

Multisensor Image Registration by Feature Consensus

Chandra Shekhar Venu Govindu Rama Chellappa

Computer Vision Laboratory
Center for Automation Research
University of Maryland
College Park, MD 20742-3275
Email: shekhar@cfar.umd.edu

Summary

In this paper, we address the problem of registering two images obtained using different sensors, fields of view and/or lighting conditions, where conventional approaches relying on feature correspondence or area correlation are likely to fail. The approach presented in this paper eliminates the need for feature matching, and is robust to variations in sensor characteristics, imaging conditions, and fields of view. It is similar in spirit to methods based on the Generalized Hough Transform (GHT), but it eliminates many of the problems (such as lack of robustness and high computational cost) associated with GHT-style methods.

We make two basic assumptions: (1) the characteristics of the scene give rise to detectable features such as points and lines in both images, and at least a part of these features are common to both images and (2) the two images can be at least approximately aligned by global 2-D transformation. For a given problem, we select an appropriate transformation (Euclidean, similarity or affine) based on sensor geometry and other criteria. We first decompose this transformation into a sequence of elementary stages. At each stage, we select an appropriate image feature class, and estimate the value of one transformation parameter by a *feature consensus* mechanism in which each feature pair is allowed to select the value of the parameter that is consistent with it. The value of the parameter that is maximally consistent with respect to all the feature pairs is considered to be its best estimate. We introduce the concept of parameter observability to formalize this process. A very useful notion, parameter separability, makes it possible in most cases to completely eliminate the need for feature pairings, and instead work with aggregate properties of features determined from each image separately.

The global registration achieved by feature consensus should be sufficient for many applications such as those employing registration for performing focus of attention. If a more accurate global registration is needed, as in medical applications, the feature consensus result may be used as an initial condition for more elaborate schemes that use feature correspondence (e.g. [10]) or multi-dimensional search (e.g. [14]), which require a good initial guesses for the transformation parameters. Methods like deformable template matching could also be invoked for local refinement of the registration.

Multisensor Image Registration by Feature Consensus

Abstract

This paper presents an approach for registering images obtained using different sensors, viewpoints or lighting conditions. This approach does not require feature correspondence or area correlation. The geometric transformation between the images is reparametrized into a sequence of elementary stages. At each stage, a single transformation parameter is estimated using a *feature consensus* mechanism wherein the value of the parameter that is maximally consistent with all possible feature pairings is determined. The concepts of parameter observability and separability are introduced to guide the choice of features, attributes and transformation parameters. Experimental results on real data are provided.

KEYWORDS:

**Registration, Multisensor, Transformations, Parameters,
Features, Attributes**

1 Introduction

There are many image understanding applications where it is beneficial to integrate data from different types of sensors. In general, different sensors respond to scene characteristics in different, and often complementary ways. Integration or fusion of this multisensor information is possible only if the image data are registered, or “positioned” with respect to a common coordinate system [1, 9]. Most traditional methods for image registration based on area correlation or feature matching can handle only minor geometric and photometric variations. In a multisensor context, however, the images to be registered may be of widely different types, obtained by disparate sensors

with different resolutions, noise levels and imaging geometries. The common or “mutual” information, which is the basis of automatic image registration, may manifest itself in a very different way in each image [14]. This is because different sensors record different physical phenomena in the scene. For instance, an infra-red sensor responds to the temperature distribution of the scene, whereas a radar responds to material properties such as dielectric constant, electrical conductivity and surface roughness.

Since the underlying scene giving rise to the shared information is the same in all images of the scene, certain qualitative statements can be made about the manner in which information is preserved across multisensor data. Although the pixels corresponding to the same scene region may have different values depending on the sensor, *pixel similarity* and *pixel dissimilarity* are usually preserved. A region that appears homogeneous to one sensor is likely to appear homogeneous to another, local textural variations apart. Regions that can be clearly distinguished from one another in one image are likely to be distinguishable from one another in other images, irrespective of the sensor used. Although this is not true in all cases, it is generally valid for most types of sensors and scenes. Man-made objects such as buildings and roads in aerial imagery and implants, prostheses, metallic probes etc. in medical imagery also give rise to features that are likely to be preserved in multisensor images. Feature-based methods that exploit the information contained in region boundaries and in man-made structures are therefore useful for multisensor registration [4, 5, 10, 11].

Feature-based methods traditionally rely on establishing feature correspondence between the two images. Such correspondence-based methods

first employ feature matching techniques to determine corresponding feature pairs from the two images, and then compute the geometric transformation relating them, typically using a least-squares approach. Their primary advantage is that the transformation parameters can be computed in a single step, and are accurate if the feature matching is reliable. Their drawback is that they require feature matching, which is difficult to accomplish in a multisensor context, and is computationally expensive due to the well-known *correspondence problem*: Given N features in each image, the number of possible one-to-one feature mappings is $N!$, out of which only one is correct. Some heuristics can be employed to reduce the number of potential correspondences, but this problem still remains intractable, unless the two images are already approximately registered, or the number of features is small.

Some correspondence-less registration methods based on moments of image features have been proposed (e.g [17]), but these techniques, although mathematically elegant, work only if the two images contain exactly the same set of features. This requirement is rarely met in real images. Another class of methods based on the generalized Hough transform (GHT) [3, 8] have been proposed [6, 12]. These methods map the feature space into the parameter space, by allowing each feature pair to vote for a subspace of the parameter space. Clusters of votes in the parameter space are then used to estimate parameter values. These methods, although far more robust and practical than moment-based methods, have some limitations. Methods based on the GHT tend to produce a large number of false positives [15]. They also tend to be computationally expensive, since the dimensionality of the problem is equal to the number of transformation parameters.

In this paper, we propose a practical approach to multisensor registration that eliminates the need for feature matching, and is capable of dealing with large photometric and geometric variations, occlusions and differences in fields of view. It is similar in spirit to the GHT-style methods, but employs a different search strategy to eliminate the problems associated with them. We first decompose the original transformation¹ into a sequence of elementary stages. At each stage, we estimate the value of one transformation parameter by a *feature consensus* mechanism in which each feature pair is allowed to select the value of the parameter that is consistent with it. The value of the parameter that is maximally consistent with respect to all the feature pairs is considered to be its best estimate. We introduce the concept of parameter observability to formalize this process. A very useful notion, parameter separability, makes it possible in most cases to completely eliminate the need for feature pairings, and instead work with aggregate properties of features determined from each image separately.

The global registration achieved by feature consensus should be sufficient for many applications such as those employing registration for performing focus of attention. If a more accurate global registration is needed, as in medical applications, the feature consensus result may be used as an initial condition for more elaborate schemes that use feature correspondence (e.g. [10]) or multi-dimensional search (e.g. [14]), which require a good initial guesses for the transformation parameters. Methods like deformable template

¹We assume that the images do not contain large 3-D effects, and can therefore be registered by a global 2-D transformation. If 3-D effects are present, the registration obtained using the proposed method can be used as an initial condition for more elaborate 3-D positioning schemes.

matching could also be invoked for local refinement of the registration.

The rest of this paper is organized as follows. Section 2 defines the problem and introduces the notation used in the paper. The feature consensus scheme for multisensor registration is described in Section 3, and its application to different transformation models in Section 4. Practical issues are discussed in Section 5. Experimental results are presented in Section 6.

2 Statement of the problem

Registration of two images is defined to be the determination of the transformation that maps one image to the other. We assume that the scene being imaged is approximately planar, and that there are two sensors, S and \tilde{S} , which transform the visual information in their fields of view into 2-D images I and \tilde{I} . (We shall use the notation x or \tilde{x} to denote that the quantity x belongs to I or \tilde{I} , respectively.)

Since we are interested in exploiting the information present in geometric image features, we shall assume that this visual information consists of features on a 3-D plane, and hence that I and \tilde{I} are sets of features in the image plane. These features can be of various types, such as points, lines, edges, curves, regions, etc. Features of the k th type constitute a *feature class* $\mathbf{f}^{(k)}$. The images are assumed to be composed of subsets $\mathcal{F}^{(k)}$ of geometric features

$$\mathcal{F}^{(k)} = \cup_i \{f_i^{(k)}\}$$

where $f_i^{(k)}$ denotes the i th feature of the k th class. Every feature has a geometric *attributes* associated with it. Typical attributes are position, slope,

curvature, length, area, etc. Attributes of the l th type associated with the feature class $\mathbf{f}^{(k)}$ constitute an *attribute class*, denoted by $\alpha^{(k,l)}$. We define the set of attributes associated with an image

$$\mathcal{A} = \cup_k \cup_l \Gamma^{(k,l)}$$

where

$$\Gamma^{(k,l)} = \cup_i \{ \alpha_i^{(k,l)} \}$$

is the set of attributes of the l th type associated with the set of features of class $\mathbf{f}^{(k)}$. The superscripts k and l will henceforth be dropped for convenience.

The geometric transformation mapping I to \tilde{I} can be written as

$$\hat{\tilde{I}} = TI$$

where $\hat{\tilde{I}}$ is the image obtained by mapping the points in I to the coordinate system of \tilde{I} and T is a 2-D to 2-D transformation of the form

$$T : \mathcal{R}^2 \longrightarrow \mathcal{R}^2$$

A number of choices are available for the 2-D transformation between two images, as discussed in [16, 18]. In practice, the form of T is chosen based on experience and knowledge of sensor geometries. A general transformation of a 2-D point set I into the coordinate system of another, \tilde{I} , can be written as

$$\hat{\tilde{I}} = T_{(a_1, a_2, \dots, a_n)} I,$$

where a_1, a_2, \dots, a_n are the n parameters of the transformation. The image registration problem is the determination of the parameters a_1, a_2, \dots, a_n

given the input images I and \tilde{I} , or alternatively, as the transformation of the image I to the image \hat{I} .

In order to study the relationship between feature attributes and transformation parameters, consider the feature pair (f, \tilde{f}) of class \mathbf{f} , where f is from I and \tilde{f} from \tilde{I} . Let α and $\tilde{\alpha}$ be corresponding attributes, of class α . Let us denote the relationship between them as

$$\tilde{\alpha} = g_{(a_1, a_2, \dots, a_n)}(\alpha)$$

where $g_{(a_1, a_2, \dots, a_n)}(\cdot)$ is some (known) function which depends on the form and the parameters of the transformation T . If there exists such a function $g_\theta(\cdot)$ parameterized by a single transformation parameter θ , we consider θ to be *observable* with respect to the feature class \mathbf{f} and the attribute class α . Mathematically, this can be written as

$$\exists g_\theta(\cdot) \ni \tilde{\alpha} = g_\theta(\alpha) \tag{1}$$

The class \mathbf{f} is called the *observing feature* class, and α is called the *observing attribute* class. Equation (1) is called the observability equation for θ .

3 Feature consensus

The feature consensus approach simplifies the parameter estimation problem by decomposing the original problem into a set of simpler problems each involving a single unknown parameter to be determined. The basic voting unit is a feature pair (f, \tilde{f}) , where f is from I and \tilde{f} from \tilde{I} , and each voting unit casts a single vote for a single unknown parameter.

If θ is the parameter being estimated, the features f that vote should possess attributes α that are related by

$$\tilde{\alpha} = g_{\theta}(\alpha)$$

where $g_{\theta}(\cdot)$ is a bijective function that depends only on the parameter being determined. In other words, θ should be observable with respect to some feature and attribute classes. We assume that the function is of a form that permits the parameter to be written as

$$\theta = h_{\theta}(\alpha, \tilde{\alpha})$$

The *consensus function* C_{θ} is defined as

$$C_{\theta}(\theta) = \sum_{i,j} \delta(\theta - h_{\theta}(\alpha_i, \tilde{\alpha}_j)), \alpha_i \in \Gamma, \tilde{\alpha}_j \in \tilde{\Gamma}$$

where $\delta(\cdot)$ is the discrete impulse function. The consensus function C_{θ} is defined over the range of the parameter θ . Feature consensus is simply the process of determining the value of θ that maximizes the consensus function:

$$\theta_{max} = \arg \max_{\theta} C_{\theta}(\theta) \quad (2)$$

For convenience, we will henceforth drop the subscript in C_{θ} . In order to estimate the transformation parameters a_i by feature consensus, we need to reparametrize the transformation into a set of stages in such a way that at each stage there is a single unknown parameter, and that this parameter is observable. In mathematical terms, we decompose the original transformation T into a sequence of transformations

$$T_{(a_1, a_2, \dots, a_n)} I = T_{(b_n)} T_{(b_{n-1})} \dots T_{(b_i)} \dots T_{(b)} I \quad (3)$$

where b_1, b_2, \dots, b_n are functions of the original transformation parameters a_1, a_2, \dots, a_n . In the simplest case, the b parameters are identical to the a parameters. At each stage i , there should be some feature attribute which is transformed in a manner that is dependent only on b_i :

$$\exists g_{b_i}(\cdot) \ni \tilde{\alpha} = g_{b_i}(\alpha) \quad (4)$$

The parameters b_1, b_2, \dots, b_n are determined sequentially by feature consensus, and at each stage i the transformation $T_i(b_i)$ is applied to the first image I , leaving us then with the task of estimating the remaining parameters b_{i+1}, \dots, b_n between the transformed first image and the second. This process is repeated until all the b parameters have been determined. It is straightforward to estimate the original a parameters of the transformation T . In most cases, this may not be necessary, since the first image can be aligned with the second simply by applying the last stage of the reparametrized transformation.

It may be possible to estimate the new transformation parameters b_i in parallel, rather than in sequence as the decomposition in (3) implies. However, the sequential approach is more intuitive, and it enables us to progressively reduce the number of candidate feature pairings, as explained in Sec. 5.2.

3.1 Separability

If the observability equation for a parameter is of the form

$$\tilde{\alpha} = \alpha + \theta$$

where α and $\tilde{\alpha}$ are feature attributes from class α for a potential match, then θ , the parameter under scrutiny, is said to be *separable* with respect to α . In such a case, using $\theta = \tilde{\alpha} - \alpha$, the consensus function is

$$C(\theta) = \sum_{i,j} \delta(\theta - (\tilde{\alpha}_j - \alpha_i)), \alpha_i \in \Gamma, \tilde{\alpha}_j \in \tilde{\Gamma}$$

and the value of θ that receives the maximum number of votes is given by (2).

Theorem 1 *The consensus function $C(\theta)$ is equal to the cross-correlation of the distributions of the attributes of the class α obtained separately from the two images.*

Proof:

Let $C(t) = \sum_i \delta(t - \alpha_i)$ and $\tilde{C}(t) = \sum_j \delta(t - \tilde{\alpha}_j)$ be the attribute distributions from the two images. Their cross-correlation $D = C \otimes \tilde{C}$ is given by

$$\begin{aligned} D(\theta) &= \sum_k C(k) \tilde{C}(k + \theta) \\ &= \sum_k \sum_i \delta(k - \alpha_i) \sum_j \delta(k + \theta - \tilde{\alpha}_j) \\ &= \sum_i \sum_j \sum_k \delta(k - \alpha_i) \delta(k - (\tilde{\alpha}_j - \theta)) \\ &= \sum_i \sum_j \delta(\alpha_i - (\tilde{\alpha}_j - \theta)) \\ &= \sum_i \sum_j \delta(\theta - (\tilde{\alpha}_j - \alpha_i)) \\ &= C(\theta) \end{aligned}$$

using the fact that $\sum_k \delta(k - a) \delta(k - b) = \delta(a - b)$. The value of this result is that by suitably choosing the bin size for the distributions of α and $\tilde{\alpha}$, the computational complexity of determining θ_{max} can be drastically reduced.

Further, it is simple to extend the separability principle to accommodate observability equations of the form

$$f(\tilde{\alpha}) = g(\alpha)h(\theta)$$

where $h(\cdot)$ is an invertible function.

Although the approach proposed in this paper is oriented towards attributes of discrete features, the concept of separability enables us to extend it to differential feature attributes (point slope, curvature, etc). This is useful for registering images of scenes that contain few discrete features (natural scenery, for instance). Due to lack of space details of this generalization of the feature consensus method are not presented here. The interested reader is referred to [7].

4 Examples

In this section, we illustrate the application of the feature consensus method for some common transformations. In a real application, the choice of transformation would be determined by factors such as accuracy required, prior knowledge of the scene, sensor geometries, etc. Given a transformation T , we need to find a decomposition of the form (3), and select features and attributes satisfying the observability constraint (4). We assume that lines and points have been extracted from the two images.

4.1 Similarity transformation

This transformation is characterized by four parameters (rotation β , translation t_x , t_y and scale s). Under this transformation, the point $\mathbf{p} = (x, y)$

maps to the point $\tilde{\mathbf{p}} = (\tilde{x}, \tilde{y})$ according to

$$\tilde{\mathbf{p}} = sR_\beta\mathbf{p} + \mathbf{t}. \quad (5)$$

Where

$$R_\beta = \begin{pmatrix} \cos \beta & -\sin \beta \\ \sin \beta & \cos \beta \end{pmatrix}$$

and

$$\mathbf{t} = \begin{pmatrix} t_x \\ t_y \end{pmatrix}$$

This transformation can be expressed as a sequence of four stages:

$$T\mathbf{p} = T_{t_y}T_{t_x}T_sT_\beta\mathbf{p}$$

where

$$T_\beta\mathbf{p} = R_\beta\mathbf{p} \quad (6)$$

$$T_s\mathbf{p} = s\mathbf{p} \quad (7)$$

$$T_{t_x}\mathbf{p} = \mathbf{p} + \begin{pmatrix} t_x \\ 0 \end{pmatrix} \quad (8)$$

$$T_{t_y}\mathbf{p} = \mathbf{p} + \begin{pmatrix} 0 \\ t_y \end{pmatrix} \quad (9)$$

In this case, the parameters of the new transformation (β, s, t_x, t_y) are identical to the parameters of the original transformation. The first parameter to be determined is β , the angle of rotation. This parameter is observable from the slopes of line features in the images. If l and \tilde{l} are corresponding line features with slope angles ϕ and $\tilde{\phi}$ respectively, we have

$$\tilde{\phi} = \phi + \beta$$

Thus β can be determined by consensus of line features, and I can be transformed according to (6). The scale s can be determined by consensus of pairs of point features, with the distance d between the two points being the observing attribute:

$$\tilde{d} = sd$$

The translational shifts $\mathbf{t} = (t_x, t_y)$ are observed using point feature location as the observing attribute class:

$$\tilde{\mathbf{p}} = \mathbf{p} + \mathbf{t}$$

All four parameters are separable if the features and attributes are chosen as indicated.

4.2 Semi-affine transformation

Let us now consider a more complex model with five parameters (rotation β , translation t_x, t_y and scales s_x, s_y). The transformation can be written as

$$\tilde{\mathbf{p}} = \begin{pmatrix} s_x & 0 \\ 0 & s_y \end{pmatrix} R_\beta \mathbf{p} + \mathbf{t} \quad (10)$$

In this case, the rotation β and scales s_x, s_y cannot be observed independently of each other from the slopes and lengths of the segments, as in the previous case. The solution to this problem is to reparametrize the transformation with respect to a new set of parameters that are observable with respect to other feature/attribute classes. One such set of parameters is $(\beta, t_x, t_y, \Delta, \rho)$, where Δ is the scale factor, given by

$$\Delta = \sqrt{s_x s_y}$$

and ρ is the square root of the scale ratio, given by

$$\rho = \sqrt{\frac{s_y}{s_x}}$$

The decomposition can be written as:

$$T\mathbf{p} = T_{t_y}T_{t_x}T_{\Delta}T_{\rho}T_{\beta}\mathbf{p}$$

where

$$T_{\rho}\mathbf{p} = \begin{pmatrix} 1/\rho & 0 \\ 0 & \rho \end{pmatrix} \mathbf{p}$$

$$T_{\Delta}\mathbf{p} = \Delta\mathbf{p}$$

and T_{β} , T_{t_x} and T_{t_y} are the same as before. The rotation β is observable (but not separable) using pairs of lines as the observing feature class, and the ratio of line slopes as the observing attribute class. Consider a unit vector \mathbf{v} at an angle ϕ , given by

$$\mathbf{v} = \begin{pmatrix} \cos \phi \\ \sin \phi \end{pmatrix}$$

Under the transformations T_{β} and T_{ρ} it becomes

$$\begin{aligned} \tilde{\mathbf{v}} &= T_{\rho}T_{\beta}\mathbf{v} \\ &= \begin{pmatrix} 1/\rho & 0 \\ 0 & \rho \end{pmatrix} R_{\beta} \begin{pmatrix} \cos \phi \\ \sin \phi \end{pmatrix} \\ &= \begin{pmatrix} (1/\rho) \cos(\phi + \beta) \\ \rho \sin(\phi + \beta) \end{pmatrix} \end{aligned}$$

Let $\tilde{m} = \tan \tilde{\phi}$ be the slope of $\tilde{\mathbf{v}}$. By dividing the y -component of $\tilde{\mathbf{v}}$ by its x -component, we get

$$\tan \tilde{\phi} = \rho^2 \tan(\phi + \beta) \tag{11}$$

In order to observe β , we need to eliminate the ρ term from the above equation. This can be achieved by taking the ratio of two line slopes. Let lines with slope angles ϕ and ψ be transformed into lines with slope angles $\tilde{\phi}$ and $\tilde{\psi}$ as a result of transformations T_β and T_ρ . Then,

$$\frac{\tan \tilde{\phi}}{\tan \tilde{\psi}} = \frac{\tan(\phi + \beta)}{\tan(\psi + \beta)} \quad (12)$$

Given any pair of line pairs, one from each image, the only unknown in (12) is the rotation angle β . After some simple manipulations, we obtain the following expression for β

$$\beta = (1/2) \sin^{-1}(k \sin(\phi - \psi)) - \phi/2 - \psi/2$$

where

$$k = \frac{\tan \tilde{\phi} + \tan \tilde{\psi}}{\tan \tilde{\phi} - \tan \tilde{\psi}}.$$

After observing β , image I can be rotated accordingly, and then (11) reduces to

$$\tan \tilde{\phi} = \rho^2 \tan \phi \quad (13)$$

Using (13), we can thus observe ρ , which turns out to be a separable parameter.

The remaining transformation between the two images consists of a scaling and a translation, which can be determined as in the case of the similarity transformation.

4.3 Affine transformation

We are now ready to look at the affine transformation model, given by

$$\tilde{\mathbf{p}} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mathbf{p} + \mathbf{t}$$

Using the QR transformation from linear algebra, this can be written in terms of six parameters (rotation β , translation t_x, t_y , scale ratio ρ , scale Δ and skew α).

$$\tilde{\mathbf{p}} = \Delta \begin{pmatrix} 1/\rho & 0 \\ \alpha & \rho \end{pmatrix} R_\beta \mathbf{p} + \mathbf{t} \quad (14)$$

Proceeding as in the previous case, it can be shown that the rotation angle is observable using triplets of lines as features. If a feature pair consists of lines at slope angles (ϕ, ψ, λ) and $(\tilde{\phi}, \tilde{\psi}, \tilde{\lambda})$, we can show that the rotation angle that is consistent with the feature pair is given by

$$\beta = \tan^{-1} \left(\frac{-\cos \lambda + k \cos \psi}{\sin \lambda - k \sin \psi} \right)$$

where

$$k = \frac{(\tan \tilde{\phi} - \tan \tilde{\psi}) \sin(\phi - \lambda)}{(\tan \tilde{\phi} - \tan \tilde{\lambda}) \sin(\phi - \psi)}$$

Once rotation is compensated for, the scale ratio is separably observable using line pairs as features, according to

$$\tan \tilde{\phi} - \tan \tilde{\psi} = \rho^2 (\tan \phi - \tan \psi)$$

After compensating for the scale ratio, the skew is separably observable from line slopes, according to

$$\tan \tilde{\phi} = \tan \phi + \alpha$$

The scale and translation are determined as in the previous cases.

5 Practical issues

In this section, we discuss issues relating to the practicality of the proposed approach.

5.1 Computational complexity and SNR

In the non-separable case with N features in each image, the complexity is $O(N^2)$. In the separable case, if b bins are used for each attribute distribution, the complexity is $O(b^2)$ or $O(N)$, whichever is higher. (Usually, $b \ll N$). In the general case, for each parameter there are a total of N^2 entries in the consensus function, of which only N can possibly be correct. Although this is an improvement over the $N!$ complexity associated with correspondence-based methods, the “signal” component of the consensus function is small compared to the “noise” component. However, the signal does not get lost in the noise, because the $N^2 - N$ incorrect votes will typically be dispersed over a wide range of values, whereas the N correct votes will be clustered around the true parameter value at the mode of the distribution of votes. A more formal discussion of voting schemes can be found in [15]. As a simple example, if we are estimating rotation by comparing the slope angles of line features, the incorrect votes will be distributed evenly in the range $-\pi/2 : \pi/2$, whereas the correct matches will vote for the true rotation angle β .

5.2 Progressive feature filtering

Even though feature consensus has a lower complexity than correspondence-based methods, it is desirable to reduce the number of incorrect votes, since it would improve both the efficiency as well as the robustness of the approach. One way is to employ a simple matching scheme to restrict each feature to have a maximum of $m \ll N$ possible matches, thereby reducing the total number of votes to Nm . Initially, this is possible only if the feature class

under consideration has an attribute that is invariant with respect to the transformation. However, when one or more transformation parameters have been determined, they can be used to progressively filter out unlikely feature pairings. For instance, in a similarity transformation, once rotation has been estimated and compensated for, only segments at similar orientations are considered as potential matches for estimating scale, and only the corner points having approximately the same bisector orientations are considered for translation estimation. This “feature filtering” technique greatly improves the performance of the system. For a detailed discussion of the robustness of feature-matching techniques, see [2].

5.3 Iterative overlap analysis

In cases where the images to be registered have only a small overlap, the initial parameter estimates may be corrupted by the features from the non-overlapping parts of the images. In such a situation, the initial estimates are used to determine the initial overlap, and features in this overlapping region are then used to refine the parameter estimates. This process is iterated (typically, once or twice) until stable estimates are obtained.

5.4 Parameter visibility

Parameter observability is the necessary mathematical condition for determining the transformation using feature consensus. However, it is not a sufficient condition, because it is possible for a parameter to be observable according to the transformation model, but not be *visible* from the consensus function *for a given data set*. In the separable case, this happens when

the individual attribute distributions (and hence the consensus function) are essentially flat. For instance, if we have two images consisting of circles, the rotation angle will not be visible from slope distributions. In our experience, such cases do not occur very often. A more common problem is the presence of multiple peaks in the consensus function, due to the peculiarities of the data or ambiguities inherent in the features. For instance, when line slope angles are used as attributes for estimating rotation, we get false peaks 180 degrees away from the true rotation angle, since it is impossible to distinguish a line with a slope angle of β from lines with slope angles $\beta \pm 180$. Further, if the images contain rectangular objects (such as buildings), there may be additional spurious peaks 90 degrees away from the true peak.

Ideally, the consensus function should have a single sharp peak. As mentioned earlier, there are mainly two ways in which the function may deviate from this ideal: there may be too many peaks of nearly the same strength (lack of peak distinctness), and/or the function may be too flat (lack of parameter visibility). We therefore use two criteria to quantify the deviation of the consensus function from the ideal. The first criterion, termed the peak distinctness Q_{pd} , is the ratio of the strength h_n of the n th most significant peak to the strength h_1 of the most significant peak :

$$Q_{pd} = h_n/h_1$$

Typically, we use $n = 2$. If the n th peak does not exist, $Q_{pd} = 0$, which is the ideal case. The second criterion, termed parameter visibility Q_{pv} , is related to the entropy of the consensus function viewed as a probability function. First, the consensus function $C(\theta)$ is normalized to obtain a corresponding

probability function $P(\theta)$:

$$P(\theta) = \frac{C(\theta)}{\sum_{\theta} C(\theta)}$$

The entropy is then determined using the standard formula:

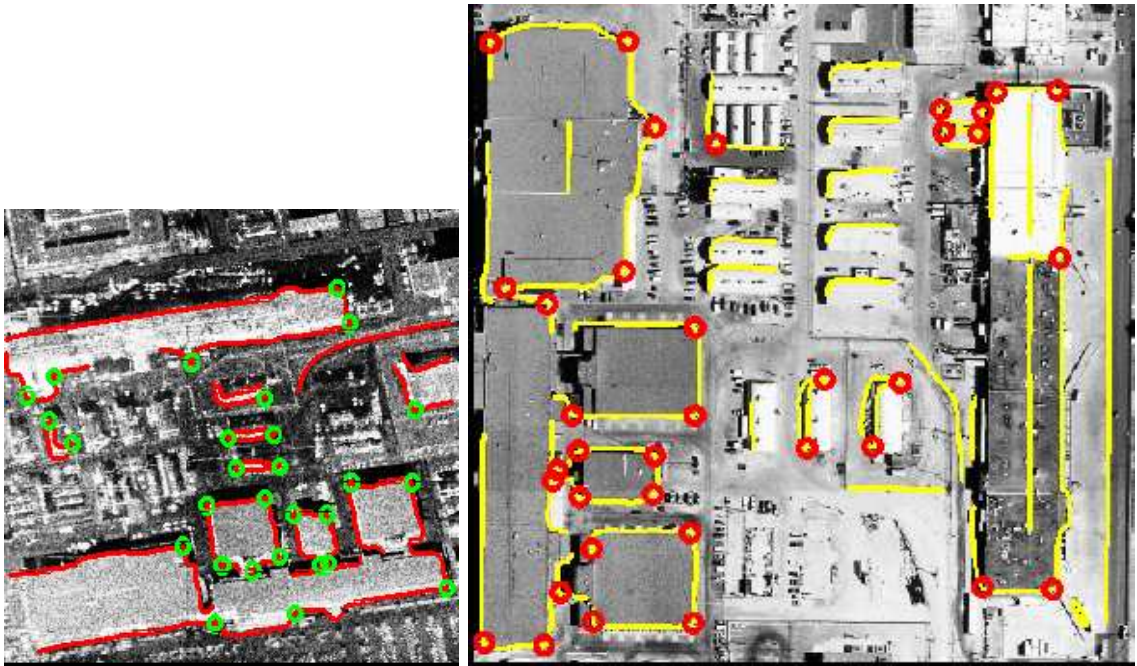
$$E = \sum_{\theta} -P(\theta) \ln P(\theta)$$

The parameter visibility is then defined as

$$Q_{pv} = \exp \left(-\sqrt{1 - \frac{E}{\ln N}} \right),$$

where N is the total number of bins used for computing the consensus function.

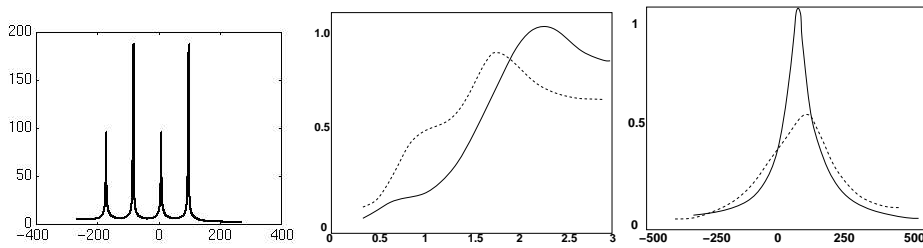
The parameter visibility measure favors peaked distribution over flat ones, but does not distinguish between unimodal and multimodal distributions. The peak distinctness measure favors distributions that have a single dominant peak. Together, they give an idea of how accurate the registration will be. If a large false peak leads to an incorrect initial choice for the parameter, it will manifest itself later in the poor shapes of other consensus functions, and therefore higher values of Q_{pd} and Q_{pv} . A conservative policy would be to use a tree descent strategy to scan all possible paths to determine the one that has the lowest values of Q_{pd} and Q_{pv} for all parameters, and is therefore likely to be the most accurate. Our approach is to scan one parameter ahead—i.e. to try all the promising peaks for the i_{th} parameter, and choose the one that gives the best Q_{pd} and Q_{pv} for the $(i + 1)_{st}$ parameter. An example of this is given in Appendix A.



(a) radar

(b) visual

Figure 1: Multisensor (radar and visual) images of the Kirtland AFB, with superimposed features. Locations of feature points are indicated by circles.



(a) Rotation

(b) Scaling

(c) Translation

Figure 2: Consensus functions for the multisensor images in Fig. 1. The x - and y - components are drawn with solid and dashed curves.

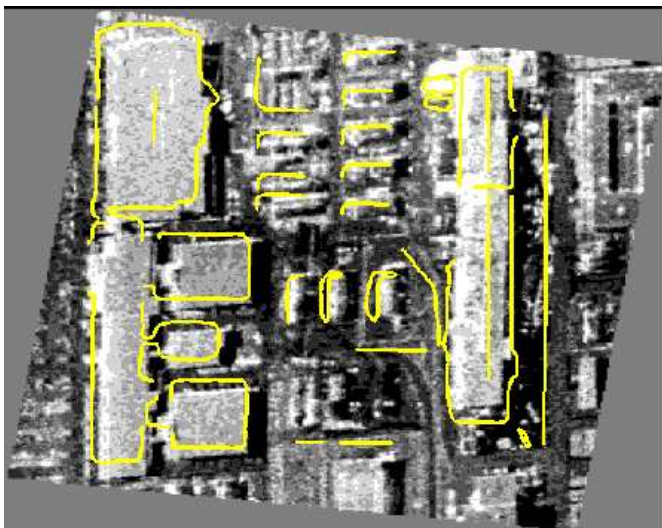


Figure 3: Final registration result for the images in Fig. 1, shown by overlaying the visual contours on the transformed radar image.

6 Implementation and experimental results

We are in the process of implementing and testing the method for the affine model in 4.3. Currently, we have results using a simplified version the method in Sec. 4.2 for semi-affine transformations. Rather than computing the rotation angle using ratios of line slopes, as in (11), it is directly observed from line slope angles, as in the case of the similarity transformation. This simplification gives acceptable results if the scale ratio is not too large or too small.

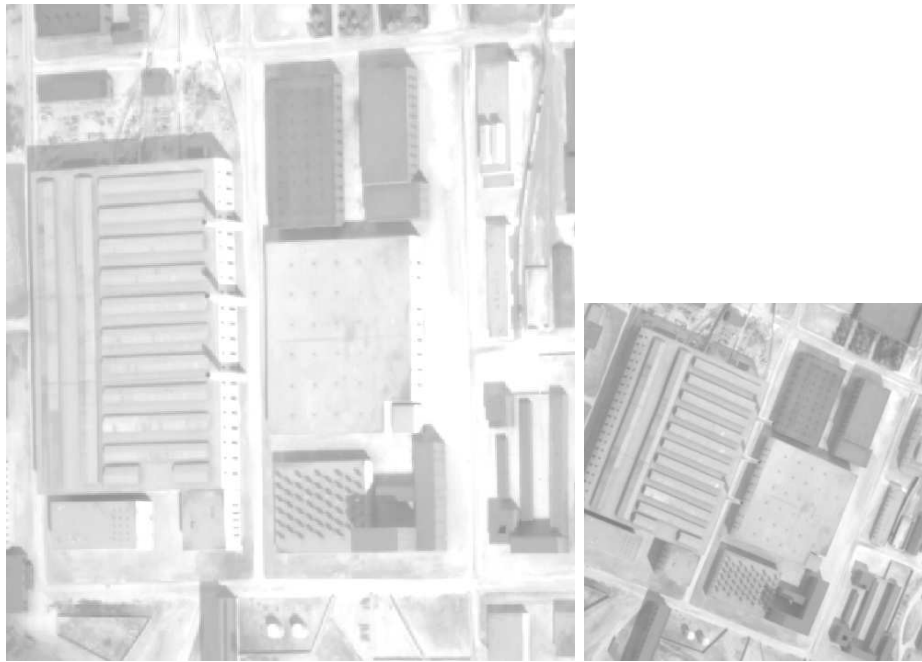
Contours are extracted from the images using the Canny operator, with the thresholds set using the method proposed in [13]. Lines are obtained by a polygonal approximation of contours. Feature points are extracted from contours using curvature as the criterion. Rotation is determined by

consensus of polygonal segments, with slope angle as attribute. Each pair (s, \tilde{s}) of segments, s from I and \tilde{s} from \tilde{I} , produces one vote for the angle of rotation. The vote is weighted by the product of the lengths of the segments. For scale estimation, the observing feature class consists of pairs of feature points from the same image. The horizontal and vertical distances between them, denoted by Δx and Δy , are the observing attributes. The observation equations for the scale factors are simply

$$\begin{aligned}\tilde{\Delta x} &= s_x \Delta x \\ \tilde{\Delta y} &= s_y \Delta y\end{aligned}$$

Once the rotation and scale have been determined and compensated for, the translation is directly observed from the positions of feature points in the two images.

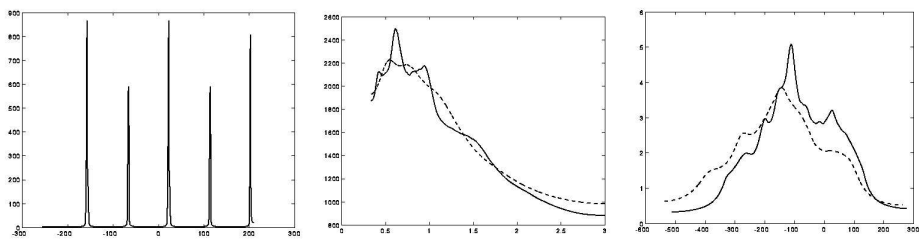
We illustrate the performance of the method on two real data sets, one containing a radar-visual image pair of the Kirtland AFB area (Fig. 1), and the other containing an visual image pair the Model Board data set with viewpoint and photometric differences (Fig. 4). The consensus functions for the radar-visual data are shown in Fig. 2. There is a sharp peak in the consensus function at the true rotation angle of $+100^\circ$. The s_x and s_y scale factors are approximately 2.1 and 1.6. The peaks in the consensus functions for t_x and t_y are unambiguous. The final result after applying all the stages of transformation is shown in Fig. 3. Registration results for the visual image pair are shown in Fig. 6.



(a) I

(b) \tilde{I}

Figure 4: Two aerial images (from the Model Board set) to be registered.

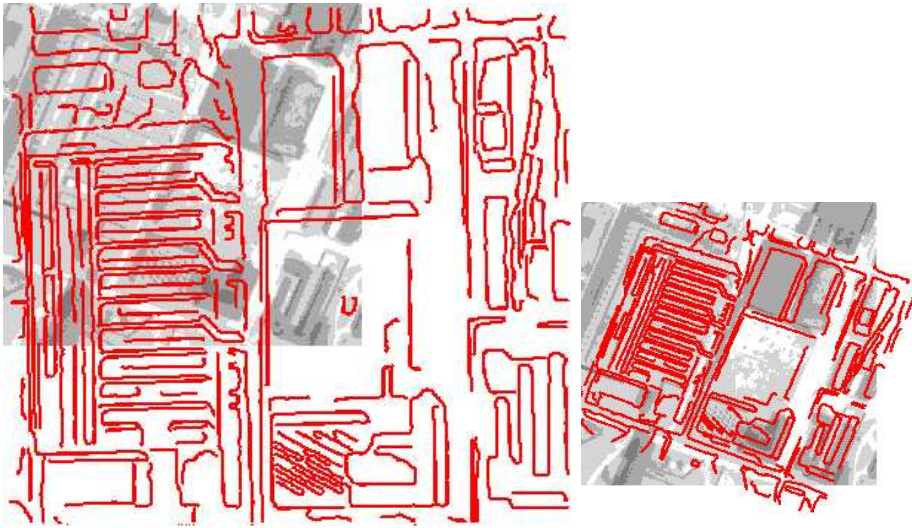


(a) Rotation

(b) Scaling

(c) Translation

Figure 5: Consensus functions for the Model Board set.



(a) Initial alignment

(b) Final alignment

Figure 6: Initial and final alignment for the Model Board images showing \tilde{I} superimposed with original and transformed contours from I .

7 Conclusion

We have proposed an approach to multisensor registration that does not rely on feature correspondence as its primary mechanism. We argue that feature correspondence cannot be reliably determined in images obtained from disparate sensors, and hence we have proposed a method that is based on feature consensus. By considering all pairs of features as potential matches, and by allowing them to select the transformation parameters, we eliminate the need for correspondence. By decomposing the transformation into a sequence of elementary stages and using the progressive feature filtering technique, we avoid the complexity associated with GHT-style methods. We have introduced the notion of parameter observability to analyze the relationship between features, attributes and transformation parameters. We have presented results on real data to validate our approach. Further work is needed on developing better feature detectors for multisensor imagery, and in developing a comprehensive taxonomy of features and attributes for various transformation models, especially for the projective transformation.

A Resolving ambiguities: An example

We illustrate the ambiguity resolution method proposed in Section 5.4 on the Kirtland EO-SAR pair, shown in Fig. 1. The consensus functions shown in Fig. 2 were obtained by selecting the true peaks for the rotation and scaling. Here we demonstrate what happens if a false peak is chosen, and how we can avoid the cascading errors that would normally result from this initial error. In Fig.7, we show the consensus functions for the scale factors that

result from selecting each of the four peaks in the consensus function for rotation. The values of the distinctness and visibility criteria, introduced in Section 5.4 are shown below the corresponding consensus functions. (There are two values in each case corresponding to the x - and the y -components.) In each case, the consensus functions and the criteria for the translation are shown corresponding to the choice of the most prominent peak for the scaling.

By visual examination of the consensus functions in Fig. 7, it is apparent that an incorrect choice the rotation angle leads to poorer quality of the consensus functions for the scaling and translation. This observation can be verified by comparing the values of the Q_{pd} and Q_{pv} criteria. Only the true peak in the rotation (at 100 degrees) leads to single-peak consensus functions for scaling and translation, thus scoring $(0, 0)$ on the Q_{pd} criterion. In the other three cases, corresponding to the false peaks at -170, -80 and +10 degrees, there are multiple peaks in the scaling and/or the translation. The consensus function for translation obtained by a correct choice of rotation (and therefore the correct choice of scale factors) has the best score on the parameter visibility (Q_{pv}) criterion. Only the y -component of the true scaling seems to score rather poorly on the Q_{pv} criteria.

B Refining the transformation

The alignment obtained by the feature consensus technique, although approximate in the general case of a general 3-D transformation, is sufficient for many applications. However, if further accuracy is required, we can perform a simple nearest-neighbor matching of lines or feature points, and recompute

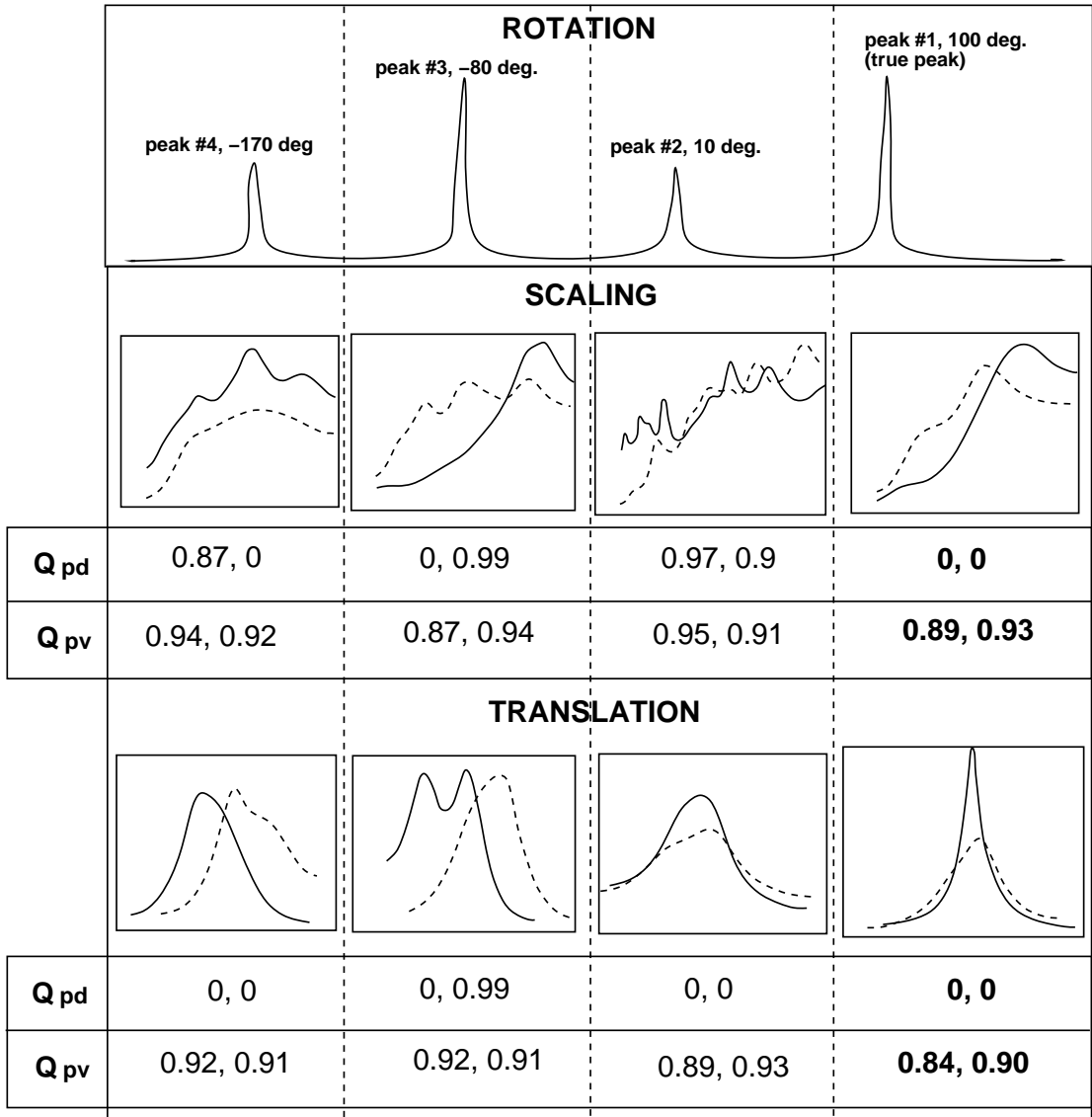


Figure 7: Effect of peak selection on the peak distinctness (Q_{pd}) and parameter visibility (Q_{pv}) criteria. Incorrect peak selection leads in most cases to consensus functions that have a worse (i.e. higher) score on one or both criteria.

the transformation parameters. Since this is a correspondence-based technique, we can use a more complex transformation model than was used for the feature consensus approach. In our system, we use a projective transformation model for this stage, in which a point (x, y) is transformed to (X, Y) according to

$$\begin{pmatrix} X \\ Y \\ 1 \end{pmatrix} = cA \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

where c is an arbitrary scalar, and

$$A = \begin{pmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{13} & a_{23} & 1 \end{pmatrix}$$

For each point correspondence, we can write two equations:

$$X = a_{11}x + a_{21}y + a_{31} - a_{13}xX - a_{23}yX$$

$$Y = a_{12}x + a_{22}y + a_{32} - a_{13}xY - a_{23}yY$$

Thus, we get two equations for each point correspondence $\langle (x, y), (X, Y) \rangle$. A minimum of four such correspondences are therefore needed to solve for the eight transformation parameters:

$$\begin{bmatrix} x_1 & y_1 & 1 & 0 & 0 & 0 & -x_1X_1 & -y_1X_1 \\ 0 & 0 & 0 & x_1 & y_1 & 1 & -x_1Y_1 & -y_1Y_1 \\ \vdots & & & & & & & \vdots \end{bmatrix} \mathbf{a} = P \quad (15)$$

where

$$\mathbf{a} = \left(a_{11} \ a_{21} \ a_{31} \ a_{12} \ a_{22} \ a_{32} \ a_{13} \ a_{23} \right)^T$$

and

$$P = (X_1 \ Y_1 \ \cdots \ X_4 \ Y_4)^T$$

Similarly, given a pair of matching lines $l = (a \ b \ c)^T$ and $L = (A \ B \ C)^T$, with

$$\begin{aligned} (a \ b \ c)(x \ y \ 1)^T &= 0 \\ (A \ B \ C)(X \ Y \ 1)^T &= 0 \end{aligned}$$

we can write the following two equations:

$$\begin{bmatrix} 0 & Ac & -Ab & 0 & Bc & -Bb & 0 & Cc \\ -Ac & 0 & Aa & -Bc & 0 & Ba & -Cc & 0 \end{bmatrix} \mathbf{a} = \begin{pmatrix} Cb \\ -Ca \end{pmatrix}$$

Four line correspondences are therefore required to solve for the parameters of the projective transformation.

In our implementation, we use a nearest-neighbor approach to obtain potential point correspondences, and then estimate the transformation parameters using (15). Almost invariably, there will be some false matches that will lead to errors in the estimates. A number of methods are available to detect and prune out such “outliers” [2]. We use the iterative refinement approach developed in [18]. In this approach, transformation parameters are first estimated using the available candidate point correspondences. The computed transformation parameters are used to project points in the first image onto the second. A match (p_1, p_2) is considered to be correct if the projection of p_1 does not lie too far from p_2 . Matches that fail to satisfy this constraint are eliminated, and the transformation parameters are recomputed. This estimate-and-prune step is repeated until all matches satisfy the constraint.

Acknowledgments

The authors wish to thank Srinu Raghavan, Radha Poovendran, Naresh Gupta and Shridhar Shrinivasan for many helpful discussions on multisensor registration, and Philippe Burlina for suggestions for improving to the mathematical notation. Prof. Azriel Rosenfeld provided many helpful suggestions on improving the style of the paper.

References

- [1] R. Chellappa *et al.*, “On the positioning of multiple sensors for image exploitation and target recognition,” *Proceedings of the IEEE*, Vol. 5, pp. 120–138, Jan. 1997.
- [2] C.V.Stewart, “A new robust operator for computer vision: Theoretical analysis,” in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, (Seattle, Washington), pp. 1–8, 1994.
- [3] R. O. Duda and P. E. Hart, “Use of the Hough transformation to detect lines and curves in picture,” *Communications of the ACM*, Vol. 15, pp. 11–15, Jan. 1972.
- [4] L. Fonseca and B. S. Manjunath, “Registration techniques for multisensor remotely-sensed imagery,” *Photogrammetric Engineering and Remote Sensing*, Sept. 1996. In press.
- [5] A. Goshtasby, “Image registration by local approximation,” *Image and Vision Computing*, Vol. 6, pp. 255–261, Nov. 1988.
- [6] A. Goshtasby and G. Stockman, “Point pattern matching using convex hull edges,” *IEEE Transactions on Systems, Man and Cybernetics*, Vol. SMC-15, pp. 631–637, September/October 1985.
- [7] V. Govindu, C. Shekhar, and R. Chellappa, “Correspondence-less alignment using a geometric framework,” in *NASA Image Registration Workshop*, (Greenbelt, MD), Nov. 1997. Accepted for publication(also available from the URL [http://www.cfar.umd.edu/ venu/IRW97.ps](http://www.cfar.umd.edu/venu/IRW97.ps)).

- [8] P. V. Hough, "Method and Means for Recognizing Complex Patterns." U.S. Patent 3069654, Dec. 1962.
- [9] L.G.Brown, "A survey of image registration techniques," *ACM Computing Surveys*, Vol. 24, pp. 325–376, Dec. 1992.
- [10] H. Li, B. Manjunath, and S. Mitra, "A contour-based approach to multisensor image registration," *IEEE Transactions on Image Processing*, Vol. 4, pp. 320–334, Mar. 1995.
- [11] J. M. Rignot *et al.*, "Automated multisensor registration: requirements and techniques," *Photogrammetric Engineering and Remote Sensing*, Vol. 57, pp. 1029–1038, Aug. 1991.
- [12] G. C. Stockman, S. Kopstein, and S. Bennet, "Matching images to models for registration and object detection via clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-4, pp. 229–41, May 1982.
- [13] S. Venkatesh and P. L. Rosin, "Dynamic threshold determination by local and global edge evaluation," *Graphical Models and Image Processing*, Vol. 75, No. 2, pp. 146–160, 1995.
- [14] P. Viola and W.M.Wells, "Alignment by maximization of mutual information," in *Proc. International Conference on Computer Vision*, (Cambridge, MA), pp. 16–23, 1995.
- [15] W.E.L.Grimson and D.P.Huttenlocher, "On the sensitivity of the Hough transform for object recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 12, pp. 255–274, Mar. 1990.

- [16] G. Wolberg, *Digital Image Warping*, Los Alamitos, CA: IEEE Computer Society Press, 1990.
- [17] R. Y. Wong, "Scene Matching with Invariant Moments," *Computer Graphics and Image Processing*, Vol. 8, pp. 16–24, 1978.
- [18] Q. Zheng and R. Chellappa, "A computational vision approach to image registration," *IEEE Transactions on Image Processing*, Vol. 2, pp. 311–326, July 1993.