Mutual-Information-Based Registration of TerraSAR-X and Ikonos Imagery in Urban Areas

Sahil Suri and Peter Reinartz

***Abstract*—The launch of high-resolution remote sensing satellites like TerraSAR-X, WorldView, and Ikonos has benefited the combined application of synthetic aperture radar (SAR) and optical imageries tremendously. Specifically, in case of natural calamities or disasters, decision makers can now easily use an old archived optical with a newly acquired (postdisaster) SAR image. Although the latest satellites provide the end user already georeferenced and orthorectified data products, still, registration differences exist between different data sets. These differences need to be taken care of through quick automated registration techniques before using the images in different applications. Specifically, mutual information (MI) has been utilized for the intricate SAR–optical registration problem. The computation of this metric involves estimating the joint histogram directly from image intensity values, which might have been generated from different sensor geometries and/or modalities (e.g., SAR and optical). Satellites carrying high-resolution remote sensing sensors like TerraSAR-X and Ikonos generate enormous data volume along with fine Earth observation details that might lead to failure of MI to detect correct registration parameters. In this paper, a solely histogram-based method to achieve automatic registration within TerraSAR-X and Ikonos images acquired specifically over urban areas is analyzed. Taking future sensors into a perspective, techniques like compression and segmentation for handling the enormous data volume and incompatible radiometry generated due to different SAR–optical image acquisition characteristics have been rightfully analyzed. The findings indicate that the proposed method is successful in estimating large global shifts followed by a fine refinement of registration parameters for high-resolution images acquired over dense urban areas.**

***Index Terms*—High resolution, image matching, remote sensing.**

I. Introduction

DURING the last decades, remote sensing sensors have undergone a rapid development in terms of both data quantity and characteristics. Recently, there has been a significant increase in the number of high-resolution sensors placed in remote sensing satellites like Ikonos, Quickbird, TerraSAR-X, Cosmo-Skymed, and WorldView-1. With this enormous increase in availability and quality of remote sensing data products, remote sensing imagery and techniques have found applications in diverse areas like traffic studies, treaty and border monitoring, agricultural studies, generation of 3-D models and topographic maps, early-warning systems, urban growth monitoring, damage assessment, and, specifically, disaster mitigation. In general, images acquired both by the passive optical sensors and active synthetic aperture radar (SAR) sensors (alone and in combination) are major sources for crisis information management. In particular, the SAR sensor's active nature using microwaves gives them the capability to see through clouds and to acquire images at night, which might be the only possible option during a catastrophic event. However, images acquired by SAR sensors have very different characteristics from normally used optical sensor images. On top of the very different geometry (sideways looking and measuring distances) from their optical counterparts (measuring angles), images acquired by SAR sensors show a high amount of speckle influence caused by coherent source of imaging. Normally, remote sensing applications might need to accommodate images from different sensors/modalities, depending upon specific application demands or data unavailability. For example, in case of a natural calamity, decision makers might be forced to use old archived optical data with a newly acquired (postdisaster) SAR image. Combined application of data from different sensors requires georeferenced and fine coregistered images for an accurate and successful analysis. Although latest satellites provide the end user already georeferenced and orthorectified data products, still, registration differences exist between various data sets acquired from different sources even after a digital elevation model (DEM)-supported orthorectification process. These differences need to be taken care of through quick automated registration techniques before using the images in different applications.

Image registration refers to the task of aligning two or more images acquired at different times, from different sensors or from different view points. Image registration can roughly be classified into categories, namely, feature- and intensity-based techniques. Feature-based techniques depend upon detecting and matching landmark features within the images, and on the other hand, in intensity-based techniques, images are registered based on a relation between pixel intensity values of two images. An extensive overview and survey of various image registration methods used in the aforementioned fields can be found in the literature [1], [2]. Recent developments in remote sensing image registration are available in [3] and [4].

The complexity of registering SAR and optical data from high-resolution sensors particularly in urban areas can be

Manuscript received February 26, 2009; revised May 8, 2009. First published November 24, 2009; current version published January 20, 2010.

S. Suri is with the Remote Sensing Technology Institute, German Aerospace Centre (DLR), 82234 Wessling, Germany, and also with the Technical University of Munich, 80333 Munich, Germany (e-mail: sahil.suri@dlr.de).

P. Reinartz is with the Remote Sensing Technology Institute, German Aerospace Centre (DLR), 82234 Wessling, Germany (e-mail: peter.reinartz@dlr.de).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2009.2034842

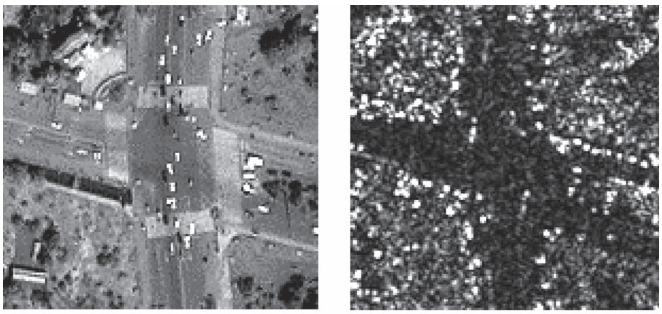


Fig. 1. Road intersection in the city of Sichuan, China, as imaged by (a) Ikonos and (b) TerraSAR-X sensors.

visualized in Fig. 1, showing a road intersection imaged by Ikonos-2 and TerraSAR-X sensors. Due to fine and very different minute details now available with the high-resolution sensors, the idea of finding the conjugate features (for registration) becomes much more challenging than ever before. Keeping in mind the upcoming new challenges and driven by joint application demands of SAR and optical imagery, the development of automatic image registration techniques suitable for high-resolution imagery is encouraged.

Considering the difficult task of extracting common features (both automatically and manually) for SAR and optical image pairs, intensity-based registration techniques have been preferred over the feature-based ones. For intensity-based techniques, the problem of registration is generally mapped as an optimization problem (1), where the spatial transformation function T is the argument of the optimum of some similarity metric S, applied to reference image I_R and transformed input image I_{TI}. This can be expressed as

$$T = \arg\left(\mathrm{opt}\left(S(I_R, I_{\mathrm{TI}})\right)\right). \tag{1}$$

Registration based on the aforementioned equation has been successfully utilized for various multimodal satellite image pairs. The similarity metric "S" that has been found useful for registration of SAR and optical images are, namely, mutual-information (MI) and cluster-reward algorithms [5]–[9]. Mentioned work mainly done for 5–10-m spatial resolution imagery indicates the usefulness of MI as a multimodal registration similarity metric. Still the performance of MI for data sets from latest high-resolution sensors like TerraSAR-X and Ikonos has not been explored to develop registration strategies for different industrial and academic applications. Certain recent efforts in registering urban-area images from airborne SAR sensors and optical satellite images like Ikonos have been made by combining feature- and intensity-based techniques [10], [11]. However, considering the complicated nature of SAR imagery, registration techniques dependent upon line and point detectors might not be a robust alternative as, to date, there is no proven feature detector, generating similar features in SAR and optical images. An MI-based combination of optical and SAR imagery for image interpretation applications in urban areas can be found in [12]. The cited reference highlights the occurrence of multiple peaks (representing building and on-ground registrations) in MI registration surfaces generated from 1-m urban-area images. We present an approach to achieve a completely automated MI-based global (large shifts) followed by a fine registration of images acquired by TerraSAR-X and Ikonos sensors over dense urban areas. Furthermore, we analyze the similarity metric behavior in the context of ever increasing sensor spatial resolution, a chief characteristic of future Earth observation missions.

In general, the challenges in the registration of high-resolution SAR and optical imagery are as follows.

1) Data volume. MI computation depends upon the joint histogram (JH) of the images being registered. Therefore, as the spatial resolution approaches metric and submetric levels, the data volume becomes just huge and leads to undesirable large processing time. This is an important limitation as, in the future, processing large volumes of Earth observation data is set to become a normal daily activity.
2) Geometric and radiometric differences. The increase in spatial resolution further widens the existing gap of sensor geometry and radiometry between the two sensors [11]. For a detailed understanding of high-resolution radar images in urban areas and the difference from their optical counterparts, interested readers are referred to [13].

Objectives of this contribution might be enumerated as follows:

1) to highlight the importance and usefulness of image compression and suitable feature selection for MI-based registration of high-resolution imagery acquired even over plain areas;
2) to depict the adverse influence of different SAR and optical image acquisition mechanisms on intensity-based registration particularly for imagery acquired over urban/semiurban areas;
3) to analyze a solely histogram-based segmentation step for a global followed by a local fine registration of images acquired over dense urban areas, considering that MI is a JH-based matching technique.

II. MI-Based Registration

MI has evolved from the field of information theory [14], [15]. MI describes a statistical dependence between two random variables (e.g., A and B) expressed in terms of variable entropies. In case Shannon entropy (additive in nature) is selected to represent the individual variable information, MI between two variables A and B is defined as

$$MI(A,B) = H(A) + H(B) - H(A,B) \tag{2}$$

where $H(A)$ and $H(B)$ are the Shannon entropies of A and B, respectively, and $H(A,B)$ is the joint variable entropy. Considering A and B as two remote sensing images, their

registration is based on the maximization of $MI(A,B)$ (2). The entropies (marginal and joint) are from [8]

$$H(A) = \sum_{a} -p_A(a)\log p_A(a) \tag{3}$$

$$H(B) = \sum_{b} -p_B(b)\log p_B(b) \tag{4}$$

$$H(A,B) = \sum_{a,b} -p_{A,B}(a,b)\log p_{A,B}(a,b) \tag{5}$$

where $p_A(a)$ and $p_B(b)$ are the marginal probability mass functions and $p_{A,B}(a,b)$ is the joint probability mass function. These probability mass functions can be obtained from

$$p_{A,B}(a,b) = h(a,b) \Big/ \sum_{a,b} h(a,b) \tag{6}$$

$$p_A(a) = \sum_{b} p_{A,B}(a,b) \tag{7}$$

$$p_B(b) = \sum_{a} p_{A,B}(a,b) \tag{8}$$

where h is a JH of the two images involved. It is a 2-D matrix with the intensity values of one image along one axis and the intensity values of the other image along the other axis. Thus, it can be seen from (2)–(8) that the JH is the only requirement for MI computation between any two images. Here, we have employed the normalized MI implementation proposed in [16]. This reduces the sensitivity of MI toward changes in image overlap

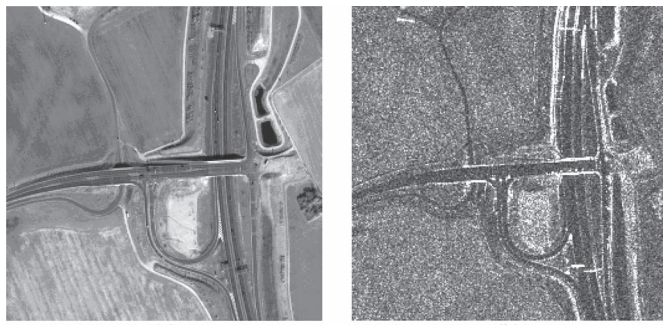
$$MI(A,B) = \left(H(A) + H(B)\right)/H(A,B). \tag{9}$$

An MI-based registration includes a careful selection of a joint histogramming technique and an optimizer to find the registration parameters. For optimizing the registration function, an input image might be transformed several times over the reference image grid. In many cases, the transformed input image might not coincide with the target reference image grid. Therefore, an exact JH may not be obtained, and some approximation becomes inevitable. For JH estimation, one- and two-step histogramming techniques have been utilized in the past [7], [8]. In general, the interpolation step of an intensity value has been held responsible for introducing interpolation-induced artifacts [17], [18] in an MI function which makes the function optimization a tedious task. For the presented work, a one-step joint histogramming technique, namely, generalized partial volume estimation [19], is utilized. For selecting an optimizer, [20] presents a survey of optimization techniques utilized in MI-based registration.

III. Evaluation for High-Resolution Imagery

A. Data Volume

Fortunately, MI, when combined with simple compression techniques, can achieve meaningful registration results in quick time, and this has been demonstrated in this section of this paper. We present here the metric performances for an image pair extracted from TerraSAR-X and Ikonos-2 scenes acquired over west of Munich, Germany (Fig. 2). The selected data set is extracted from geometrically corrected scenes but georeferencing errors exist (TerraSAR-X scene orthorectified using a DEM). Due to the unavailability of true shift differences, we present here a consistency analysis to show obtained shift parameters at different image pixel spacings (images compressed using averaging block filter). For experimentation, we fix our reference image (TerraSAR-X) and move our input image (Ikonos-2) over the reference grid from $[-20\ 20]$ pixels both in the x- and y-directions at different pixel spacings. At every image movement (subpixel), the MI value has been plotted, and the parameters producing the peak of the surface generated are assumed to be the correct shift parameters (theoretical assumption). The results of the carried exercise can be visualized in Table I and Fig. 3. A typical behavior of intensity-based techniques is observed by tabulated registration processing time which shows a drastic increase with an increase in image pixels. The key point to be noted here is that quite accurate registration estimates (in far less time) can be obtained by utilizing a simple image compression technique (block mean filter).



(a) (b)

Fig. 2. (Left) Ikonos-2 and (right) TerraSAR-X images of data set 1 (Ikonos image courtesy of European Space Imaging).

TABLE I
Registration Peaks Along With Their Computation Time Obtained for TerraSAR-X and Ikonos Image Pairs at Different Resolution Levels

Image Size (pixels)	MI Peak	No. of Evaluations	Time
64 x 64	(14.0, -2.0)	25921	13 mins
128 x 128	(12.0, -3.0)	25921	55 mins
256 x 256	(12.50, -5.0)	25921	4 hrs
512 x 512	(11.5, -5.5)	6561	4 hrs

*All the experiments have been done on Genuine Intel Pentium D CPU (2.8GHz) with 2 GB RAM

Furthermore, an important characteristic behavior of MI that has not been explored so far is the fact that it might not require the complete image information to estimate accurate registration parameters. This can be used to decrease the processing time considerably as precisely selecting the necessary pixels for metric computation would lower down the required interpolation steps in the process. Taking the images shown in Fig. 2 (512×512 pixels), we analyze the metric performance; here, we leave out pixels on the basis of their intensity values by introducing thresholds. Pixels falling outside the threshold criteria are marked by a background value in the reference SAR image and are not further used for MI computations. By utilizing the SAR image histogram thresholds in increments of 10% (from higher end), MI performance has been analyzed

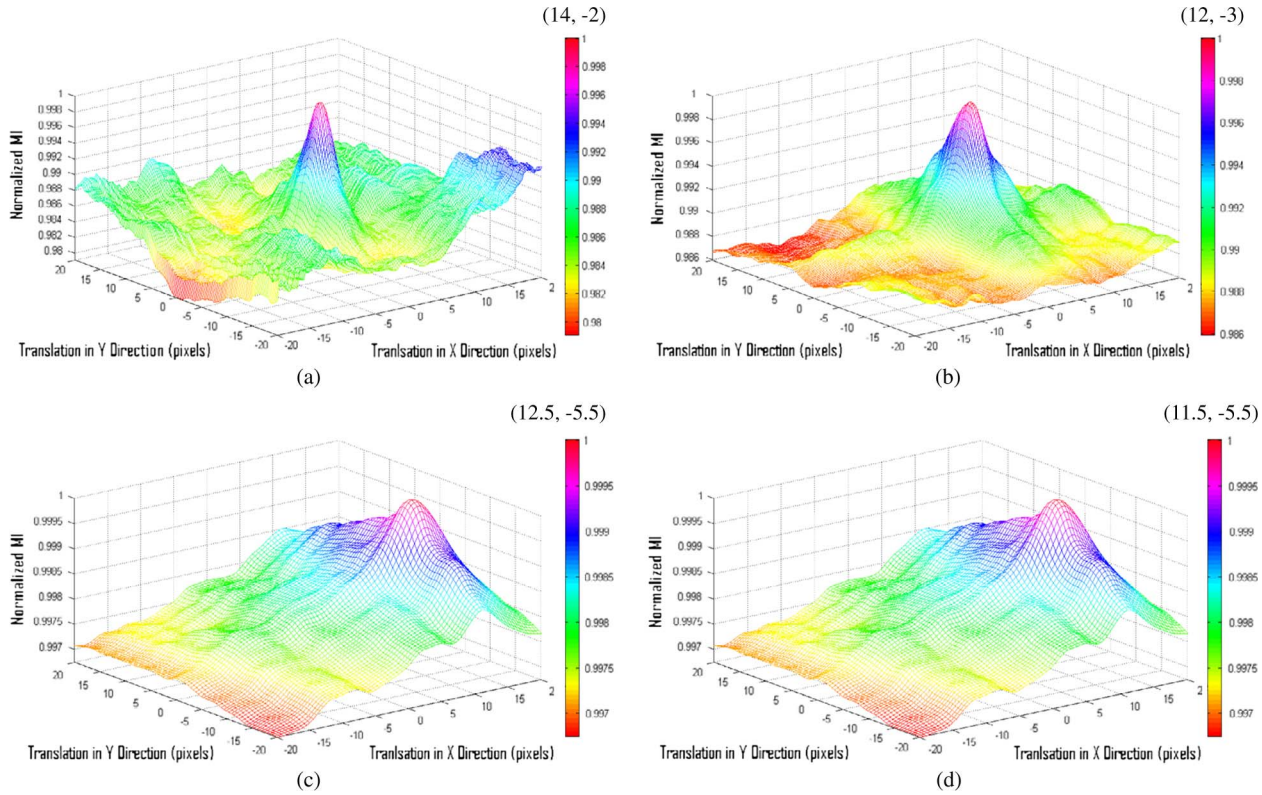


Fig. 3. MI search spaces generated for different resolution image pairs. It is observed that (top right) fairly accurate registration errors present within the image pairs could be estimated from compressed coarser resolution image. (a) Resolution: 64 × 64. (b) Resolution: 128 × 128. (c) Resolution: 256 × 256. (d) Resolution: 512 × 512.

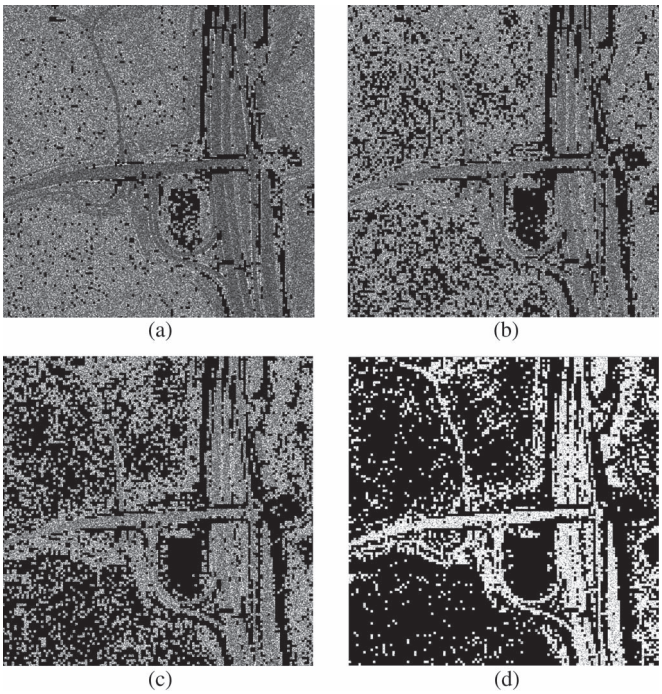


Fig. 4. Effect of high thresholds on SAR imagery. All the SAR images shown produced the same registration peak (Table II) with the corresponding optical image, suggesting that complete information from both the images is not required to retrieve accurate registration parameters. (a) 10% threshold. (b) 30% threshold. (c) 50% threshold. (d) 70% threshold.

for thresholded SAR and optical image pairs. For visualization, SAR images thresholded by different levels have been shown in Fig. 4 (excluded pixels have been marked by a value zero). The

TABLE II
MI REGISTRATION PEAKS OBTAINED BETWEEN THE REFERENCE THRESHOLDED SAR IMAGE AND THE ORIGINAL OPTICAL IMAGE

Threshold	Peak	Pixels in Ref Image (%)	Ref Image Entropy (%)	Time (mins)
0%	(12, -5)	100	100	36
10%	(12, -6)	90.0	97.4	32
30%	(12, -6)	71.3	95.4	25
50%	(12, -6)	51.3	93.6	18
70%	(12, -7)	30.2	91.2	11
90%	(-20, 14)	10.2	87.6	4

metric performance for thresholded SAR (different levels) and optical images is tabulated in Table II.

Taking future sensors into perspective, we study two important parameters, namely, the number of pixels in the reference image for each threshold level and their entropy content relative to the original image entropy value. The registration time for an MI-based process mainly depends upon the reference image size N ($\mathrm{O}(N^2)$ algorithm), and this can be visualized from Table II through the considerable drop in processing time as an increase in image thresholds reduces the number of pixels participating in the registration process.

A strong high-resolution data characteristic is also observed here; removing the higher end pixels from a high-resolution SAR image does not influence much of the original image entropy (Shannon) content. It is worthwhile to note that almost 90% of the image entropy is contained in 30% of the image pixels which can be utilized to achieve a registration speedup factor of three. This phenomenon can mainly be attributed to the radiometric properties of high-resolution SAR sensors

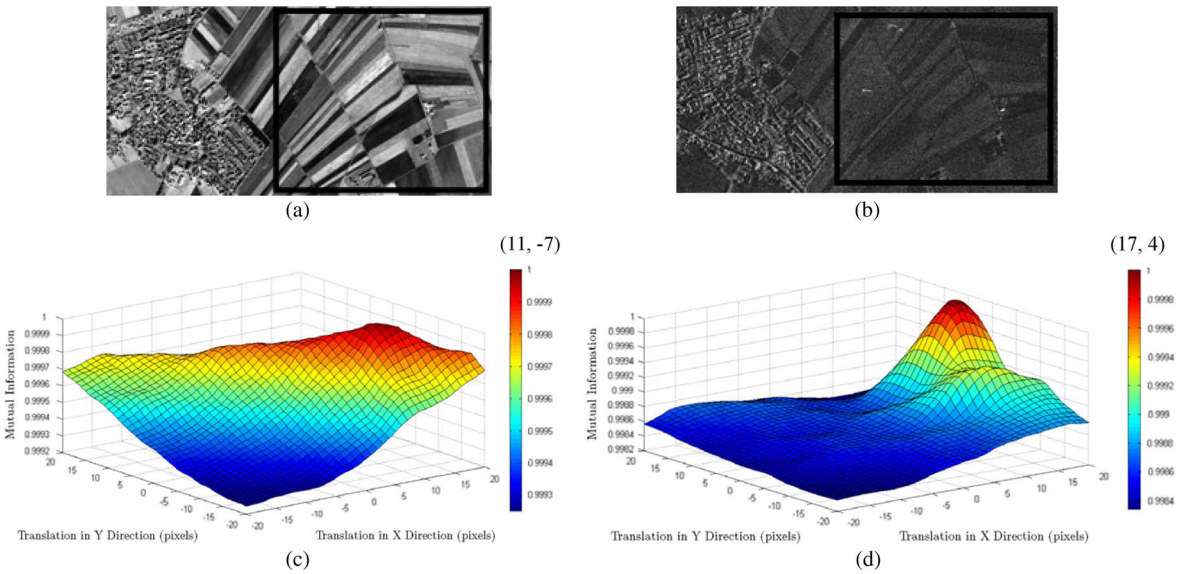


Fig. 5. (a) Ikonos and (b) TerraSAR-X imageries for data set 2. (c) Plain-field pixels lead to a registration peak of (11, −7) for MI. (d) Introduction of urban-area pixels shifted the registration peak for MI to (17, 4).

where most of the image content toward the higher end of the histogram is redundant (for intensity-based registration metrics) and mainly generated by real-world 3-D targets. An abundant number of these targets can actually derail the registration process (shown in subsequent sections) and, therefore, should be avoided. With this observation and understanding, the selection of fewer relevant pixels from the lower end of the SAR image histogram can be taken for an efficient and effective MI-based registration. This step cannot only improve the processing time dramatically but is an important prerequisite to produce accurate registration results for images from future sensors which are expected to provide much higher spatial resolution (generating more redundant information in the SAR image histogram) Earth observation data than the considered TerraSAR-X and Ikonos sensors. For this case, MI reported the same registration peak of (12, −6) in decreasing registration processing time between the thresholded SAR (10%–50%) and the optical image, indicating that this feature selection is useful for fast and robust MI performance.

B. Different Image Acquisition Mechanisms

To show this effect, we perform our analysis with high-resolution imagery acquired over a suburban area in the west of Munich, Germany. The images can be shown in Fig. 5(a) and (b); the imaged scene has an urban settlement situated very next to vast agricultural fields, providing an opportunity to analyze the similarity metric performance for two different land covers. The following two scenarios are considered.

1) For case 1, we select pixels only from the plain fields (roughly demarcated with a rectangle in Fig. 5) in both the images for computing the registration parameters (size: 953 × 1096 pixels). For the plain-field pixels, the sideways-looking SAR sensor and the downward-looking optical sensor are not expected to have much of their geometric influence so favorable registration results as in the previous case are expected.
2) For case 2, we select the entire image scene for registration parameter computation and hence analyze the influence of the suburban establishments on similarity metric performance (size: 1001 × 2001 pixels). The urban establishments due to their strong radar backscatter cause great changes in SAR image radiometry (histogram information) and thus offer a testing case for the similarity metric.

The search spaces shown in Fig. 5 represent the MI response obtained by moving the input Ikonos image over the reference image grid in a predefined range of [−20 20] pixels in both x- and y-directions. Fig. 5(c) and (d) shows the generated search spaces for MI while utilizing pixels belonging only to the land cover class fields and the complete image region, respectively. For the two cases, MI reported a peak at (11, −7) and (17, 4) in Fig. 5(c) and (d), respectively. A visual analysis using an overlay tool clearly indicates the present misalignment within the imagery after using the obtained registration parameters from case 2. Although the land cover fields constitute more than 60% of the total image area, still, the introduction of the bright urban-area pixels have derailed the registration process. This shift in the registration peak can directly be related to the introduction of a region generating incompatible radiometric information (3-D real-world objects).

Practically, different land cover classes are hardly as segregated as available in the analyzed data set. Therefore, the problem of mixed land cover classes asks for a segmentation step before MI-based image registration. The segmentation should be targeted to incorporate only those pixels in the registration process which are not influenced by different sensor geometries (like the plain-field pixels). The idea of introducing a segmentation step prior to registration has the following concerns to be addressed.

1) Supervised or unsupervised. Ideally, unsupervised would be preferred to avoid any kind of manual intervention in the registration process.

2) The accuracy and speed of the segmentation. The degree of accuracy in segmentation that is mandatory for robust performances needs to be established.
3) Segmentation required only in one image or both the images involved in the registration process.

In [12], an urban-area mask is estimated based on directional variances [21] and MI is computed, utilizing only "on-ground" pixels to solve the problem of undesired registration peak shown in Fig. 5(d). In this approach, the final mask is obtained by combining individual urban-area masks from both SAR and optical imageries. This combination is achieved with an assumption that images participating in the registration scheme are already roughly registered. Rough registration has been suggested by utilizing MI on undersampled data or can be achieved if a sensor geometrical model and its *a priori* parameters give a satisfactory localization. In general, the mentioned criteria may not always yield fruitful results (see results in Table IV), and this possibility might totally be extinct in the near future with images having much higher spatial details. Considering this, we propose a method to successfully adapt MI for heterogeneous land cover scenes to achieve a first global rough registration and then followed by a fine refinement by estimating local deformations. The proposed method is unsupervised, very fast, and easy to implement and requires segmentation in only one of the images being registered (reference image preferably to reduce computation time). Inspired from the results obtained for data set 1, the idea of the proposed solution lies in the histogram of a SAR image acquired over urban/semiurban areas. Normally, the pixels generated by the double/triple bounce phenomenon result into a very strong backscatter to the radar sensor, and thus, most of these pixels would be located toward the higher end of a SAR image histogram. Here, this has to be kept in mind that certain other pixels (not influenced by SAR geometry), due to constructive interference of the radar waves, can also produce high intensity values (strong backscatter). However, by histogram thresholds, it still might be possible to bin out most of the pixels explicitly produced by the SAR sensor working principle.

For data set 1, MI recorded the same registration peak for different threshold levels; therefore, as long as the numbers of false pixels being binned out represent a minority of the total pixel population, the true registration peak is not expected to change. The number of such false pixels can definitely be reduced by speckle filtering but the intensity-based registration of SAR and optical imagery does not require any mandatory image smoothing step; hence, we refrain to perform the same in the presented approach.

The results of the proposed segmentation scheme can be shown in Fig. 7. First, the SAR image [Fig. 4(b)] is downsampled to one-fourth of its original resolution. The histogram of the obtained downsampled image is now used to generate thresholds for binning out possible pixels affected by the SAR sensor geometry in the original resolution image. To realize the goal of the segmentation process, thresholds are made from the higher end of the image histogram. It can be clearly observed that, as the threshold limit is relaxed (from 5% to 30% in Fig. 6), more and more pixels from the plain fields start coming into the undesired pixel category, and this might have an adverse influence on similarity metric performance.

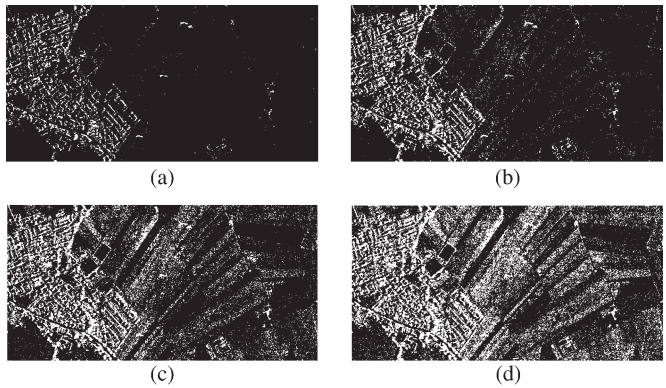


Fig. 6. White pixels are left out (in SAR image) of the registration process after introducing thresholds of 5%–30%.

To analyze the segmentation influence, we register the segmented SAR image (different thresholds analyzed) with the corresponding optical image. In this scenario, all the pixels from optical imagery would contribute to the similarity metric statistics but, from the SAR imagery, only those pixels which are within the threshold limits (assigned the value zero in the masks of Fig. 6) would participate in the registration process. The MI surfaces generated in the search space of $[-20\ 20]$ pixels for the segmented SAR images and the original optical image are shown in Fig. 7. The segmentation of the SAR image using the mask shown in Fig. 7(a) (5% threshold) influenced the registration peaks shown in Fig. 5(c). The registration peaks obtained by MI (17, 4) has been shifted to (15, 1). Further segmentation of the SAR image, i.e., using a threshold on the order of 10%, 20%, and 30% yielded the same registration peaks as were reported by the similarity metrics using only from plain-field pixels [Fig. 5(c)]. The MI peaks obtained for the segmented SAR and the optical imagery are the same as the peaks obtained earlier using only the plain-field pixels; these are assumed true registration parameters.

IV. Registration in Urban Areas

In this section, we present our analysis using a scenario which end users are confronted with, while utilizing high-resolution images acquired over urban areas. We consider TerraSAR-X and Ikonos imageries acquired over the city of Sichuan in China (post-earthquake) (see Table III).

The image pairs have been procured from geometrically corrected scenes and have significant georeferencing differences of approximately 90 m in the x-direction and 45 m in the y-direction (rough estimates). The Ikonos image here is the standard geometrically corrected scene which is expected to have a planar accuracy of 50 m [22]. The TerraSAR-X image is the standard GEC product which has been corrected to universal transverse mercator projection using a constant ellipsoidal height [23]. For consistency and analysis of the possible local-shift magnitudes, we retrieve both local and global shift parameters between two pairs of images (zero overlap) extracted from the same complete scene (image pairs referred to as data sets 3a and 3b hereafter are shown in Fig. 8). The Ikonos image has been acquired nearly a month after the disastrous earthquake; therefore, there are considerable urban changes in

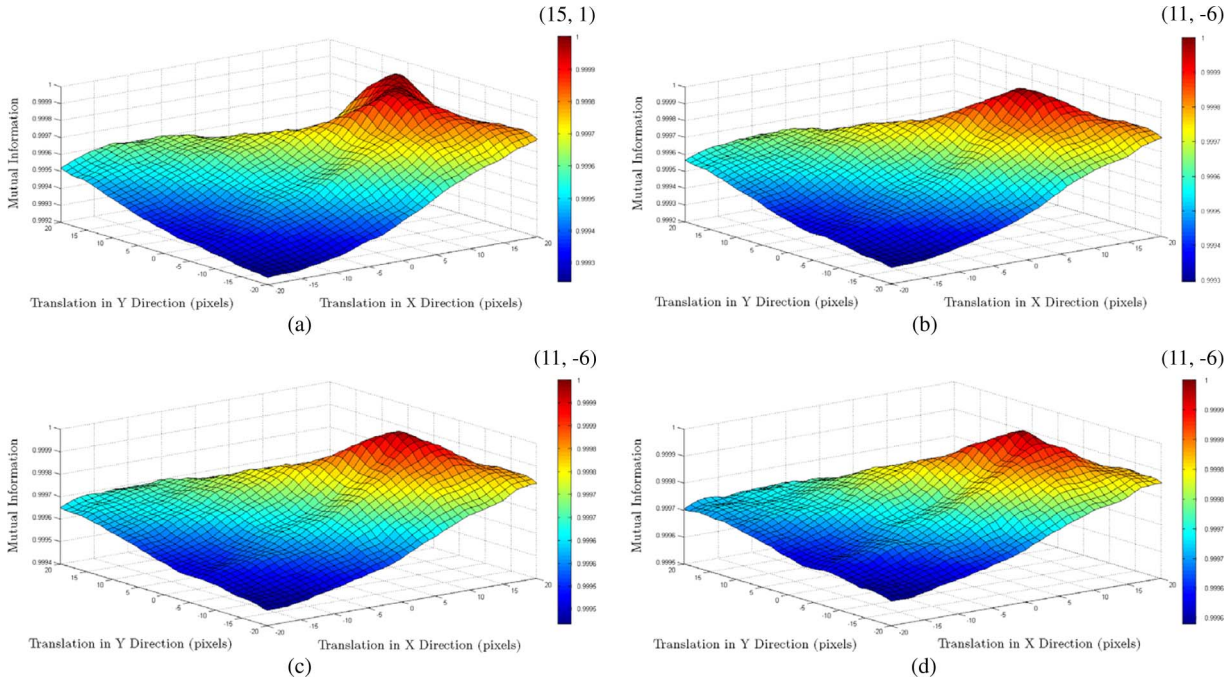


Fig. 7. Registration surfaces generated by MI between segmented SAR (using masks in Fig. 5) and the original optical image. (a) 5% threshold. (b) 10% threshold. (c) 20% threshold. (d) 30% threshold.

TABLE III
DETAILS OF TERRASAR-X AND IKONOS-2 IMAGERIES ACQUIRED OVER SICHUAN, CHINA

	TerraSAR-X	**Ikonos-2**
Mode	High resolution spot light (HS)	Forward Scanning
Pixel Spacing	1m	1m (pan)
Radiometry	16 bit	11 bit
Acquisition Date	15/05/08	28/06/8
Processing Level	GEC Product	Ikonos Geo

the two images. Specifically, the images of data set 3b have a lot of new urban settlements observed in the postdisaster Ikonos image; a subset of the scene which has undergone a significant urban change is demarcated by rectangles in both the TerraSAR-X and Ikonos imageries. The influence of this urban change on MI performance in determining both local and global shift parameters is shown later in this section.

A. Determining Global Shift Parameters

According to the registration errors observed, we translate the input Ikonos image over the reference TerraSAR-X image in a range of -120 to -80 pixels (meters) in the x-direction and -65 to -25 pixels (meters) in the y-direction. For both the data sets, we first test the metric performance for the original images (Fig. 8) and then test the proposed segmentation scheme. As observed earlier, in this case also, the registration peak obtained for the original imagery shifts as the pixels influenced from the SAR geometry are removed from similarity metric computation (Fig. 9). Results in Table IV indicate a consistent MI performance, for data set 3a, after recording a peak at $(-80, -47)$ for original images, and the registration peak stabilizes at $(-96, -47)$ for 10%–40% threshold introduced in the SAR image (search spaces in Fig. 10).

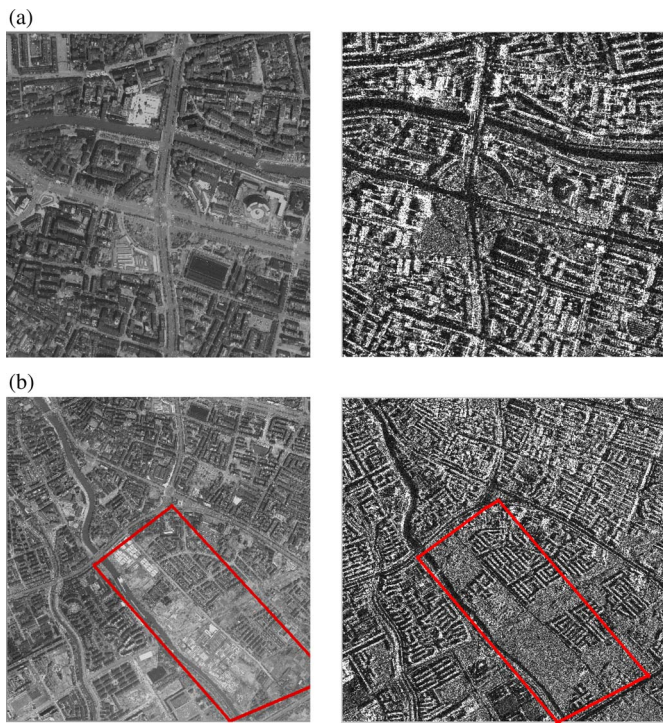
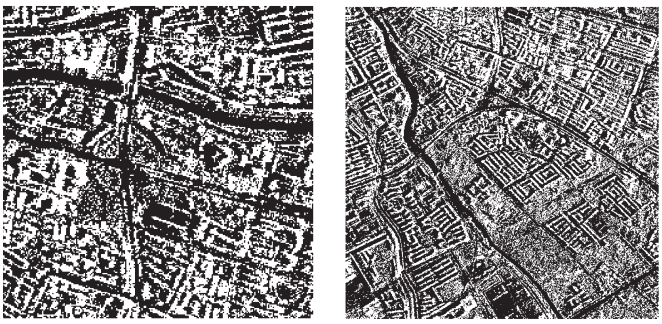


Fig. 8. (a) (Left) Ikonos and (right) TerraSAR-X image for data set 3a (original size: 1000 × 1000 pixels). (b) (Left) Ikonos and (right) TerraSAR-X image for data set 3b (original size: 2000 × 2000 pixels). The demarcated rectangles in images of data set 3b highlight the city area with a lot of new urban settlements in the Ikonos image acquired one month after the earthquake.

This trend is also reflected in data set 3b, where the registration peak remains at $(-80, -43)$ until 20% thresholding and then stabilizes at $(-91, -46)$ pixels until 50% thresholding. Specifically, introducing thresholds from 5% to 40% indicates a systematic shift of MI registration peak from $(-80, -46)$ to $(-96, -47)$ pixels in the x- and y-directions for data set 3a.



(a) (b)

Fig. 9. White pixels left out (in SAR image) of the registration process after introducing 40% threshold in (left) data set 3a and (right) data set 3b.

TABLE IV
REGISTRATION PEAKS AT DIFFERENT PIXEL SPACINGS BETWEEN ORIGINAL AND SEGMENTED SAR–OPTICAL IMAGE PAIRS

		L2	L1	L0
Pixel Spacing		4m	2m	1m
Search Space (pixels)		x: -30 to -20 Y: -17 to -7	x: -60 to -40 y: -33 to -13	x: -120 to -80 y: -65 to -25
0% TH	3a	(-80, -44)	(-80, -46)	(-80, -46)
	3b	(-80, -42)	(-80, -42)	(-80, -42)
5% TH	3a	(-104, -48)	(-96, -48)	(-80, -46)
	3b	(-100, -45)	(-99, -45)	(-80, -43)
10% TH	3a	(-96, -48)	(-96, -48)	(-96, -47)
	3b	(-100, -45)	(-99, -45)	(-80, -43)
20% TH	3a	(-96, -48)	(-96, -48)	(-96, -47)
	3b	(-101, -45)	(-99, -45)	(-80, -43)
30% TH	3a	(-96, -48)	(-96, -48)	(-96, -47)
	3b	(-101, -45)	(-100, -45)	(-91, -46)
40% TH	3a	(-96, -48)	(-100, -48)	(-96, -47)
	3b	(-101, -45)	(-100, -45)	(-91, -46)
50% TH	3a	(-96, -48)	(-100, -48)	(-96, -47)
	3b	(-103, -45)	(-100, -45)	(-91, -46)

TH – Threshold introduced in SAR image. 3a and 3b represent the two datasets respectively.

After the 40% threshold, as further information in the form of SAR image pixels are removed (threshold 50% and 60%), the MI registration search space is observed to be rough (artifacts), and the peak also shows small shifts in both the x- and y-directions. Similar results are also obtained for data set 3b (not shown for brevity).

For a multiresolution analysis, we present results from images reduced to 2- and 4-m pixel spacing (estimation of rough registration parameters through compressed images was highlighted from data set 1). The results have been tabulated in Table IV. Here also, MI could estimate rough shift parameters from compressed and thresholded images. For data set 3b, the parameters at 1-m pixel spacing differ by 10 m in the x-direction from the parameters retrieved at 2- and 4-m pixel spacing. Considering the fact that we do not have access to any other ancillary information, the determination of correct registration parameters for data set 3b is a tedious task. As already mentioned in Section I, measuring the control points manually may not be reliable, and observing an error of around 10 m is very subjective to visual interpretation.

Considering that the global shift parameters from the two data sets differ only by 5 m in the x-direction and 1 m in the y-direction, it can be safely concluded that the significant urban changes observed in data set 3b have not adversely affected MI performance at least globally. Assuming the obtained parameters to be the correct registration errors for the two data sets, we have shifted the input Ikonos image over the reference image grid. Images registered through global shift parameters from data sets 3a and 3b have been check squared in Fig. 1 for visualization.

B. Determining Local-Shift Parameters

After aligning the images globally, the dominant ground features like the river looks very well aligned but, considering the nature of high-resolution images, local distortions are still very much expected. Furthermore, the utilized images have been orthorectified without using a DEM; therefore, local-shift distortions can be anticipated to have certain effects. To further corroborate the applicability of the presented approach, we now compute the local-shift deformations between the globally aligned images (Fig. 11). For this, we mark a regular grid of points on both the SAR and the Ikonos images and then compute shift parameters using a window size of 300 $\times$ 300 pixels through multiresolution simultaneous perturbation stochastic approximation optimization [24]. The shifts obtained for all regularly arranged input Ikonos image tie points for the two data sets are shown in Fig. 12.

After the refinement process, outlier elimination here has been performed using the RANSAC [25] scheme, and the result statistics for local shifts estimated both with and without segmentation are provided in Table V. It is observed that the presented histogram thresholding method improves the similarity metric performance compared to the scheme without any segmentation. This is indicated by the fact that, for the same RANSAC threshold, a lesser number of tie point pairs are rejected using segmentation in both the data sets. Due to significant urban change in images of data set 3b (Fig. 8), MI does not deliver the optimal results locally, and a large number of tie point pairs are rejected in both the schemes. For data set 3a, the local-shift estimation has yielded better results (pixel-level consistency) using the histogram threshold scheme but this cannot be said very certainly for data set 3b where local-shift estimation with and without the SAR image segmentation has yielded very similar performance (consistency wise). Here, the absence of ground truth limits us only to a consistency analysis of both the schemes for MI-based local-shift deformations (we obtain different first-order coefficients using the two schemes).

The influence of different sensor (TerraSAR-X and Ikonos) geometries on MI performance both globally and locally has been analyzed. For future registration scenarios, estimating local variations within very high resolution imagery can also suffer due to incompatible 3-D object (like houses) shadows generated by different sensor geometries. In particular, the pixels generated by radar shadow are very much similar to the desired "on-ground" feature pixels, and therefore, it might be difficult to segregate these just from the image histogram. Overall, for global shift estimation, these pixels have not been found to have an adverse influence on MI performance but, for very precise local-shift deformations, ideally, these should be

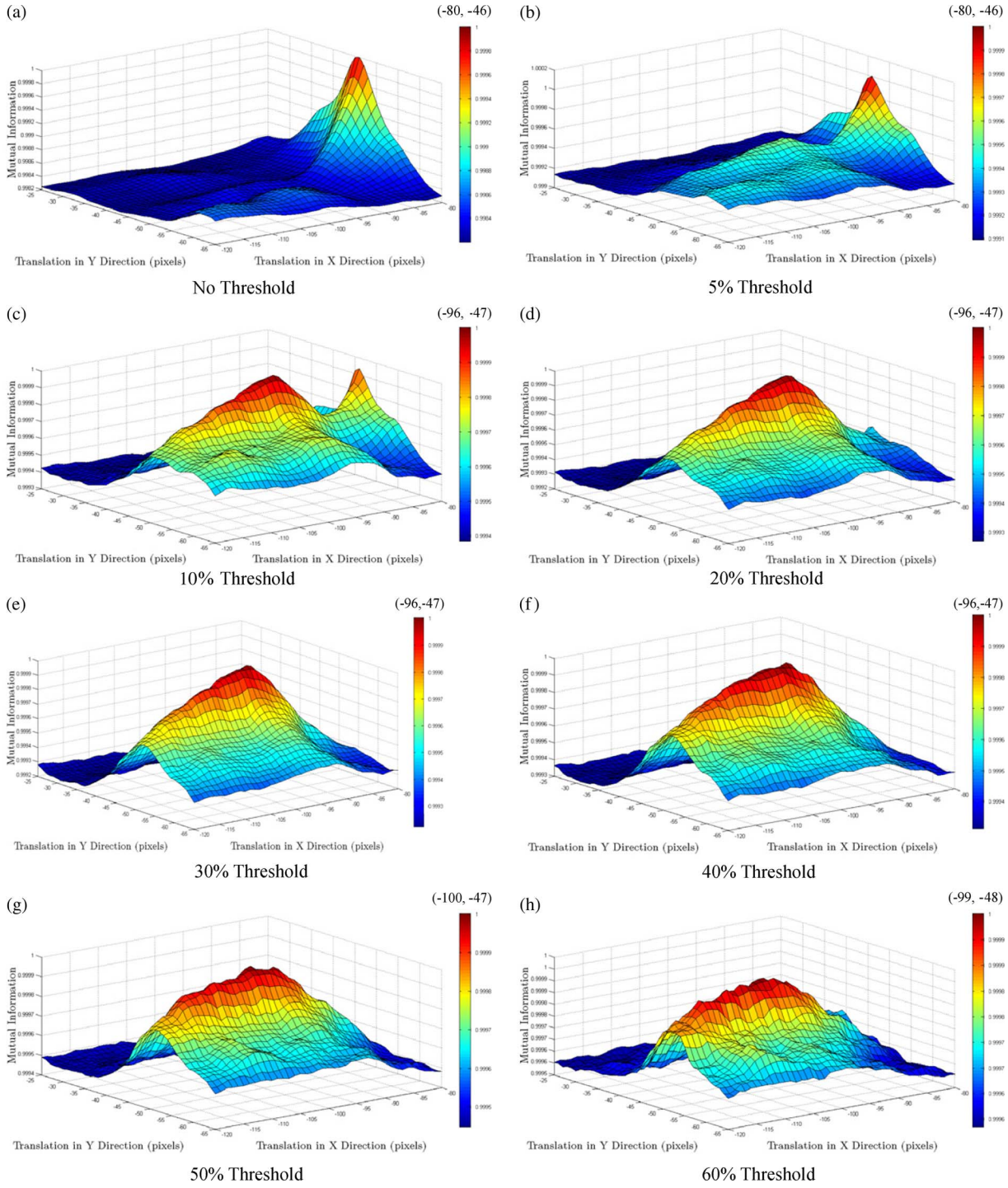


Fig. 10. Registration surfaces generated by MI between original SAR–optical image [Fig. 7(a)] pairs and the segmented SAR (masks in Fig. 8) and the original optical images for data set 3a.

binned out which actually, without a precise DSM, look like a far distant task. In general, we expect an accuracy of around 5 m for the registration of SAR and optical images in dense urban areas using this presented approach.

V. Discussion and Conclusion

For the combined utilization of data from diverse-natured sensors, automatic coregistration methods are a must, as satellite images may have georeferencing errors of magnitude that might affect any combined analysis or decision-making process. Considering the meticulous task of extracting and matching conjugate features in SAR and optical imageries (particularly metric-resolution imagery), a general feature-based image registration technique for various scenarios might be difficult to develop and implement. On the other hand, the intensity-based metrics for medium-resolution images (5–15-m ground sampling distance) have shown enough potential of suitably being modified and extended for different registration scenarios. In this paper, we have investigated the

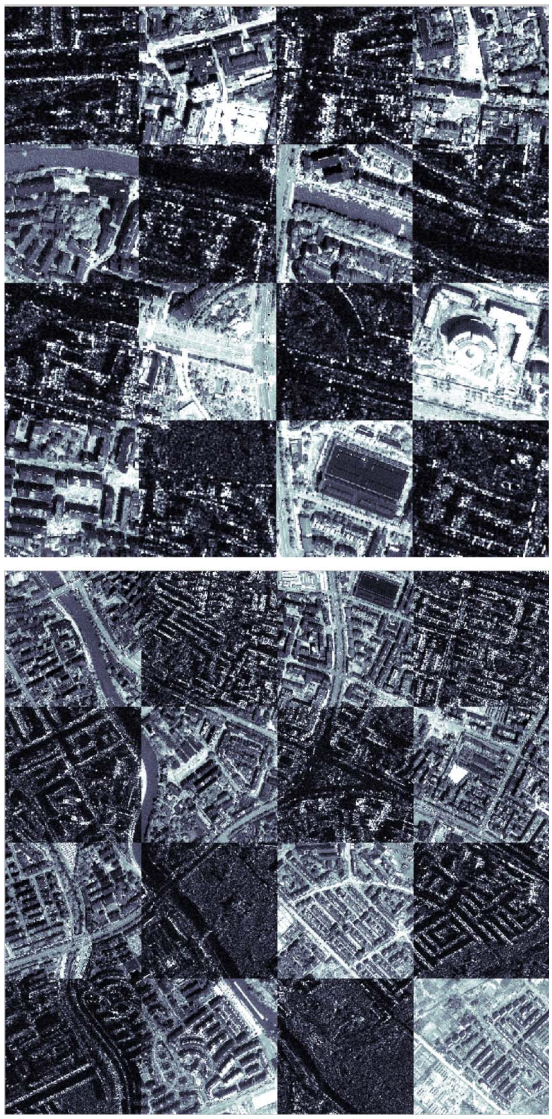


Fig. 11. Registered (dark patches) TerraSAR-X and Ikonos imagery from (top) data sets 3a and (bottom) 3b.

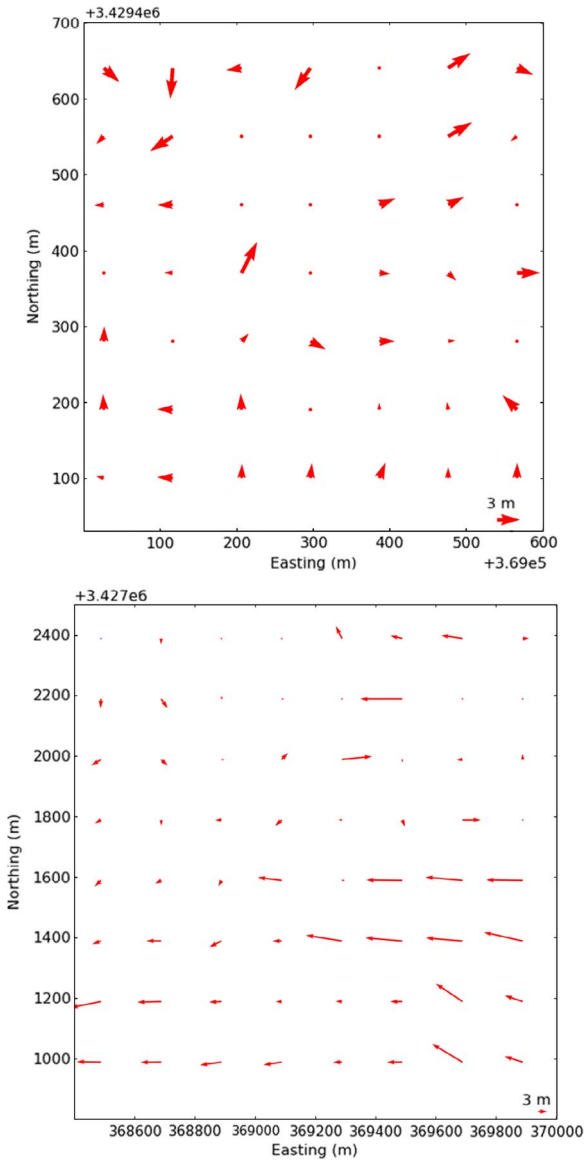


Fig. 12. Local shifts obtained for Ikonos image tie points arranged in a regular grid on images of (top) data sets 3a and (bottom) 3b.

performance of MI for metric resolution imagery acquired by TerraSAR-X and Ikonos-2 satellites. In accordance to the laid down objectives in Section I, the needful analysis has been made to report the following conclusions.

Considering the histogram nature of high-resolution SAR imagery, the idea of relevant feature selection should be utilized even for images acquired over flat terrain (as in data set 1) to further improve the registration turn around time and remove any possible influence of over ground pixels. MI has shown capabilities to estimate quite accurate registration parameters from downsampled images. The presented analysis only considered two translations as the registration parameters but this property is also observed even with a present rotational difference [26].

As expected, the influence of different sensor geometries is more profound for urban settlements as compared with the plain areas. MI has shown enough capabilities of handling this difference through segmentation steps introduced only in the SAR image. Considering the similarity metric nature, the analyzed approach has been found to be efficient and very simple to incorporate in the registration process. As the registration peaks were observed to be consistent for different threshold levels, it can be safely inferred that the accuracy of segmentation is not a very strict criterion for MI performance but excessive thresholding can introduce artifacts in the registration search space, effecting adverse registration results.

TABLE V
COMPUTED LOCAL-SHIFT STATISTICS (WITH AND WITHOUT SEGMENTATION) FOR TIE POINTS OF DATA SETS 3a AND 3b

FP	Mean Shift X	Std Dev X	Mean Shift Y	Std Dev Y	1st Order X residual	1st Order Y residual
Dataset 3a (49 points)						
39	2.02	1.76	1.74	1.38	2.02	2.13
49*	1.03	1.02	0.96	1.05	1.31	1.16
Dataset 3b (64 points)						
35	4.16	2.60	1.22	0.94	1.96	0.91
47*	2.69	2.57	0.86	0.96	1.93	1.05

FP: Filtered Points. * After Thresholding (20%)

We have presented a detailed analysis of a two-step MI-based procedure for automatic registration of high-resolution images acquired over dense urban areas. The first step involves estimating the rough shift parameters that are normally present within georeferenced imagery procured from different sources. Following the rough alignment, local deformations can be obtained to achieve a fine image registration. The results presented here for local deformation computations have utilized a window of size 300×300 pixels but, in general, almost very similar results for both the data sets were obtained from a range of window size of 280×280 to 400×400 pixels. Furthermore, the work presented in this paper has an important implication for the improvement of sensor orientation and orthorectification of very high resolution optical sensors. Considering the achievable geometrical accuracy of the TerraSAR-X products, the finely matched tie points can be utilized as ground control points to improve the optical sensor's rational polynomial coefficients. An application of this sort would require a precisely orthorectified TerraSAR-X image (enhanced ellipsoid corrected product), and this will be our future endeavor in the field of high-resolution SAR and optical image matching.

REFERENCES

[1] L. G. Brown, "A survey of image registration techniques," *ACM Comput. Surv.*, vol. 24, no. 4, pp. 325–376, Dec. 1992.

[2] B. Zitová and J. Flusser, "Image registration methods: A survey," *Image Vis. Comput.*, vol. 21, no. 11, pp. 977–1000, Oct. 2003.

[3] A. Wong and D. A. Clausi, "ARRSI: Automatic registration of remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 5, pt. II, pp. 1483–1493, May 2007.

[4] G.-J. Wen, J.-J. Lv, and W.-X. Yu, "A high-performance feature-matching method for image registration by combining spatial and similarity information," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 4, pp. 1266–1277, Apr. 2008.

[5] J. Inglada, "Similarity measures for multisensor remote sensing images," in *Proc. IGARSS*, Toronto, ON, Canada, 2002, pp. 104–106, [CD-ROM].

[6] J. Inglada and A. Giros, "On the possibility of automatic multisensor image registration," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 10, pp. 2104–2120, Oct. 2004.

[7] H. Chen, P. K. Varshney, and M. K. Arora, "Mutual information-based image registration for remote sensing data," *Int. J. Remote Sens.*, vol. 24, no. 18, pp. 3701–3706, Sep. 2003.

[8] H. Chen, P. K. Varshney, and M. K. Arora, "Performance of mutual information similarity measure for registration of multitemporal remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 11, pp. 2445–2454, Nov. 2003.

[9] A. A. Cole Rhodes, K. L. Johnson, J. LeMoigne, and I. Zavorin, "Multiresolution registration of remote sensing imagery by optimization of mutual information using a stochastic gradient," *IEEE Trans. Image Process.*, vol. 12, no. 12, pp. 1495–1511, Dec. 2003.

[10] W. Yang, C. Han, H. Sun, and Y. Cao, "Registration of high resolution SAR and optical images based on multiple features," in *Proc. IGARSS*, Seoul, Korea, 2005, pp. 3542–3544.

[11] G. Lehureau, F. Tupin, C. Tison, G. Oller, and D. Petit, "Registration of metric resolution SAR and optical images in urban areas," in *Proc. 7th Eur. Conf. Synthetic Aperture Radar*, Friedrichshafen, Germany, 2008.

[12] G. Oller, D. Petit, and J. Inglada, "On the use of SAR and optical images combination for scene interpretation," in *Proc. ISPRS Symp.—Commission I, WGI/3, From Sensors to Imagery*, 2006, pp. 87–91.

[13] U. Stilla, "High resolution radar imaging of urban areas," in *Photogrammetric Week*, D. Fritsch, Ed. Heidelberg, Germany: Wichmann, 2007, pp. 149–158. [Online]. Available: http://www.ifp.uni-stuttgart.de/publications/phowo07/180Stilla.pdf

[14] P. Viola and W. M. Wells, "Alignment by maximization of mutual information," *Int. J. Comput. Vis.*, vol. 24, no. 2, pp. 137–154, 1997.

[15] A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens, and G. Marchal, "Automated multimodality image registration based on information theory," in *Information Processing in Medical Imaging*, Y. Bizais, C. Barillot, and R. Di Paola, Eds. Dordrecht, The Netherlands: Kluwer, 1995, pp. 263–274.

[16] C. Studholme, D. L. G. Hill, and D. J. Hawkes, "An overlap invariant entropy measure of 3D medical image alignment," *Pattern Recognit.*, vol. 32, no. 1, pp. 71–86, Jan. 1999.

[17] J. Tsao, "Interpolation artifacts in multimodality image registration based on maximization of mutual information," *IEEE Trans. Med. Imag.*, vol. 22, no. 7, pp. 854–864, Jul. 2003.

[18] J. Inglada, V. Muron, D. Pichard, and T. Feuvrier, "Analysis of artifacts in subpixel remote sensing image registration," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 1, pp. 254–264, Jan. 2007.

[19] H. Chen and P. K. Varshney, "Mutual information-based CT-MR brain image registration using generalized partial volume joint histogram estimation," *IEEE Trans. Med. Imag.*, vol. 22, no. 9, pp. 1111–1119, Sep. 2003.

[20] J. Pluim, J. Maintz, and M. Viergever, "Mutual-information-based registration of medical images: A survey," *IEEE Trans. Med. Imag.*, vol. 22, no. 8, pp. 986–1004, Aug. 2003.

[21] A. Lorette, "Texture analysis through Markov random fields: Urban area extractions," in *Proc. IEEE ICIP*, Kobe, Japan, Oct. 1999, pp. 430–434.

[22] Space Imaging, Supplement, Space Imaging, Catalog of Products and Services, Thornton, CO, 2000.

[23] T. Fritz and M. Eineder, TerraSAR-X ground segment, SAR basic product specification document (TX-GS-DD-3302), Feb. 2008.

[24] J. C. Spall, "Multivariate stochastic approximation using a simultaneous perturbation gradient approximation," *IEEE Trans. Autom. Control*, vol. 37, no. 3, pp. 332–341, Mar. 1999.

[25] M. A. Fischler and R. C. 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, Jun. 1981.

[26] S. Suri, P. Schwind, P. Reinartz, and J. Uhl, "Combining mutual information and scale invariant feature transform for fast and robust multisensor SAR image registration," in *Proc. 75th ASRPS Conf.*, Baltimore, MD, Mar. 2009.

Sahil Suri received the Bachelor of Information Technology (BIT) from Hamdard University, New Delhi, India, in 2004 and the M.Tech. degree with specialization in geomatics engineering from the Indian Institute of Technology Roorkee, Roorkee, India, in 2006. He is currently working toward the Ph.D. degree with specialization in remote sensing image processing at the Technical University of Munich, Munich, Germany. His dissertation is on developing feature- and intensity-based registration algorithms for multimodal remote sensing images (SAR–SAR and SAR–optical).

He has been with the Department of "Photogrammetry and Image Analysis," Remote Sensing Technology Institute (IMF), German Aerospace Centre (DLR), Wessling, Germany, since September 2006. His research interests include high-resolution image analysis, matching, segmentation, and registration. He also holds interest in traffic monitoring and application of digital photogrammetry to DEM and 3-D model generation.

Mr. Suri is the recipient of the German Academic Exchange Service (DAAD) fellowship (September 2005–May 2006) and a two-year Government of India scholarship for his M.S. degree (August 2004–June 2006).

Peter Reinartz received the Diploma (Dipl.-Phys.) degree in theoretical physics from the University of Munich, Munich, Germany, in 1983 and the Ph.D. (Dr. Ing.) degree in civil engineering from the University of Hanover, Hanover, Germany, in 1989. His dissertation was on statistical optimization of classification methods for multispectral image data.

He is the Department Head of the Department of "Photogrammetry and Image Analysis," Remote Sensing Technology Institute (IMF), German Aerospace Centre (DLR), Wessling, Germany. He has more than 20 years of experience in image processing and remote sensing and over 150 publications in his fields of interest. His main interests are in direct georeferencing, stereo photogrammetry, and data fusion of spaceborne and airborne data, generation of digital elevation models, and interpretation of very high resolution data from sensors like Ikonos and Quickbird. He is also engaged in using remote sensing data for disaster management and using high-frequency time series of airborne image data for real-time operations in case of disasters as well as for traffic monitoring.

Dr. Reinartz was the recipient of the Heinz–Maier–Leibnitz Award from the German Ministry of Research and Technology in 1989.