

# ARRSI: Automatic Registration of Remote-Sensing Images

Alexander Wong, *Student Member, IEEE*, and David A. Clausi, *Senior Member, IEEE*

**Abstract**—This paper presents the Automatic Registration of Remote-Sensing Images (ARRSI); an automatic registration system built to register satellite and aerial remotely sensed images. The system is designed specifically to address the problems associated with the registration of remotely sensed images obtained at different times and/or from different sensors. The ARRSI system is capable of handling remotely sensed images geometrically distorted by various transformations such as translation, rotation, and shear. Global and local contrast issues associated with remotely sensed images are addressed in ARRSI using control-point detection and matching processes based on a phase-congruency model. Intensity-difference issues associated with multimodal registration of remotely sensed images are addressed in ARRSI through the use of features that are invariant to intensity mappings during the control-point matching process. An adaptive control-point matching scheme is employed in ARRSI to reduce the performance issues associated with the registration of large remotely sensed images. Finally, a variation on the Random Sample and Consensus algorithm called Maximum Distance Sample Consensus is introduced in ARRSI to improve the accuracy of the transformation model between two remotely sensed images while minimizing computational overhead. The ARRSI system has been tested using various satellite and aerial remotely sensed images and evaluated based on its accuracy and computational performance. The results indicate that the registration accuracy of ARRSI is comparable to that produced by a human expert and improvement over the baseline and multimodal sum of squared differences registration techniques tested.

**Index Terms**—Image registration, intersensor, intrasensor, invariant descriptor, remote sensing.

## I. INTRODUCTION

IMAGE registration is the process of aligning different images of the same scene acquired at different times, different viewing angles, and/or different sensors. Image registration is frequently used in remote sensing for a wide variety of tasks such as change detection, image fusion, and image overlay. Traditional image-registration techniques in remote sensing required the manual selection of ground control points (GCPs) at significant landmarks of the images. These GCPs are then used to estimate the transformation model that aligns one image to another. The primary drawback to this approach is that a trained expert is needed to manually select each individual GCP in the remotely sensed images. This is very laborious and time consuming, especially when dealing with the large volumes of remote-sensing data available today. Therefore, an automatic

method of aligning such images is highly desired. A number of methods have been proposed to automate the process of image registration. These methods can be generalized into the following categories.

- 1) *Methods based on pixel intensities [1]–[6]*: In these methods, the similarity between pixel intensities is used to determine the alignment between two images. Similarity measures used in these algorithms include maximum likelihood [1] and mutual information [2].
- 2) *Methods based on frequency-domain characteristics [7]–[9]*: Such algorithms attempt to find an optimal alignment match between two images based on characteristics in the frequency domain. A common frequency-domain technique is phase correlation, which is based on the Fourier Shift Theorem. The Fourier coefficients of first image are divided by the Fourier coefficients of the second image, and the inverse of the result is an image with a single peak. This peak indicates the translation between the two images. This technique has also been extended to account for rotation and scaling [8].
- 3) *Methods based on low-level features [10]–[13]*: Such techniques extract low-level features such as edges, ridges, and corners from the images and use the correlation between these features to determine the optimal alignment between the images. These techniques are useful for situations where the distinctive details are prominent and when complex distortions exist.
- 4) *Methods based on high-level features [14], [15]*: Such algorithms attempt to extract high-level features such as regions and specific objects (e.g., roads, buildings, and rivers) and attempt to find the optimal alignment of the images by matching the features from one image to another based on specific feature characteristics. For example, regions may be matched based on their area, perimeter, and centroid. These techniques are useful if the structural characteristics of specific object types are well known.

While these methods differ in their specific approach to the automatic image-registration problem, the majority of these methods, particularly for nonrigid registration, can be broken down into the following steps.

- 1) *Control-point (CP) detection*: A set of potential CPs is selected automatically from the set of images.
- 2) *CP matching*: Similarity analysis is performed to determine a set of matching CPs from the CP candidates.

Manuscript received June 22, 2006; revised December 15, 2006.

The authors are with the University of Waterloo, Waterloo, ON N2L 3G1, Canada.

Digital Object Identifier 10.1109/TGRS.2007.892601

- 3) *Transformation estimation*: Based on the set of matching CPs, the transformation model is estimated to provide the best alignment between the images.
- 4) *Transformation and resampling*: The images are transformed based on the determined model and are resampled using an interpolation method.

Each of these steps plays an important part in the image-registration procedure. Therefore, the following criteria are important in the design of an automatic image-registration system.

- 1) *Efficiency*: Given the large size of remotely sensed images, it is important to minimize the computational effort required to perform each of these steps while maintaining alignment accuracy.
- 2) *Robustness*: Differences in remotely sensed images of the same scene often exist due to factors such as environmental noise, differences in illumination and contrast, and differences in viewpoint. Therefore, it is important to minimize the effect of such image variances on image-registration accuracy.
- 3) *Accuracy*: Visualization and analysis of remotely sensed data require that a reasonable level of accuracy be achieved during the registration process. Therefore, it is important that the registration process produces an image that is visually and numerically correct.

The main problem with previous work, including those applied to remotely sensed images, is that they do not take into account the registration issues and difficulties associated with the characteristics of remotely sensed images. Therefore, they are not well suited for the registration of intersensor and intrasensor images in many situations. First, due to differing environmental conditions and the fact that images are often acquired at different times, remotely sensed images often have contrast and illumination variations in both a global and local sense. The performance of CP detection schemes employed by many registration schemes are heavily affected by such contrast and illumination conditions. Second, remotely sensed images are often acquired using different modalities. In this context, the term modalities refer to different sensor devices that capture information at different frequency bands or using different techniques (optical versus radar). As such, images from different modalities have very different intensity mappings. This is problematic for most registration schemes, as they rely on the exact pixel-intensity values to find matching CP pairs. Third, remotely sensed images are often very large in size and can result in performance issues, particularly in practical systems with fast registration requirements, such as on-the-fly visualization systems. Finally, the complex nature of remotely sensed images often results in a high number of mismatched CP pairs. Such outliers have a significant impact on determining the transformational model required in aligning the images together. The goal of this paper is to take each of these criteria into consideration in the design of each of the steps in the image-registration process to provide a practical automatic nonrigid image-registration algorithm for such real-time systems.

The main contribution of this paper is the Automatic Registration of Remote Sensing Images (ARRSI), an automatic registration system designed for registering satellite and aerial remotely sensed images. The system is designed specifically to address the registration problems related to the characteristics of remotely sensed images. ARRSI employ techniques based on a phase-congruency model in the CP-detection process to address global and local contrast and illumination conditions that may affect the accuracy of the detected CPs. ARRSI utilize phase-congruency moment-based patches as local feature descriptors that are invariant to intensity-mapping conditions during the CP-matching process. An adaptive CP-matching scheme is employed in ARRSI to reduce the performance overhead associated with the registration of large remotely sensed images. Finally, an advancement on the Random Sample and Consensus (RANSAC) algorithm called Maximum Distance Sample Consensus (MDSAC) is introduced in ARRSI to provide an efficient and effective method of reducing the effect of mismatched CPs and improving the accuracy of the transformation model between two remotely sensed images.

In this paper, the theory underlying the design process of each step in the image-registration process is presented and explained in Section II along with an outline of the proposed algorithm. The testing methods and test data are outlined in Section III. The computational performance and registration accuracy for intrasensor and intersensor and interband images are discussed in Section IV. Finally, conclusions are drawn based on the results in Section V.

## II. THEORY

The ARRSI system utilizes various techniques to address the issues related to computational complexity and robustness to variations in the remotely sensed images. Hence, it is important to discuss the theory behind the design process before outlining the proposed algorithm. For each step of the automatic registration process outlined in Section I, the issues pertaining to efficiency, robustness, and accuracy will be presented, and the methods used to address these issues will be described and explained in detail. The reference image refers to the image to which sensed images are aligned. The reference and sensed images are assumed to be at the same spatial resolution. If the reference image and the sensed image are at different spatial resolutions, then the image with the finer resolution is down-sampled to that of the coarser resolution.

### A. CP Candidate Detection

The first step in nonrigid automatic image registration is to select a set of potential CPs from the reference image and the sensed image. This can be accomplished using a number of feature-detection methods [16]–[19]. One particularly interesting approach to feature perception is that based on a phase-congruency model [20]–[22]. The phase-congruency model for determining feature significance is based on the local-energy model [23], which postulates that perceptually significant features are situated at locations of the image where the Fourier components are maximally in phase. This approach is further reinforced by psychophysical evidence [24] showing

that the human visual perception system is highly responsive to visual features with high phase congruency. It was shown that the phase-congruency model provides a measure of feature significance that is invariant to illumination and contrast conditions [22]. This is in contrast to gradient-based methods, which are highly sensitive to such conditions, particularly suffering in situations where illumination and contrast conditions vary within a single image. Given the significant advantages of phase congruency over conventional gradient-based approaches, the ARRSI system utilizes a candidate CP-detection algorithm based around a phase-congruency model. The phase-congruency measure used in the ARRSI candidate CP-selection algorithm is based on that presented in [22], which was shown to provide good feature localization and noise compensation. Local frequency information at a particular point in the image is obtained using logarithmic Gabor filters over multiple scales and orientations. The phase congruency at each point in the image is

$$P(x, y) = \frac{\sum_n W(x, y) [A_n(x, y) \Delta\Phi(x, y) - T]}{\sum_n A_n(x, y) + \varepsilon} \quad (1)$$

$$\Delta\Phi(x, y) = \cos(\phi_n(x, y) - \bar{\phi}(x, y)) - |\sin(\phi_n(x, y) - \bar{\phi}(x, y))| \quad (2)$$

where  $(x, y)$  indicates the coordinate of the point,  $W(x, y)$  is the weighting factor based on frequency spread,  $A_n(x, y)$  and  $\phi_n(x, y)$  are the amplitude and phase at  $(x, y)$  at wavelet scale  $n$ , respectively,  $\bar{\phi}(x, y)$  is the weighted mean phase at  $(x, y)$ ,  $T$  is the noise threshold, and  $\varepsilon$  is a small constant to avoid division by zero. The parameter values used in the ARRSI system ( $W(x, y)$ ,  $A_n(x, y)$ ,  $\phi_n(x, y)$ ,  $\bar{\phi}(x, y)$ , and  $T$ ) are the same as those found in [21]. The only exception to the parameters presented in [21] is that the ARRSI system makes use of five discrete scales. Feature significance is then derived using the principle moments of phase congruency. The minimum moments provide a good representation of corner feature significance and can be calculated as the expression in (3), shown at the bottom of the page, where  $P(\theta)$  is the phase congruency at orientation  $\theta$ . High values of  $m$  indicate high corner strength. The resultant minimum moment map can then be used to locate candidate CP locations with strong feature significance. There are a number of issues that need to be considered in terms of efficiency and robustness with regard to using the phase-congruency detector for candidate CP detection. Remotely sensed images are typically large in size,

resulting in high computational cost in the feature-detection process. This is problematic for practical situations with fast registration requirements, such as on-the-fly visualization. Furthermore, remotely sensed images are generally complex by nature and, hence, result in the detection of insignificant features. One method of addressing these issues is by performing image subsampling. A number of different image-subsampling techniques have been introduced over the years, including some more recent techniques designed for better quality and/or performance [25], [26]. By downsampling the images, the computational effort required in the CP-detection process is reduced. Furthermore, the downscaling process help reduce the detection of insignificant candidates that may affect the accuracy during the matching process. The downsampled image size is selected based on the performance and accuracy requirements of the underlying system. For example, an on-the-fly visualization system would require fast image-registration performance and, therefore, the images would be downsampled to more than that of a system that requires a higher degree of accuracy and has lower speed requirements. The ARRSI system utilizes bicubic downsampling as it provides high overall speed and quality.

Another technique that can be used to improve registration performance and accuracy in large and complex images is to allow for the selection of regions of interest (ROIs) from which CP candidates are selected. For example, in the case of image mosaicking, it is easy to select rough ROIs around the areas in which the images overlap each other. Not only will this improve overall performance but it will also improve overall accuracy by reducing the nonoverlapping regions present during the matching process. Both subsampling- and ROI-selection techniques are employed in the ARRSI system. In the ARRSI system, if the user does not manually select a ROI, an estimated ROI is automatically determined based on *a priori* georeferencing information if it is available. The automatic ROI-estimation process is performed by determining the overlapping region between the two images based on available georeferencing information. The estimated overlapping region is then increased by 20% in each dimension to account for potential errors in the georeferencing information. This new estimated overlapping region is then set as the ROI in the corresponding images. This helps to reduce total search size during the matching process as well as improve the robustness of the system.

Another issue that needs to be considered is the effect of the quantity of detected candidate CPs on the performance and accuracy of image registration. Too few candidates result in poor registration accuracy. However, too many candidates result in high computational cost during the matching process. If a

$$m = \frac{1}{2} \left( \sum_{\theta} [(P(\theta) \sin(\theta))^2 + (P(\theta) \cos(\theta))^2] - \sqrt{4 \left( \sum_{\theta} (P(\theta) \sin(\theta)) (P(\theta) \cos(\theta)) \right)^2 + \left( \sum_{\theta} [(P(\theta) \cos(\theta))^2 - (P(\theta) \sin(\theta))^2] \right)^2} \right) \quad (3)$$



Fig. 1. Extracted CP candidates (as indicated by crosses). (Left) Reference image. (Right) Sensed image.

traditional CP-matching algorithm is employed, then each CP candidate from the reference image is correlated with each CP candidate from the sensed image. This results in the number of correlations to be in the order of  $O(n^2)$ . Even though ARRSI introduces an efficient search strategy that achieves sub- $O(n^2)$  performance, the correlation computations performed in the matching process remains the computational bottleneck of the system. Therefore, it is important to ensure that only significant CPs are detected. After determining the minimum moment map for the image, the threshold is set to a low threshold  $t_o$ , and preliminary candidates are selected as points that are a local maximum within a fixed radius  $r$  and have a corner strength greater than  $t_o$ . Finally, only the strongest  $n$  candidates are selected as the final set of CP candidates, where  $n$  is the number of candidate points desired. Hence, the effective threshold is adaptive to allow for the desired number of candidate points to be retained and thus varies based on the image. Once the candidate points have been selected, the position of a CP candidate is readjusted for subpixel accuracy by fitting a 2-D quadratic to the corner strength in its local neighborhood and then finding the maximum of the quadratic. This candidate-selection method will be referred to as the adaptive phase-congruency feature detector (APCFD) and provides an effective solution to the above issues. Therefore, by reducing the problem-size through image subsampling and ROI selection, as well as introducing a feature-detection method that is illumination and contrast invariant, the efficiency, robustness, and accuracy criteria are satisfied. An example of CP candidates extracted using this method is shown in Fig. 1.

### B. CP Candidate Matching

After the CP candidates have been selected, the candidates from the sensed image need to be matched with the candidates

from the reference image. The first issue in matching candidates is the selection of an appropriate local-feature descriptor. The simplest local-feature descriptor is to use the image intensity within a close neighborhood. This is commonly known as an intensity patch. However, there are a number of problems with using such a local-feature descriptor. First, as with the case of CP-candidate detection, one of the biggest problems with intensity patches is variation in image illumination and contrast. Remotely sensed images are often acquired at different times and/or by different sensors. Therefore, corresponding patches may appear different due to differing illumination conditions and sensor sensitivities. The situation is made more difficult in the case of intersensor images, where images acquired with different modalities may have the same content mapped to significantly different intensity values. Therefore, a local-feature descriptor that is invariant to illumination and contrast as well as intensity mappings is highly desired to improve the robustness of the system. A popular local-feature descriptor that is invariant to all of the above situations is the scale-invariant-feature-transform descriptor [27], [28], which is also invariant to scale and rotation. However, it is relatively computationally expensive for situations with a large number of CP candidates. Furthermore, since scale is typically known for remotely sensed images and adjusted for prior to registration, simpler and more efficient approaches can be used with similar results. To address this issue in a more efficient manner, the proposed algorithm makes use of phase-congruency moment-based patches as local-feature descriptors. The maximum moment of phase congruency provides a good representation of structural feature significance within an image. The maximum moments can be computed as the expression in (4), shown at the bottom of the page, where  $P(\theta)$  is the phase congruency at orientation  $\theta$ . High values of  $M$  indicate high structural feature significance. There are a number of advantages in using this approach in obtaining local-feature descriptors. This approach is invariant to illumination and contrast variations. Furthermore, maximum moments can be calculated in an efficient manner as its formula is almost identical to that used to calculate minimum moments. As such, most of the calculations made in determining the minimum moment map during candidate CP detection can be reused to calculate the maximum moment map. Finally, it is invariant to intensity mappings. This is important as images from different modalities can have very different intensity mappings. Therefore, it is a suitable local-feature descriptor for both intrasensor and intersensor images.

Another issue faced in candidate matching is the computational cost of the matching process itself. The matching process

$$M = \frac{1}{2} \left( \sum_{\theta} \left[ (P(\theta) \sin(\theta))^2 + (P(\theta) \cos(\theta))^2 \right] + \sqrt{4 \left( \sum_{\theta} (P(\theta) \sin(\theta)) (P(\theta) \cos(\theta)) \right)^2 + \left( \sum_{\theta} \left[ (P(\theta) \cos(\theta))^2 - (P(\theta) \sin(\theta))^2 \right] \right)^2} \right) \quad (4)$$

is typically done through an exhaustive search, where each CP candidate from the sensed image is matched with every candidate CP from the reference image. The number of correlations that needs to be performed using this approach is in the order of  $O(n^2)$ . This method is computationally expensive, particularly for large sets of CP candidates. Therefore, a method that reduces the search size while maintaining registration accuracy is desired for improving the computational efficiency of the system. In devising an efficient search strategy, it is important to take two key factors into account: 1) the center of the search window and 2) the size of the search window.

To determine the center and size of the search window for a particular CP candidate in the sensed image, it is useful to exploit the spatial relationships of neighboring CP candidates. First, let us define a mapping vector as a vector that maps the coordinates of a CP candidate in the sensed image to its corresponding CP candidate in the reference image. For example, the mapping vector between a point  $A$  at  $(x, y) = (1, 1)$  in the sensed image and its corresponding point  $A'$  at  $(x', y') = (3, 5)$  is computed as  $(\Delta x, \Delta y) = (2, 4)$ . It is intuitive that there is a high probability that the candidates within a close neighborhood in the sensed image will have corresponding candidates that are spatially close to each other in the reference image. Therefore, a good center point for the search window of a particular candidate can be determined based on the mean mapping vector of its matched neighboring candidates. For example, if the mean vector is  $(\Delta x_{\text{mean}}, \Delta y_{\text{mean}}) = (2, 4)$ , then the center point of the search window for a point  $B$  at  $(x, y) = (2, 2)$  in the sensed image is computed as  $(x', y') = (4, 6)$  in the reference image. Similarly, the search-window size can be determined based on the mean deviation of the mapping vectors of neighboring candidates. A large mean deviation indicates that the corresponding candidates are not localized near the selected center point and so a larger search size is required to maintain matching accuracy. Based on these statistical relationships, a novel adaptive search strategy is devised for the proposed algorithm.

First, the sensed image is divided into  $k$  nonoverlapping partitions and the CP candidates are then grouped based on the partition where they reside. For each partition,  $m$  candidates are selected at random from the sensed image and are compared against all candidates in the reference image that fall within a radius of  $r_{\text{max}}$  around its location. In the case where georeferencing information is not known, the value of  $r_{\text{max}}$  is set as 100% of the maximum dimension and the value of  $m$  is set to 50% of candidates in a partition. If *a priori* georeferencing information is available, then each of the selected candidates from the sensed image is instead compared against those in the reference image that fall within a radius of  $r_{\text{max}}$  around the estimated corresponding coordinate in the reference image. In this case, the value of  $r_{\text{max}}$  is set to 33% of the minimum dimension to account for errors due to distortion and inaccuracies in the georeferencing information. Furthermore, the value of  $m$  is set to 30% of candidates in a partition due to better initial localization as a result of georeferencing information. These default parameters were determined based on testing with a variety of different remote-sensing images and can be changed by the user to improve results for a specific type of image. The similarity

between a candidate in the sensed image and a candidate in the reference image is calculated by finding the correlation of maximum moments of phase congruency within a  $w \times w$  neighborhood centered on the candidates. The value of  $w$  depends on the characteristics of the remote-sensing images being registered and can be determined experimentally based on the application.

The normalized cross-correlation measure is

$$\text{corr}(\text{CP1}, \text{CP2}) = \frac{\sum_x \sum_y (\bar{M}_{\text{CP1}}(x, y) \bar{M}_{\text{CP2}}(x, y))}{\sqrt{\sum_x \sum_y (\bar{M}_{\text{CP1}}(x, y))^2 (\bar{M}_{\text{CP2}}(x, y))^2}} \quad (5)$$

where CP1 and CP2 are CP candidates in the sensed image and the reference image, respectively, and  $\bar{M}_{\text{CP1}}(x, y)$  and  $\bar{M}_{\text{CP2}}(x, y)$  are the maximum moments of phase congruency as per (4) at  $(x, y)$  within a neighborhood centered at CP1 and CP2, respectively. Based on the matched candidate pairs, the mean vector and mean deviation of the mapping vectors are computed. For each remaining candidate in a partition, the search-window size is adjusted adaptively based on the equation

$$r = r_{\text{min}} + (\min(\text{MD}/r_{\text{min}}, 1)) (r_{\text{max}} - r_{\text{min}}) \quad (6)$$

where  $r_{\text{min}}$  and  $r_{\text{max}}$  are the minimum and maximum allowable search radii, respectively, and MD is the computed mean deviation in the current partition. The relationship between the radii is

$$r_{\text{min}} < r_{\text{max}} < r_{\text{max}}. \quad (7)$$

By default,  $r_{\text{min}}$  is set to 13% of minimum dimension and  $r_{\text{max}}$  is set to 30% of minimum dimension unless otherwise specified by the user. These values were chosen based on testing with a variety of different remote-sensing images and can be changed by the user to improve results for a specific type of image. The center point of the search window for the remaining CPs in the partition is set based on the computed mean vector in the current partition as described above. An example of this decision strategy for the remaining CPs is shown in Fig. 2.

Finally, the matching CP candidate pairs are determined by searching for maximal correlation peaks between candidates that are two-way consistent. For example, if the maximal correlation peak of CP1 corresponds to CP2 and the maximal correlation peak of CP2 corresponds to CP1, then they are considered a matching pair. Hence, the total amount of correlation matching performed is significantly reduced when compared to an exhaustive search and so the computational efficiency of the algorithm is improved to achieve sub- $O(n^2)$  performance. Therefore, the efficiency and robustness criteria are satisfied in ARRSI by the invariant local-feature descriptor and the adaptive search strategy used. The coordinates of the corresponding pairs are then scaled back to that of the original images.

### C. Outlier Rejection and Transformation Model Estimation

After the matching CP candidate pairs have been found, it is necessary to determine a spatial transformation model

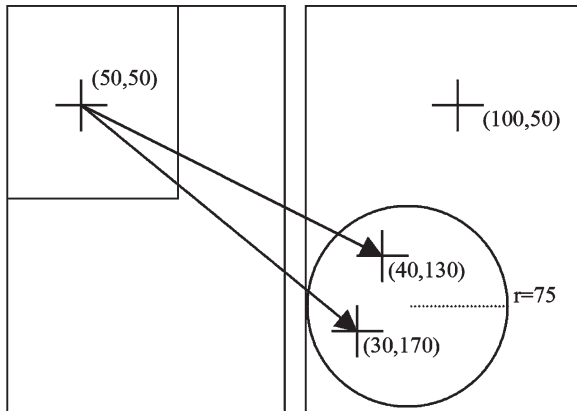


Fig. 2. This illustrates an example of deciding what candidates in the reference image are compared against a candidate in the sensed image in ARRSI. Based on matched coordinates of previously selected candidates within the partition, the center and size of the search window for a selected candidate located at  $(x, y) = (50, 50)$  within the partition are determined to be  $(x, y) = (50, 150)$  and  $r = 75$ , respectively. Therefore, all candidates in the reference image that reside within a circle that is centered at  $(x, y) = (50, 150)$  and have a radius of 75 are compared against the selected candidate. In this case, the candidate at  $(x, y) = (50, 50)$  in the sensed image is compared against the two candidates at  $(x, y) = (40, 130)$  and  $(x, y) = (30, 170)$  but not against the candidate at  $(x, y) = (100, 50)$ .

that maps the CP candidates from the sensed image to corresponding CP candidates in the reference image. There are a number of issues that need to be considered in order to establish a final transformation model. First, the type of models that represents the spatial transform between the sensed image and the reference image needs to be established. Common spatial transformation models used in image registration include translation, rigid, similarity, affine, projective, polynomial, and piecewise linear. The model used will depend on the types of relative geometric distortions exhibited in the sensed image. For the purpose of fast image registration of remotely sensed images, ARRSI support both the affine and projective transformation models due to low model-estimation complexity and the ability to handle most common geometric distortions found in remotely sensed imagery such as affine transformations (translation, rotation, scale, and shear) and perspective transformations. A more complex transformation model would be necessary for handling more complex cases such as spatially variant terrain relief images.

The second issue that needs to be considered is the method of estimating the parameters of the selected transformation model. One common approach of model-parameter estimation is least squares estimation as it often provides optimal estimates. However, the main problem with least squares estimation methods is the fact that they perform poorly in the presence of outliers. This is particularly problematic in the case of automatic image registration due to the high probability of candidate mismatches during the candidate-matching process. One approach to improving the robustness of the model-estimation process is to eliminate outlier candidate pairs from the set of candidate pairs. Two commonly used techniques for performing outlier rejection are least median of squares (LMS) [29] and RANSAC [30], which have proven popular due to their effectiveness and efficiency. In both algorithms, a subset of candidate pairs is selected at random from the total set of candidate pairs.

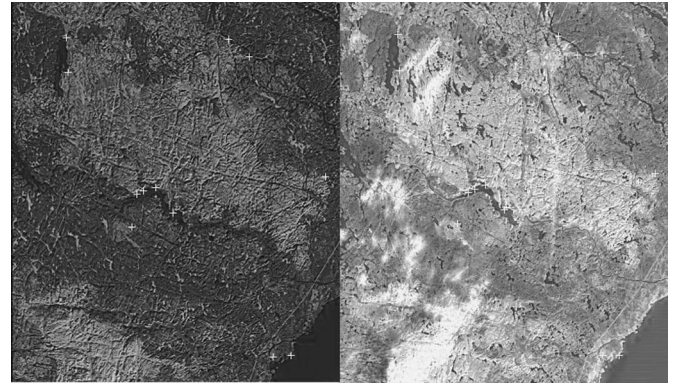


Fig. 3. Final set of candidate CPs (as indicated by crosses). (Left) Reference image. (Right) Sensed image.

These pairs are then used to estimate a candidate transformation model. In the case of LMS, the process is repeated for  $K$  iterations, and the candidate solution with the lowest median squared residual error is selected as the final transformation model. In the case of RANSAC, the process is repeated for  $K$  iterations, and the candidate solution with the highest number of inliers based on squared residual error is selected as the best transformation model. In the case of RANSAC, the number of iterations can be determined as

$$K = \frac{\log(1-p)}{\log(1-(n_{\text{inliers}}/N)^s)} \quad (8)$$

where  $p$  is the desired probability of selecting at least one transformation model that is free of outliers within  $K$  iterations,  $n_{\text{inliers}}$  is the number of candidates that fit the current estimated model,  $N$  is the total number of candidates, and  $s$  is the minimum number of candidates needed to fit the transformation model. The key advantage of RANSAC over LMS is that RANSAC also yields a final set of inliers that can then be used to determine a more refined transformation model using least squares estimation.

One issue with the use of the above algorithms is that all randomly selected candidate subsets are evaluated equally. Therefore, candidate subsets that consist of closely spaced candidates are treated the same as candidate subsets that consist of well-spaced candidates. However, subsets with closely spaced candidates do not provide a good representation of the overall transformation between the sensed image and the reference image. This problem can be illustrated with a simple example. Imagine that the transformation between the sensed image and the reference image is a clockwise rotation. Therefore, the candidates at the top right side of the sensed image are mapped to corresponding candidates that are positioned below and to the right of it, and the candidates at the bottom-left side of the sensed image are mapped to corresponding candidates that are positioned above and to the left of it. Therefore, if a subset consists of candidates that are clustered to the top right of the image, the resultant transformation-model estimate would reflect a translation if a similarity model was fitted. However, if the subset consists of candidates from both the bottom-left and top right side of the image, then the resultant transformation

TABLE I  
REGISTRATION ACCURACY

Test Set	No. of CPs matched automatically by ARRSI	RMSE <sup>1</sup>			
		Manual Selection	Baseline	Multimodal SSD [3]	ARRSI
<b>Intra-sensor and Intra-band images</b>					
INTRA1	41	0.5325	2.0501	0.6325	2.2816
INTRA2	67	1.3224	1.9790	4.1110	1.9097
INTRA3	40	1.3057	1.5694	45.332	1.5027
<b>Inter-sensor and Inter-band images</b>					
INTER1	26	2.1352	- <sup>2</sup>	55.6633	3.7815
INTER2	40	1.4428	- <sup>2</sup>	2.3664	1.6426
INTER3	36	1.6612	1.8174	6.5345	1.2184
INTER4	23	1.1073	- <sup>2</sup>	3.7815	2.7906
INTER5	10	8.4443	- <sup>2</sup>	- <sup>2</sup>	8.5854

1: The No. of CPs detected and RMSE are computed as the average of 5 test trials given the randomness in the registration process. The RMSE is computed in pixels relative to the reference image.

2: Fails to register images within RMSE < 100.

model would reflect a rotation if a similarity model was fitted. To address this problem, a variation on the RANSAC algorithm is introduced. It will be referred to as MDSAC. Rather than selecting a single subset per iteration, a number of subsets are selected and the sum of squared distances between the candidates within each subset is computed. The subset with the maximum sum-squared distance is used to produce the model estimate. This provides an effective and efficient method of encouraging subsets with better spaced candidates, thus improving the robustness of the system. An example of the final set of CPs for the sensed image and the reference image determined using MDSAC is shown in Fig. 3. This set of CPs, now largely free of outliers, is then used to estimate the final transformation model using a least squares algorithm such as the normalized direct-linear-transformation (DLT) algorithm described in [31].

#### D. Transformation and Resampling

Finally, the sensed image is transformed based on the final transformation model and then resampled using an interpolation method such as bilinear and bicubic interpolation depending on the level of image quality and the level of computational performance required.

#### E. Registration Algorithm

Based on the above theory, the ARRSI automatic registration algorithm is summarized as follows for a sensed image  $f$  and a reference image  $g$ .

- 1) Detect a set of CP candidates from  $f$  and  $g$  using the APCFD algorithm described in Section II-A. ROI may be specified manually or determined automatically using georeferencing information to improve performance and accuracy of the system.
- 2) Determine a set of CP candidate pairs between  $f$  and  $g$  by matching the maximum moments of phase congruency around the candidates using the efficient search strategy described in Section II-B.
- 3) Apply the MDSAC algorithm described in Section II-C on the set of CP candidates to determine a final set of CPs.

TABLE II  
REGISTRATION ACCURACY OF INTER3 AT VARIOUS ROTATIONS

Degree of Rotation	RMSE <sup>1</sup>		
	Manual Selection	Baseline	ARRSI
-30°	6.2157	- <sup>2</sup>	- <sup>2</sup>
-15°	2.6526	4.4173	2.4151
-5°	2.3801	2.3967	2.1200
5°	1.8267	1.8533	1.9570
15°	3.7068	6.4016	3.3826
30°	3.7805	- <sup>2</sup>	- <sup>2</sup>

1: The RMSE is computed as the average of 5 test trials given the randomness in the registration process. The RMSE is computed in pixels relative to the reference image.

2: Fails to register images within RMSE < 100.

- 4) Use the DLT algorithm to estimate a final transformation model using the final set of CPs.
- 5) Use the final transformation model to transform  $f$  into an aligned image  $f'$ .

### III. TESTING METHODS

The ARRSI system was implemented in MATLAB and was tested using six sets of images derived from the U.S. Geological Survey (USGS) Global Visualization Viewer project, one set derived from the Spaceborne Imaging Radar-C/X-Band Synthetic Aperture Radar (SIR-C/X-SAR) project and one set from Intermap Technologies Inc. The test sets can be divided into two categories: 1) intrasensor and intraband images and 2) intersensor and interband images. Intersensor and interband images are typically more difficult to register due to the differing intensity mappings of remotely sensed images acquired using different modalities and frequencies. Therefore, intersensor and interband images are more important in testing the effectiveness of the ARRSI system. All test images are eight-bit grayscale images. Each test set consists of a reference image and a sensed image. A description of each test case is described below.

- 1) *INTRA1*: Set of two  $902 \times 1131$  orthorectified air-photos of Highlands Ranch, CO, 1-m resolution. This test set was provided by Intermap Technologies Inc.
- 2) *INTRA2*: Set of two  $774 \times 750$  Landsat-7 ETM+ images from USGS project, 240-m resolution. Band: 3. Reference image: Lat/Long:  $64.2 / -81.5$ . Date:

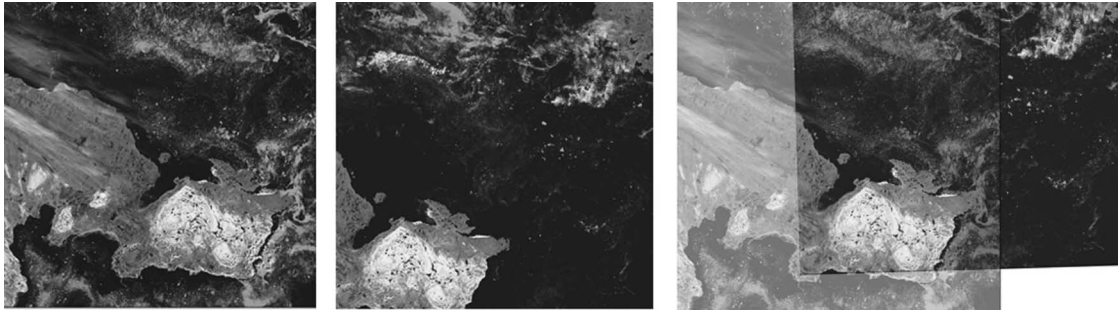


Fig. 4. Image registration from INTRA2 test set. (Left) Reference image. (Center) Sensed image. (Right) Aligned images.

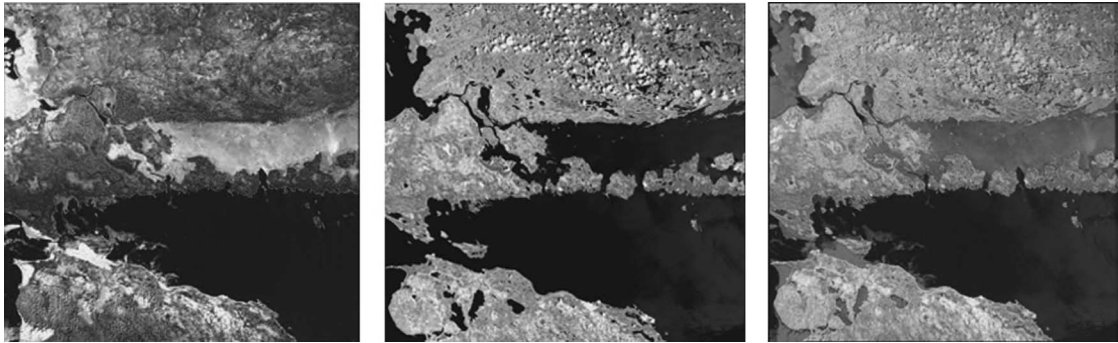


Fig. 5. Image registration from INTER1 test set. (Left) Reference image. (Center) Sensed image. (Right) Aligned images.

2002/7/26. Sensed image: Lat/Long: 64.2/  $-79.9$ . Date: 2002/7/17.

- 3) *INTRA3*: Set of two  $771 \times 749$  Landsat-7 ETM+ images from USGS project, 240-m resolution. Band: 3. Reference image: Lat/Long: 53.1/  $-82.4$ , Date: 2003/3/18. Sensed image: Lat/Long: 53.1/  $-80.8$ , Date: 2003/3/27.
- 4) *INTER1*: Set of two  $761 \times 748$  images from USGS project, Lat/Long: 46.0/  $-83.8$ , 240-m resolution. Reference image: Sensor: Landsat-7 ETM+, Band: 3, Date: 2003/4/12. Sensed image: Sensor: Landsat 4–5 TM, Band: 5, Date: 2006/06/15.
- 5) *INTER2*: Set of two  $761 \times 748$  images from USGS project, Lat/Long: 69.6/  $-92.7$ , 240-m resolution. Reference image: Sensor: Landsat-7 ETM+, Band: 5, Date: 2000/7/24. Sensed image: Sensor: Landsat 4–5 TM, Band: 3, Date: 1999/7/6.
- 6) *INTER3*: Set of two  $761 \times 748$  images from USGS project, Lat/Long: 46.0/  $-113.1$ , 240-m resolution. Reference image: Sensor: Landsat-7 ETM+, Band: 5, Date: 2001/8/17. Sensed image: Sensor: Landsat 4–5 TM, Band: 3, Date: 2006/2/12.
- 7) *INTER4*: Set of two  $761 \times 748$  images from USGS project, Lat/Long: 48.9/  $-68.8$ , 240-m resolution. Reference image: Sensor: Landsat-7 ETM+, Band: 5, Date: 1999/12/14. Sensed image: Sensor: Landsat 4–5 TM, Band: 3, Date: 2005/10/13.
- 8) *INTER5*: Reference image: Sensor: Landsat-7 ETM+, Band: 3, Lat/Long: 48.9/  $-68.8$ , 240-m resolution, Date: 1999/12/14. Sensed image: Sensor: SIR-C/X-SAR, Lat/Long: 41.8/  $-70.3$ , 25-m resolution, Date: 1994/04/15.

Each sensed image in a test case is registered with the reference image in the same test case. For all test cases, the number of initial CP candidates was set to a maximum of 300 candidates in each image and the sensed image was split into 16 nonoverlapping partitions. However, it is possible to utilize overlapping partitions to provide better local estimates for the adaptive search strategy used during the CP-candidate-matching process. The downscaling algorithm used in ARRSI is set up such that the maximum dimension is 400 pixels for all test cases except for INTRA1, where the images are downsampled such that the maximum dimension is 500 pixels. This is done to improve the overall performance of the system.

A set of 30 CP pairs was selected manually at the original resolution for each test case. From the set of manually selected GCP pairs, 20 pairs were selected for training purposes and the remaining ten pairs were selected for testing purposes. To judge the registration accuracy of the proposed algorithm, the root-mean-squared error (rmse) is computed for the ten test pairs. The unit used for determining rmse is based on pixels at original resolution. The rmse computed for a transformation model that was fitted using the 20 training pairs is used as the reference baseline standard. The baseline automatic registration algorithm described in [6] and the multimodal registration algorithm proposed in [3] were used for comparison. Finally, to test the robustness of ARRSI and the baseline algorithm under rotation conditions, the sensed image of the INTER3 test case was rotated around its center by  $\pm\{5^\circ, 15^\circ, 30^\circ\}$ , and the rmse was computed for each rotation. The multimodal registration algorithm was not evaluated for rotation conditions because it does not handle rotations.



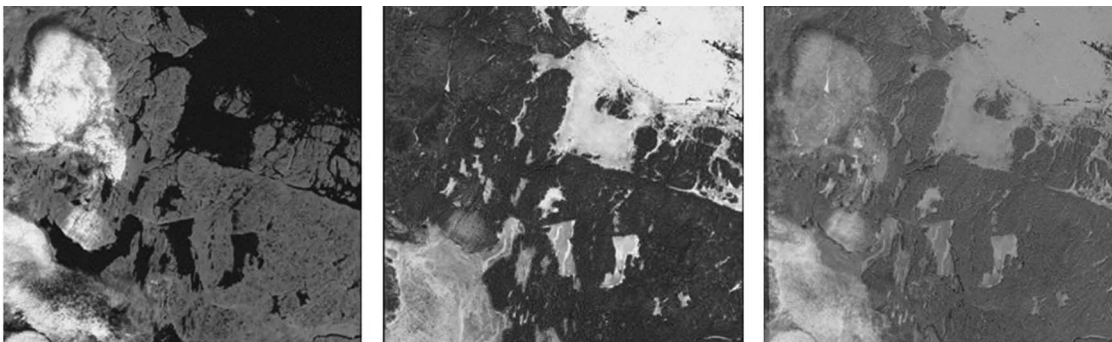


Fig. 6. Image registration from INTER2 test set. (Left) Reference image. (Center) Sensed image. (Right) Aligned images.



Fig. 7. Image registration from INTER5 test set. (Left) Reference image. (Center) Sensed image. (Right) Aligned images.

#### IV. EXPERIMENTAL RESULTS

The registration accuracy results are shown in Table I. It can be observed that the ARRSI system achieved rmse that are higher but within the range of that achieved by the manual selection of CP pairs in seven of the eight test cases. In the INTER3 test case, the proposed algorithm slightly outperforms the results achieved using the manually selected CP pairs. It can also be observed that the ARRSI system outperformed the multimodal registration algorithm proposed in [3] in seven of the eight test cases. Furthermore, the ARRSI system significantly outperformed the baseline algorithm in four of the eight test cases. The registration accuracy for the INTER3 test case at various rotations is shown in Table II. It can be observed that the ARRSI system achieved rmse that is comparable to that produced by manually selected CP pairs for  $\pm\{5^\circ, 15^\circ\}$ . For rotations of  $\pm 30^\circ$ , the ARRSI system performs poorly and is unable to produce a correct alignment. This illustrates the limitations of the ARRSI system in terms of robustness under rotation conditions. An example of the registration achieved for four of the test cases is shown in Figs. 4–7. By visual inspection, it appears that the registration is valid and accurate in all test cases. These results illustrate the effectiveness of the ARRSI system for registering intrasensor and intersensor remotely sensed images.

To analyze the effectiveness of the aforementioned MDSAC algorithm, motion-model estimation was performed for two test sets TEST1 and TEST2. The number of outlier pairs is pruned such that only 20% of the total pairs left are outliers.

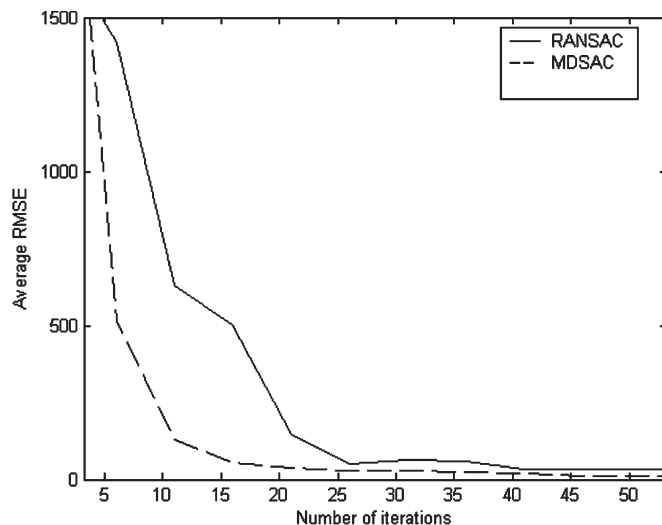


Fig. 8. Average rmse versus number of iterations for TEST1.

The algorithm was run for 100 iterations and the rmse of the resultant estimated model at various iterations was measured using a set of 30 ground-truth test CP pairs. Given the random nature of the algorithm, a total of 30 test trials were performed and the results were used to obtain an average rmse at various iterations. Figs. 8 and 9 show the average rmse plot for TEST1 and TEST2, respectively. The results for the standard RANSAC algorithm are provided as a comparison. It can be observed that the MDSAC algorithm provides noticeable performance

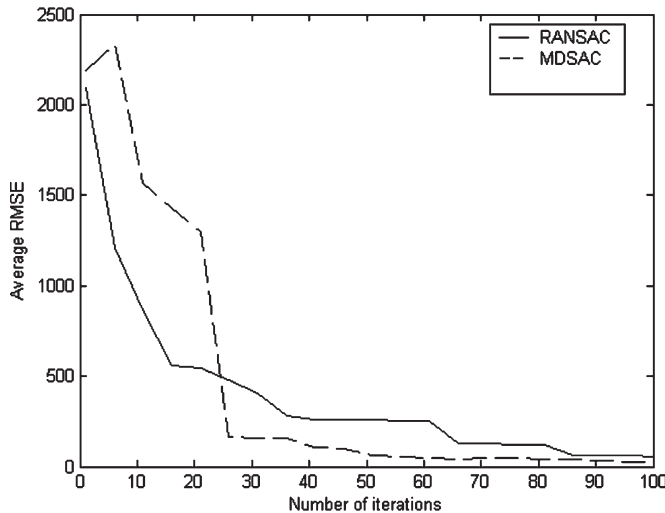


Fig. 9. Average rmse versus number of iterations for TEST2.

improvements over the standard RANSAC method in both cases. In the case of TEST1, it took RANSAC 26 iterations to achieve similar results as MDSAC after 16 iterations. In the case of TEST2, it took RANSAC 90 iterations to achieve the similar results as MDSAC after 50 iterations. Therefore, MDSAC is able to arrive at an estimated motion model with reasonably low rmse in fewer iterations than RANSAC. This is important as it indicates that the resultant pruned set of pairs is largely free of outliers and can be used to produce a more accurate estimation of the global motion using DLT.

## V. CONCLUSION AND FUTURE WORK

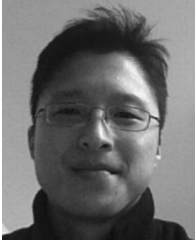
In this paper, we have introduced ARRSI, a new system for performing efficient and robust registration for remotely sensed images. Experimental results show that overall registration performance and accuracy are relatively high. It is our belief that this method can be successfully implemented for intrasensor and intersensor image registration and rectification purposes. Future work includes investigating the effectiveness of the ARRSI system using different similarity metrics such as sum of square differences and mutual information, as well as image-analysis techniques for automating the setting of parameters for the registration system.

## ACKNOWLEDGMENT

The authors would like to thank Intermap Technologies Inc., USGS, and NASA/JPL for the test data.

## REFERENCES

- [1] W. Li and H. Leung, "A maximum likelihood approach for image registration using control point and intensity," *IEEE Trans. Image Process.*, vol. 13, no. 8, pp. 1115–1127, Aug. 2004.
- [2] H. Chen, P. Varshney, and M. Arora, "Mutual information based image registration for remote sensing data," *Int. J. Remote Sens.*, vol. 24, no. 18, pp. 3701–3706, 2003.
- [3] J. Orchard, "Efficient global weighted least-squares translation registration in the frequency domain," in *Proc. Int. Conf. Image Anal. and Recog.*, 2005, pp. 116–124.
- [4] Q. Zheng and R. Chellappa, "Computational vision approach to image registration," *IEEE Trans. Image Process.*, vol. 2, no. 3, pp. 311–326, Jul. 1993.
- [5] J. Kybic and M. Unser, "Fast parametric elastic image registration," *IEEE Trans. Image Process.*, vol. 12, no. 11, pp. 1427–1442, Nov. 2003.
- [6] D. Capel and A. Zisserman, "Automated mosaicing with super-resolution zoom," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. and Pattern Recog.*, 1998, pp. 885–891.
- [7] E. Castro and C. Morandi, "Registration of translated and rotated images using finite Fourier transforms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-9, no. 5, pp. 700–703, Sep. 1987.
- [8] B. Reddy and B. Chatterji, "An FFT-based technique for translation, rotation and scale invariant image registration," *IEEE Trans. Image Process.*, vol. 5, no. 8, pp. 1266–1271, Aug. 1996.
- [9] I. Zavorin and J. Le Moigne, "Use of multiresolution wavelet feature pyramids for automatic registration of multisensor imagery," *IEEE Trans. Image Process.*, vol. 14, no. 6, pp. 770–782, Jun. 2005.
- [10] H. Li, B. Manjunath, and S. Mitra, "A contour-based approach to multisensor image registration," *IEEE Trans. Image Process.*, vol. 4, no. 3, pp. 320–334, Mar. 1995.
- [11] F. Eugenio, F. Marques, and J. Marcello, "A contour-based approach to automatic and accurate registration of multitemporal and multisensor satellite imagery," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2002, vol. 6, pp. 3390–3392.
- [12] C. Hsu and R. Beuker, "Multiresolution feature-based image registration," in *Proc. SPIE—Visual Communications and Image Processing*, 2002, vol. 4067, pp. 1490–1498.
- [13] V. Govindu and C. Shekhar, "Alignment using distributions of local geometric properties," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 10, pp. 1031–1043, Oct. 1999.
- [14] D. Xiaolong and S. Khorram, "A feature-based image registration algorithm using improved chain-code representation combined with invariant moments," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 5, pp. 2351–2362, Sep. 1999.
- [15] M. Ali and D. Clausi, "Automatic registration of SAR and visible band remote sensing images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2002, vol. 3, pp. 1331–1333.
- [16] S. Smith and J. Brady, "SUSAN—A new approach to low level image processing," *Int. J. Comput. Vis.*, vol. 23, no. 1, pp. 45–78, May 1997.
- [17] C. Harris and M. Plessey, "A combined corner and edge detector," in *Proc. 4th Alvey Vis. Conf.*, 1988, pp. 147–151.
- [18] M. Trajkovic and M. Hedley, "Fast corner detection," *Image Vis. Comput.*, vol. 16, no. 2, pp. 75–87, 1998.
- [19] A. Noble, "Descriptions of image surfaces," Ph.D. dissertation, Oxford Univ., Oxford, U.K., 1989.
- [20] P. Kovess, "Phase congruency: A low-level image invariant," *Psychol. Res.*, vol. 64, no. 2, pp. 136–148, 2000.
- [21] —, "Image features from phase congruency," *J. Comput. Vis. Res.*, vol. 1, no. 3, pp. 2–26, 1999.
- [22] —, "Phase congruency detects corners and edges," in *Proc. Australian Pattern Recog. Soc. Conf.*, 2003, pp. 309–318.
- [23] M. Morrone and R. Owens, "Feature detection from local energy," *Pattern Recognit. Lett.*, vol. 6, no. 5, pp. 303–313, Dec. 1987.
- [24] M. Morrone and D. Burr, "Feature detection in human vision: A phase-dependent energy model," *Proc. R. Soc. London B*, vol. 235, no. 1280, pp. 221–245, 1988.
- [25] S. Young, "Alias-free image subsampling using Fourier-based windowing methods," *Opt. Eng.*, vol. 43, no. 4, pp. 843–855, 2004.
- [26] A. Dumitras and F. Kossentini, "Fast and high performance image subsampling using feedforward neural networks," *IEEE Trans. Image Process.*, vol. 9, no. 4, pp. 720–728, Apr. 2000.
- [27] D. Lowe, "Object recognition from local scale-invariant features," in *Proc. Int. Conf. Comput. Vis.*, 1999, pp. 1150–1157.
- [28] —, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
- [29] C. Stewart, "Robust parameter estimation in computer vision," *SIAM Rev.*, vol. 41, no. 3, pp. 513–537, 1999.
- [30] 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, Jun. 1981.
- [31] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge, U.K.: Cambridge Univ. Press, 2001.



**Alexander Wong** (S'05) received the B.A.Sc. degree in computer engineering from the University of Waterloo, Waterloo, ON, Canada, in 2005.

In 2004, he was with Epson Canada Ltd., ON, Canada, in the area of commercial computer vision and image processing. During his time there, he was responsible for several patent applications in the areas of automatic panoramic generation and automatic media-management systems. He is currently with the University of Waterloo. He has published refereed journal and conference papers in various

fields such as computer vision, graphics, image processing, and multimedia systems. His current research encompasses computer vision, image and video processing, remote sensing, medical imaging, multisensor systems, and multimedia retrieval systems.



**David A. Clausi** (S'93–M'96–SM'03) received the B.A.Sc., M.A.Sc., and Ph.D. degrees from the University of Waterloo, Waterloo, ON, Canada, in 1990, 1992, and 1996, respectively, all in systems design engineering.

After completing his doctorate, he worked in the medical-imaging field at Mitra Imaging Inc., Waterloo. He started his academic career in 1997 as an Assistant Professor in Geomatics Engineering at the University of Calgary, AB, Canada. In 1999, he returned to the University of Waterloo, and was

awarded tenure and promotion to Associate Professor in 2003. He is an active interdisciplinary and multidisciplinary researcher. He has an extensive publication record, publishing refereed journal and conference papers in the diverse fields of remote sensing, computer vision, algorithm design, and biomechanics. His primary research interest is the automated interpretation of synthetic aperture radar sea-ice imagery, in support of operational activities of the Canadian Ice Service. The research results have led to successful commercial implementations.

Dr. Clausi is the recipient of numerous scholarships, conference paper awards, and two Teaching Excellence Awards.