and D. Hawkes, 2000. Voxel similarity measures for 3-D serial MR brain image registration. *IEEE Transactions on Medical Imageing,* 19(2), 94–102.

Hong, G., and Y. Zhang, 2008. Wavelet-based image registration technique for high-resolution remote sensing images.

Computers & Geosciences, 34(12):1708–1720.
Kennedy, R. E., and W. B. Cohen, 2003. Automated designation of tie-points for image-to-image coregistration. *International Journal of Remote Sensing*, 24(17):3467–3490.
Li, Q., G. Wang, J. Liu, and S. Chen, 2009. Robust scale-invariant feature matching for remote sensing image registration. *IEEE Geoscience and Remote Sensing Letters,* 6(2):287–291.
Liu, D., P. Gong, M. Kelly, and Q. Guo, 2006. Automatic registration of airborne images with complex local distortion. *Photogrammetric Engineering and Remote Sensing*, 72(9):1049–1059.
Liu, J., and H. Yan, 2008. Phase correlation pixel-to-pixel image co-registration based on optical flow and median shift propagation. *International Journal of Remote Sensing,* 29(20):5943–5956.
Lowe, D. G., 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110.
Schowengerdt, R. A., 2007. *Remote sensing: Models and methods for image processing*, 3rd edition, San Diego, Academic Press.
Song, Z., S. Li, and T. F. George, 2010. Remote sensing image registration approach based on a retrofitted SIFT algorithm and Lissajous-curve trajectories. *Optics. Express*, 18, 513–522.
Xiong, Z., and Y. Zhang, 2009. A novel interest-point-matching algorithm for high-resolution satellite images. *IEEE Transactions on Geoscience and Remote Sensing*, 47(12):4189–4200.
Yang, G., C. V. Stewart, M. Sofka, and C. L. Tsai, 2007. Registration of challenging image pairs: Initialization, estimation, and decision. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 29(11), 1973–1989.
Yu, L., D. Zhang, and E. J. Holden, 2008. A fast and fully automatic registration approach based on point features for multi-source remote-sensing images. *Computers & Geosciences*, 34(7):838–848.
Zhang, C., and C. S. Fraser, 2007. Automated registration of high-resolution satellite images. *The Photogrammetric Record*, 22(117):75–87.
Zitová, B., and J. Flusser, 2003. Image registration methods: A survey. *Image and Vision Computing*, 21(11):977–1000.



(a) (b)

Figure 5. Visual assessment of SIFT and proposed method: (a) magnified mosaic results from the SIFT method; and (b) magnified mosaic results from the proposed method.