

# Multisensor Image Registration via Implicit Similarity

Yosi Keller and Amir Averbuch

**Abstract**—This paper presents an approach to the registration of significantly dissimilar images, acquired by sensors of different modalities. A robust matching criterion is derived by aligning the locations of gradient maxima. The alignment is achieved by iteratively maximizing the magnitudes of the intensity gradients of a set of pixels in one of the images, where the set is initialized by the gradient maxima locations of the second image. No explicit similarity measure that uses the intensities of both images is used. The computation utilizes the full spatial information of the first image and the accuracy and robustness of the registration depend only on it. False matchings are detected and adaptively weighted using a directional similarity measure. By embedding the scheme in a “coarse to fine” formulation, we were able to estimate affine and projective global motions, even when the images were characterized by complex space varying intensity transformations. The scheme is especially suitable when one of the images is of considerably better quality than the other (noise, blur, etc.). We demonstrate these properties via experiments on real multisensor image sets.

**Index Terms**—Global motion estimation, multisensor registration, multimodality image alignment.

## 1 INTRODUCTION

THE registration of images acquired by sensors of different modalities is of special interest to remote sensing and medical imaging applications as the information gained from such sets of images is of a complementary nature. Proper fusion of the data obtained from separate images requires accurate spatial alignment. This issue was extensively studied in the context of remote-sensing (SAR, FLIR, IR, and optical sensors) [1], [2] and medical image registration (CT, MRI) [3], [4], [5]. Due to the different physical characteristics of various sensors, the relationship between the intensities of matching pixels is often complex and unknown a priori. Features present in one image might appear only partially in the other image or do not appear at all. Contrast reversal may occur in some image regions while not in others, multiple intensity values in one image may map to a single intensity value in the other image and vice versa. Furthermore, imaging sensors may produce considerably dissimilar images of the same scene when configured with different imaging parameters.

Geometrical matching techniques [2], [6], [7] align geometrical primitives such as feature points, contours, and corners using vectorial representations. These matchings are invariant to intensity changes once the geometrical primitives are detected in both input images. Since the geometrical primitives are matched one by one, such algorithms are not used within a global optimization scheme, where the registration is derived by optimizing a functional over the

entire set of features. This results in poor global alignment, sensitivity to outliers, and the need to apply random sampling and cross-validation techniques, such as RANSAC [8].

Thus, for the monomodal as well as the multimodal case, the general approach to image registration consists of defining an *explicit similarity measure* [9], a functional  $\phi = \phi(I_1, I_2)$  applied to the intensities of the input images  $I_1$  and  $I_2$  which measures their similarity. The choice of the similarity measure corresponds to an assumption regarding the relationship between the intensities of the images (e.g., functional, statistical, etc.). The registration is then derived by optimizing  $\phi$  with respect to the motion parameters vector  $\mathbf{p}$ . For instance, for monomodal images,  $\phi$  is often chosen to be the  $L_2$  norm

$$\mathbf{p}^* = \arg \min_{\mathbf{p}} \sum (I_1 - I_2(\mathbf{p}))^2, \quad (1)$$

where  $I_2(\mathbf{p})$  is a replica of  $I_2$  warped according to  $\mathbf{p}$  and the summation is conducted over the common parts of  $I_1$  and  $I_2(\mathbf{p})$ . For significantly dissimilar (multisensor) images, solving (1) does not result in image registration and other similarity measures, such as correlation coefficient [1], correlation ratio [10], mutual information [3], [4], [11], and statistical divergence [12], are used. Further robustness can be achieved by registering transformed representations of  $I_1$  and  $I_2$  which are invariant to certain brightness dissimilarities. These representations include feature points [13], edge-maps [6], oriented edge-maps [1], and edge contours [2]. Matching is achieved by applying gradient methods [6], robust gradient methods [1], chain-code correlation [2], or geometrical hashing [7] to the invariant representations. The computation of invariant representations results in important image information being lost and the optimization process might fail to converge.

The Mutual Information (MI) similarity measure [3], [4], [11], [14] originates from information theory and is considered the state-of-the-art registration framework for multimodality images. It is a measure of statistical dependency between two data sets. Given two random variables  $X$  and  $Y$ , their MI is

- Y. Keller is with the Department of Mathematics, Yale University, PO Box 208283, New Haven, CT 06520-8283. E-mail: yosi.keller@yale.edu.
- A. Averbuch is with the Department of Computer Science, School of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, Israel. E-mail: amir@math.tau.ac.il.

Manuscript received 27 Apr. 2005; revised 14 Aug. 2005; accepted 30 Sept. 2005; published online 13 Mar. 2006.

Recommended for acceptance by A. Rangarajan.

For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-0216-0405.

$$MI(X, Y) = \sum_{ij} P(X_i, Y_j) \log \left( \frac{P(X_i, Y_j)}{P(X_i)P(Y_j)} \right), \quad (2)$$

where  $P(X)$  is the probability of  $X$  and  $P(X, Y)$  is the joint probability of  $X$  and  $Y$ .  $MI(I_1, I_2)$  was shown to be maximized for aligned images [3]. The optimization can be implemented by either Powell's method [14] or a faster coarse-to-fine Marquardt-Levenberg scheme [15].

However, MI-based matchings can result in misregistration [14], due to several issues. First, the MI measure is highly nonconvex with respect to the transformation parameters  $p$ . Thus, the optimization procedure can be easily attracted to local maxima. Moreover, the nonconvexity of the criterion is enhanced by the interpolation used in subpixel accurate motion estimation [16]. Second, when the images are of low resolution, contain little information, or there is only a small overlapping region, they become less statistically dependent and the maximization of the MI measure becomes ill-posed. Third, the assumption of a global statistical relationship between the image intensities may be violated by the presence of various image artifacts. For instance, intensity inhomogeneities are found in several image modalities [5], while, in remote-sensing applications, the image structures being matched result from different physical properties.

The work of this paper, preliminarily presented in [17], suggests an implicit similarity measure to address some of the limitations of the approaches discussed above. Thus, no functional  $\phi(I_1, I_2)$  which uses the intensities of both images is used. Instead, the algorithm aligns the set of pixels having large gradient magnitudes in both images in two steps; first, the pixels are detected in  $I_1$  (using only  $I_1$ 's intensity value). Second, the pixels are aligned to  $I_2$  using iterative optimization, where the optimization is carried out using only  $I_2$ 's intensity values. Hence, the convergence only depends on the properties of  $I_2$  which can be chosen to be the better quality input image. Dominant edges can be detected in noisy images by the smoothing edge detection schemes discussed in Section 2.1. This approach allows feature matching using a global optimization scheme, where, in prior works, only one of these attributes was achieved. In this work, we improve the robustness of the preliminary scheme given in [17] by introducing a weight function based on the angle between the gradient vectors of matching pixels. This allows adaptive handling of false edge correspondences.

Compared to previous schemes that combined edge-maps representations and gradient-based optimization [2], [6], [13], our scheme utilizes the full spatial information of  $I_2$  and not just its edge map. Moreover, nonlinear intensity mappings might result in significantly different edge map values for corresponding pixels such that minimizing their differences does not result in image alignment. The proposed scheme does not minimize discrepancies. It iteratively maximizes the gradient energy over a single image and is thus invariant to the intensities variations between the input images.

The paper is organized as follows: Section 2 presents the proposed algorithm and its convergence is analyzed in Section 3. Experimental results and concluding remarks are given in Sections 4 and 5, respectively.

## 2 THE PROPOSED ALGORITHM

In this section, we formulate the proposed scheme illustrated in Fig. 1. Let  $I_1$  and  $I_2$  be the input images, we

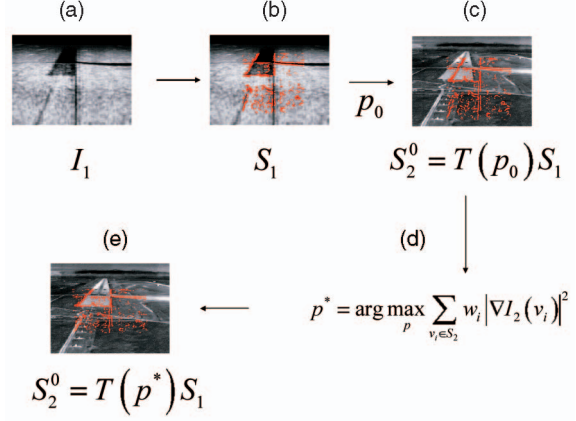


Fig. 1. The flow of the multisensor registration algorithm. (a) Input image  $I_1$ . (b)  $S_1$ , the set of pixels with the largest gradient magnitude is overlaid in red. (c)  $S_1$  is warped toward  $I_2$  using the initial estimate of the motion  $p_0$  and denoted  $S_2^0$ . (d)  $S_2^0$  is used to initialize an iterative maximization of the gradient magnitudes of  $S_2$ . (e) For the maximal gradient energy of  $S_2$ ,  $p^*$  corresponds to the motion aligning  $I_1$  and  $I_2$ .

aim to compute the parametric motion  $T(p)$ , where  $p$  is a parameters vector that maps  $S_1$ , a set of large gradient magnitude pixels in  $I_1$ , to the corresponding set  $S_2$  in  $I_2$ .  $T(p)$  will also align the remainder of the images  $I_1$  and  $I_2$ .

The sets are aligned by maximizing the gradient energy of the set  $S_2$  as a function of  $p$

$$F(p) \triangleq \sum_{v_i \in S_2} w_i |\nabla I_2(v_i)|^2, \quad (3)$$

where  $v_i = (x_i, y_i)$  and  $0 \leq w_i \leq 1$  are adaptive weights whose computation is detailed in Section 2.4.

The scheme first detects the set  $S_1$  in  $I_1$  and then maximizes  $F(p)$  iteratively. At each iteration  $n$ ,  $S_2$  is given by

$$S_2^n = T(p_n)S_1, \quad (4)$$

where  $p_n$  is the parameters vector at iteration  $n$  and  $S_2^n$  is computed by applying the transform  $T(p_n)$  to every pixel in  $S_1$ .

The detection of  $S_1$  requires the intensity values of  $I_1$  only and (4) uses the intensities of  $I_2$  only as  $S_1$  is just a set of coordinates. Thus, at no step of the alignment process do we have to directly minimize the discrepancies between the intensities of  $I_1$  and  $I_2$  or any function of them, such as directional derivatives or edge maps. Hence, we consider our approach an *implicit solution* of the registration of  $I_1$  and  $I_2$  as no explicit similarity measure (similar to  $\phi(I_1, I_2)$ ) is used. The accuracy and robustness of the solution mainly depend on the properties of  $I_2$  and, by choosing it to be the better quality image, the performance of our approach is limited by the properties of the better quality image rather than the worst quality one as in other schemes. Note that  $I_1$  is only needed to detect the initial set of pixels  $S_1$  and that edges can be reliably detected in noisy and low quality images. Equation (3) is maximized using a global optimization scheme that utilizes the entire set  $S_2$  and the full intensity information of  $I_2$  over  $S_2$ .

The underlying assumption, that pixels with large gradient magnitudes can be found in both images, stems from the fact that such pixels were shown to correspond to salient structures in natural images [18]. A similar approach was used in [19] to introduce edge information into the Mutual

Information framework. In particular, due to the inherent resolution limitations of imaging sensors, multisensor and multimodality images are often untextured (similar to the ones shown in Fig. 1); thus, the similarity between such images is strictly related to the locations of their edges.

We note that, while the choice of image gradients as the features used for registration proved efficient and straight forward, one can use other local features such as directional derivatives [20]. Such a feature vector,

$$\phi(\mathbf{v}_i) = [\nabla_x I(\mathbf{v}_i) \quad \nabla_y I(\mathbf{v}_i) \quad \nabla_{xy} I(\mathbf{v}_i) \quad \dots],$$

can be used within our framework by defining a scalar measure  $\psi(\mathbf{v}_i)$  on it. Thus, (3) is a particular choice of such a measure

$$\psi(\mathbf{v}_i) = \nabla_x I(\mathbf{v}_i)^2 + \nabla_y I(\mathbf{v}_i)^2.$$

An optimal measure  $\psi$  can be chosen by maximizing the mutual information between the measure  $\psi$  computed in both images as in [20]. Such an approach can be further extended by applying the Information Bottleneck feature selection scheme [21]. Alternatively,  $\psi$  can be a local PCA measure such as the metric tensor.

## 2.1 Initial Pixel Set Detection

The initial pixel set  $S_1$  is computed by detecting the set of pixels in  $I_1$  with the highest gradient magnitudes where the gradients are computed using centralized finite differences

$$\begin{aligned} |\nabla I(i, j)| &= \sqrt{I_x^2(i, j) + I_y^2(i, j)}, \\ I_x(i, j) &= 0.5(I(i, j+1) - I(i, j-1)), \\ I_y(i, j) &= 0.5(I(i+1, j) - I(i-1, j)). \end{aligned} \quad (5)$$

Contrary to higher-level geometric primitives, such as contours and corners, no preset thresholds are needed. For instance, there is no need to determine the parameter  $k$  for the Harris corner detector [22] or chain together edge pixels to identify contours.  $|S_1|$ , the number of pixels used, is typically 25 percent of the total number of pixels in  $I_1$ . An estimate of the optimal set size  $|S_1|$  is derived in Section 3.

$S_1$  can be detected even if the image  $I_1$  is noisy or blurred. In such a case, we employ a higher level edge detector that incorporate image smoothing. Possible choices for such schemes include (among others) Canny's edge detector, the DoG operator, and regularized edge detection schemes [23].

An accurate global transformation can be found only if the selected points are properly distributed across the image and are not clustered along a few edges. In order to ensure a uniform distribution,  $I_1$  is divided into  $10 \times 10$  regions, where each region is of size  $\frac{M}{10} \times \frac{N}{10}$ , where  $(M, N)$  are the dimensions of  $I_1$ . In each region, the top 25 percent of the pixels are chosen. Using a different number of regions did not change the registration results and  $10 \times 10$  regions were used in all of our experiments. This procedure is summarized in Algorithm 1.

### Algorithm 1. Detecting the Initial Set of Pixels

- 1:  $I_1$  is divided into  $10 \times 10$  subimages, each of size  $\frac{M}{10} \times \frac{N}{10}$ , where  $(M, N)$  are the dimensions of  $I_1$ .
- 2: Equation (5) is used to compute the edges in each subimage.
- 3: In each subimage, the top 25 percent of the edges are chosen.

## 2.2 Iterative Optimization

In this section, we present a numerical scheme for the maximization of (3).  $F(\mathbf{p})$  is nonlinear with respect to  $\mathbf{p}$  and the motion parameters are computed using Newton's iterative optimization method [24]. The iterative refinement is given by

$$\mathbf{p}_{n+1} = \mathbf{p}_n + (H^n)^{-1} \nabla_{\mathbf{p}} F_n, \quad (6)$$

where  $\nabla_{\mathbf{p}} F_n$  and  $H^n$  are the gradient and Hessian of  $F$ , respectively, with respect to  $\mathbf{p}$  at iteration  $n$ .  $\nabla_{\mathbf{p}} F_n(\mathbf{v}_i)$  and  $H^n(\mathbf{v}_i)$ , the gradient and Hessian evaluated at a pixel  $\mathbf{v}_i$ , are computed by the derivative chain rule. The adaptive weights,  $w_i$ , are computed in each iteration according to (11). Thus, for the affine and projective motion models, we have [25]

$$\nabla_{\mathbf{p}} F_n = \sum_{\mathbf{v}_i \in S_2^n} w_i X_i^T \nabla_{\mathbf{v}} F_n(\mathbf{v}_i) \quad (7)$$

and

$$H_{\mathbf{p}} = \sum_{\mathbf{v}_i \in S_2^n} w_i X_i^T H_v^n(\mathbf{v}_i) X_i, \quad (8)$$

where  $\nabla_{\mathbf{v}} F_n(\mathbf{v}_i)$  and  $H_v^n(\mathbf{v}_i)$  are the gradient and Hessian of  $F_n$ , respectively, with respect to  $(x, y)$ . For the affine and projective motion models,  $X_i$  is given by

$$X_{Affine} \triangleq \begin{bmatrix} x_i & y_i & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & x_i & y_i & 1 \end{bmatrix} \quad (9)$$

and

$$X_{Projective} \triangleq \begin{bmatrix} x_1 & y & 1 & 0 & 0 & 0 & x_i^2 & x_i y_i \\ 0 & 0 & 0 & x_i & y_i & 1 & x_i y_1 & y_1^2 \end{bmatrix}. \quad (10)$$

Equation (6) is iterated until a maximal predefined number of iteration is reached or the update of the parameters in (6) becomes smaller than a predefined threshold  $\varepsilon$ ,  $\|\mathbf{p}_{n+1} - \mathbf{p}_n\| < \varepsilon$ .

## 2.3 Multiscale Extension

In order to improve the convergence properties, a resolution pyramid is constructed. The alignment is initiated at the coarsest resolution scale of the pyramid and then follows the subsequent levels in a coarse-to-fine approach. The computation given in Section 2.2 is repeated at each resolution scale, where the result of each scale serves as the initial estimate of the solution at the finer scale. Finally, when the procedure stops at the finest resolution scale, the final motion parameters are obtained. The set  $S_1$  is initially detected at the finest resolution scale, then, given the set at a certain resolution scale, the coarser level set  $\tilde{S}_1$  is computed using Algorithm 2.

### Algorithm 2. Coarse Scale Pixel Set Detection

- 1: The coordinates of each pixel in  $S_1$  are scaled down by the scale factor  $S$  and the new set is denoted  $\tilde{S}_1$ .
- 2: A local search in an area of  $S \times S$  around each pixel in  $\tilde{S}_1$  is conducted to find the pixel with the largest gradient magnitude. That pixel is inserted into  $\tilde{S}_1$ .
- 3:  $\tilde{S}_1$  is scanned to eliminate duplicate pixels.



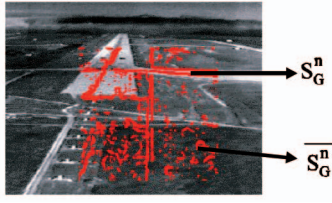


Fig. 2. Definition of the pixel sets used in the convergence analysis.  $S_2$  is the set of all the pixels in the image.  $S_2^n$  is the set of pixels marked in red. These pixels are used by the registration algorithm.  $S_G^n$  are the pixels within  $S_2^n$  which have high gradient magnitudes. The rest of the pixels in  $S_2^n$  are referred to as  $\overline{S}_G^n$ .

The coarse-to-fine refinement allows the alignment process to lock on a single motion denoted the *dominant motion* [25], even when multiple motions are present. This property is essential when outlier feature points are present.

## 2.4 Adaptive Weight Computation

The introduction of the adaptive weights  $w_i$  into (3) improves the handling of false matchings of points in  $S_1$  and  $S_2^n$ . The weights measure the correspondence of the gradient directions by computing the cosine of the angle between corresponding gradients in  $S_1$  and  $S_2^n$

$$w_i = |\cos \theta| = \frac{|\nabla I_1(\mathbf{u}_i) \cdot \nabla I_2(\mathbf{v}_i)|}{|\nabla I_1(\mathbf{u}_i)| |\nabla I_2(\mathbf{v}_i)|}, \quad (11)$$

where  $\mathbf{u}_i \in S_1$  and  $\mathbf{v}_i \in S_2^n$  are the  $i$ th corresponding pixel in  $I_1$  and  $I_2$ , respectively. The absolute value of  $\cos \theta$  is used as the gradients may point in opposite directions due to contrast reversal.

## 3 CONVERGENCE ANALYSIS

The convergence properties of the proposed scheme are similar to those of gradient methods [25]. Hence, a sufficient convergence condition can be derived by considering the refinement term in (6):

$$\delta \mathbf{p}_n \triangleq (H^n)^{-1} \nabla_p F_n. \quad (12)$$

The asymptotic result ( $n \rightarrow \infty$ ) of the optimization does not depend on  $H^n$ , which can be any positive-definite matrix and whose choice, only affects the convergence rate. Hence, the asymptotic results depend on  $\nabla_p F_n$ . Define the following pixel sets, depicted in Fig. 2.

- $S_2$ . All the pixels in  $I_2$ .
- $S_2^n$ . A subset of  $S_2$ —the set of pixels used at iteration  $n$  in  $I_2$  to evaluate (3).
- $S_G^n$ . A subset of  $S_2^n$ —the set of pixels in  $S_2^n$  having large gradient magnitudes and corresponding to true alignment.
- $\overline{S}_G^n$ . A subset of  $S_2^n$ —the set of pixels in  $S_2^n$  having small gradient magnitudes and corresponding to false alignment.

$E\{\nabla_p F_n\}$ , the expectancy of  $\nabla_p F_n$ , is shown in the Appendix to be

$$E\{\nabla_p F_n\} \sim (E\{S_G^n\} - E\{\overline{S}_G^n\})P(\mathbf{v}_2 \in S_G^n)(1 - P(\mathbf{v}_1 \in S_G^n)) \quad (13)$$

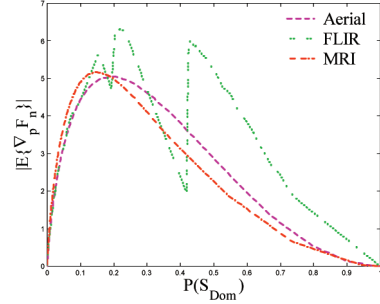


Fig. 3.  $|E\{\nabla_p F_n\}|$  as a function of the size of the initial pixel set. Choosing the pixel set size to be 25 percent of the images size proved to be a reasonable choice for the test image set. The images used for the computation of  $|E\{\nabla_p F_n\}|$  are shown in Section 4.

and, in order to achieve robustness to outliers and image noise, we require  $|E\{\nabla_p F_n\}| \gg 0$ .

For a given pair of input images, the optimal value of  $P(\mathbf{v}_2 \in S_G^n)$  can be numerically estimated by evaluating (13). Fig. 3 shows  $|E\{\nabla_p F_n\}|$  as a function of  $P(\mathbf{v}_2 \in S_G^n)$  for several images and setting  $\frac{|S_G^n|}{|S_2^n|} = 25\%$  turns out to be a reasonable choice. While  $P(\mathbf{v}_2 \in S_G^n)$  cannot be set directly, we note that

$$P(\mathbf{v}_2 \in S_G^n) = \frac{|S_G^n|}{|S_2^n|} \leq \frac{|S_2^n|}{|S_2|} \quad (14)$$

and an upper limit to  $S_G^n$  can be placed by choosing the size of  $S_2^n$ . Furthermore, it follows that, for a given value of  $|E\{\nabla_p F_n\}|$ , the higher the difference in  $I_2$  between  $E\{S_G^n\}$  and  $E\{\overline{S}_G^n\}$ , the better the convergence properties.

## 4 EXPERIMENTAL RESULTS

The proposed Implicit Similarity (IS) approach was experimentally evaluated by registering image pairs acquired in medical and remote sensing applications. The optimal size of  $S_1$  was computed according to (13) and a multiresolution pyramid (Section 2.3) with three resolution scales was used. One hundred iterations were used at each resolution scale and the initial pixel set was computed by dividing the image into  $10 \times 10$  subimages and finding the top 25 percent of the gradient maxima in each subimage. Similar results were achieved for subdivisions of  $5 \times 5$  and  $20 \times 20$ .

The initial estimate of the motion was given by manually picking a pair of matching pixels in both images. The registration accuracy was visually verified by overlaying the initial and final pixel sets on both images. For comparison, the images were also registered by a Mutual Information (MI) based scheme, where the densities were estimated by histograms with a fixed number of 100 bins for both images when estimating the marginal densities and  $100 \times 100$  bins when estimating the joint density. The intensity values were linearly mapped to the bins and Powell's method was used for optimization. The images were resampled by bilinear interpolation. As the ground truth was unknown for the sets of real image, it was estimated by manually picking 10 corresponding points and computing a least squares estimate of an affine motion model.

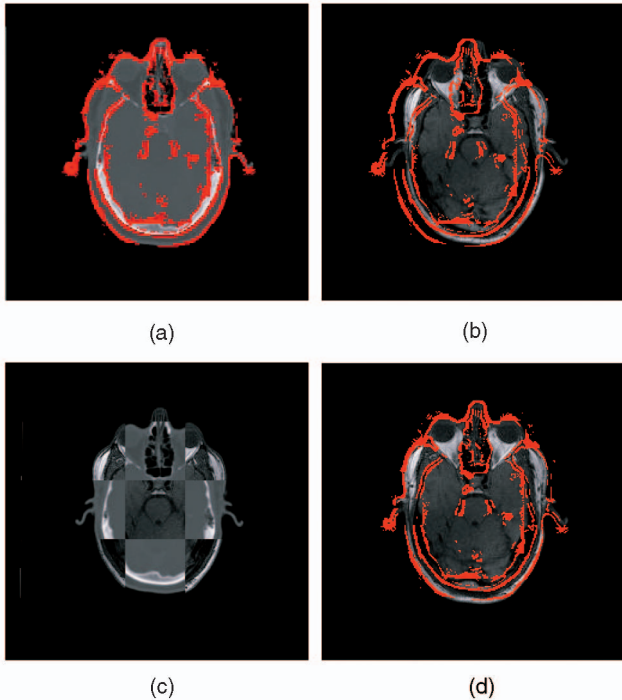


Fig. 4. Registration results of CT and MRI multimodality images using the affine motion model: (a) Original CT image overlaid with the set of high-gradient pixels. (b) Initial estimate of the overlaid set in the MRI image. (c) Composite image. (d) Final alignment of the pixel set after registration.

We provide three test cases for which we relate the properties of the registered images and the intensity mappings between them to the performance of the alignment schemes. The first deals with medical multimodality images that are well registered using MI. We compare the accuracy and noise sensitivity of the IS to MI in that case. Second, we register multisensor images where the MI fails and the proposed scheme is still applicable. Last, we provide a simulated example (Fig. 8) where the proposed scheme outperformed the normalized correlation [1] which can be used to register the images in the second example.

Fig. 4 shows the registration of MRI and CT images using an affine motion model

$$\begin{aligned} x_1 &= a \cdot x_2 + b \cdot y_2 + c, \\ y_1 &= d \cdot x_2 + e \cdot y_2 + f. \end{aligned} \quad (15)$$

The composite mosaic and aligned overlay are given in Figs. 4c and 4d, respectively. The registration is accurate as it is visually evident. The accuracy was quantitatively assessed by manually aligning the input images, applying 50 random rotations and translations to the CT image, and registering it to the MRI image using the IS and MI schemes. The motions parameters were normally distributed around  $(\Delta x, \Delta y, \theta) = (5, 5, 0^\circ)$  with standard deviations of  $(5, 5, 10^\circ)$  and the results are shown in Table 1. The MI was more accurate than the IS, which was implemented with and without the directional weights, which were shown to improve the accuracy.

The sensitivity to noise was tested by adding different levels of White Gaussian noise (WGN) to the images. We added noise levels of  $\sigma = 0, \dots, 10$  with steps of  $\Delta\sigma = 0.1$  and searched for the noise level for which the algorithms converged to local minima in more than 20 percent of the

TABLE 1  
Accuracy Comparison of the Different Registration Schemes when Applied to the Multimodality Images in Fig. 4

Scheme	$\sigma_\varepsilon$		
	$\Delta x$	$\Delta y$	$\theta$
MI	0.22	0.19	0.053
IS	0.24	0.21	0.071
IS+DW	0.23	0.17	0.062

Mutual Information (MI), Implicit Similarity (IS), and Implicit Similarity with adaptive directional weights (IS + DW).  $\sigma_\varepsilon$  is the standard deviation of the errors.

TABLE 2  
The Convergence Range of the Different Registration Schemes

	MI	IS	IS+DW
$\sigma_n$	5.5	3.0	3.6

We compare the Mutual Information (MI), Implicit Similarity (IS), and Implicit Similarity with adaptive directional weights (IS + DW).

trials. These values are depicted in Table 2, where we see that the MI was more robust, and using the directional weights improved the performance of the IS as it rejects outlier correspondences.

The MI scheme outperformed the IS approach in this test and we attribute that to the global intensity mapping between the registered images: All pixels with certain intensity  $x$  in Fig. 4a correspond to pixels having the same intensity value  $y$  in Fig. 4b. This intensity injection (one-to-one mapping) results in a well-defined mutual information function, which is statistically spatially stationary. Moreover, due to the black (empty) background, there are no large nonoverlapping regions which might result in noninjective mappings. This is the preferred setting for applying the MI framework and the registration results in a well-defined maximum of the mutual information function. The injection property does not hold for the remote sensing images in Figs. 5, 6, and 7, degrading the performance of the MI registration. Figs. 5 and 7 were registered using an affine motion model, while a projective motion was used in Fig. 6. In all cases, the proposed algorithm provided accurate results, while the MI scheme diverged. We attribute this to the spatially-varying, noninjection intensity mapping characterizing the images, compared to the injection mapping in the MRI/CT example. This makes the global mutual information function ill-posed and causes the MI to diverge. It is further exemplified in Fig. 5e where we show the MI at each iteration of the IS registration process, we see that the measure does not significantly increase as we align the images. Moreover, by comparing the MI measure in Fig. 5e over different resolution levels, we notice its ineffectiveness in low resolution scales. This property was previously mentioned in [26] and makes the implementation of multi-scale MI schemes difficult.

The accuracy of the proposed scheme was estimated by registering the images in Fig. 8. First, a reference solution was estimated by applying gradient methods [25] to Figs. 8a and 8c. Second, Figs. 8b and 8c were registered by the proposed algorithm, where the pixel set of dominant edge maxima in Fig. 8b was projected onto  $I_2$  and shown in Fig. 8c. This set was used to initialize the proposed algorithm. Note that, while Figs. 8a and 8b are very different visually, both result in the same initial pixel set shown in Fig. 8c. Hence, the same

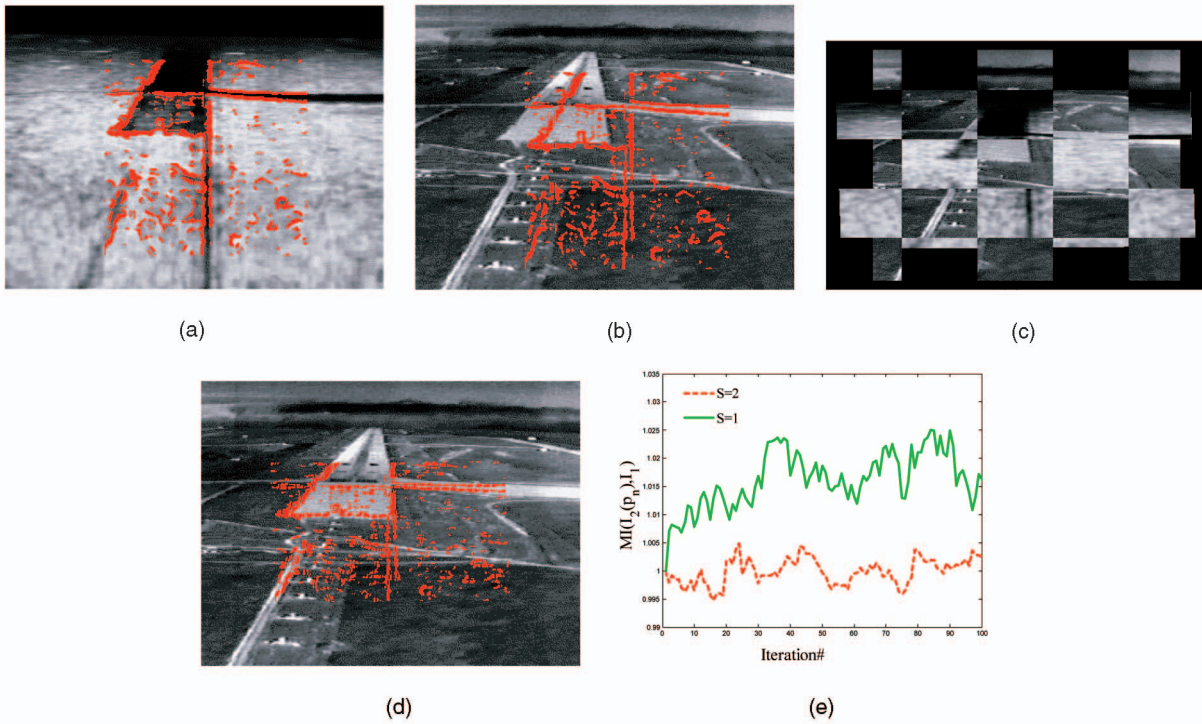


Fig. 5. Registration results of SAR and FLIR images using the affine motion model: (a) Original SAR image overlaid with the set of high-gradient pixels. (b) Initial estimate of the overlaid set in the FLIR image. (c) Composite image. (d) Final alignment of the registered pixel set. (e) The Mutual Information of the registered images computed over the Implicit Similarity registration process.  $S = 1$  corresponds to the finest resolution scale.

overlay (depicted in Fig. 8d) was computed. These were the most accurate results over the entire set. In particular, we note that, for the proposed scheme, there is no difference between the registration of Fig. 8c to either Figs. 8a or 8b. The quality of the first image is less significant since it is only used to detect the initial locations of the gradient maxima. As the images are not related by global intensity mapping, the MI scheme could not be effectively applied and it diverged. This is exemplified by Fig. 8e, where we computed the MI measure between the images in each iteration over the registration process and we see that the MI is not increased as the images are registered in the different resolution scales.

To conclude, the proposed scheme was able to register several classes of multimodality and multisensor images and is especially suitable for the alignment of image sets where one of the images is of considerably better quality than the other is.

## 5 CONCLUDING REMARKS

In this paper, we presented an approach to multimodality image registration. The algorithm uses an implicit similarity measure that is invariant to intensity dissimilarities. The measure is optimized by a fast multiscale gradient scheme. We introduced a weighting measure that enables the automatic detection and adaptive weighting of false pixel matchings. The convergence properties of the scheme were analyzed and an approximation of the optimal size of the initial pixel set size was derived. The experimental results demonstrate its applicability and it was shown to be especially suited for the registration of image sets, where one of the images is of better quality (noise, blur, etc.) than the other was. In the future, we intend to extend the proposed framework to nonrigid registration by maximizing the gradient energy with respect to an elastic, locally varying motion model similar to the work in [27]. Another issue is to derive a robust initial registration measure, similar to the role of Fourier techniques in single-modality image registration.

## APPENDIX

This section analyzes  $E\{\nabla_p F_n\}$  concerning the probabilities of the pixel sets  $S_G^n$  and  $S_G^n$ , defined in Section 3. Following (7),  $E\{\nabla_p F_n\}$ , the expectancy of  $\nabla_p F_n$  with respect to the motion parameters, is given by

$$E\{\nabla_p F_n\} = E\{X_i \nabla_v F_n(\mathbf{v}_i)\},$$

where  $X_i$  was defined in Section 2.2.  $\nabla_p F_n$  is evaluated at the pixels which are the projections of the initial pixel set  $S_1$ . As  $S_1$  is chosen such that it is uniformly distributed over  $I_1$  (see Section 2.1), its projection is also uniformly distributed in  $I_2$ . This statement is based on the assumption that the

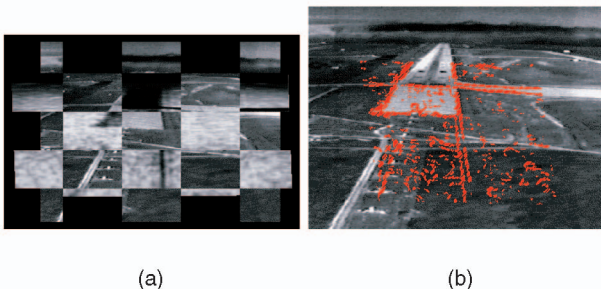


Fig. 6. Registration results of SAR and FLIR images using the projective motion model: (a) Composite image. (b) Final alignment of the pixel set after registration. The initial pixel set and its initial projection are the same as in Fig. 5.



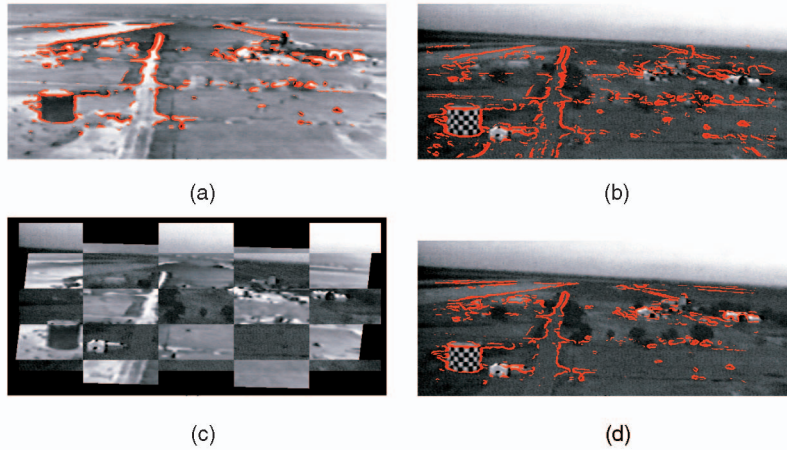


Fig. 7. Registration results of EO and IR images presented in [1] using the affine motion model: (a) Original IR image overlaid with the set of high-gradient pixels. (b) Initial estimate of the overlaid set in the EO image. (c) Composite image. (d) Final alignment of the pixel set after registration.

affinity and scale coefficients of the estimated motion are small. For large affinity and scalings, one can think of a transformation that maps the set  $S_1$  into a small region of  $I_2$ .

Yet, the assumption of small motion is valid as registration schemes based on nonlinear optimization have a limited basin of convergence. Thus,  $\nabla_{\mathbf{v}} F_n(\mathbf{v}_i)$  is statistically independent of  $X_i$  and

$$E\{\nabla_{\mathbf{p}} F_n\} = E\{X_i\} E\{\nabla_{\mathbf{v}} F_n(\mathbf{v}_i)\}.$$

$E\{\nabla_{\mathbf{v}} F_n(\mathbf{v}_i)\}$  is computed by analyzing  $E\{\frac{\partial F_n}{\partial x}\}$  and the results apply to  $E\{\frac{\partial F_n}{\partial y}\}$  due to symmetry. Using a centralized approximation of  $\frac{\partial F_n}{\partial x}$ , we get

$$2E\left\{\frac{\partial F_n}{\partial x}\right\} \simeq E\{F(x+1, y)\} - E\{F(x-1, y)\}$$

and denote  $\mathbf{v}_1 = (x+1, y)$  and  $\mathbf{v}_2 = (x-1, y)$

$$\begin{aligned} 2E\left\{\frac{\partial F_n}{\partial x}\right\} &= E\left\{\frac{\partial F_n}{\partial x} \mid \mathbf{v}_1, \mathbf{v}_2 \in S_G^m\right\} P(\mathbf{v}_1, \mathbf{v}_2 \in S_G^m) \\ &\quad + E\left\{\frac{\partial F_n}{\partial x} \mid \mathbf{v}_1, \mathbf{v}_2 \in \overline{S_G^m}\right\} P(\mathbf{v}_1, \mathbf{v}_2 \in \overline{S_G^m}) \\ &\quad + E\left\{\frac{\partial F_n}{\partial x} \mid \mathbf{v}_1 \in S_G^m, \mathbf{v}_2 \in \overline{S_G^m}\right\} P(\mathbf{v}_1 \in S_G^m, \mathbf{v}_2 \in \overline{S_G^m}) \\ &\quad + E\left\{\frac{\partial F_n}{\partial x} \mid \mathbf{v}_1 \in \overline{S_G^m}, \mathbf{v}_2 \in S_G^m\right\} P(\mathbf{v}_1 \in \overline{S_G^m}, \mathbf{v}_2 \in S_G^m). \end{aligned} \quad (16)$$

For  $\mathbf{v}_1$  and  $\mathbf{v}_2$  of the same class (either  $S_G^m$  or  $\overline{S_G^m}$ ),

$$E\left\{\frac{\partial F_n}{\partial x} \mid \mathbf{v}_1, \mathbf{v}_2 \in S_G^m\right\} = E\{S_G^m\} - E\{S_G^m\} = 0$$

and

$$E\left\{\frac{\partial F_n}{\partial x} \mid \mathbf{v}_1, \mathbf{v}_2 \in \overline{S_G^m}\right\} = E\{\overline{S_G^m}\} - E\{\overline{S_G^m}\} = 0.$$

Due to symmetry, it suffices to consider one of the last two terms

$$E\left\{\frac{\partial F_n}{\partial x} \mid \mathbf{v}_1 \in S_G^m, \mathbf{v}_2 \in \overline{S_G^m}\right\} = E\{S_G^m\} - E\{\overline{S_G^m}\} > 0. \quad (17)$$

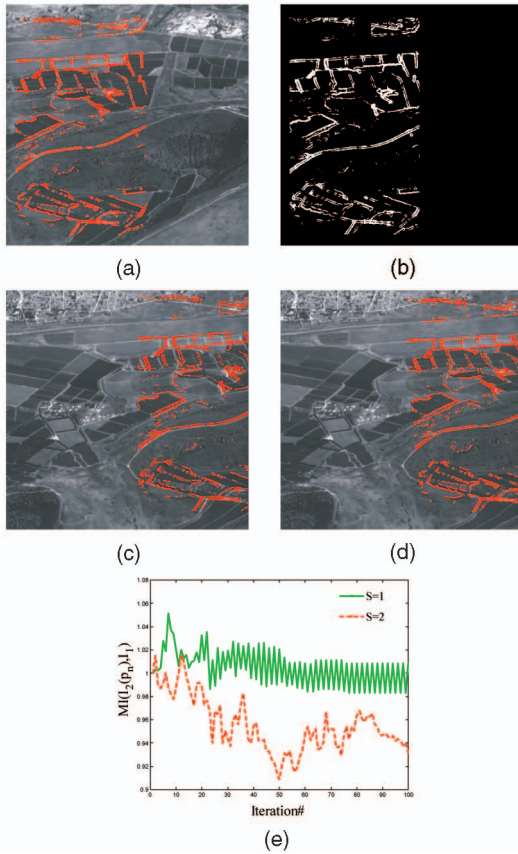


Fig. 8. Registration of real aerial images. (a) The pixel set with high gradient magnitude initially detected (this image acts as  $I_1$ ). (b) A binary edge map of (a), the initial pixel set is the same as in (a). (c) The initial pixel set projected onto the second image and used as an initial estimate for the registration. Note the misalignments in the lower and upper parts of the image. (d) The pixel set after alignment. For the proposed algorithm there is no difference in the registration of (c) to either (a) or (b). (e) The Mutual Information of the registered images computed over the Implicit Similarity registration process.  $S = 1$  corresponds to the finest resolution scale.

Hence,

$$\begin{aligned} E\left\{\frac{\partial F_n}{\partial x}\right\} &\approx (E\{S_G^m\} - E\{\overline{S_G^m}\})P(\mathbf{v}_1 \in \overline{S_G^m}, \mathbf{v}_2 \in S_G^m)v \\ &= (E\{S_G^m\} - E\{\overline{S_G^m}\})P(\mathbf{v}_2 \in S_G^m)P(\mathbf{v}_1 \in \overline{S_G^m}) \\ &= (E\{S_G^m\} - E\{\overline{S_G^m}\})P(\mathbf{v}_2 \in S_G^m)(1 - P(\mathbf{v}_1 \in S_G^m)) \end{aligned}$$

and

$$\begin{aligned} E\{\nabla_p F_n\} &\sim \\ (E\{S_G^m\} - E\{\overline{S_G^m}\})P(\mathbf{v}_2 \in S_G^m)(1 - P(\mathbf{v}_1 \in S_G^m)). \end{aligned} \quad (18)$$

## ACKNOWLEDGMENTS

The authors would like to thank Professor Michal Irani for allowing the use of the multisensor imagery in this paper. The authors also thank *TPAMI*'s associate editor and the anonymous reviewers for their constructive feedback on an earlier version of this paper.

## REFERENCES

- [1] M. Irani and P. Anandan, "Robust Multisensor Image Alignment," *Proc. IEEE Int'l Conf. Computer Vision*, pp. 959-966, Jan. 1998.
- [2] H. Li, B.S. Manjunath, and S.K. Mitra, "A Contour-Based Approach to Multisensor Image Registration," *IEEE Trans. Image Processing*, vol. 4, no. 3, pp. 320-334, Mar. 1995.
- [3] W.M. Wells, P. Viola, H. Atsumi, and S. Nakajima, "Multimodal Volume Registration by Maximization of Mutual Information," *Medical Image Analysis*, vol. 1, no. 1, pp. 35-51, 1996.
- [4] J. Pluim, J. Maintz, and M. Viergever, "Image Registration by Maximization of Combined Mutual Information and Gradient Information," *IEEE Trans. Medical Imaging*, vol. 19, no. 8, pp. 809-814, Aug. 2000.
- [5] J.B.A. Maintz and M.A. Viergever, "A Survey of Medical Image Registration," *Medical Image Analysis*, vol. 2, no. 1, pp. 1-36, 1998.
- [6] R. Sharma and M. Pavel, "Registration of Video Sequences from Multiple Sensors," *Proc. Image Registration Workshop*, pp. 361-366, 1997.
- [7] A. Guezic, X. Pennec, and N. Ayache, "Medical Image Registration Using Geometric Hashing," *IEEE Computational Science and Eng.*, vol. 4, no. 4, pp. 29-41, Oct.-Dec. 1997.
- [8] M. Fischler and R. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography," *Comm. ACM*, vol. 24, no. 6, pp. 381-395, 1981.
- [9] A. Roche, G. Malandain, N. Ayache, and S. Prima, "Towards a Better Comprehension of Similarity Measures used in Medical Image Registration," *Proc. Second Int'l Conf. Medical Image Computing and Computer-Assisted Intervention (MICCAI '99)*, pp. 555-566, Sept. 1999.
- [10] A. Roche, G. Malandain, X. Pennec, and N. Ayache, "The Correlation Ratio as a New Similarity Measure for Multimodal Image Registration," *Proc. First Int'l Conf. Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 1115-1124, Oct. 1998.
- [11] A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens, and G. Marchal, *Information Processing in Medical Imaging*. pp. 263-274, Dordrecht: Kluwer Academic, 1995.
- [12] Y. He, A. Ben-Hamza, and H. Krim, "A Generalized Divergence Measure for Robust Image Registration," *IEEE Trans. Signal Processing*, vol. 51, no. 5, pp. 1211-1220, May 2003.
- [13] L.G. Brown, "A Survey of Image Registration Techniques," *ACM Computing Surveys*, vol. 24, no. 4, pp. 325-376, Dec. 1992.
- [14] J. Pluim, J. Maintz, and M. Viergever, "Mutual Information Based Registration of Medical Images: A Survey," *IEEE Trans. Medical Imaging*, vol. 22, no. 8, pp. 986-1004, Aug. 2003.
- [15] P. Thevenaz and M. Unser, "Spline Pyramids for Inter-Modal Image Registration Using Mutual Information," *Proc. SPIE Conf. Mathematical Imaging: Wavelet Applications in Signal and Image Processing V*, vol. 3169, pp. 236-247, July-Aug. 1997.
- [16] J. Pluim, J. Maintz, and M. Viergever, "Interpolation Artifacts in Mutual Information-Based Image Registration," *Computer Vision and Image Understanding*, vol. 77, no. 2, pp. 211-232, Feb. 2000.
- [17] Y. Keller and A. Averbuch, "Implicit Similarity: A New Approach to Multisensor Image Registration," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 2, pp. 543-548, June 2003.
- [18] E.P. Simoncelli and E.H. Adelson, "Noise Removal via Bayesian Wavelet Coring," *Proc. IEEE Int'l Conf. Image Processing*, pp. 379-382, Sept. 1996.
- [19] J. Pluim, J. Maintz, and M. Viergever, "Image Registration by Maximization of Combined Mutual Information and Gradient Information," *IEEE Trans. Medical Imaging*, vol. 19, no. 8, pp. 809-814, 2000.
- [20] J. Zhang and A. Rangarajan, "A Unified Feature-Based Registration Method for Multimodality Images," *Proc. IEEE Int'l Symp. Biomedical Imaging (ISBI)*, pp. 724-727, 2004.
- [21] N. Tishby and N. Slonim, "Data Clustering by Markovian Relaxation and the Information Bottleneck Method," *Neural Information Processing Systems*, pp. 640-646, 2000.
- [22] C.G. Harris and M. Stephens, "A Combined Corner and Edge Detector," *Proc. Fourth Alvey Vision Conf.*, pp. 147-151, 1988.
- [23] R. Kimmel and A.M. Bruckstein, "Regularized Laplacian Zero Crossings as Optimal Edge Integrators," *Int'l J. Computer Vision*, vol. 53, no. 3, pp. 225-243, 2003.
- [24] P. Gill, *Practical Optimization*. Academic Press, 1982.
- [25] M. Irani and P. Anandan, "Vision Algorithms: Theory and Practice," *Proc. Int'l Workshop Vision Algorithms*, Sept. 1999.
- [26] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality Image Registration by Maximization of Mutual Information," *IEEE Trans. Medical Imaging*, vol. 16, no. 2, pp. 187-198, 1997.
- [27] J. Kybic and M. Unser, "Fast Parametric Elastic Image Registration," *IEEE Trans. Image Processing*, vol. 12, no. 11, pp. 1427-1442, Nov. 2003.



analysis, signal processing, and motion estimation.



department of Computer Science. In 1987, he joined the Department of Computer Science, School of Mathematical Sciences, Tel Aviv University, where he is now a professor of computer science. His research interests include wavelets, signal/image processing, multiresolution analysis, numerical computation for the solutions of PDEs, scientific computing (fast algorithms), and parallel and supercomputing (software and algorithms).

► For more information on this or any other computing topic, please visit our Digital Library at [www.computer.org/publications/dlib](http://www.computer.org/publications/dlib).

**Yosi Keller** received the BSc degree in electrical engineering in 1994 from The Technion-Israel Institute of Technology, Haifa. He received the MSc and PhD degrees in electrical engineering from Tel Aviv University, Tel Aviv, in 1998 and 2003, respectively. From 1994 to 1998, he was an R&D officer in the Israeli Intelligence Force. He is a visiting assistant professor with the Department of Mathematics, Yale University. His research interests include statistical pattern

**Amir Averbuch** received the BSc and MSc degrees in mathematics from the Hebrew University in Jerusalem, Israel, in 1971 and 1975, respectively. He received the PhD degree in computer science from Columbia University, New York, in 1983. From 1966-1970 and 1973-1976, he served in the Israeli Defense Forces. From 1976-1986, he was a research staff member at the IBM T.J. Watson Research Center, Yorktown Heights, New York, Department of Computer Science. In 1987, he joined the Department of Computer Science, School of Mathematical Sciences, Tel Aviv University, where he is now a professor of computer science. His research interests include wavelets, signal/image processing, multiresolution analysis, numerical computation for the solutions of PDEs, scientific computing (fast algorithms), and parallel and supercomputing (software and algorithms).