# A Survey of Image Registration Techniques

LISA GOTTESFELD BROWN

*Department of Computer Science, Columbia University, New York, NY 10027*

Registration is a fundamental task in image processing used to match two or more pictures taken, for example, at different times, from different sensors, or from different viewpoints. Virtually all large systems which evaluate images require the registration of images, or a closely related operation, as an intermediate step. Specific examples of systems where image registration is a significant component include matching a target with a real-time image of a scene for target recognition, monitoring global land usage using satellite images, matching stereo images to recover shape for autonomous navigation, and aligning images from different medical modalities for diagnosis.

Over the years, a broad range of techniques has been developed for various types of data and problems. These techniques have been independently studied for several different applications, resulting in a large body of research. This paper organizes this material by establishing the relationship between the variations in the images and the type of registration techniques which can most appropriately be applied. Three major types of variations are distinguished. The first type are the variations due to the differences in acquisition which cause the images to be misaligned. To register images, a spatial transformation is found which will remove these variations. The class of transformations which must be searched to find the optimal transformation is determined by knowledge about the variations of this type. The transformation class in turn influences the general technique that should be taken. The second type of variations are those which are also due to differences in acquisition, but cannot be modeled easily such as lighting and atmospheric conditions. This type usually effects intensity values, but they may also be spatial, such as perspective distortions. The third type of variations are differences in the images that are of interest such as object movements, growths, or other scene changes. Variations of the second and third type are not directly removed by registration, but they make registration more difficult since an exact match is no longer possible. In particular, it is critical that variations of the third type are not removed. Knowledge about the characteristics of each type of variation effect the choice of feature space, similarity measure, search space, and search strategy which will make up the final technique. All registration techniques can be viewed as different combinations of these choices. This framework is useful for understanding the merits and relationships between the wide variety of existing techniques and for assisting in the selection of the most suitable technique for a specific problem.

Categories and Subject Descriptors: A.1 [**General Literature**]: Introductory and Survey; I.2.10 [**Artificial Intelligence**]: Vision and Scene Understanding; I.4 [**Computing Methodologies**]: Image Processing; I.5 [**Computing Methodologies**]: Pattern Recognition

General Terms: Algorithms, Design, Measurement, Performance

Additional Key Words and Phrases: Image registration, image warping, rectification, template matching

CONTENTS



## 1. INTRODUCTION

A frequent problem arises when images taken, at different times, by different sensors or from different viewpoints need to be compared. The images need to be aligned with one another so that differences can be detected. A similar problem occurs when searching for a prototype or template in another image. To find the optimal match for the template in the image, the proper alignment between the image and template must be found. All of these problems, and many related variations, are solved by methods that perform image registration. A transformation must be found so that the points in one image can be related to their corresponding points in the other. The determination of the optimal transformation for registration depends on the types of variations between the images. The objective of this paper is to provide a framework for solving image registration tasks and to survey the classical approaches.

Registration methods can be viewed as different combinations of choices for the following four components:

(1) a feature space,
(2) a search space,
(3) a search strategy, and
(4) a similarity metric.

The *feature space* extracts the information in the images that will be used for matching. The *search space* is the class of transformations that is capable of aligning the images. The *search strategy* decides how to choose the next transformation from this space, to be tested in the search for the optimal transformation. The *similarity metric* determines the relative merit for each test. Search continues according to the search strategy until a transformation is found whose similarity measure is satisfactory. As we shall see, the types of variations present in the images will determine the selection for each of these components.

For example, consider the problem of registering the two x-ray images of chest taken of the same patient at different times shown in Figure 1. Properly aligning the two images is useful for detecting, locating, and measuring pathological and other physical changes. A standard approach to registration for these images might be as follows: the images might first be reduced to binary images by detecting the edges or regions of highest contrast using a standard edge detection scheme. This removes extraneous information and reduces the amount of data to be evaluated. If it is thought that the primary difference in acquisition of the images was a small translation of the scanner, the search space might be a set of small translations. For each translation of the edges of the left image onto the edges of the right image, a measure of similarity would be computed. A typical similarity measure would be the correlation between the images. If the similarity measure is computed for all translations then the search strategy is simply exhaustive. The images are registered using the translation which optimizes the similarity criterion. However, the choice of using edges for features, translations for the search space, exhaustive search for the search strategy and correlation for the similarity metric will influence the outcome of this registration. In fact, in this case, the registration will undoubtably be unsatisfactory since the images are misaligned in a more complex

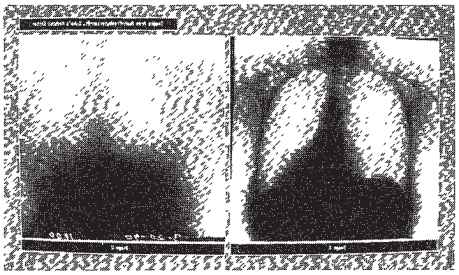


**Figure 1.** X-ray images of a patient's chest, taken at different times. (Thanks to A. Goshtasby.)

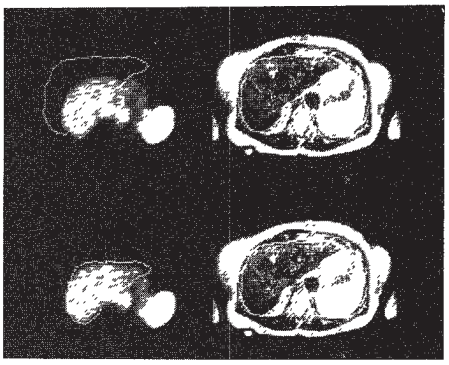


**Figure 2.** The top left image is a SPECT image of a patient's liver. The top right shows the same region viewed by MRI. A contour was manually drawn around the liver in the MRI image. The location of this contour in the SPECT image shows the mismatch between the two images. At the bottom right the MRI image has been registered to the SPECT image, and the location of the transformed contour is shown on the SPECT image, bottom left. A brief description of the registration method employed is in Section 3.3.3. (Courtesy of QSH, an image display and processing toolkit [Noz 1988] and New York University; I would like to thank B. A. Birnbaum, E. L. Kramer, M. E. Noz, and J. J. Sanger of New York University, and G. Q. Maguire, Jr. of Columbia University.)

way than translation. By establishing the relationship between the variations between the images and the choices for the four components of image registration, this paper provides a framework for understanding the exisiting registration techniques and also a methodology for assisting in the selection of the appropriate technique for a specific problem. By establishing the relationship between the variations among the images and the choices for the four components of image registration, this paper provides a framework for understanding the existing registration techniques and also a methodology for assisting in the selection of the appropriate technique for a specific problem.

The need to register images has arisen in many practical problems in diverse fields. Registration is often necessary for (1) integrating information taken from different sensors, (2) finding changes in images taken at different times or under different conditions, (3) inferring three-dimensional information from images in which either the camera or the objects in the scene have moved, and (4) for model-based object recognition [Rosenfeld and Kak 1982].

An example of the first case is shown in Figure 2. In this figure the upper right image is a Magnetic Resonance Image (MRI) of a patient's liver. From this image it is possible to discern the anatomical structures. Since this image is similar to what a surgeon will see during an operation, this image might be used to plan a medical procedure. The upper left image is from single photon emission computed tomography (SPECT). It shows the same anatomical region after intravenous administration of a Tc-99m (a radionuclide) labeled compound. This image depicts some of the functional behavior of the liver (the Tc-99m compound binds to red blood cells) and can more accurately distinguish between cancers and other benign lesions. Since the two images are taken at different resolutions, from different viewpoints, and at different times, it is not possible to simply overlay the two images. However, if the images can be registered, then the functional information of the SPECT image can be structurally localized using the MRI image. Indeed, the registration of images which show anatomical structures such as MRI, CT (computed tomography) and ultrasound, and images which show functional and metabolic activity such as SPECT, PET (positron emission

tomography), and MRS (magnetic resonance spectroscopy) has led to improved diagnosis, better surgical planning, more accurate radiation therapy, and countless other medical benefits [Maguire et al. 1990].

In this survey, the registration methods from three major research areas are studied:

(1) Computer Vision and Pattern Recognition—for numerous different tasks such as segmentation, object recognition, shape reconstruction, motion tracking, stereomapping, and character recognition.
(2) Medical Image Analysis—including diagnostic medical imaging, such as tumor detection and disease localization, and biomedical research including classification of microscopic images of blood cells, cervical smears, and chromosomes.
(3) Remotely Sensed Data Processing—for civilian and military applications in agriculture, geology, oceanography, oil and mineral exploration, pollution and urban studies, forestry, and target location and identification.

For more information specifically related to each of these fields, the reader may consult Katuri and Jain [1991] or Horn [1989] in computer vision, Stytz et al. [1991] and Petra et al. [1992] in medical imaging, and Jensen [1986] and Thomas et al. [1986] in remote sensing. Although these three areas have contributed a great deal to the development of registration techniques, there are still many other areas which have developed their own specialized matching techniques, for example, in speech understanding, robotics and automatic inspection, computer-aided design and manufacturing (CAD/CAM), and astronomy. The three areas studied in this paper include many instances from the four classes of problems mentioned above and a good range of distortion types including:

- sensor noise
- perspective changes from sensor viewpoint or platform perturbations
- object changes such as movements, deformations, or growths
- lighting and atmospheric changes including shadows and cloud coverage
- different sensors.

Tables 1 and 2 contain examples of specific problems in registration for each of the four classes of problems taken from computer vision and pattern recognition, medical image analysis, and remotely sensed data processing. The four classes are (1) multimodal registration, (2) template matching, (3) viewpoint registration, and (4) temporal registration. In classes (1), (3), and (4) the typical objective of registration is to align the images so that the respective changes in sensors, in viewpoint, and over time can be detected. In class (2), template matching, the usual objective is to find the optimal location and orientation, if one exists, of a template image in another image, often as part of a larger problem of object recognition. Each class of problems is described by its typical applications and the characteristics of methods commonly used for that class. Registration problems are by no means limited by this categorization scheme. Many problems are combinations of these four classes of problems; for example, frequently images are taken from different perspectives *and* under different conditions. Furthermore, the typical applications mentioned for each class of problems are often applications in other classes as well. Similarly, method characteristics are listed only to give an idea of some of the more common attributes used by researchers for solving these kinds of problems. In general, methods are developed to match images for a wide range of possible distortions, and it is not obvious exactly for which types of problems they are best suited. One of the objectives of these tables is to present to the reader the wide range of registration problems. Not surprisingly, this diversity in problems and their applications has been the cause for the de-

**Table 1.** Registration Problems — Part I

| MULTIMODAL REGISTRATION |
|---|
| *Class of Problems:* Registration of images of the same scene acquired from different sensors.<br>*Typical Application:* Integration of information for improved segmentation and pixel classification.<br>*Characteristics of Methods:* Often use sensor models; need to preregister intensities; image acquisition using subject frames and fiducial markers can simplify problem.<br>*Example 1*<br>*Field:* Medical Image Analysis<br>*Problem:* Integrate structural information from CT or MRI with functional information from radionucleic scanners such as PET or SPECT for anatomically locating metabolic function.<br>*Example 2*<br>*Field:* Remotely Sensed Data Processing<br>*Problem:* Integrating images from different electromagnetic bands, e.g., microwave, radar, infrared, visual, or multispectral for improved scene classification such as classifying buildings, roads, vehicles, and type of vegetation. |
| TEMPLATE REGISTRATION |
| *Class of Problems:* Find a match for a reference pattern in an image.<br>*Typical Application:* Recognizing or locating a pattern such as an atlas, map, or object model in an image.<br>*Characteristics of Methods:* Model-based approaches, preselected features, known properties of objects, higher-level matching.<br>*Example 1*<br>*Field:* Remotely Sensed Data Processing<br>*Problem:* Interpretation of well-defined scenes such as airports; locating positions and orientations of known features such as runways, terminals, and parking lots.<br>*Example 2*<br>*Field:* Pattern Recognition<br>*Problem:* Character recognition, signature verification, and waveform analysis. |

**Table 2.** Registration Problems — Part II

| VIEWPOINT REGISTRATION |
|---|
| *Class of Problems:* Registration of images taken from different viewpoints.<br>*Typical Application:* Depth or shape reconstruction.<br>*Characteristics of Methods:* Need local transformation to account for perspective distortions; often use assumptions about viewing geometry and surface properties to reduce search; typical approach is feature correspondence, but problem of occlusion must be addressed.<br>*Example 1*<br>*Field:* Computer Vision<br>*Problem:* Stereomapping to recover depth or shape from disparities.<br>*Example 2*<br>*Field:* Computer Vision<br>*Problem:* Tracking object motion; image sequence analysis may have several images which differ only slightly, so assumptions about smooth changes are justified. |
| TEMPORAL REGISTRATION |
| *Class of Problems:* Registration of images of same scene taken at different times or under different conditions.<br>*Typical Applications:* Detection and monitoring of changes or growths.<br>*Characteristics of Methods:* Need to address problem of dissimilar images, i.e., registration must tolerate distortions due to change, best if can model sensor noise and viewpoint changes; frequently use Fourier methods to minimize sensitivity to dissimilarity.<br>*Example 1*<br>*Field:* Medical Image Analysis<br>*Problem:* Digital Subtraction Angiography (DSA)—registration of images before and after radio isotope injections to characterize functionality, Digital Subtraction Mammiography to detect tumors, early cataract detection.<br>*Example 2*<br>*Field:* Remotely Sensed Data Processing<br>*Problem:* Natural resource monitoring, surveillance of nuclear plants, urban growth monitoring. |

velopment of enumerable independent registration methodologies.

This broad spectrum of methodologies makes it difficult to classify and compare techniques since each technique is often designed for specific applications and not necessarily for specific types of problems or data. However, most registration techniques involve searching over the space of transformations of a certain type to find the optimal transformation for a particular problem. In Figure 3, an example of several of the major transformation classes are shown. In the top left of Figure 3, an example is shown in which images are misaligned by a small shift due to a small change in the camera's position. Registration, in this case, involves a search for the direction and amount of translation needed to match the images. The transformation class is thus the class of small translations. The other transformations shown in Figure 3 are a rotational, rigid body, shear, and a more general global transformation due to terrain relief. In general, the type of transformation used to register images is one of the best ways to categorize the methodology and assist in selecting techniques for particular applications. The transformation type depends on the cause of the misalignment which may or may not account for all the variations between the images. This will be discussed in more detail in Section 2.3.

A few definitions and important distinctions about the nomenclature used throughout this survey may prevent some confusion; see Table 3.

The distinctions to be clarified are between global/local *transformations*, global/local *variations*, and global/local *computations*. In addition, we will define what we mean by transformation, variation, and computation in the context of registration.

A transformation is a mapping of locations of points in one image to new locations in another. Transformations used to align two images may be global or local. A global transformation is given by a single equation which maps the entire image. Examples (to be described in Section 2.2) are the affine, projective, perspective, and polynomial transformations. Local transformations map the image differently depending on the spatial location and are thus much more difficult to express succinctly. In this survey, since we classify registration *methods* according to their transformation type, a method is global or local according to the transformation type that it uses. This is not always the case in other papers on this subject.

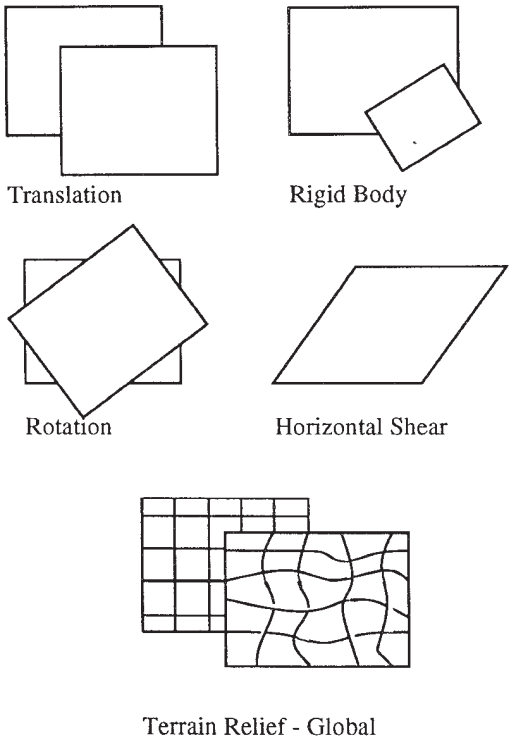



**Figure 3.** Examples of typical geometric transformations.

Variations refer to the differences in values and locations of pixels (picture elements) between the two images. We refer to differences in values as *valumetric* differences. Typically, value changes are differences in intensity or radiometry, but we use this more general term in order to include the wide variety of existing sensors whose values are not intensities, such as many medical sensors which

**Table 3.** Important Distinctions for Image Registration Methods

| | |
|---|---|
| **TRANSFORMATION:** a mapping of locations of points in one image to new locations of points in another. | |
| **GLOBAL:** | map is composed of a single equation that maps each point in the first image to new location in the second image. The equation is a function of the locations of the first image, but it is the same function for all parts of the image, i.e., the parameters of the function do not depend on the location. |
| **LOCAL:** | mapping of points in the image depends on their location—the map is composed of several smaller maps (several equations) for each piece of the image that is considered. |
| **VARIATIONS:** the differences in the values of pixels and their location between the images including distortions which have corrupted the true measurements. | |
| GLOBAL: | the images differ similarly throughout the entire image. For example, variations due to additive white noise affect the intensity values of all pixels in the same way. Each pixel will be affected differently, but the difference does not depend on the location of the pixel. |
| LOCAL: | the variation between images depends on the location in the image. For example, distortions due to perspective depend on the depth of the objects projected onto the image. Regions in the image which correspond to objects which are farther away are distorted in a different way than regions which correspond to closer objects. |
| **COMPUTATION:** refers to the set of calculations performed to determine the parameters of the registration transformation. | |
| GLOBAL: | uses all parts of the image to compute the parameters of the transformation. If a local *transformation* is being calculated, then each set of local parameters is computed using the entire image. This is generally a costly method but has the advantage of using more information. |
| LOCAL: | uses only the relevant local parts of the image for each set of local parameters in determining a local transformation. By using only local parts of the image for each calculation, the method is faster. It can also have the advantage of not being erroneously influenced by other parts of the image. |

measure everything from hydrogen density (magnetic resonance imaging) to temperature (thermography). Some of the variations between the images are *distortions*. Distortions refer to the noise that has corrupted or altered the true intensity values and their locations in the image. What is a distortion and what is not depend on what assumptions are made about the sensor and the conditions under which the images are taken. This will be discussed in more detail in Section 2.3. The variations in the image may be due to changes in the scene or the changes caused by a sensor and its position and viewpoint. We would like to remove some of these changes via registration; but others may be difficult to remove (such as the effects of illumination changes), or we are not interested in removing them, i.e., there may be changes that we would like to detect. When we describe a set of variations as global or local, we are referring to whether or not the variations can be removed by a global or a local transformation. However, since it is not always possible to remove all the distortions between the images, and because we do not want to remove some of the variations, it is critical for the understanding of registration methods to recognize the difference between whether certain variations are global or local and whether the selected transformation is global or local. For example, images may have local variations, but a registration method may use a global transformation to align them because some of the variations are differences to be detected after registration. The important distinctions between the various types of variations will be explained in more detail in Section 2.3.

The final definition and distinction we address are with respect to the registration computation. The registration computation refers to the calculations performed to determine the parameters of the transformation. When a computation is described as global or local this refers to whether the calculations needed to determine the parameters of the transformation require information from the entire image or whether each subset of

parameters can be computed from small local regions. This distinction only makes sense when a local transformation is used for registration, since when a global transformation is required only one set of parameters are computed. However, this is again distinct from the type of transformation used. For example, registration methods which search for the optimal local transformation may be more accurate and slower if they require global computations in order to determine local parameters since they use information from the entire image to find the best alignment.

One further comment is in order. In this paper the registration techniques reviewed were developed for images which are two dimensional. With the advent of cheaper memory, faster computers, and improved sensor capability, it has become more and more common to acquire three-dimensional images, for example, with laser range finders, motion sequences, and the latest 3D medical modalities. Registration problems abound in both 2D and 3D cases, but in this paper only 2D techniques are examined. Although many of the 2D techniques can be generalized to higher-dimensional data, there are several additional aspects that inevitably need to be considered when dealing with the immense amount of data and the associated computational cost in the 3D case. Furthermore, many of the problems arising from the projection of 3-space onto a 2D image are no longer relevant. Techniques developed to overcome the unique problems of 3D registration are not surveyed in this paper.

In the next section of this paper the basic theory of the registration problem is given. Image registration is defined mathematically as are the most commonly used transformations. Then image variations and distortions and their relationship to solving the registration problem are described. Finally the related problem of rectification, which refers to the correction of geometric distortions produced by the projection of a flat plane, is detailed.

In Section 3 of this paper the major approaches to registration are described based on the complexity of the type of transformation that is searched. In Section 3.1, the traditional technique of the cross-correlation function and its close relatives, statistical correlation, matched filters, the correlation coefficient, and sequential techniques are described. These methods are typically used for small well-defined affine transformations, most often for a single translation. Another class of techniques used for affine transformations, in cases where frequency-dependent noise is present, are the Fourier methods described in Section 3.2. If an affine transformation is not sufficient to match the images then a more general global transformation is required. The primary approach in this case requires feature point mapping to define a polynomial transformation. These techniques are described in 3.3. However, if the source of misregistration is not global, i.e., the images are misaligned in different ways over different parts of the image, then a local transformation is needed. In the last section of 3.3, the techniques which use the simplest local transformation based on piecewise interpolation are described. In the most complex cases, where the registration technique must determine a local transformation when legitimate local distortions are present, i.e., distortions that are *not* the cause of misregistration, techniques based on specific transformation models such as an elastic membrane are used. These are described in Section 3.4.

The methods described in Section 3 are used as examples for the last section of this survey. Section 4 offers a framework for the broad range of possible registration techniques. Given knowledge of the kinds of variations present, and those which need to be corrected, registration techniques can be designed, based on the transformation class which will be sufficient to align the images. The transformation class may be one of the classical ones described in Section 2.2 or a specific class defined by the parameters of the problem. Then a feature space and similarity measure are selected which are least sensitive to remaining variations

and are most likely to find the best match. Lastly, search techniques are chosen to reduce the cost of computations and guide the search to the best match given the nature of the remaining variations. In Section 4, several alternatives for each component of a registration method are discussed using the framework developed, in particular, with respect to the characteristics of the variations between the images as categorized in Section 2.3.

## 2. IMAGE REGISTRATION IN THEORY

### 2.1 Definition

Image registration can be defined as a mapping between two images both spatially and with respect to intensity. If we define these images as two 2D arrays of a given size denoted by $I_1$ and $I_2$ where $I_1(x, y)$ and $I_2(x, y)$ each map to their respective intensity (or other measurement) values, then the mapping between images can be expressed as:

$$I_2(x, y) = g(I_1(f(x, y)))$$

where $f$ is a 2D spatial-coordinate transformation, i.e., $f$ is a transformation which maps two spatial coordinates, $x$ and $y$, to new spatial coordinates $x'$ and $y'$,

$$(x', y') = f(x, y)$$

and $g$ is a 1D intensity or radiometric transformation.

The registration problem is to find the optimal spatial and intensity transformations so that the images are matched either for the purposes of determining the parameters of the matching transformation or to expose differences of interest between the images. The intensity transformation is not always necessary, and often a simple lookup table determined by sensor calibration techniques is sufficient [Bernstein 1976]. An example where an intensity transformation is used is in the case where there is a change in sensor type (such as optical to radar [Wong 1977]). Another example when an intensity transformation is needed is when objects in the scene are highly specular (their reflectance is mirror-like) and when there is a change in viewpoint or surface orientation relative to the light source. In the latter case, although an intensity transformation is needed, in practice it is impossible to determine the necessary transformation since it requires knowing the reflectance properties of the objects in the scene and their shape and distance from the sensor. Notice, that in these two examples, the intensity variations are due to changes in the acquisition of the images of the scene: in the first case by the change in sensors and in the second by the change in reflectance seen by the sensor. In many other instances of variations in intensity, the changes are due to differences in the scene that are not due to how the scene was projected by the sensor onto an image, but rather the changes are intrinsic differences in the scene, such as movements, growths, or differences in relative depths, that are to be exposed by the registration process—not removed. After all, if the images are matched exactly, then besides learning the parameters of the best transformation, what information is obtained by performing the registration?

Finding the parameters of the optimal spatial or geometric transformation is generally the key to any registration problem. It is frequently expressed parametrically as two single-valued functions, $f_x, f_y$:

$$I_2(x, y) = I_1(f_x(x, y), f_y(x, y))$$

which may be more easily implemented.

### 2.2 Transformations

The fundamental characteristic of any image registration technique is the type of spatial transformation or mapping used to properly overlay two images. Although many types of variations may be present in each image, the registration technique must select the class of transformation which will remove only the spatial distortions between images due to differences in acquisition and scene characteristics which affect acquisition. Other differences in scene characteristics that are to be *exposed* by registration should

not be used to select the class of transformation. In this section, we will define several types of transformations and their parameters, but we defer our discussion of how the transformation type is selected for a specific problem and what procedures are used to find its parameters until later.

The most common general transformations are rigid, affine, projective, perspective, and global polynomial. Rigid transformations account for object or sensor movement in which objects in the images retain their relative shape and size. A rigid-body transformation is composed of a combination of a rotation, a translation, and a scale change. An example is shown in Figure 3. Affine transformations are more general than rigid and can therefore tolerate more complicated distortions while still maintaining some nice mathematical properties. A shear transformation, also shown in Figure 3, is an example of one type of affine transformation. Projective transformations and the more general perspective transformations account for distortions due to the projection of objects at varying distances to the sensor onto the image plane. In order to use the perspective transformation for registration, knowledge of the distance of the objects of the scene relative to the sensor is needed. Polynomial transformations are one of the most general global transformations (of which affine is the simplest) and can account for many types of distortions so long as the distortions do not vary too much over the image. Distortions due to moderate terrain relief (see the bottom example in Figure 3) can often be corrected by a polynomial transformation. The transformations just described are all well-defined mappings of one image onto another. Given the intrinsic nature of imagery of nonrigid objects, it has been suggested (personal communication, Maguire, G. Q., Jr., 1989) that some problems, especially in medical diagnosis, might benefit from the use of fuzzy or probabilistic transformations.

In this section we will briefly define the different transformation classes and their properties. A transformation $T$ is *linear* if,

$$T(\mathrm{x}_1 + \mathrm{x}_2) = T(\mathrm{x}_1) + T(\mathrm{x}_2)$$

and for every constant $c$,

$$cT(\mathrm{x}) = T(c\mathrm{x}).$$

A transformation is *affine* if $T(\mathrm{x}) - T(0)$ is linear. Affine transformations are linear in the sense that they map straight lines into straight lines. The most commonly used registration transformation is the affine transformation which is sufficient to match two images of a scene taken from the same viewing angle but from a different position, i.e., the camera can be moved, and it can be rotated around its optical axis. This affine transformation is composed of the cartesian operations of a scaling, a translation, and a rotation. It is a global transformation which is *rigid* since the overall geometric relationships between points do not change, i.e., a triangle in one image maps into a similar triangle in the second image. It typically has four parameters, $t_x, t_y, s, \theta$, which map a point $(x_1, y_1)$ of the first image to a point $(x_2, y_2)$ of the second image as follows:

$$\begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = \begin{pmatrix} t_x \\ t_y \end{pmatrix} + s \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}.$$

This can be rewritten as

$$\bar{p}_2 = \bar{t} + sR\bar{p}_1$$

where $\bar{p}_1, \bar{p}_2$ are the coordinate vectors of the two images; $\bar{t}$ is the translation vector; $s$ is a scalar scale factor, and $R$ is the rotation matrix. Since the rotation matrix $R$ is orthogonal (the rows or columns are perpendicular to each other), the angles and lengths in the original image are preserved after the registration. Because of the scalar scale factor $s$, the rigid-body transformation allows changes in length relative to the original image, but it is the same in both $x$ and $y$. Without the addition of the translation vector, the transformation becomes linear.

The general 2D affine transformation

$$\begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = \begin{pmatrix} a_{13} \\ a_{23} \end{pmatrix} + \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}$$

does not have the properties associated with the orthogonal rotation matrix. Angles and lengths are no longer preserved, but parallel lines do remain parallel. The general affine transformation can account for more general spatial distortions such as shear (sometimes called skew) and changes in aspect ratio. Shear, which can act either along the $x$-axis, $Shear_x$, or along the $y$-axis, $Shear_y$, causes a distortion of pixels along one axis, proportional to their location in the other axis. The shear component of an affine transformation is represented by

$$Shear_x = \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}, \quad Shear_y = \begin{pmatrix} 1 & 0 \\ b & 1 \end{pmatrix}.$$

Another distortion which can occur with an affine transformation is a change in aspect ratio. The aspect ratio refers to the relative scale between the $x$ and $y$ axes. By scaling each axis independently,

$$Scale = \begin{pmatrix} s_x & 0 \\ 0 & s_y \end{pmatrix}$$

the ratio between the $x$ and $y$ scale is altered. By applying any sequence of rigid-body transformations, shears and aspect ratio changes, an affine transformation is obtained which describes the cumulative distortions.

The *perspective* transformation accounts for the distortion which occurs when a 3D scene is projected through an idealized optical image system as in Figure 4. This is a mapping from 3D to 2D. In the special case, where the scene is a flat plane such as in an aerial photograph, the distortion is accounted for by a projective transformation. Perspective distortions cause imagery to appear smaller the farther it is from the camera and more compressed the more it is inclined away from the camera. The latter effect is sometimes called foreshortening. If the coordinates of the objects in the scene are known, say $(x_o, y_o, z_o)$, then the corresponding point in the image $(x_i, y_i)$ is given by

$$x_i = \frac{-fx_o}{z_o - f}, \quad y_i = \frac{-fy_o}{z_o - f}$$

where $f$ is the position of the center of the camera lens. (If the camera is in focus for distant objects, $f$ is the focal length of the lens.) In the special case where the scene is composed of a flat plane tilted with respect to the image plane, the *projective* transformation is needed to map the scene plane into an image which is tilt free and of a desired scale [Slama 1980]. This process, called *rectification*, is described in more detail in Section 2.4. The projective transformation maps a coordinate on the plane $(x_p, y_p)$ to a coordinate in the image $(x_i, y_i)$ as follows:

$$x_i = \frac{a_{11}x_p + a_{12}y_p + a_{13}}{a_{31}x_p + a_{32}y_p + a_{33}}$$

$$y_i = \frac{a_{21}x_p + a_{22}y_p + a_{23}}{a_{31}x_p + a_{32}y_p + a_{33}}$$

where the $a$ terms are constants which depend on the equations of the scene and image plane.

If these transformations do not account for the distortions in the scene or if not enough information is known about the camera geometry, global alignment can be determined using a polynomial transformation. This is defined in Section 3.3.3. For perspective distortion of complex 3D scenes, or nonlinear distortions due to the sensor, object deformations and movements and other domain-specific factors, local transformations are necessary. These can be constructed via piecewise interpolation, e.g., splines when matched features are known, or model-based techniques such as elastic warping and object/motion models.

If the geometric transformation $f(x, y)$ can be expressed as a pair of separable functions, i.e., such that two consecutive 1D (scanline) operations can be used to

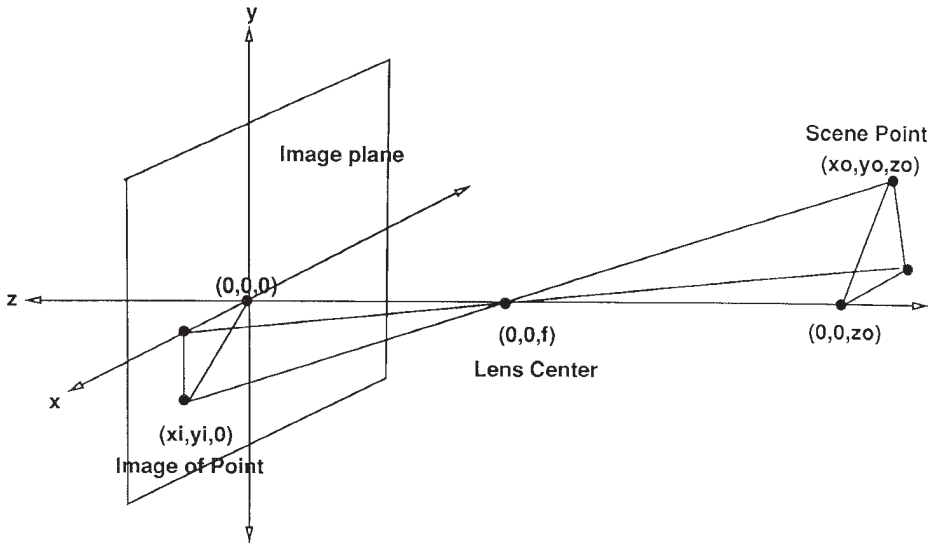



**Figure 4.** Imaging system.

compute the transformation,

$$f(x, y) = f_1(x) \circ f_2(y),$$

then significant savings in efficiency and memory usage can be realized during the implementation. Generally, $f_2$ is applied to each row; then $f_1$ is applied to each column. In classical separability the two operations are multiplied, but for practical purposes any compositing operation can offer considerable speedup [Wolberg and Boult 1989].

### 2.3 Image Variations

Since image registration deals with the removal of distortions and the detection of changes between images, knowledge about the types of variations between images plays a fundamental role in any registration problem. We have found it useful to categorize these variations in the images into three groups based on their different roles in registration problems. These categories are described in Table 4.

First, it is important to distinguish between distortions and other variations. Distortions are variations which are the *source of misregistration*. By this, we mean they are variations which have caused the images to be misaligned and have obscured the true measurement values. It is the distortions between images which we would like to remove by registration. The other variations are usually changes that we are interested in detecting after registration has been performed; they are therefore not distortions. Distortions may be due to a change in the sensor viewpoint, noise introduced by the sensor or its operation, changes in the subject's position, and other undesirable changes in the scene or sensor. They almost always arise from differences in the way or the circumstances under which the images are acquired. This is in contrast to variations of interest which stem from intrinsic differences in the scene, such as physical growths or movements.

Second, we distinguish two categories of distortions. In any registration problem, we would like to remove all the distortions possible. However, this is seldom possible or practical. What is typically done instead is remove the primary spatial discrepancies and to limit the influence of valumetric and small local errors. This is accomplished by choosing a

**Table 4.** Categorization of Variations between Images to be Registered

| |
|---|
| I. Corrected Distortions<br>These are distortions which can be modeled. The model of these distortions determines the class of transformations that will register the images. These distortions are typically geometric, due to simple viewpoint changes and sensor noise. |
| II. Uncorrected Distortions<br>These are distortions which are difficult to model. They are often dependent on the scene and are often valumetric. Typical examples are lighting and atmospheric variations, shadows, and valumetric sensor noise. |
| III. Variations of Interest<br>These are differences between the two images which we would like to detect. For some problems, the purpose of registration is to expose these differences. Common examples include changes in the scene such as object movements, growths or deformations, and differences in sensor measurements when using sensors with varying sensitivities, or using sensors which measure different qualities. |

viable spatial transformation class and by ignoring other variations by choosing the appropriate feature space, similarity measure, and search strategy. This effectively splits the distortions into two categories. The first category is the spatial distortions which can be satisfactorily modeled by a practical transformation class. We call these the *corrected distortions*. The remaining distortions are often caused by lighting and atmospheric changes. This is because their effects depend on the characteristics of the physical objects in the scene, and hence they are difficult to model effectively.

In summary, there are three categories of variations that play important roles in the registration of images. The first type (Type I) are the variations, usually spatial, which are used to determine an appropriate transformation. Since the application of an optimal transformation in this class will remove these distortions, they are called *corrected distortions*. The second type of variations (Type II) are also distortions, usually valumetric, but distortions which are not corrected by the registration transformation. We call these *uncorrected distortions*. Finally, the third type (Type III) are *variations of interest*, differences between the images which may be spatial or valumetric but are not to be removed by registration. Both the uncorrected distortions and the variations of interest, which together we call *uncorrected variations*, affect the choice of feature space, similarity measure, and search strategy that make up the final registration method. The distinction between uncorrected distortions and variations of interest is important, especially in the case where both the distortions and the variations of interest are local, because the registration method must address the problem of removing as many of the distortions as possible while leaving the variations of interest intact.

Table 5 decomposes registration methods based on the type of variations present in the images. This table shows how registration methods can be classified based first on the transformation class (Type I variations) and then subclassified based on the other variations (Types II and III). This table serves as an outline for Section 3.

All variations can be further classified as either static/dynamic, internal/external, and geometric/photometric. Static variations do not change for each image and hence can be corrected in all images in the same procedure via calibration techniques. Internal variations are due to the sensor. Typical internal geometric distortions in earth observation sensors [Bernstein 1976] are centering, size, skew, scan nonlinearity, and radially (pin-cushion) or tangentially symmetric errors. Internal variations which are partially photometric (effect intensity values) include those caused by camera-shading effects (which effectively limit the viewing window), detector gain variations and errors, lens distortions, sensor

**Table 5.** Registration Methods Categorized by the Type of Variations Present between Images

| **Transformation Class** (Type I Variations) | **Type of Variations** (Type II & III Variations) | **Appropriate Methods** (Section 3) |
|---|---|---|
| **Global** | | |
| Small Rigid/Affine | Frequency Independent Valumetric | Correlation-Based, Sequential |
| Translation/Rotation | Frequency Dependent Valumetric | Fourier-Based |
| Small Rigid/Affine | Local | Point Mapping with Feedback |
| General/ | Valumetric | Point Mapping without Feedback |
| Global Polynomial | Few Accurate Control Points (manual, specific domain) | Interpolation |
| | Local Many Inaccurate Control Points (automatic, general domain) | Approximation |
| **Local** | | |
| Local Basis Functions | Global and Valumetric, Control Points Available | Piecewise Interpolation |
| Elastic Model | Local | Elastic Model Based |

imperfections, and sensor-induced filtering (which can cause blemishes and banding).

External errors, on the other hand, arise from continuously changing sensor operations and individual scene characteristics. These might be due to platform perturbations (i.e., changes in viewing geometry) and scene changes due to movement or atmospheric conditions. External errors can similarly be broken down into spatial and value (intensity) distortions. The majority of internal errors and many of the photometric ones are static and thus can be removed using calibration.

Since registration is principally concerned with spatially mapping one image onto another, external geometric distortions play the most critical role in registration. Internal distortions typically do not cause misalignment between images, and the effect on intensity from either internal or external distortions is either of interest or difficult to remove. Intensity distortions that are not static usually arise from a change in sensor, which might be of interest, or from varied lighting and atmospheric conditions. In the cases where intensity distortions are corrected, the intensity histogram and other statistics about the distribution of intensities are used. An example is presented by the method developed by Wong [1977] to register radar and optical data using the Karhunen-Loeve transformation. In Herbin [1989], intensity correction is performed simultaneously with geometric correction.

Since a common objective of registration is to detect changes between images, it is important that images are matched only with regards to the misregistration source. Otherwise the changes of interest will be removed at the same time. Images which contain variations of interest (Type III variations) are sometimes referred to as dissimilar images because the images remain substantially different after they are registered. Registration of dissimilar images often has a special need to model the misregistration source. In hierarchical search techniques described by Hall [1979], for example, matching rules are selected which are more invariant to natural or even man-made changes in scenery. In general, registration of images obtained at different times or under different scene conditions is performed to extract changes in the

scene. Examples are the detection of the growth of urban developments in aerial photography or of tumors in mammiograms. Registration of images acquired from different sensors integrates the different measurements in order to classify picture points for segmentation (whereby regions of the image can be found that correspond to meaningful objects in the scene) and for object recognition (so that these regions can be labeled according to what they correspond to in the scene). In both these instances, of multimodal or temporal registration, variations exist which are not to be removed by registration. This presents additional constraints on matching which must find similarities in the face of irrelevant variations.

Not surprisingly, the more that is known about the type of distortions present in a particular system, the more effective registration can be. For example, Van Wie [1977] decomposes the error sources in Landsat multispectral imagery into those due to sensor operation, orbit and altitude anomalies, and earth rotation. Errors are also categorized as global continuous, swath continuous, or swath discontinuous. Swath errors are produced by differences between sweeps of the sensor mirror in which only a certain number of scan lines are acquired. This decomposition of the sources of misregistration is used in the generation of a registration system with several specialized techniques which depend on the application and classes of distortions to be rectified. For example, a set of control points can be used to solve an altitude model, and swath errors can be corrected independent of other errors, thus reducing the load of the global corrections and improving performance.

In computer vision, images with different viewing geometries, such as stereo image pairs, are "registered" to determine the depth of objects in the scene or their three-dimensional shape characteristics. Surveys in stereomapping include Barnard and Fischler [1982] and Dhond and Aggarwal [1989]. This requires matching features in the images and finding the disparity between them; this is often called the correspondence problem. In this case, the majority of the variations are corrected by the mapping between images, but on the other hand the resulting mapping is highly complex. Consider the problems of occlusion, the different relative position of imaged objects and the complete unpredictability of the mapping because of the unknown depths and shapes of objects in the scene. Hence, problems of stereo matching and motion tracking also have a real need to model the source of misregistration. By exploiting camera and object model characteristics such as viewing geometry, smooth surfaces, and small motions, these registration-like techniques become very specialized. For example, in stereomapping, images differ by their imaging viewpoint, and therefore the source of misregistration is due to differences in perspective. This greatly reduces the possible transformations and allows registration methods to exploit the properties of stereo imagery. Because of the geometry imposed by the camera viewpoints, the location of any point in one image constrains the location of the point in the other image (which represents the same point in the 3D scene) to a line. This is called the epipolar constraint, and the line in which the matching point must lie is called the epipolar line. If the surfaces in the scene are opaque, continuous and if their scanlines (the rows of pixels in the image) are parallel to the baseline (the line connecting their two viewpoints), then an ordering constraint is also imposed along corresponding epipolar lines. See Figure 5. Furthermore, the gradient of the disparity (the change in the difference in position between the two images of a projected point) is directly related to the smoothness of surfaces in the scene. By using these constraints instead of looking for an arbitrary transformation with a general registration method, the stereo correspondence problem can be solved more directly, i.e., search is more efficient and intelligent.

When sufficient information about the misregistration source is available, it may

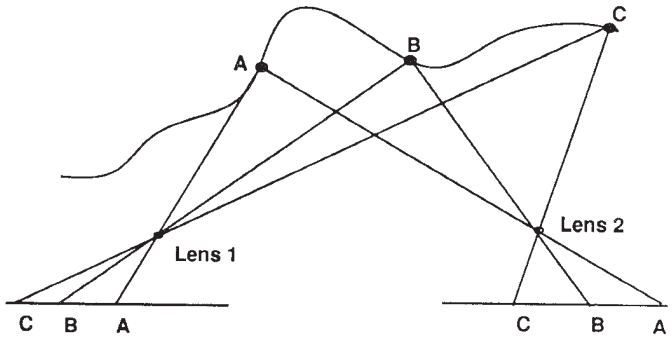



**Figure 5.** Typical stereo imaging system, showing the epipolar constraint. If the surface is opaque and continuous, then the ordering of points along the corresponding epipolar lines of the two images is guaranteed to be the same (Based on a figure in Horn [1989].)

be possible to register images analytically and statically. For example, if the scene is a flat plane and if the two images of the scene differ only in their viewing geometries, and this relative difference is known, then an appropriate sequence of elementary cartesian transformations (namely, a translation, rotation, and scale change or rigid transformation) can be found to align the two images. It may be possible to determine the difference in the viewing geometry for each image, i.e., the position, orientation, and scale of one coordinate system relative to the other, from orbit ephemerides (star maps), platform sensors, or backwards from knowing the depth at three points. This is only possible if there is a simple optical system without optical abberations, i.e., the viewing sensor images a plane at a constant distance from the sensor at a constant scale factor. Registration in this case is accomplished through image rectification which will now be described in detail. Although this form of registration is closely related to calibration (where the distortion is static and hence measurable), it is a good example of the typical viewing geometry and the imaging properties that can be used to determine the appropriate registration transformation. This is the only example that will be given however, where the source of misregistration is completely known and leads directly to an analytical solution for registration.

### 2.4 Rectification

One of the simplest types of registration can be performed when the scene under observation is relatively flat and the viewing geometry is known. The former condition is often the case in remote sensing if the altitude is sufficiently high. This type of registration is accomplished by rectification, i.e., the process which corrects for the perspective distortion in an image of a flat scene. Perspective distortion has the effect of compressing the image of scene features the farther they are from the camera. Rectification is often performed to correct images so that they conform to a specific map standard such as the Universal Transverse Mercator projection. But it can also be used to register two images of a flat surface taken from different viewpoints.

Given an imaging system in which the image center $O$ is at the origin and the lens center $L$ is at $(0, 0, f)$, any scene point $P_o = (x_o, y_o, z_o)$ can be mapped to an image point $P_i = (x_i, y_i)$ by the scale factor $f/(z_o - f)$. This can be seen from the similar triangles in the viewing geometry illustrated in Figure 4. If the scene is a flat plane which is perpendicular to the camera axis (i.e., $z$ is constant) it is already rectified since the scale factor is now constant for all points in the image. For any other flat plane $S$, given by

$$Ax_o + By_o + z_0 = C$$

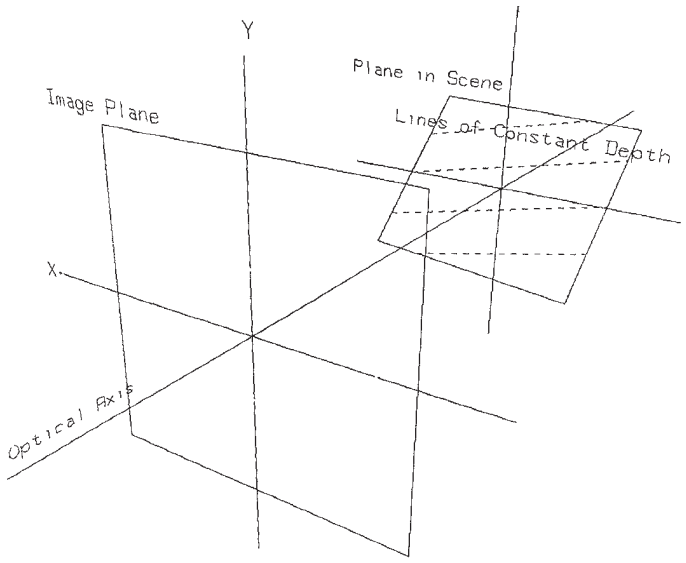



**Figure 6.** Any plane can be decomposed into lines parallel to the image plane.

where $A$, $B$, and $C$ are constants, rectification can be performed by mapping the intensity of the image point at $(x_i, y_i)$ into the new rectified image point location $(fx_i/Z, fy_i/Z)$ where $Z = f - Ax_i - By_i$ [Rosenfeld 1982]. This is because the scene plane can be decomposed into lines $Ax_o + By_o = C'$ each at a constant distance $(z_o = C - C')$ from the image plane. Each line then maps to a line in the image plane, and since its perspective distortion is related to its distance from the image, all points on this line must be scaled accordingly by $f/(C - C' - f)$. Figure 6 shows how a plane is decomposed into lines that are each parallel to the image plane.

Two pictures of the flat plane taken from different viewpoints can be registered by the following steps. First, the scene coordinates $(x_1, y_1, z_1)$ are related to their image coordinates in image 1 of a point with respect to camera 1 by a scale factor $(z_1 - f)/f$ dependent on their depth (the $z_1$ coordinate) and the lens center $f$ because of similar triangles. This gives us two equations. Since they must also satisfy the equation of the plane, we have three equations from which we can derive the three coordinates of each scene point using its corresponding image point with respect to coordinate system of camera 1. The scene coordinates are then converted from the coordinate system with respect to camera 1 to a coordinate system with respect to camera 2 to obtain $(x_2, y_2, z_2)$. Lastly, these can be projected onto image 2 by the factor $f/(z_2 - f)$, again by similar triangles. Of course, if these are discrete images, there is still the problem of interpolation if the registered points do not fall on grid locations. See Wolberg [1990] for a good survey of interpolation methods.

## 3. REGISTRATION METHODS

### 3.1 Correlation and Sequential Methods

Cross-correlation is the basic statistical approach to registration. If is often used for template matching or pattern recognition in which the location and orientation of a template or pattern is found in a picture. By itself, cross-correlation is not a registration method. It is a similarity measure or match metric, i.e., it gives a measure of the degree of similarity between an image and a template. However, there are several registration methods for which it is the primary tool, and it is these methods and the closely re-

lated sequential methods which are discussed in this section. These methods are generally useful for images which are misaligned by small rigid or affine transformations.

For a template $T$ and image $I$, where $T$ is small compared to $I$, the two-dimensional normalized cross-correlation function measures the similarity for each translation:

$$C(u,v) = \frac{\Sigma_x \Sigma_y T(x,y) I(x-u, y-v)}{\sqrt{\left[\Sigma_x \Sigma_y I^2(x-u, y-v)\right]}}.$$

If the template matches the image exactly, except for an intensity scale factor, at a translation of $(i,j)$, the cross-correlation will have its peak at $C(i,j)$. (See Rosenfeld and Kak [1982] for a proof of this using the Cauchy-Schwarz inequality.) Thus, by computing $C$ over all possible translations, it is possible to find the degree of similarity for any template-sized window in the image. Notice the cross-correlation must be normalized since local image intensity would otherwise influence the measure.

The cross-correlation measure is directly related to the more intuitive measure which computes the sum of the differences squared between the template and the picture at each location of the template:

$$D(u,v) = \sum_x \sum_y (T(x,y) - I(x-u, y-v))^2.$$

This measure decreases with the degree of similarity since, when the template is placed over the picture at the location $(u,v)$ for which the template is most similar, the differences between the corresponding intensities will be smallest. The template energy defined as $\Sigma_x \Sigma_y T^2(x,y)$ is constant for each position $(u,v)$ that we measure. Therefore, we should normalize, as before, using the local image energy $\Sigma_x \Sigma_y I^2(x-u, y-v)$. Notice that if you expand this intuitive measure $D(u,v)$ into its quadratic terms, there are three terms: a template energy term, a product term of template and image, and an image energy term. It is the product term or correlation $\Sigma_x \Sigma_y T(x,y) I(x-u, y-v)$ which when normalized, determines the outcome of this measure.

A related measure, which is advantageous when an absolute measure is needed, is the correlation coefficient

$$\frac{covariance(I,T)}{\sigma_I \sigma_t} = \frac{\Sigma_x \Sigma_y (T(x,y) - \mu_T)(I(x-u, y-v) - \mu_I)}{\sqrt{\Sigma_x \Sigma_y (I(x-u, y-v) - \mu_I)^2 \Sigma_x \Sigma_y (T(x,y) - \mu_T)^2}}$$

where $\mu_T$ and $\sigma_T$ are mean and standard deviation of the template and $\mu_I$ and $\sigma_I$ are mean and standard deviation of the image.[1] This statistical measure has the property that it measures correlation on an absolute scale ranging from $[-1, 1]$. Under certain statistical assumptions, the value measured by the correlation coefficient gives a linear indication of the similarity between images. This is useful in order to quantitatively measure confidence or reliability in a match and to reduce the number of measurements needed when a prespecified confidence is sufficient [Svedlow et al. 1976].

Consider a simple example of a binary image and binary template, i.e., all the pixels are either black or white, for which it is possible to predict with some probability whether or not a pixel in the image will have the same binary value as a pixel in the template. Using the correlation coefficient, it is possible to compute the probability or confidence that the image is an instance of the template. We assume the template is an ideal repre-

[1] The mean $\mu$ of an image is the average intensity value; if the image $I$ is defined over a region $x = 1, N$; $y = 1, M$ then $\mu_I = \Sigma_{x=1}^{N} \Sigma_{y=1}^{M} (I(x,y)/(N * M))$. The standard deviation is a measure of the variation there in the intensity values. It is defined as $\sigma_I^2 = \Sigma_{x=1}^{N} \Sigma_{y=1}^{M} ((I(x,y) - \mu_I)^2/(N * M))$.

sentation of the pattern we are looking for. The image may or may not be an instance of this pattern. However, if we can statistically characterize the noise that has corrupted the image, then the correlation coefficient can be used to quantitatively measure how likely it is that the image is an instance of the template.

Another useful property of correlation is given by the Correlation theorem. The Correlation theorem states that the Fourier transform of the correlation of two images is the product of the Fourier transform of one image and the complex conjugate of the Fourier transform of the other. This theorem gives an alternate way to compute the correlation between images. The Fourier transform is simply another way to represent the image function. Instead of representing the image in the spatial domain, as we normally do, the Fourier transform represents the same information in the frequency domain. Given the information in one domain we can easily convert to the other domain. The Fourier transform is widely used in many disciplines, both in cases where it is of intrinsic interest and as a tool, as in this case. It can be computed efficiently for images using the Fast Fourier Transform or FFT. Hence, an important reason why the correlation metric is chosen in many registration problems is because the Correlation theorem enables it to be computed efficiently, with existing, well-tested programs using the FFT (and occasionally in hardware using specialized optics). The use of the FFT becomes most beneficial for cases where the image and template to be tested are large. However there are two major caveats. Only the cross-correlation before normalization may be treated by FFT. Second, although the FFT is faster it also requires a memory capacity that grows with the log of the image area. Last, both direct correlation and correlation using FFT have costs which grow at least linearly with the image area.

Solving registration problems like template matching using correlation has many variations [Pratt 1978]. Typically the cross-correlation between the image and the template (or one of the related similarity measures given above) is computed for each allowable transformation of the template. The transformation whose cross-correlation is the largest specifies how the template can be optimally registered to the image. This is the standard approach when the allowable transformations include a small range of translations, rotations, and scale changes; the template is translated, rotated, and scaled for each possible translation, rotation, and scale of interest. As the number of transformations grows, however, the computational costs quickly become unmanageable. This is the reason that the correlation methods are generally limited to registration problems in which the images are misaligned only by a small rigid or affine transformation. In addition, to reduce the cost of each measurement for each transformation instance, measures are often computed on features instead of the whole image area. Small local features of the template which are more invariant to shape and scale, such as edges joined in a $Y$ or a $T$, are frequently used.

If the image is noisy, i.e., there are significant distortions which cannot be removed by the transformation, the peak of the correlation may not be clearly discernible. The Matched Filter Theorem states that for certain types of noise such as additive white noise, the cross-correlation filter that maximizes the ratio of signal power to the expected noise power of the image, i.e., the information content, is the template itself. In other cases however, the image must be prefiltered before cross-correlation to maintain this property. The prefilter and the cross-correlation filter (the template) can be used to produce a single filter which can simultaneously perform both filtering operations. The prefilter to be used can sometimes be determined if the noise in the image satisfies certain statistical properties. These techniques, which prefilter based on the properties of the noise of the image in order to maximize the peak correlation with respect to this noise

(using the Matched Filter Theorem) and then cross-correlate, are called *matched filter* techniques [Rosenfeld and Kak 1982]. The disadvantages of these techniques are that they can be computationally intensive, and in practice the statistical assumptions about the noise in the image are difficult to satisfy.

A far more efficient class of algorithms than traditional cross-correlation, called the sequential similarity detection algorithms (SSDAs), was proposed by Barnea and Silverman [1972]. Two major improvements are offered. First, they suggest a similarity measure $E(u, v)$, which is computationally much simpler, based on the absolute differences between the pixels in the two images,

$$E(u,v) = \sum_x \sum_y |T(x,y) - I(x-u, y-v)|.$$

The normalized measure is defined as

$$E(u,v) = \sum_x \sum_y |T(x,y) - \hat{T} - I(x-u, y-v) + \hat{I}(u,v)|$$

where $\hat{T}$ and $\hat{I}$ are the average intensities of the template and local image window respectively. This is significantly more efficient than correlation. Correlation requires both normalization and the added expense of multiplications. Even if this measure is unnormalized a minimum is guaranteed for a perfect match. Normalization is useful, however, to get an absolute measure of how the two images differ, regardless of their intensity scales.

The second improvement Barnea and Silverman [1972] introduce is a sequential search strategy. In the simplest case of translation registration this strategy might be a sequential thresholding. For each window of the image (determined by the translation to be tested and the template size), one of the similarity measures defined above is accumulated until the threshold is exceeded. For each window the number of points that were examined before the threshold was exceeded is recorded. The window which examined the most points is assumed to have the lowest measure and is therefore the best registration.

The sequential technique can significantly reduce the computational complexity with minimal performance degradation. There are also many variations that can be implemented in order to adapt the method to a particular set of images to be registered. For example, an ordering algorithm can be used to order the windows tested which may depend on intermediate results, such as a coarse-to-fine search or a gradient technique. These strategies will be discussed in more detail in Section 4.3. The ordering of the points examined during each test can also vary depending on critical features to be tested in the template. The similarity measure and the sequential decision algorithm might vary depending on the required accuracy, acceptable speed, and complexity of the data.

Although the sequential methods improve the efficiency of the similarity measure and search, they still have increasing complexity as the degrees of freedom of the transformation is increased. As the transformation becomes more general the size of the search grows. On the one hand, sequential search becomes more important in order to maintain reasonable time complexity; on the other hand it becomes more difficult not to miss good matches.

In comparison with correlation, the sequential similarity technique improves efficiency by orders of magnitude. Tests conducted by Barnea and Silverman [1972], however, also showed differences in results. In satellite imagery taken under bad weather conditions, clouds needed to be detected and replaced with random noise before correlation would yield a meaningful peak. Whether the differences found in their small study can be extended to more general cases remains to be investigated.

A limitation of both of these methods is their inability to deal with dissimilar images. The similarity measures described so far, the correlation coefficient, and the

sum of absolute differences are maximized and minimized, respectively for identical matches. For this reason, feature-based techniques and measures based on the invariant properties of the Fourier transform are preferable when images are acquired under different circumstances, e.g., varying lighting or atmospheric conditions. In the next section the Fourier methods will be described. Like the correlation and sequential methods, the Fourier methods are appropriate for small translations, rotations, or scale changes. The correlation methods can be used sometimes for more general rigid transformations but become inefficient as the degrees of freedom of the transformation grows. The Fourier methods can only be used where the Fourier transform of an image which has undergone the transformation is related in a nice mathematical way to the original image. The methods to be described in the next section are applicable for images which have been translated or rotated or both. They are specifically well suited for images with low frequency or frequency-dependent noise; lighting and atmospheric variations often cause low-frequency distortions. They are not appropriate for images with frequency-independent noise (white noise) or for more general transformations.

### 3.2 Fourier Methods

The methods to be described in this section register images by exploiting several nice properties of the Fourier Transform. Translation, rotation, reflection, distributivity, and scale all have their counterpart in the Fourier domain. Furthermore, as mentioned in the previous section, the transform can be efficiently implemented in either hardware or using the Fast Fourier Transform. These methods differ from the methods in the last section because they search for the optimal match according to information *in the frequency domain*. The method described in Section 3.1 used the Fourier Transform as a tool to perform a spatial operation, namely correlation.

By using the frequency domain, the Fourier methods achieve excellent robustness against correlated and frequency-dependent noise. They are applicable, however, only for images which have been at most rigidly misaligned. In this section we will first describe the most basic method which uses Fourier Analysis. It is called phase correlation and can be used to register images which have been shifted relative to each other. Then we will describe an extension to this method and several related methods which handle images which have been both shifted and rotated with respect to each other.

Kuglin and Hines [1975] proposed an elegant method, called *phase correlation*, to align two images which are shifted relative to one another. In order to describe their method, we will define a few of the terms used in Fourier Analysis which we will need. The Fourier transform of an image $f(x, y)$ is a complex function; each function value has a real part $R(\omega_x, \omega_y)$ and an imaginary part $I(\omega_x, \omega_y)$ at each frequency $(\omega_x, \omega_y)$ of the frequency spectrum:

$$F(\omega_x, \omega_y) = R(\omega_x, \omega_y) + iI(\omega_x, \omega_y)$$

where $i = \sqrt{-1}$. This can be expressed alternatively using the exponential form as

$$F(\omega_x, \omega_y) = |F(\omega_x, \omega_y)| e^{i\phi(\omega_x, \omega_y)}$$

where $|F(\omega_x, \omega_y)|$ is the magnitude or amplitude of the Fourier transform and where $\phi(\omega_x, \omega_y)$ is the phase angle. The square of the magnitude is equal to the amount of energy or power at each frequency of the image and is defined as:

$$|F(\omega_x, \omega_y)|^2 = R^2(\omega_x, \omega_y) + I^2(\omega_x, \omega_y).$$

The phase angle describes the amount of phase shift at each frequency and is defined as

$$\phi(\omega_x, \omega_y) = \tan^{-1}\left[I(\omega_x, \omega_y)/R(\omega_x, \omega_y)\right].$$

Phase correlation relies on the translation property of the Fourier transform,

sometimes referred to as the Shift Theorem. Given two images $f_1$ and $f_2$ which differ only by a displacement $(d_x, d_y)$, i.e.,

$$f_2(x, y) = f_1(x - d_x, y - d_y),$$

their corresponding Fourier transforms $F_1$ and $F_2$ will be related by

$$F_2(\omega_x, \omega_y) = e^{-j(\omega_x d_x + \omega_y d_y)} F_1(\omega_x, \omega_y).$$

In other words, the two images have the same Fourier magnitude but a phase difference directly related to their displacement. This phase difference is given by $e^{j(\phi_1 - \phi_2)}$. It turns out that if we compute the cross-power spectrum of the two images defined as

$$\frac{F_1(\omega_x, \omega_y) F_2^*(\omega_x, \omega_y)}{|F_1(\omega_x, \omega_y) F_2^*(\omega_x, \omega_y)|} = e^{(\omega_x d_x + \omega_y d_y)}$$

where $F^*$ is the complex conjugate of $F$, the Shift Theorem guarantees that the phase of the cross-power spectrum is equivalent to the phase difference between the images. Furthermore, if we represent the phase of the cross-power spectrum in its spatial form, i.e., by taking the inverse Fourier transform of the representation in the frequency domain, then we will have a function which is an impulse, that is, it is approximately zero everywhere except at the displacement which is needed to optimally register the two images.

The Fourier registration method for images which have been displaced with respect to each other therefore entails determining the location of the peak of the inverse Fourier transform of the cross-power spectrum phase. Since the phase difference for every frequency contributes equally, the location of the peak will not change if there is noise which is limited to a narrow bandwidth, i.e., a small range of frequencies. Thus this technique is particularly well suited to images with this type of noise. Consequently, it is an effective technique for images obtained under differing conditions of illumination since illumination changes are usually slow varying and therefore concentrated at low-spatial frequencies. Similarly, the technique is relatively scene independent and useful for images acquired from different sensors since it is insensitive to changes in spectral energy. This property of using only the phase information for correlation is sometimes referred to as a *whitening* of each image. Among other things, whitening is invariant to linear changes in brightness and makes the phase correlation measure relatively scene independent.

On the other hand, if the images have significant white noise, noise which is spread across all frequencies, then the location of the peak will be inaccurate since the phase difference at each frequency is corrupted. In this case, the methods described in the last section which find the peak of the spatial cross-correlation are optimal. Kuglin and Hines [1975] suggest introducing a generalized weighting function to the phase difference before taking the inverse Fourier transform to create a family of correlation techniques, including both phase correlation and conventional cross-correlation. In this way, a weighting function can be selected according to the type of noise immunity desired.

In an extension of the phase correlation technique, De Castro and Morandi [1987] have proposed a technique to register images which are both translated and rotated with respect to each other. Rotational movement, by itself without translation, can be deduced in a similar manner as translation using phase correlation by representing the rotation as a translational displacement with polar coordinates. But rotation and translation together represent a more complicated transformation. De Castro and Morandi [1987] present the following two-step process to first determine the angle of rotation and then determine the translational shift.

Rotation is invariant with the Fourier transform. Rotating an image rotates the Fourier transform of that image by the same angle. If we know the angle, then we can rotate the cross-power spectrum and determine the translation according

to the phase correlation method. However, since we do not know the angle, we compute the phase of the cross-power spectrum as a function of the rotation angle estimate $\phi$ and use polar coordinates $(r, \theta)$ to simplify the equation. This gives us a function

$$G(r, \theta; \phi) = \frac{F_1(r, \theta)F_2^*(r, \theta - \phi)}{|F_1(r, \theta)F_2^*(r, \theta - \phi)|}$$

which at the true angle of rotation should have the form expected for images which have only been translated. Therefore, by first determining the angle $\phi$ which makes the inverse Fourier transform of the phase of the cross-power spectrum the closest approximation to an impulse, we can then determine the translation as the location of this pulse.

In implementing the above method, it should be noted that some form of interpolation must be used to find the values of the transform after rotation since they do not naturally fall in the discrete grid. Although this might be accomplished by computing the transform after first rotating in the spatial domain, this would be too costly. De Castro and Morandi [1987] applied the transform to a zero-padded image thus increasing the resolution and improving the approximation of the transform after rotation. The method is also costly because of the difficulty in testing for each $\phi$. Alliney and Morandi [1986] presented a method which only requires one-dimensional Fourier transformations to compute the phase correlation. By using the x- and y-projections of each image, the Fourier transforms are given by the projection slice theorem. The 1D transform of the x- and y-projections is simply the row of the 2D transform where $\omega_x = 0$ and the column where $\omega_y = 0$ respectively. Although substantial computational savings are gained, the method is no longer robust except for relatively small translations.

The Fourier methods, as a class, offer advantages in noise sensitivity and computational complexity. Lee et al. [1987] developed a similar technique which uses the power cepstrum of an image (the power spectrum of the logarithm of the power spectrum) to register images for the early detection of glaucoma. First the images are made parallel by determining the angle which minimizes the differences in their power spectra (which should theoretically be zero if there is only a translational shift between them). Then the power spectrum is used to determine the translational correspondence in a similar manner to phase correlation. This has the advantage over De Castro and Morandi [1987] of the computational savings gained by adding images instead of multiplying them due to the use of logarithms. The work of De Castro and Morandi [1987] summarizes previous work published in Italy before 1987, but no direct comparison with Lee et al. [1987] has yet been undertaken. Both methods achieve better accuracy and robustness than the primary methods mentioned in Section 3.1 and for less computational time than classical correlation. However, because the Fourier methods rely on their invariant properties, they are only applicable for certain well-defined transformations such as rotation and translation. In the following section a more general technique is described based on a set of matched control points. These techniques can be used for arbitrary transformations including polynomial and piecewise local. Furthermore, even in the case of a small rigid transformation, it is not always possible to register images using the techniques described so far. If there exists a significant amount of spatially local variation (even though the misalignment is only due to a small rigid transformation), then the correlation and Fourier techniques break down. The sophisticated use of feature detection can help to overcome some of these local variations, but in Section 3.3.2 we will introduce th e more powerful methods which use point mapping with feedback in order to tolerate these types of variations.

### 3.3 Point Mapping

The point- or landmark-mapping technique is the primary approach currently taken to register two images whose type

of misalignment is unknown. This occurs if the class of transformations cannot be easily categorized such as by a set of small translations or rigid-body movements. For example, if images are taken from varying viewpoints of a scene with smooth depth variations, then the two images will differ depending on the perspective distortion. We cannot determine the proper perspective transformation because in general we do not know the actual depths in the scene, but we can use the landmarks that can be found in both images and match them using a general transformation. However, if the scene is not composed of smooth surfaces, but has large depth variations, then the distortions will include occlusions which differ between images, objects which appear in different relative positions between images, and other distortions which are significantly more local. As these distortions become more local, it will become progressively more difficult for a global point-mapping method to account for the misalignment between the images. In this case, methods which use a local transformation, such as the local point-mapping methods, would be preferable.

The general method for point mapping consists of three stages. In the first stage features in the image are computed. In the second stage, feature points in the reference image, often referred to as *control points*, are corresponded with feature points in the data image. In the last stage, a spatial mapping, usually two 2D polynomial functions of a specified order (one for each coordinate in the registered image) is determined using these matched feature points. Resampling of one image onto the other is performed by applying the spatial mapping and interpolation.

The above description of point mapping as a three-stage process is the standard point-mapping technique used for images which are misaligned by an unknown transformation. However, there is also a group of point-mapping methods which are used for images whose misalignment is a small rigid or affine transformation, but which contain significant amounts of local uncorrected variations. The techniques in Sections 3.1 and 3.2 are not adequate in this case because the relative measures of similarity between the possible matches become unreliable. Point-mapping methods can overcome this problem by the use of feedback between the stages of finding the correspondence between control points and finding the optimal transformation.

In the following four sections, we will describe (1) the different types of control points and how they are matched, (2) point mapping with feedback for small rigid or affine transformations with local variations, (3) the global point-mapping methods which find a general transformation from the matched control points, and (4) more recent work in local mapping that uses image-partitioning techniques and local piecewise transformations.

#### 3.3.1 Control Points

Control points for point matching play an important role in the efficacy of this approach. After point matching, the remaining procedure (of the three-stage point mapping) acts only to interpolate or approximate. Thus the accuracy of the point matching lays the foundation for accurate registration. In this section, we will describe the various features used as control points, how they are determined, and how the correspondence between control points in the reference and data image is found.

There are many registration methods other than point-mapping techniques which also perform feature detection. The features which are used in all of these techniques, including point mapping, are described in Section 4.1. In this section the emphasis is on the aspects of features which are used as control points for point mapping and how they are matched prior to the determination of the optimal transformation.

Control points can either be intrinsic or extrinsic. Intrinsic control points are markers in the image which are not rele-

vant to the data itself. They are often placed in the scene specifically for registration purposes and are easily identified. If marks are placed on the sensor, such as reseau marks which are small crossbars inscribed on the faceplate of the sensor, then they aid registration insofar as they independently calibrate each image according to sensor distortions. In medical imaging, identifiable structures, called fiducial markers, are placed in known positions in the patients to act as reference points. In magnetic resonance imaging (MRI) systems [Evans et al. 1988], chemical markers, such as plastic N-shaped tubing filled with $CuSO_4$, are strategically placed to assist in registration. In positron emission tomography (PET) [Bergström et al. 1981; Bohm et al. 1983; Bohm and Greitz 1988; Fox et al. 1985] stereotactic coordinate frames are used so that a three-dimensional coordinate reference frame can be identified. Although intrinsic control points are preferable for obvious reasons, there are not always intrinsic points that can be used. For example, precisely placing markers internally is not always possible in diagnostic images [Singh et al. 1979].

Control points that are extrinsic are determined from the data, either manually or automatically. Manual control points, i.e., points recognized by human intervention, such as identifiable landmarks or anatomical structures, have several advantages. Points can be selected which are known to be rigid, stationary, and easily pin-pointed in both data sets. Of course, they require someone who is knowledgeable in the domain. In cases where there is a large amount of data this is not feasible. Therefore many applications use automatic location of control points. Typical features that are used are corners, line intersections, points of locally maximum curvature on contour lines, centers of windows having locally maximum curvature, and centers of gravity of closed-boundary regions [Goshtasby 1988]. Features are selected which are likely to be uniquely found in both images (a more delicate issue when using multisensor data) and more tolerant of local distortions. Since computing the proper transformation depends on these features, a sufficient number must be detected to perform the calculation. On the other hand too many features will make feature matching more difficult. The number of features to use becomes a critical issue since both the accuracy and the efficiency of point-matching methods will be strongly influenced. This will be discussed in more detail in Section 3.3.3 for the case where a global polynomial transformation is used.

After the set of features has been determined, the features in each picture must be matched, i.e., each feature in one image is matched with its corresponding feature in the other image. For manually identified landmarks, finding the points and matching them are done simultaneously. For most cases, however, a small-scale registration requiring only translation such as template matching is applied to find the match for each feature. Commonly, especially with manual or intrinsic landmarks, if they are not matched manually, this is done using cross-correlation since high accuracy is desired at this level and since the template size is small enough so the computation is feasible. For landmarks which are found automatically, matches can be determined based on the properties of these points, such as curvature or the direction of the principal axes. Other techniques combine the matching of features and the determination of the optimal transformation; these involve clustering, relaxation, matching of minimum spanning trees of the two sets, and matching of convex hull edges of the two sets [Goshtasby 1988]. Some of these techniques are described in Section 3.3.2. Instead of mapping each point individually, these techniques map the set of points in one image onto the corresponding set in the second image. Consequently the matching solution uses the information from all points and their relative locations. This results in a registration technique which matches control points and determines the best spatial

transformation simultaneously. In cases where it is hard to match the control points, e.g., ambiguous features such as corners that are found automatically or local variations which make matching unreliable, this has the advantage that the transformation type is used to constrain the match. However, in the cases where an accurate set of point matches can be determined a priori, an optimal global transformation can be found directly using standard statistical techniques. The latter is the major approach to registration that has been taken historically because control points were often manually determined and because of its computational feasibility.

#### 3.3.2 Point Mapping with Feedback

In this section, we will briefly describe a few examples of methods which have been developed for rigid and affine transformations for cases where feature detection and feature matching are difficult. Through the use of feedback between the stages of finding control point matches and finding the optimal transformation, these techniques can successfully register images where automatically acquired features are ambiguous or when there are significant amounts of uncorrected local variations. These techniques rely on more sophisticated search strategies including relaxation, cooperation, clustering, hierarchical search, and graph matching. Search strategies are described in more detail in Section 4.3.

An example of a point-mapping technique with feedback is the relaxation technique described by Ranade and Rosenfeld [1980], which can be used to register images under translation. Point matching and the determination of the best spatial transformation are accomplished simultaneously. Each possible match for a feature point defines a displacement which is given a rating according to how closely other pairs would match under this displacement. The procedure is iterated, adjusting in parallel the weights of each pair of points based on their ratings until the optimal translation is found. Each match whose displacement is close to the actual displacement will tend to have a higher rating, causing it to have a larger influence as the procedure iterates. This type of technique can tolerate global and local uncorrected variations. It was able to find the correct translation for cases in which shifted patterns were also rotated and scaled and in cases where feature points were each independently displaced using a uniformly distributed direction and jump size. However, the computational complexity is $O(n^4)$ where $n$ is the number of control points. This was improved on by Ton and Jain [1989], who performed experiments on LANDSAT images, by taking advantage of the distinguishing properties of the features (in addition to their relative displacements) and by the use of two-way matching in which points in both images initiate the matching process. The time complexity of their improved relaxation algorithm was $O(n^3)$.

The clustering technique described by Stockman et al. [1982] is another example of a point-mapping method with feedback, or, in other words, a method which determines the optimal spatial transformation between images by an evaluation of all possible pairs of feature matches. In this case the transformation is rigid, i.e., a rotation, scaling, and translation, although it could be extended to other simple transformations. For each possible pair of matched features, the parameters of the transformation are determined which represent a point in the cluster space. By finding the best cluster of these points, using classical statistical methods, the transformation which most closely matches the largest number of points is found. This technique, like the relaxation method described above, can tolerate uncorrected local variations but also has a time complexity of $O(n^4)$. Since this becomes prohibitive as the number of points grows, Goshtasby and Stockman [1985] suggested selecting a subset of the points to reduce the search domain. Subsets were selected as points on the boundary of the convex hulls of the

point sets. (The convex hull is the smallest convex shape that includes all the points in the set.) Although the sets may not be the same (if the images are noisy), the expectation is that there will be some common points.

Another refinement which was used by [Goshtasby et al. 86] was to use the center of gravity of closed-boundary regions as control points that are iteratively updated based on the current optimal rigid transformation. Using a simple segmentation scheme based on iteratively thresholding the image, closed-boundary regions are found. The centers of gravity of these regions are used as control points. The correspondence between control points in the two images was determined based on a clustering scheme like the one used in [Stockman et al. 82]. These matches were used to find the best rigid transformation based on a least squares error analysis. This rigid transformation was in turn used to improve the segmentation of each region until it was the most similar to its corresponding region in the other image (based on its shape, independent of its position, orientation or scale.) In this way, the segmented regions become optimally similar to their corresponding region in the other image. Furthermore, the centers of gravity of each region can be computed with subpixel accuracy with the expectation that they are reasonably immune to random noise and that this information can be used to improve the registration.

An example of satellite images registered by this technique are shown in Figure 7. Figure 7(a) shows a Heat Capacity Mapping Mission Satellite Day-Visible (HCMM Day-Vis) image from an area over Michigan acquired 9/26/79. Figure 7(b) shows an HCMM Night-IR image of about the same area acquired on 7/4/78. Figures 7(c) and (d) show the closed boundary regions (whose perimiters are not too large or small) found from the initial segmentation of these images. Figure 7(e) shows the regions of (d), the Night-IR image, after the application of the refinement algorithm. Notice how the regions in (e) are significantly more similar to the regions of (c) than before refinement. Using the centers of gravity of these regions as the final set of control points, the results of mapping Figure 7(b) to match Figure 7(a) are shown in (f). To see the registration more clearly, the result of negating (f) and overlaying it onto (a) are shown in (g). The mean square error between the centers of gravity of the corresponding regions in the two images was slightly less than one pixel.

Schemes of this type allow for global matching which is less sensitive to uncorrected local variations because (1) they use control points and local similarity measures, (2) they use information from spatial relationships between control points in the image, and (3) they are able to consider possible matches based only on supporting evidence. The implementation of these techniques requires more sophisticated search algorithms because the local uncorrected variations make the search space more difficult to traverse. Hence, these methods take advantage of the more subtle information available based on partial matches and the relationships between matches and by testing more possible combinations of matches. However, without the use of additional constraints, such as imposed by camera geometry or the semantics of the scene, these techniques are limited to small affine transformations because otherwise the search space becomes too large and unmanageable.

#### 3.3.3 *Point Mapping without Feedback — Global Polynomial Methods*

Standard point-mapping techniques, i.e., point mapping without feedback, can be used to register images for which the transformation necessary to align the images is unknown. Since it is often very difficult to categorize and model the source of misregistration, these techniques are widely used.

Global methods based on point matching use a set of matched points to generate a single optimal transformation. Given a sufficient number of points we can derive the parameters of any trans-

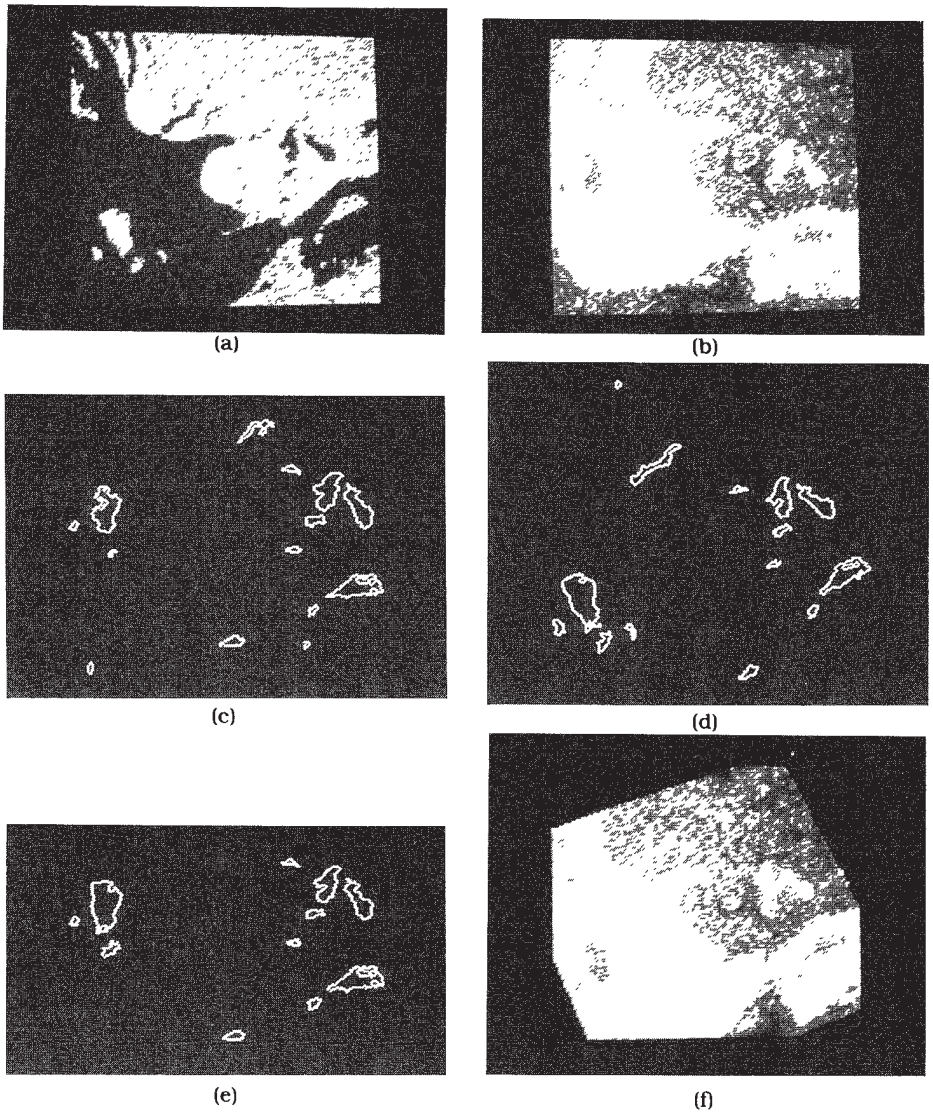
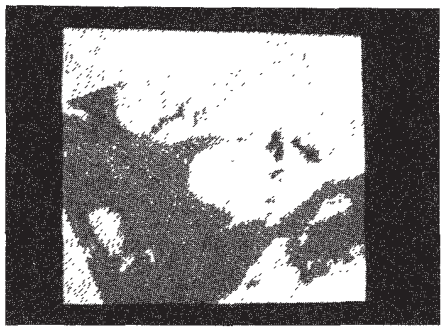



**Figure 7.** An example of satellite images registered by a point-mapping technique with feedback. (a) and (b) are the original images; (c) and (d) show the closed-boundary regions found in the first iteration; (e) shows the regions in (d) after they have been refined to make them as similar to the regions in (c) as possible; (f) shows (b) transformed, based on these final regions to match (a); and (g) shows the final registration, i.e., (f) is overlaid onto (a). (Reprinted with permission from Goshtasby et al. [1986], copyright 1986, IEEE.)

formation either through approximation or interpolation. In approximation, parameters of the transformation are found so the matched points satisfy it as nearly as possible. This is typically done with a statistical method such as least squares regression analysis or clustering. The approximation approach assumes that the matches are distorted by local noise. This noise cannot be removed by the transformation, either because the transformation cannot account for it or because the



(g)

**Figure 7.** Continued

images contain differences of interest. Therefore the transformation to be found does not match the control points exactly, but finds the best approximation. The number of matched points must be sufficiently greater than the number of parameters of the transformation so that sufficient statistical information is available to make the approximation reliable. For large numbers of automatic control points, approximation makes the most sense since the matches are likely to be inaccurate; but taken together they contain a lot of statistical information. For intrinsic or manual control points, there are usually fewer but more accurate matches, suggesting that interpolation may be more applicable. Interpolation finds the transformation which matches the two images so that the matches found for control points are exactly satisfied. There must be precisely one matched point for each independent parameter of the transformation to solve the system of equations. The resulting transformation defines how the image should be resampled. However, if there are too many control points then the number of constraints to be satisfied also increases. If polynomial transformations are used, this causes the order of the polynomial to grow and the polynomial to have large unexpected undulations. In this case, least squares approximation or splines and other piecewise interpolation methods are preferable.

In many registration problems, the precise form of the mapping function is unknown, and therefore a general transformation is needed. For this reason, bivariate polynomial transformations are typically used. They can be expressed as two spatial mappings

$$u = \sum_{i=0}^{m} \sum_{j=0}^{i} a_{ij} x^i y^{j-i}$$

$$v = \sum_{i=0}^{m} \sum_{j=0}^{i} b_{ij} x^i y^{j-i}$$

where $(x, y)$ are indices into the reference image; $(u, v)$ are indices into the image to be mapped onto, and $a_{ij}$ and $b_{ij}$ are the constant polynomial coefficients to be determined. The order of the polynomial, $m$, depends on the trade-off between accuracy and speed needed for the specific problem. For many applications, second or third order is sufficient [Nack 1977; Van Wie and Stein 1977]. In general, however, polynomial transformations are only useful to account for low-frequency distortions because of their unpredictable behavior when the degree of the polynomial is high. A famous example of this, discovered by C. Runge in 1901, is shown in Figure 9 [Forsythe et al. 1977].

If interpolation is used, the coefficients of the polynomials are determined by a system of $N$ equations determined by the mapping of each of the $N$ control points. In least squares approximation, the sum over all control points of the squared difference between the left- and right-hand side of the above equations is minimized. In the simplest scheme, the minimum can be determined by setting the partial derivatives to zero, giving a system of $T = (m = 2)(m = 1)/2$ linear equations known as the normal equations. These equations can be solved if the number of control points is much larger than $T$.

Bernstein [1976] uses this method to correct satellite imagery with low-frequency sensor-associated distortions as well as for distortions caused by earth curvature and camera attitude and alti-

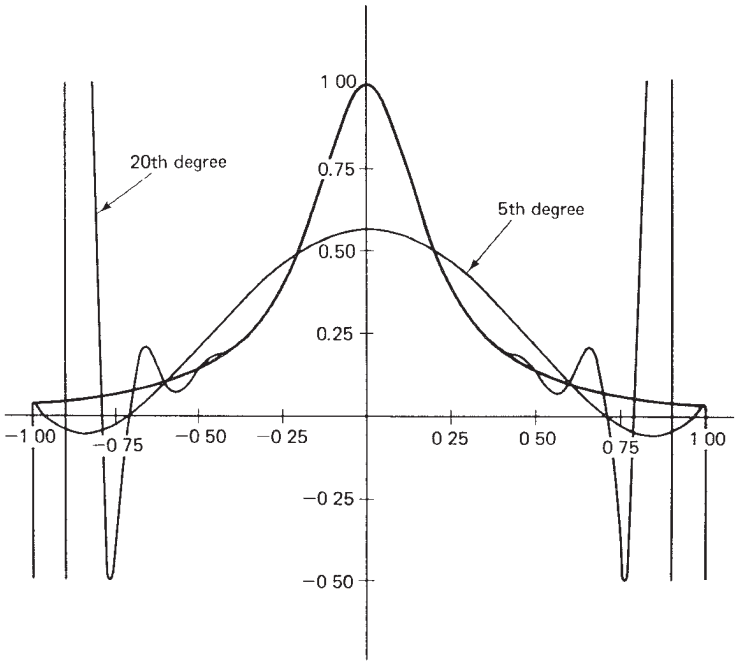



**Figure 8.** The simple function $1/(1 + 25x^2)$ is interpolated by a 5th and 20th degree polynomial. (Reprinted from Forsythe et al. [1977], with permission, Prentice-Hall, Inc.)

tude deviations. Maguire et al. [1990] used this method for registration of medical images from different modalities. The CT and SPECT images shown in Figure 2 are an example of their method. Anatomic landmarks, found manually, in the first image, are cross-correlated with pixels near the corresponding landmarks in the second image to create a set of matched control points. Linear regression is used to fit a low order polynomial transformation which is applied to map one image onto the other. In the bottom of the figure, the MRI image (on the right) is transformed to fit the SPECT image. A contour drawn by hand around the liver on the MRI image is shown on the SPECT image to show the results of the registration. Registration was able to correct for affine distortions such as translation, rotation, scale and shear, in addition to other global distortions which are more difficult to categorize. However, in cases where more information is known about the differences in acquisition of the two images, then a general polynomial transformation may not be needed. Merickel [1988] registers successive serial sections of biological tissue for their 3D reconstruction using a linear least squares fitting of feature points to a transformation composed directly of a rotation, translation, and scaling.

As the order of the polynomial transformation increases and hence the dependencies between the parameters multiplies, using the normal equations to solve the least squares approximation can become computationally costly and inaccurate. This can be alleviated by using orthogonal polynomials for the terms of the polynomial mapping. This basically involves representing the original polynomial mapping as a combination of orthogonal polynomials which are in turn constructed from linearly independent functions. Because the polynomials are orthogonal, the dependencies between parameters are simplified, and the parameters of the new representation can

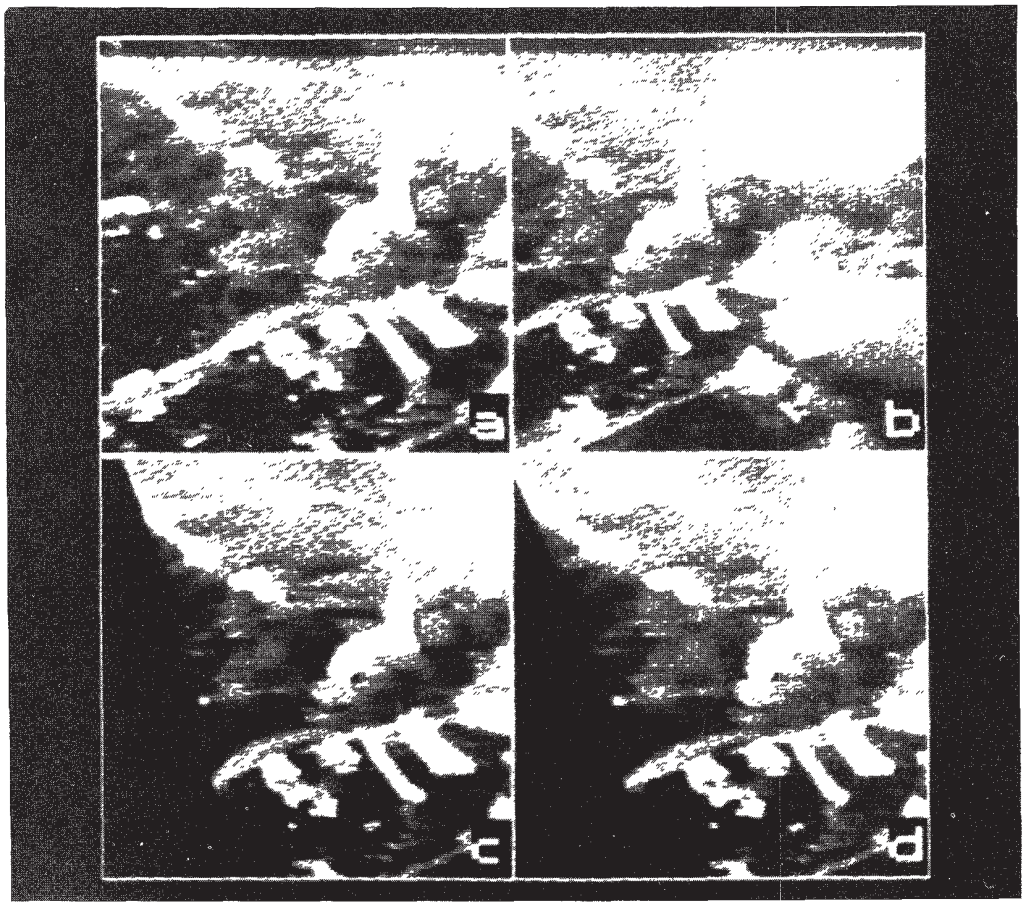



**Figure 9.** An example of images that could not be satisfactorily registered using a global polynomial mapping (bottom right), but required finding a local transformation (bottom left). Original aerial images taken at different times and from different positions are shown at the top. The local registration technique used was based on an approximation to the surface spline, but was much faster. (Reprinted from Flusser et al. [1992] with permission from Pergamon Press, thanks to J. Flusser.)

be found analytically, i.e., there is no longer any need to solve a system of linear equations. The orthogonal polynomials have the additional nice property that the accuracy of the transformation can be increased as desired without recalculating all the coefficients by simply adding new terms until the error is sufficiently small [Goshtasby 1988].

The major limitation of the global point-mapping approach is that a global transformation cannot account for local geometric distortions such as sensor nonlinearities, atmospheric conditions, and local three-dimensional scene features observed from different viewpoints. In the next section, we will describe how to overcome this drawback by computing local transformations which depend only on the control points in their vicinity.

### *3.3.4 Local Methods — Piecewise Interpolation*

The global point-mapping methods discussed above cannot handle local distortions. Approximation methods spread local distortions throughout the image, and polynomial interpolation methods used

with too many control points require high-order polynomials which behave erratically. These methods are characterized as global because a single transformation is used to map one image onto the other. This transformation is generally found from a single computation using all the control points equally.

In the local methods to be discussed in this section, multiple computations are performed, either for each local piece or iteratively, spreading computations to different neighborhoods. Only control points sufficiently close, or perhaps, weighted by their proximity, influence each part of the mapping transformation. In other words, the mapping transformation is no longer a single mapping with one set of parameters independent of position. The parameters of the local mapping transformation vary across the different regions of the image, thus accounting for distortions which differ across the image. Local methods are more powerful and can handle many distortions that global methods cannot; examples include complex 3D scenes taken from different viewpoints, deformable objects or motions, and the effects of different sensors or scene conditions. On the other hand, there is a trade-off between the power of these methods and their corresponding computational cost.

An example is shown in Figure 9 of aerial images which could not be satisfactorily registered using polynomial mapping. In the top of the figure are two images taken at different times from different positions of the aircraft. Using 17 control points, a 2nd-order, polynomial, mapping function was fit using least squares approximation. The results of using this mapping are shown at the bottom right. Because of the local distortions, the average error is more than 5 pixels. Using a local method proposed by Flusser [1992] the images were registered (bottom left) with less than 1 pixel accuracy.

The class of techniques which can be used to account for local distortion by point matching is piecewise interpolation. In this methodology, a spatial mapping transformation for each coordinate is specified which interpolates between the matched coordinate values. For $N$ control points whose coordinates are mapped by

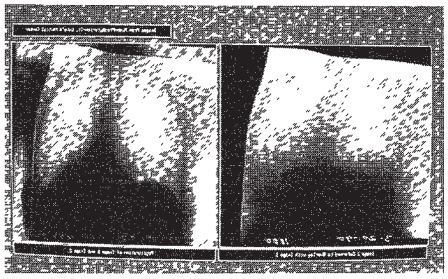


**Figure 10.** The chest x-rays shown in Figure 1 are shown here after registration using surface splines. (Thanks to A. Goshtasby.)

$$X_i = F_x(x_i, y_i)$$
$$Y_i = F_y(x_i, y_i) \quad i = 1, \ldots, N$$

two bivariate functions (usually smooth) are constructed which take on these values at the prescribed locations. Methods which can be applied in this instance must be designed for irregularly spaced data points since the control points are inevitably scattered. A study of surface approximation techniques conducted by Franke [1979] compared these methods exactly, testing each on several surfaces and evaluating their performance characteristics. As will be seen, the methods used in Franke's [1979] study, although not designed for this purpose, underlie much of the current work in local image registration.

Most of the methods evaluated by Franke [1979] use the general spline approach to piecewise interpolation. This requires the selection of a set of basis functions, $B_{i,j}$, and a set of constraints to be satisfied so that solving a system of linear equations will specify the interpolating function. In particular, the spline surface $S(x, y)$ can be defined as

$$S(x, y) = \sum_{i,j} V_{i,j} B_{i,j}(x, y)$$

where $V_{i,j}$ are the control points. For most splines, the basis functions are constructed from low-order polynomials, and the coefficients are computed using constraints derived by satisfying end conditions and various orders of spatial continuity. In the simplest case, a weighted sum of neighboring points is computed where the weights are related inversely with distance such as in linear interpolation. These methods are called inverse-distance weighted interpolation. Another alternative is to have the set of neighboring points determined from some partitioning of the image, such as triangulation. In this case, the weights depend on the properties of the subregions. These methods are called triangle-based methods. Another set of methods considered in Franke's study are the global basis function type methods. These methods are characterized by global basis functions $G_{i,j}(x, y)$ and coefficients $A_k$ which are determined by enforcing that the equation, $S(x, y) = \Sigma_{i,j} A_{i,j} G_{i,j}(x, y)$, interpolates the data. These techniques include the so called surface spline which interpolates the data by representing it as the surface of an infinite plate under the imposition of point loads, i.e., the data. Several variations of each method were examined, altering the basis functions, the weighting system, and the type of image partitioning. This comprehensive study is a good reference for comparing the accuracy and complexity of surface interpolation techniques for scattered data.

Although all these methods compute local interpolation values, they may or may not use all points in the calculation. Those which do are generally more costly and may not be suitable for large data sets. However, because global information can be important, many local methods (i.e., methods which look for a local registration transformation) employ parameters computed from global information. The global basis function type methods are an instance of this. The chest x-ray images shown in Figure 1 which have significant local distortions between them, were registered by Goshtasby [1988] using the surface spline variety of this technique. The results of this registration are shown in Figure 10. Among the methods studied by Franke, most of the global basis function methods, including the surface spline, were among the most accurate, albeit they were also among the slowest.

Flusser [1992] has modified this approach to make it faster while maintaining satisfactory accuracy by adaptively subdividing the image, computing each subregion with a simpler, i.e., faster transformation but only using this transformation if the error between it and the results of using the surface spline are sufficiently small. (The error is computed by sampling and evaluating a subset of random points in the region.) If the error is too large, the region is recursively subdivided until the error criterion is met. The images in Figure 9 were registered using this technique. While global methods are often the most accurate, local methods which rely only on local computations are not only more efficient but they can also be locally controllable. These methods can be very useful for manual registration in a graphics environment. Regions of the image can be registered without influencing other portions which have already been matched. Furthermore, in some cases, e.g., if local variations of interest exist which should not influence the transformation, then local computations may actually be preferable.

From the set of surface interpolation techniques discussed in the study, many registration techniques are possible. For instance, Goshtasby [1986] proposed using "optimal" triangulation of the control points to partition the image into local regions for interpolation. Triangulation decomposes the convex hull of the control points of the image into triangular regions; in "optimal" triangulation, the points inside each triangular region are closer to one of its vertices than to the vertices of any other triangle. The mapping transformation is then computed for each point in the image from interpolation of the vertices in the triangular patch

to which it belongs. Later, Goshtasby [1987] extended this method so that mapping would be continuous and smooth ($C^1$) by using piecewise cubic polynomial interpolation. To match the number of constraints to the number of parameters in the cubic polyonmials, Goshtasby [1987] decomposed each triangle into three subtriangles (using any point inside the triangle as the third vertex for each subtriangle) as suggested by Clough and Tocher [1965]. By ensuring that the partial derivatives match at each point and for each edge which belongs to two triangular patches, the method can solve for the parameters of each cubic patch and provide a smooth transition between patches.

The piecewise cubic polynomial method can successively register images with local geometric distortion assuming the difference between images is continuous and smooth. However, where discontinuous geometric differences exist, such as in motion sequences where occlusion has occurred, the method fails. Also, the Franke [1979] study concluded that methods that use triangulation are problematic when long thin triangles occur and that estimation of partial derivatives can prove difficult. The cost of this technique is composed of the cost of the triangulation, the cost of solving a system of linear equations for each triangular patch, and the cost of computing the value of each registered point from the resulting polynomial. Triangulation is the preliminary "global" step whose complexity grows with the number of control points. Of the various algorithms that can be used for triangulation, Goshtasby [1987]selected one of the fastest and easiest to implement. It is based on a divide-and-conquer algorithm with complexity $O(N \log N)$ where $N$ is the number of control points. Since the remaining computation is purely local, it is relatively efficient, but its success is strictly limited by the number, location, and proximity of the control points which completely control the final registration.

For many registration problems, both local and global distortions exist, and it is useful to take a hierarchical approach in finding the optimal transformation. Ratib [1988] suggests that it is sufficient for the "elastic" matching of PET images of the heart to match the images globally by the best rigid transformation and then improve this by a local interpolation scheme which perfectly matches the control points. From the rigid transformation, the displacement needed to perfectly align each control point with the nearest control point in the other image is computed. Each image point is then interpolated by the weighted average of the displacements of each of the control points, where the weights are inversely proportional to its distance to each control point. This is very simple; however the latter is still a global computation and hence expensive. Franke [1979] mentions several ways to make such computations local by using disk-shaped regions around each control point which specifies its area of influence. Weights are computed either as a parabolic function which decreases to zero outside the disk or using a simpler function which varies inversely with the distance relative to the disk size and decreases in a parabolic-like manner to zero outside the disk. These methods are all examples of inverse-distance weighted interpolation. They are efficient and simple, but according to Franke's [1979] study, they generally do not compare well with many of the other surface interpolation techniques such as the triangulation or finite-element-based methods. However, based on tests of accuracy, appearance, time, and storage costs conducted on six data sets, a quadratic least squares fit at each data point in conjunction with localization of the weights (using a simple function which is zero outside the disk centered at the interpolation point) was found to be one of the best methods of all.

Another registration technique proposed by Goshtasby [1988], which is also derived from the interpolation methods discussed in Franke's [1979] study is called the local weighted-mean method. Although Franke's [1979] study is very useful to compare the interpolation

methods, the application of these techniques to real registration problems exposes other types of limitations and dependencies. In the local weighted-mean method, a polynomial of order $n$ is found for each control point which fits its $n - 1$ nearest control points. A point in the registered image is then computed as the weighted mean of all these polynomials where the weights are chosen to correspond to the distance to each of the neighboring control points and to guarantee smoothness everywhere. The computational complexity of the local weighted method depends linearly on the product of the number of controls points, $P$, the square of the order of the polynomial, $M^2$, and the size of the image, $N^2$, i.e., it complexity is $O(PM^2N2)$. Again, the method relies on an entirely local computation, each polynomial is based on local information, and each point is computed using only local polynomials. Thus, the efficiency is good compared with methods whose parameters rely on global computations, but the procedure's success is limited by the accuracy and selection of the control points. In fact, during implementation, only a subset of the known control points was used so that each polynomial's influence would be spread far enough to cover image locations without points.

Notice that in comparison to the global point-mapping methods of the previous section, the complexity of local interpolation methods is vastly slower. Because the parameters of the transformation depend on the location in the image, a separate calculation, which may or may not be global, is effectively performed, for each subregion, to determine its parameters. Alternatively, when the mapping transformation is computed, a more complicated calculation must be performed for each pixel which depends on its relative location with respect to the partitioning or the other control points. This is the price we pay if we want to register images with local distortions.

Also, although these methods only use a local computation for each mapping, we have presumed that the control points have already been found and matched. This is a critical part of the registration and its final computational complexity. Furthermore, the implementation of these methods is often complicated by missing control points and insufficient information concerning how to find matches. Yet, the accuracy of these methods is highly dependent on the number, positions, and accuracy of the matches. Although they are sometimes capable of correcting local distortions, local point-mapping methods do so in a single pass; there is no feedback between the point matching and the interpolation or approximation. Nor do they take advantage of several algorithmic techniques which can improve and speed up the extraction of local distortions. Namely, these are iteration, a hierarchical approach, and cooperation. This is because these techniques (by definition) are based on control points which have been found and matched prior to the determination of the registration transformation. There is no relationship inherent in the structure of these techniques which relates the control point matches and the optimal transformation. In the next section, another class of methods is described which overcome this dependence on the accurate matching of control points by exploiting these algorithmic techniques and by the use of an elastic model to constrain the registration process.

### 3.4 Elastic Model-Based Matching

The most recent work in image registration has been the development of techniques which exploit elastic models. Instead of directly applying piecewise interpolation to compute a transformation to map the control points of one image onto another, these methods model the distortion in the image as the deformation of an elastic material. In other words, the registration transformation is the result of the deformation of an elastic material with the minimal amount of bending and stretching. The amount of bending and stretching is characterized by the energy state of the elastic mate-

rial. Nevertheless, the methods of piecewise interpolation are closely related since the energy minimization needed to satisfy the constraints of the elastic model can be solved using splines. Indeed, the forebear of the mathematical spline is the physical spline which was bent around pegs (its constraints) and assumed a shape which minimizes its strain energy.

Generally, these methods approximate the matches between images, and although they sometimes use features, they do not include a preliminary step in which features are matched. The image or object is modeled as an elastic body, and the similarity between points or features in the two images act as external forces which "stretch" the body. These are counterbalanced by stiffness or smoothness constraints which are usually parameterized to give the user some flexibility. The process is ultimately the determination of a minimum-energy state whose resulting deformation transformation defines the registration. The problems associated with finding the minimum-energy state or equilibrium usually involve iterative numerical methods.

Elastic methods, because they mimic physical deformations, register images by matching structures. Thus, it has been developed and is often used for problems in shape and motion reconstruction and medical imaging. In these domains, the critical task is to align the topological structures in image pairs removing only the differences in their details. Thus, elastic methods are capable of registering images with some of the most complex distortions, including 2D projection of 3D objects, their movements including the effects of occlusion, and the deformations of elastic objects.

One of the earliest attempts to correct for local distortions using an elastic model-based approach was called the "rubber-mask" technique [Widros 1973]. This technique was an extension of template matching for natural data and was applied to the analysis of chromosome images, chromatographic recordings, and electrocardiogram waveforms. The flexible-template technique was implemented by defining specific parameters for the possible deformations in each problem domain. For example, chromosomes had distortion parameters describing the length, width, angle, and curve for each of its four "arms," where angle and curve described how the arm was bent relative to a stereotypical chromosome. These parameters were used to iteratively modify the template until the best match was found.

However, it was not until the 1980's [Burr 1981] that automatic elastic-registration methods were developed. Burr accomplished this by an iterative technique which depends on the local neighborhood whose size is progressively smaller with each iteration. At each iteration, the distance between each edge or feature point in one image and its nearest neighbor in the second image are determined. Similarly, these distances are found starting with each edge or feature point in the second image. The images are then pulled together by a "smoothed" composite of these displacements and their neighboring displacements which are weighted by their proximity. Since after each iteration the images are closer together, the neighborhood size is decreased thus allowing for more "elastic" distortions until the two images have been matched as closely as desired. This method relies on a simple and inexpensive measure to gradually match two images which are locally distorted with respect to each other. It was applied successfully to hand-drawn characters and other images composed only of edges. For gray-scale images more costly local feature measures and their corresponding nearest neighbor displacement values needed to be computed at each iteration. Burr applied this to two images of a girl's face in which his method effectively "turned the girl's head" and "closed her mouth."

There are three aspects of this method which should be considered for any local method, i.e., a method which determines a local transformation. These techniques are particularly relevant in cases where

there is no additional knowledge to help in finding matches for control points a priori.

- Iteration: The general point-mapping method was described as a three-step procedure: (1) feature points are determined, (2) their correspondence with feature points in the second image is found, and (3) a transformation which approximates or interpolates this set of matched points is found. For iterative techniques such as this, this sequence or the latter part of it are iterated and often become intricately interrelated. In Burr's [1981] work, at each iteration step features are found, and a correspondence measure is determined which influences a transformation which is then performed before the sequence is repeated. Furthermore, the technique is dynamic in the sense that the effective interacting neighborhoods change with each iteration.
- Hierarchical Structure: Larger and more global distortions are corrected first. Then progressively smaller and more local distortions are corrected until a correspondence is found which is as finely matched as desired. Global distortions can be found as the optimal match between the two whole images perhaps at a lower resolution since for this match details are not of interest. Then, having crudely corrected for these global distortions, we find progressively more local distortions by matching smaller parts of the images at higher resolutions. This has several advantages, since for global distortions we have reduced the resolution and hence the amount of data to process, and for more local distortions, we have eliminated the more global distortions and therefore reduced the search space and the extent of the data that we need to evaluate.
- Cooperation: Features in one location influence decisions at others. This may be implemented in many ways with varying degrees of cooperation. In a typical cooperative scheme, each possible feature match is weighted by the degree to which other neighboring features agree, and then the process iterates. In this fashion, the neighboring features influence (in a nonlinear way) the determination of the best overall match by "cooperating" when they agree, or possibly, inhibiting when they disagree.

Registration techniques which use these algorithmic approaches are particularly useful for the correction of images with local distortion for basically the same reason, namely, they consider and differentiate local and global effects. Iterative updating is important for finding optimal matches that cannot be found efficiently in a single pass since distortions are locally variant but depend on neighboring distortions. Similarly, cooperation is a useful method of propagating information across the image. Most types of misregistration sources which include local geometric distortion effect the image both locally and globally. Thus hierarchical iteration is often appropriate; images misregistered by scene motion and elastic-object deformations (such as in medical or biological images) are good examples of distortions which are both local and global. Furthermore hierarchical/multiresolutional/pyramidal techniques correspond well with our intuitive approach to registration. Manual techniques to perform matching are often handled this way; images are first coarsely aligned, and then in a step-by-step procedure more detail is included. Most registration methods which correct for local distortions (except for the piecewise interpolation methods) integrate these techniques in one form or another.

One of the pioneers in elastic matching is R. Bajscy and her various collaborators [Bajscy and Broit 1982; Bajscy and Kovacic 1989; Solina and Bajscy 1990]. In their original method, developed by Broit in his Ph.D. thesis [Broit 1981], a physical model is derived from the theory of elasticity and deformation. The image is an elastic grid, theoretically an elastic membrane of a homogeneous medium, on which a field of external forces act against

a field of internal forces. The external forces cause the image to locally deform towards its most similar match while the internal forces depend on the elasticity model. From an energy minimization standpoint, this amounts to:

$$cost = deformation\ energy - similarity\ energy.$$

To find the minimum energy, a set of partial differential equations are derived whose solution is the set of displacements which register the two images. For example, to register a CT image of a human brain (the reference image) with a corresponding image from an atlas, a regular grid is placed over the reference image which can be thought of as the elastic mesh. This mesh is deformed according to the external forces, derived from the differences between the contours of the brains in the two images and the internal forces which arise from the properties of the elastic model. The deformation of this grid (and the necessary interpolation) gives a mapping between the reference image and the atlas. Bajcsy and Broit [1982] applied this technique to 2D and 3D medical images and claim greater efficiency over Burr's [1981] method. As in Burr's [1981] method iteration and cooperation are clearly utilized.

In more recent work with Bajscy and Kovacic [1989] CT scans of the human brain are elastically matched with a 3D atlas. As with many local techniques, it is necessary first to align images globally using a rigid transformation before applying elastic matching. In this way it is possible to limit the differences in the images to small, i.e., local, changes. Their work follows the earlier scheme proposed by Broit [1981], but this is extended in a hierarchical fashion. The same set of partial differential equations describing the elastic model serve as the constraint equations. The external forces, which ultimately determine the final registration, are computed as the gradient vector of a local similarity function. These forces act on the elastic grid by locally pulling it towards the maximum of the local similarity function. In particular, for each small region in one image, we measure the correlation with this region with nearby regions in the second image. The change in these measures is used as the external forces in the equations of the elastic model. This requires that the local-similarity function have a maximum that contributes unambiguous information for matching. Therefore, only forces in regions where there is a substantial maximum are used. The system of equations of the elastic model is then solved numerically by finite-difference approximation for each level, starting at the coarsest resolution. The solution at the coarsest level is interpolated and used as the first approximation to the next finer level.

The hierarchical approach has several advantages. If the elastic constants in the equation are small, the solution is controlled largely by the external forces. This causes the image to warp unrealistically and for the effects of noise to be amplified. By deforming the image step-by-step, larger elastic constants can be used, thereby producing a series of smooth deformations which guide the final transformation. The multiresolution approach also allows the neighborhoods for the similarity function to always be small and hence cheap yet also to cover both global and local deformations of various sizes. In general, the coarse-to-fine strategy improves convergence since the search for local-similarity function maxima is guided by results at coarser levels. Thus, like Burr's [1981] method, iteration, cooperation, and a hierarchical structure are exploited.

Recently, techniques similar to elastic matching have been used to recover shape and nonrigid body motion in computer vision and to make animation in computer graphics. The major difference in these techniques to the methods discussed so far is that the elastic model is applied to an object as opposed to the image grid. Hence, some sort of segmentation, i.e., a grouping of adjacent pixels in the image into meaningful units,

namely, real-world objects, must precede the analysis. The outcome is no longer a deformation to register images but parameters to match images to object models. One example can be found in Terzopoulos et al. [1987]. They proposed a system of energy constraints for elastic deformation for shape and motion recovery which was applied to a temporal sequence of stereo images of a moving finger. The external forces of the deformable model are similar to those used in elastic registration; they constrain the match based on the image data. Terzopoulos et al. [1987] use the deprojection of the gradient of occluding contours for this purpose. However, the internal forces are no longer varied with simple elastic constants but involve a more complicated model of expected object shape and motion. In their case, the internal forces induce a preference for surface continuity and axial symmetry (a sort of "loose" generalized cylinder using a rubber sheet wrapped around an elastic spine). This type of reconstruction has the advantage of being capable of integrating information in a straightforward manner. For example, although occluding boundaries in stereo image pairs correspond to different boundary curves of smooth objects, they can appropriately be represented by distinct external forces. Higher-level knowledge can similarly be incorporated. Although these techniques are not necessary for the ordinary registration of 2D images, performing intelligent segmentation of images before registration is potentially the most accurate way to match images and to expose the desired differences between them.

### 3.5 Summary

In Section 3, most of the basic registration techniques currently used have been discussed. Methods are characterized by the complexity of their corresponding transformation class. The transformation class can be determined by the source of misregistration. Methods are then limited by their applicability to this transformation class and the types of uncorrected variations they can tolerate. The early approaches using cross-correlation and other statistical measures of pointwise similarity are only applicable for small well-defined affine transformations. Fourier methods are similarly limited but can be more effective in the presence of frequency-dependent noise. If local uncorrected variations are present then the search for the affine transformation must be more sophisticated. In this case point mapping with feedback is recommended so that search space is more thoroughly investigated.

If the type of transformation is unknown but the misalignment between the images varies smoothly, i.e., the misalignment is more complex than affine but is still global, then point mapping with feedback can be used. If it is possible to find accurate matches for control points then point mapping by interpolation is sufficient. For cases with local uncorrected variation and many inaccurate control point matches then approximation is necessary.

If global transformations are not sufficient to account for the misalignment between the images, then local methods must be used. In this case, if it is possible to perform accurate feature matching, then piecewise interpolation methods can be successively applied. However, if uncorrected local variations are present, then it is often necessary to use additional knowledge to model the transformation such as an elastic membrane for modeling the possible image deformations.

## 4. CHARACTERISTICS OF REGISTRATION METHODS

The task of determining the best spatial transformation for the registration of images can be broken down into major components:

- feature space
- similarity metric
- search space
- and search strategy.

Every registration technique can be

thought of as a selection for each of these four components. As described earlier, the best available knowledge of the source of misregistration determines the transformation needed, i.e., the search space. This, in turn, determines the complexity and kind of method. Knowledge of other variations (which are not corrected for by the transformation) can then be used to decide on the best choices for the other three major components listed above. Tables 6, 7, and 8 give several examples of each of these components. In addition, these tables briefly describe the attributes for each technique and give references to works which discuss their use in more detail. In the following section, each of the components of registration is described more fully.

### 4.1 Feature Space

The first step in registering two images is to decide on the feature space to use for matching. This may be the raw pixel values, i.e., the intensities, but other common feature spaces include: edges, contours, surfaces; salient features such as corners, line intersections, and points of high curvature; statistical features such as moment invariants or centroids; and higher-level structural and syntactic descriptions. Salient features refer to specific pixels in the image which contain information indicating the presence of an easily distinguished meaningful characteristic in the scene. Statistical features refer to measures over a region (the region may be the outcome from a preprocessing segmentation step), which represent the evaluation of the region. The feature space is a fundamental aspect of image registration just as it is for almost all other high-level image processing or computer vision tasks. For image registration it influences

- which properties of the sensor and scene the data are sensitive to (often, features are chosen to reduce sensor noise or other distortions, such as illumination and atmospheric conditions, i.e., Type II variations),
- which properties of the images will be matched (e.g., more interested in matching structures than textural properties, i.e., ignore Type III variations),
- the computational cost by either reducing the cost of similarity measures or, on the other hand, increasing the precomputations necessary.

The point is, by choosing the best feature space it is possible to significantly improve registration. Features can be found on each image independently in a preprocessing step, and this in turn reduces the amount of data to be matched. It is often possible to choose a feature space which will eliminate uncorrected variations which might otherwise make matching unreliable. If there are variations of interests, the feature space can be limited to the types of structures for which these variations are not present. Similarly, features can highlight those parts of the image which represent scene elements which have undergone the expected misalignment. This is usually done by preprocessing the images in an attempt to extract intrinsic structure. By this we mean finding the pixels in the images which accurately represent significant physical locations in the world as opposed to lighting changes, shadows, or changes in reflectivity.

Extracting intrinsic structure reduces the effects of scene and sensor noise, forces matching to optimize structural similarity, and reduces the corresponding data to be matched. Image enhancement techniques which process an image to make it more suitable for a specific application [Gonzalez and Wintz 1977] can be used to emphasize structural information. Typical enhancement techniques include contrast enhancement, which increases the range of intensity values, image smoothing, which removes high-frequency noise, and image sharpening, which highlights edges. An example of an enhancement technique which is particularly suitable for registration is homomorphic filtering. This can be used to control the effects of illumination and enhance the effects of reflectance.

**Table 6.** Feature Spaces Used in Image Registration

| *Feature Spaces and Their Attributes* |
|---|
| **RAW INTENSITY**—most information |
| **EDGES**—intrinsic structure, less sensitive to noise<br>Edges [Nack 1977]<br>Contours [Medioni and Nevatia 1984]<br>Surfaces [Pelizzari et al. 1989] |
| **SALIENT FEATURES**—intrinsic structure, accurate positioning<br>Points of locally maximum curvature on contour lines [Kanal et al. 1981]<br>Centers of windows having locally maximum variances [Moravec 1981]<br>Centers of gravity of closed-boundary regions [Goshtasby 1986]<br>Line intersections [Stockman et al. 1982]<br>Fourier descriptors [Kuhl and Giardina 1982] |
| **STATISTICAL FEATURES**—use of all information, good for rigid transformations, assumptions concerning spatial scattering<br>Moment invariants [Goshtasby 1985]<br>Centroid/principal axes [Rosenfeld and Kak 1982] |
| **HIGHER-LEVEL FEATURES**—use relations and other higher-level information, good for inexact and local matching<br>Structural features: graphs of subpattern configurations [Mohr et al. 1990]<br>Syntactic features: grammars composed from patterns [Bunke and Sanfeliu 1990]<br>Semantic networks: scene regions and their relations [Faugeras and Price 1981] |
| **MATCHING AGAINST MODELS**—accurate intrinsic structure, noise in one image only<br>Anatomic atlas [Dann et al. 1989]<br>Geographic map [Maitre and Wu 1987]<br>Object model [Terzopoulos et al. 1987] |

**Table 7.** Similarity Metrics Used in Image Registration

| *Similarity Metric* | *Advantages* |
|---|---|
| Normalized cross-correlation function [Rosenfeld and Kak 1982] | accurate for white noise but not tolerant of local distortions, sharp peak in correlation space difficult to find |
| Correlation coefficient [Svedlow et al. 1976] | similar to above but has absolute measure |
| Statistical correlation and matched filters [Pratt 1978] | if noise can be modeled |
| Phase-correlation [De Castro and Morandi 1987] | tolerant of frequency-dependent noise |
| Sum of absolute differences of intensity [Barnea and Silverman 1972] | efficient computation, good for finding matches with no local distortions |
| Sum of absolute differences of contours [Barrow et al. 1977] | can be efficiently computed using "chamfer" matching, more robust against local distortions—not as sharply peaked |
| Contour/surface differences [Pelizzari et al. 1989] | for structural registration |
| Number of sign changes in pointwise intensity difference [Venot et al. 1989] | good for dissimilar images |
| Higher-level metrics: structural matching: tree and graph distances [Mohr et al. 1990], syntactic matching: automata [Bunke and Sanfeliu 1990] | optimizes match based on features or relations of interest |

**Table 8.** Search Strategies Used in Image Registration

| *Search Strategy* | *Advantages and Reference Examples* |
|---|---|
| Decision Sequencing | Improved efficiency for similarity optimization for rigid transformations [Barnea and Silverman 1972] |
| Relaxation | Practical approach to find global transformations when local distortions are present, exploits spatial relations between features [Hummel and Zucker 1983; Price 1985; Ranade and Rosenfeld 1980; Shapiro and Haralick 1990] |
| Dynamic Programming | Good efficiency for finding local transformations when an intrinsic ordering for matching is present [Guilloux 1986; Maitre and Wu 1987; Milios 1989; Ohta et al. 1987] |
| Generalized Hough Transform | For shape matching of rigidly displaced contours by mapping edge space into "dual-parameter" space [Ballard 1981; Davis 1982] |
| Linear Programming | For solving system of linear inequality constraints, used for finding rigid transformation for point matching with polygon-shaped error bounds at each point [Baird 1984] |
| Hierarchical Techniques | Applicable to improve and speed up many different approaches by guiding search through progressively finer resolutions [Bajscy and Kovacic 1989; Bieszk and Fram 1987; Davis 1982; Paar and Kropatsch 1990] |
| Tree and Graph Matching | Uses tree/graph properties to minimize search, good for inexact and matching of higher-level structures [Gmur and Bunke 1990; Sanfeliu 1990] |

Edges, contours, and boundaries, because they represent much of the intrinsic structures of an image, are frequently used as a feature space. Using the position of edges in registration has the advantages of being fast and invariant to many types of uncorrected variations. However, edge points are not typically distinguishable and therefore are not good candidates for point matching. In general using edges requires a region-based similarity measure.

Salient features are chosen to be invariant to uncorrected variations and to be highly distinguishable. Dominant points along curves are frequently used such as corners, intersections, inflection points, points of high curvature, and points along discontinuities [Katuri 1991]. Higher-level shape descriptors, such as topological, morphological, and Fourier descriptors are also used in order to be more unique and discriminating [Pavlidis 1978]. In the absence of shape or curves interesting points in regions are found. The most widely used measure of this sort is the Moravec interest operator which finds points of greatest local variance [Moravec 1981].

Statistical measures describe characteristics of regions which may or may not specify a location in the image. One possibility is to assume objects are ellipsoid-like scatters of particles uniformly distributed in space. In this case, the centers of mass and the corresponding principal axes (computed from their covariance matrices) can be used to globally register them. Another popular choice is to use moment invariants although they are computationally costly (lower-order moments are sometimes used first to guide the match and speed the process [Goshtasby 1985; Mahs and Rezaie 1987]) and can only be used to match images which have been rigidly transformed. They are one member of the class of features used because their values are independent of the coordinate system. However, as scalars they have no spatial meaning. Matching is accomplished by maximizing the similarity between the values of the moments in the two images. Mitchie and Aggarwal [1983]

suggest the use of *shape-specific points*, such as the centroid and the radius-weighted mean, for preregistration to simplify shape matching. These features are more easily computed, are similarly noise tolerant, but more importantly, they are spatially meaningful. They can be used as control points in point-mapping registration methods rather than in similarity optimization.

When sufficient information or data are available, it is useful to apply registration to an atlas, map, graph, or model instead of between two data images. In this way distortion is present in only one image, and the intrinsic structures of interest are accurately extracted.

The feature space is the representation of the data that will be used for registration. The choice of feature space determines what is matched. The similarity metric determines how matches are rated. Together the feature space and similarity metric can ignore many types of variations which are not relevant to the proper registration (Types II and III) and optimize matching for features which are important. But, while the feature space is precomputed on each image before matching, the similarity metric is computed using both images and for each test.

### 4.2 Similarity Measure

The second step made in designing or choosing a registration method is the selection of a similarity measure. This step is closely related with the selection of the matching feature since it measures the similarity between these features. The intrinsic structure, i.e., the invariance properties of the image, are extracted by both the feature space, and through the similarity measure. Typical similarity measures are cross-correlation with or without prefiltering (e.g., matched filters or statistical correlation), sum of absolute differences (for better efficiency), and Fourier invariance properties such as phase correlation. Using curves and surfaces as a feature space requires measures such as sum of squares of differences between nearest points. Structural or syntactic methods have measures highly dependent on their properties. For example, the minimum change of entropy between "random" graphs is used as a similarity criteria by Wong and You [1985] for noisy data in structural pattern recognition.

The choice of similarity metric is one of the most important elements of how the registration transformation is determined. Given the search space of possible transformations, the similarity metric may be used to find the parameters of the final registration transformation. For cross-correlation or the sum of the absolute differences the transformation is found at the peak value. Similarly, the peak value determines the best control point match for point-mapping methods. Then the set of control point matches is used to find the appropriate transformation. However, in elastic-model-based methods, the transformation is found for which the highest similarity is balanced with an acceptable level of elastic stress.

Similarity measures, like feature spaces, determine what is being matched and what is not. First the feature space extracts the information from each image which will be used for matching. Then the similarity measure evaluates this information from both images. The criteria used by the similarity measure determines what types of matches are optimal. The ability of a registration method to ignore uncorrected variations ultimately depends on both the feature space and the similarity measure. If gray values are used, instead of features, a similarity measure might be selected to be more noise tolerant since this was not done during feature detection. Correlation and its sequential counterpart are optimized for exact matches therefore requiring image preprocessing if too much noise is present. Edge correlation, i.e., correlation of edge images, is a standard approach. Fourier methods, such as phase correlation, can be used on raw images when there is frequency-dependent noise. Another possible similarity measure, suggested by Venot et al. [1984],

is based on the number of sign changes in the pointwise subtraction of the two images. If the images are aligned and noise is present, the number of sign changes is high, assuming any point is equally likely to be above zero as it is to be below. This is most advantageous in comparison to classical techniques when the images are dissimilar. Differences in the images affect the classical measures according to the gray values in the locations which differ whereas the number of sign changes decreases only by the spatial size of these differences.

The feature space and similarity metric, as discussed, can be selected to reduce the effects of noise on registration. However, if the noise is extracted in the feature space this is performed in a single step precomputed independently on each image prior to matching. Special care must be taken so that image features represent the same structures in both images when, for example, images are acquired from different sensors. On the other hand, the proper selection of a feature space can greatly reduce the search space for subsequent calculations. Because similarity measurements use both images and are computed for each transformation, it is possible to choose similarity measures which increase the desirability of matches even though distortions exist between the two correctly registered images. The method based on the number of sign differences described above is an example. Similarity metrics have the advantage that both images are used and that its measurements are relative to the measurements at other transformations. Of course, this is paid for by an increase in computational cost since it must be repeated for each test.

Lastly, using features reduces the effects of photometric noise but has little effect on spatial distortions. Similarity measures can reduce both types of distortions such as with the use of region-based correlation and other local metrics. It is important to realize, however, that the spatial distortions purposely not recognized by similarity metrics must only be those that are not part of the needed transformation. For example, when similarity metrics are chosen for finding the elastic transformation of images in which certain differences between images are of interest (such as those in the examples in the second class of problems of Table 2) they should find similarity in structure but not in more random local differences.

### 4.3 Search Space and Strategy

Because of the large computational costs associated with many of the matching features and similarity measures, the last step in the design of a registration method is to select the best search strategy. Remember, the search space is generally the class of transformations from which we would like to find the optimal transformation to align the images. We can evaluate each transformation candidate using the similarity measure on the preselected features. However, in many cases, such as with the use of correlation as the similarity measure, it is important to reduce the number of measures to be computed. The greater the complexity of the misalignment between images the more severe this requirement is. For instance, if the only misalignment is translation, a single template correlated at all possible shifts is sufficient. For more general affine transformations, many templates or a larger search area must be used for classical correlation methods. The problem gets even worse if local geometric distortions are present. Finally, if uncorrected variations have not been eliminated by the feature space and similarity metric, then the search for the optimum is also made more difficult, since there are more likely to be several local optima and a less monotonic space.

In most cases, the search space is the space of all possible transformations. Examples of common search strategies include hierarchical or multiresolution techniques, decision sequencing, relaxation, generalized Hough transforms, linear programming, tree and graph matching, dynamic programming, and heuristic search.

**Search Space.** The model of the transformation class underlying each registration technique determines the

characteristics of the search space. The model includes assumptions about the distortions and other variations present in the images. For example, if it is assumed that to register a pair of images, a translation must be performed, then the search space is the set of all translations over the range of reasonable distances. However, if after translation, it is assumed that uncorrected variations are still present (perhaps there are differences in local geometry which are of interest such as in aerial photographs taken at different times) then traversing the search space is made more difficult since determining the relative merit of each translation is more involved.

Models can be classified as allowing either global or local transformations since this directly influences the size and complexity of the search space. Global methods are typically either a search for the allowable transformation which maximizes some similarity metric or a search for the parameters of the transformation, typically a low-order polynomial which fit matched control points. By using matched control points the search costs can be significantly reduced while allowing more general transformations. In local methods, such as piecewise interpolation or elastic-model-based methods, the models become more comlex, introducing more constraints than just similarity measures. In turn they allow the most general transformations, i.e., with the greatest number of degrees of freedom. Consequently, local methods have the largest and most complex search spaces, often requiring the solution to large systems of equations.

Although most registration methods search the space of allowable transformations, other types of searches may be advantageous when other information is available. When the source of misregistration is known to be perspective distortion, Barrow et al. [1977] and Kiremedjian [1987] search the parameter space of a sensor model to map an image to a three-dimensional database. For each set of sensor parameters, the 3D database is projected onto the image, and its similarity is measured. This search space exploits knowledge of the imaging process and its effects on three-dimensional structures. Another example of very different search space is given by Mort and Srinath [1988]. He uses a stochastic model of the noise in the image to search, probabilistically, for the maximum likelihood image registration in images which have been displaced relative to each other.

**Search Strategies.** Table 8 gives several examples of search strategies and the kinds of problems for which they are used. Alternatively, specialized architectures have been designed to speed up the performance of certain registration methods. Fu and Ichikawa [1982] contains several examples of computer architectures designed for registration problems in pattern processing.

It is difficult to give a taxonomy of search strategies; each strategy has its advantages and disadvantages; some have limited domains; some can be used concurrently with others, and all of them have a wide range of variations within them. In large part, the choice of search strategy is determined by the characteristics of the search space including the form of the transformation (what type of constraints must we satisfy?) and how hard it is to find the optimum. For example, if we must satisfy linear inequalities then linear programming is advisable. If image features are composed of trees or graphs to be matched then we need search strategies which are specialized for these data structures. The Generalized Hough Transform was developed specifically for matching shapes from contours. Some things to consider are: how does the strategy deal with missing information; can the strategy be implemented in parallel; does the strategy make any assumptions, and what are the typical computational and storage costs?

For this discussion, two of the most frequently used search strategies have been chosen to exemplify the kinds of strategies used in registration: relaxation and dynamic programming. These strategies have been applied in a variety of different tasks, in a variety of different ways. They were chosen to illustrate how

the proper choice of a search strategy can make a significant difference in the ability to register certain types of images. Relaxation matching is most often used in the case where a global transformation is needed, but local distortion is present. If local distortion is not present, global transformations can typically be determined by the more standard hill-climbing or decision-sequencing techniques (see Section 3.1) to find maxima and by linear equations or regression to fit polynomials (see Section 3.3.3). Dynamic programming, on the other hand, is used to register images where a local transformation is needed. For dynamic programming the ordering properties of the problem are exploited to reduce the searching computations. Other search strategies used for local methods depend largely on the specific model used, such as the use of iterative methods for discretely solving a set of partial differential equations [Bajscy and Kovacic 1989], linear programming for solving point matching with polygonal-shaped point errors [Baird 1984], generalized Hough transforms for shape matching [Ballard 1981].

**Relaxation Matching.** Relaxation gets its name from the iterative numerical methods which it resembles. It is a bottom-up search strategy that involves local ratings (of similarity) which depend on the ratings of their neighbors. These ratings are updated iteratively until the ratings converge or until a sufficiently good match is found. It is usually used in registration to find a global maximum to a similarity criteria for rigid transformations.[2]

Several researchers have investigated the use of relaxation matching as a search strategy for registration [Hummel and Zucker 1983; Ranade and Rosenfeld 1980]. The advantage of this method lies in its ability to tolerate local geometric distortions. This is accomplished by the use of local-similarity measures. The local-similarity measures are used to assign heuristic, fuzzy, or probabilistic ratings for each location. These ratings are then iteratively strengthened or weakened, potentially in parallel, in accordance with the ratings of the neighboring measures. Although, the convergence and complexity of this approach are not always well defined, in practice it is often a good short cut over more rigorous techniques such as linear programming.

Relaxation-matching techniques have been compared by Price [1985] for the matching of regions of correspondence between two scenes. Relaxation is a preferred technique in scene matching as opposed to point matching since local distortions need to be tolerated. In their study, objects and their relations are represented symbolically as feature values and links in a semantic network. An automatic segmentation is performed to find homogeneous regions from which a few semantically relevant objects are interactively selected. Feature values of objects alone are inadequate for correctly matching objects. They require contextual information which is gradually determined by the relaxation process. The rate assignments (or probabilities) are iteratively updated based on an optimizing criteria that evaluates the compatibility of the current assignments with the assignments of their neighbors in the graph (i.e., objects linked by relations). Four relaxation techniques were compared with varying optimization criteria and updating schemes. The same general matching system is used, i.e., the same feature space and local similarity measure. Complexity and convergence are measured empirically on several aerial test images.

Price's [1985] study is representative of the studies undertaken to compare search strategies for registration problems. Relaxation is not compared with other strategies here, nor is its selection for this problem explicitly justified. It is empirically compared on aerial photographs, and thus their results cannot necessarily be generalized. One of their

[2] The related technique, called relaxation labeling, refers to the use of relaxation in the problem of assigning labels consistently to objects in a scene.

primary contributions is their description of the relative merits of the four methods. Although this would of course be useful for future work where relaxation is applied to similar problems, the larger questions of whether to apply relaxation or some other search strategy for a given problem remain unanswered.

**Dynamic Programming.** Another commonly used search strategy for image registration is dynamic programming (DP). DP is an algorithmic approach to solving problems by effectively using the solutions to subproblems. Progressively larger problems are solved by using the best solutions to subproblems thus avoiding redundant calculations and pruning the search. This strategy can only be applied when an intrinsic ordering of the data/problem exists. Several examples in which it has been applied include: signature verification [Parizeau and Plamondon 1990], the registration of geographic contours with maps [Maitre and Yu 1987], shape matching [Milios 1989], stereomapping [Ohta et al. 1987], and horizontal-motion tracking [Guilloux 1986]. Notice that in each of these examples, the data can be expressed in a linear ordering. In the shape-matching example this was done using a cyclic sequence of the convex and concave segments of contours for each shape. In stereomapping, the two images were rectified so that their scanlines were parallel to the baseline (the line connecting to the two viewpoints). Then, the scanlines become the epipolar lines, so that all the corresponding matches for points in the scanline on one image lie in the corresponding scanline of the other image. Similarly in horizontal-motion tracking, scanlines are the ordered data sets to be matched. In each of these instances, dynamic programming is used to find the correspondence between the points in the two images, i.e., the segments in the shape-matching example and feature points in the stereo or motion example.

Notice also, that the matching to be done in these problems is from many-to-many. The problem is often posed as a search for the optimal (lowest cost) path which matches each point along the ordering (scanline or contour etc.) of one image with a point along the ordering of the other image. The resulting search space is therefore very large, exponential to be precise. DP reduces this to $O(n^3)$ where $n$ is the length of the longest ordering. In practice, the cost is reduced by limiting the matches to an interval size which reflects the largest expected disparity between images. The cost of the algorithm is also proportional to the cost of the similarity measure which is the elementary cost operation which is minimized recursively. Typical measures include the absolute difference between pixel intensities or their first-order statistics. Similarity metrics often have additional factors which depend on the application in order to optimize other characteristics such as minimal path length, minimal disparity size, and interval uniformity. As a search strategy, DP offers an efficient scheme for matching images whose distortions are nonlinear including noisy features and missing matches (such as occlusions) but which can be constrained by an ordering.

### 4.4 Summary

This survey has offered a taxonomy of existing registration techniques and a framework to aid in the selection of the appropriate technique for a specific problem. Knowledge of the causes of distortions present in images to be registered should be used as much as possible in designing or selecting a method for a particular application. Distortions which are the source of misregistration can be used to decide on the class of transformations which will optimally map the images onto each other. The class of transformations and its complexity determine the general type of method to be used. Given the class of transformation, i.e., the search space, the types of variations that remain uncorrected by this transformation can be used to further specify the most suitable method.

Affine transformations can be found by Fourier methods and techniques related

to cross-correlation. When local uncorrected variations are present then point mapping with feedback is used. Polynomial transformations are generally determined by point-mapping techniques using either interpolation or approximation methods. Local transformations are either determined with piecewise interpolation techniques when matched control points can be accurately found or with model-based approaches exploiting knowledge of the possible distortions. The technique is completely specified by selecting a particular feature space, similarity metric, search space, and search strategy from the types of methods available for registration. The choices for the feature space, similarity metric, and search strategy for a registration method depend on the uncorrected variations, spatial and valumetric, which obscure the true registration.

Selecting a feature space instead of matching on the raw intensities can be advantageous when complex distortions are present. Typically, the feature space attempts to extract the intrinsic structures in the image. For small computational costs the search space is greatly reduced, and irrelevant information is removed.

The similarity metric defines the test to be made for each possible match. For white noise, cross-correlation is robust; for frequency-dependent noise due to illumination or changes in sensors, similarity metrics based on the invariant properties of the Fourier Transform are good candidates. If features are used, efficient similarity metrics which measure the spatial differences between the locations of the features in each image are available. Other measures specialize in matching higher-level structures such as graphs or grammars.

The search space and strategy also exploit the knowledge available concerning the source of distortion. Assumptions about the imaging system and scene properties can be used to determine the set of possible or most probable transformations to guide the search for the best transformation.

The most difficult registration problems occur when local variations are present. This can happen even when it is known that a global transformation is sufficient to align the two images. Feedback between feature detection, similarity measurements, and computing the optimal transformation can be used to overcome many of these problems. Iteration, cooperation, and hierarchical structures can be used to improve and speed up registration when local distortions are present by using global information without the computational and memory costs associated with global image operations. The distinctions bewteen global and local registration transformations and methods, global and local distortions, and global and local computations should be carefully considered when designing or choosing techniques for given applications.

Over the years, techniques to perform registration have become increasingly automatic, efficient, and robust. Current research efforts have begun to address the more difficult problems in which local variations, both correctable and not, are present. The need for a taxonomy of these techniques has arisen so that these methods can be properly applied, their capabilities can be more quickly assessed, and comparisons among techniques can be performed. This paper has provided this taxonomy based on the types of variations in the images. The distinctions between corrected and uncorrected, spatial and valumetric, and local and global variations have been used to develop a framework for registration methods and its four components: feature space, similarity metric, search space, and search strategy. This framework should be useful in the future evaluation, development, and practice of registration.

## ACKNOWLEDGMENTS

I would like to thank my advisor, T. Boult, for many valuable discussions. This work was supported in part by DARPA contract N00039-84-C-0165 and in part by NSF PYI award IRI-90-57951,

and with additional support from Siemens and AT & T.

## REFERENCES


ALLINEY, S., AND MORANDI, C. 1986. Digital image registration using projections. *IEEE Trans. Pattern Anal. Machine Intell. PAMI-8*, 2 (Mar.), 222–233.

BAJSCY, R., AND BROIT, C. 1982 Matching of deformed images. In *The 6th International Conference on Pattern Recognition*. pp. 351–353.

BAJSCY, R., AND KOVACIC, S. 1989. Multiresolution elastic matching. *Comput. Vision Graph. Image Process. 46*, 1–21.

BAIRD, H. S. 1984. Model-based image matching using location. In *An ACM Distinguished Dissertation*. MIT Press, Cambridge, Mass.

BALLARD, D. H. 1981. Generalizing the Hough Transform to detect arbitrary shapes. *Patt. Recog. 13*, 2, 111–122.

BARNARD, S. T., AND FISCHLER, M. A. 1982. Computational stereo. *ACM Comput. Surv. 14*, 4 (Dec.), 553–572.

BARNEA, D. I., AND SILVERMAN, H. F. 1972. A class of algorithms for fast digital registration. *IEEE Trans. Comput. C-21*, 179–186.

BARROW, H. G., TENENBAUM, J. M., BOLLES, R. C., AND WOLF, H. C. 1977. Parametric correspondence and chamfer matching: Two new techniques for image matching. In *Proceedings of the International Joint Conference in Artificial Intelligence*. pp. 659–663.

BERGSTRÖM, M., ÉTHIUS, B. J., ERIKSSON, L., GREITZ, T., RIBBE, T., AND WIDÉN, L. 1981. Head fixation device for reproducible position alignment in transmission CT and positron emission tomography. *J. Comput. Assisted Tomogr. 5*, (Feb.), 136–141.

BERNSTEIN, R. 1976. Digital image processing of Earth observation sensor data. *IBM J. Res. Devel. 20*, (Jan.), 40–67.

BERNSTEIN, R., AND SILVERMAN, H. 1971. Digital techniques for Earth resource image data processing. In *Proceedings of the American Institute of Aeronautics and Astronautics 8th Annual Meeting*, vol. 21. AIAA.

BIESZK, J. A., AND FRAM, I. 1987. Automatic elastic image registration. In *Proceedings of Computers in Cardiology* (Leuven, Belgium, Sept.). pp. 3–5.

BOHM, C., AND GREITZ, T. 1989. The construction of a functional brain atlas—Elimination of bias from anatomical variations at PET by reforming 3-D data into a standardized anatomy. In *Visualization of Brain Functions*, D. Ottoson and W. Rostène, Eds. Wenner-Gren International Symposium Series, vol. 53. pp. 137–140.

BOHM, C., GREITZ, T., KINGSLEY, D., BERGGREN, B. M., AND OLSSON, L. 1983. Adjustable computerized stereotaxic brain atlas for transmission and emission tomography. *Amer. J. Neuroradiol. 4*, (Mar.), 731–733.

BRESLER, Y., AND MERHAV, S. J. 1987. Recursive image registration with application to motion estimation. *IEEE Trans. Acoust. Speech Signal Proc. ASSP-35*, 1 (Jan.), 70–85.

BROIT, C. 1981. Optimal registrations of deformed images. Ph.D. Dissertation, Univ. of Pennsylvania.

BUNKE, H., AND SANFELIU, A., EDS. 1990. *Syntactic and Structural Pattern Recognition, Theory and Applications*. World Scientific, Teaneck, N.J.

BURR, D. J. 1981. A dynamic model for image registration. *Comput. Graphics Image Proc. 15*, 102–112.

CLOUGH, R. W., AND TOCHER, J. L. 1965. Finite element stiffness matrices for analysis of plates in bending. In *Proceedings of the Conference on Matrix Methods in Structural Mechanics* (Wright-Patterson A.F.B., Ohio). pp. 515–545.

DANN, R., HOFORD, J., KOVACIC, S., REIVICH, M., AND BAJCSY, R. 1989. Evaluation of elastic matching system for anatomic (CT, MR) and functional (PET) cerebral images, *J. Comput. Assisted Tomogr. 13*, (July/Aug.), 603–611.

DAVIS, L. S. 1982. Hierarchical Generalized Hough Transform and line segment based Generalized Hough Transforms. *Patt. Recog. 15*, 277–285.

DE CASTRO, E., AND MORANDI, C. 1987. Registration of translated and rotated images using finite Fourier Transforms. *IEEE Trans. Patt. Anal. Machine Intell. PAMI-9*, 5 (Sept.), 700–703.

DEGUCHI, K., 1986. Registration techniques for partially covered image sequence. In *Proceedings of the 8th International Conference on Pattern Recognition* (Paris, Oct.). IEEE, New York, pp. 1186–1189.

DENGLER, J. 1986. Local motion estimation with the dynamic pyramid. In *The 8th International Conference on Pattern Recognition* (Paris). pp. 1289–1292.

DHOND, U. R., AND AGGARWAL, J. K. 1989. Structure from stereo—A review. *IEEE Trans. Syst. Man Cybernetics 19*, 6 (Nov./Dec.), 1489–1510.

DUDA, R. O., AND HART, P. E. 1973. *Pattern Classification and Scene Analysis*. John Wiley & Sons, New York.

EVANS, A. C., BEIL, C., MARRETT, S., THOMPSON, C. J., AND HAKIM, A. 1988. Anatomical-functional correlation using an adjustable MRI-based region of interest atlas with positron emission tomography. *J. Cerebral Blood Flow Metabol. 8*, 513–530.

FAUGERAS, O., AND PRICE, K. 1981. Semantic description of aerial images using stochastic labeling. *IEEE Trans. Patt. Anal. Machine Intell. PAMI-3*, (Nov.), 638–642.

FLUSSER, J. 1992. An adaptive method for image registration. *Patt. Recog. 25*, 1, 45–54.

FORSYTHE, G. E., MALCOLM, M. A., AND MOLER, C. B. 1977. *Computer Methods for Mathematical Computations*. Prentice-Hall, Englewood Cliffs, N.J.

FOX, P. T., PERLMUTTER, J. S., AND RAICHLE, M. E. 1985. Stereotactic method of anatomical localization for positron emission tomography. *J. Comput. Assisted Tomogr. 9*, 141–153.

FRANKE, R. 1979. A critical comparison of some methods for interpolation of scattered data. Tech. Rep. NPS-53-79-003, Naval Postgraduate School.

FU, K. S., AND ICKIKAWA, T. 1982. *Special Computer Architectures for Pattern Processing*. CRC Press, Boca Raton, Fla.

GERLOT, P., AND BIZAIS, Y. 1987. Image registration: A review and a strategy for medical applications. In *Proceedings of the 10th International Conference on Information Processing in Medical Imaging* (Utretcht, Netherlands). pp. 81–89.

GMUR, E., AND BUNKE, H. 1990. 3-D object recognition based on subgraph matching in polynomial time. In *Structural Pattern Analysis*. World Scientific, Teaneck, N.J.

GONZALEZ, R. C., AND WINTZ, P. 1977. *Digital Image Processing*. Addison-Wesley, Reading, Mass.

GOSHTASBY, A. 1988. Image registration by local approximation. *Image Vision Comput. 6*, 4 (Nov.), 255–261.

GOSHTASBY, A. 1987. Piecewise cubic mapping functions for image registration. *Patt. Recog. 20*, 5, 525–533.

GOSHTASBY, A. 1986. Piecewise linear mapping functions for image registration. *Patt. Recog. 19*, 6, 459–466

GOSHTASBY, A. 1985. Template matching in rotated images. *IEEE Trans. Patt. Anal. Machine Intell. 7*, 3 (May), 338–344.

GOSHTASBY, A., AND STOCKMAN, G. C. 1985. Point pattern matching using convex hull edges. *IEEE Trans. Syst. Man Cybernetics SMC-15*, 5 (Sept./Oct.), 631–637.

GOSHTASBY, A., STOCKMAN, G. C., AND PAGE, C. V. 1986. A region-based approach to digital image registration with subpixel accuracy. *IEEE Trans. Geosci. Remote Sensing 24*, 3, 390–399

GUILLOUX, Y. LE 1986. A matching algorithm for horizontal motion, application to tracking. In *Proceedings of the 8th International IEEE Conference on Pattern Recognition* (Paris, Oct.). IEEE, New York, pp. 1190–1192.

HALL, E. L. 1979 *Computer Image Processing and Recognition*. Academic Press, New York.

HARALICK, R. M. 1979. Automatic remote sensor image registration. In *Topics in Applied Physics*. Vol. 11, *Digital Picture Analysis*, A. Rosenfeld, Ed. Springer-Verlag, New York, pp. 5–63.

HERBIN, M., VENOT, A., DEVAUX, J. Y., WALTER, E., LEBRUCHEC, J. F., DUBERTRET, L., AND ROUCAYROL, J. C. 1989. Automated registration of dissimilar images: Application to medical imagery. *Comput. Vision Graph Image Process. 47*, 77–88.

HORN, B. K. P. 1989. *Robot Vision*. MIT Press, Cambridge, Mass.

HUMMEL, R., AND ZUCKER, S 1983. On the foundations of relaxation labeling processes. *IEEE Trans. Patt. Anal. Machine Intell. 5*, (May), 267–287.

JENSEN, J. R. 1986. *Introductory Digital Image Processing, A Remote Sensing Perspective*. Prentice-Hall, Englewood Cliffs, N.J

KANAL, L. N., LAMBIRD, B. A., LAVINE, D., AND STOCKMAN, G. C. 1981. Digital registration of images from similar and dissimilar sensors. In *Proceedings of the International Conference on Cybernetics and Society*. pp. 347–351.

KATURI, R., AND JAIN, R. C 1991. *Computer Vision: Principles*. IEEE Computer Society Press, Los Alamitos, Calif.

KIREMIDJIAN, G. 1987. Issues in image registration. In *IEEE Proceedings of SPIE: Image Understanding and the Man-Machine Interface*, vol. 758. IEEE, New York, pp 80–87

KUGLIN, C. D., AND HINES, D. C. 1975. The phase correlation image alignment method. In *Proceedings of the IEEE 1975 International Conference on Cybernetics and Society* (Sept.). IEEE, New York, pp. 163–165.

KUHL, F. P., AND GIARDINA, C. R. 1982. Elliptic Fourier features of a closed contour. *Comput. Graph. Image Process. 18*, 236–258.

LEE, D. J., KRILE, T. F., AND MITRA, S. 1987. Digital registration techniques for sequential Fundus images. In *IEEE Proceedings of SPIE: Applications of Digital Image Processing X*, vol. 829. IEEE, New York, pp. 293–300.

MAGHSOODI, R., AND REZAIE, B. 1987. Image registration using a fast adaptive algorithm. In *IEEE Proceedings of SPIE: Methods of Handling and Processing Imagery*, vol. 757. IEEE, New York, pp. 58–63.

MAGUIRE, G. Q., JR., NOZ, M. E., LEE, E. M., AND SHIMPF, J. H. 1985. Correlation methods for tomographic images using two and three dimensional techniques. In *Proceedings of the 9th Conference of Information Processing in Medical Imaging* (Washington, D.C., June 10–14). pp 266–279.

MAGUIRE, G. Q., JR., NOZ, M. E., AND RUSINEK, H. 1990. Software tools to standardize and automate the correlation of images with and between diagnostic modalities. *IEEE Comput. Graph. Appl.*

MAITRE, H., AND WU, Y. 1987. Improving dynamic programming to solve image registration. *Patt. Recog. 20*, 4, 443–462.

MEDIONI, G., AND NEVATIA, R 1984. Matching im-

MEDIONI, G., AND NEVATIA, R. 1984. Matching images using linear features. *IEEE Trans. Patt. Anal. Machine Intell. PAMI-6*, 675–685.

MERICKEL, M. 1988. 3D reconstruction: The registration problem. *Comput. Vision Graph. Image Process. 42*, 2, 206–219.

MILIOS, E. E. 1989. Shape matching using curvature processes. *Comput. Vision Graph. Image Process. 47*, 203–226.

MITICHE, A., AND AGGARWAL, J. K. 1983. Contour registration by shape-specific points for shape matching. *Comput. Vision Graph. Image Process. 22*, 296–408.

MOHR, R., PAVLIDIS, T., AND SANFELIU, A. 1990. *Structural Pattern Analysis*. World Scientific, Teaneck, N.J.

MORAVEC, H. 1981. Rover visual obstacle avoidance. In *Proceedings of the 7th International Conference on Artificial Intelligence* (Vancouver, B.C., Canada, Aug.). pp. 785–790.

MORT, M. S., AND SRINATH, M. D. 1988. Maximum likelihood image registration with subpixel accuracy. In *IEEE Proceedings of SPIE: Applications of Digital Image Processing*, vol. 974. IEEE, New York, pp. 38–43.

NACK, M. L. 1977. Rectification and registration of digital images and the effect of cloud detection. In *Proceedings of Machine Processing of Remotely Sensed Data*. pp. 12–23.

NAHMIAS, C., AND GARNETT, E. S. 1986. Correlation between CT, NMR and PT findings in the brain. *NATO ASI Series*. Vol. F19. *Pictorial Information Systems in Medicine*, K. H. Hohne, Ed. Springer-Verlag, Berlin, pp. 507–514.

NOZ, M. E., AND MAGUIRE, G. Q., JR. 1988. QSH: A minimal but highly portable image display and processing Toolkit. *Comput. Methods Program. Biomed. 27*, 11 (Nov. 1988), 229–240.

OHTA, Y., TAKANO, K., AND IKEDA, K. 1987. A high-speed stereo matching system based on dynamic programming. In *Proceedings of the International Conference in Computer Vision* (London, England). pp. 335–342.

PAAR, G., AND KROPATSCH, W. G. 1990. Hierarchical cooperation between numerical and symbolic image representations. In *Structural Pattern Analysis*. World Scientific, Teaneck, N.J.

PARIZEAU, M., AND PLAMONDON, R. 1990. A comparative analysis of regional correlation, dynamic time warping, and skeletal tree matching for signature verification. *IEEE Trans. Patt. Anal. Machine Intell. 12*, 7 (July), 710–717.

PAVLIDIS, T. 1978. Survey: A review of algorithms for shape analysis. *Comput. Graph. Image Process. 7*, 243–258.

PRATT, W. K. 1978. *Digital Image Processing*. John Wiley & Sons, New York.

PELI, E. ET AL. 1987. Feature-based registration of retinal images. *IEEE Trans. Med. Imaging MI-6*, 3 (Sept.), 272–278.

PELIZARRI, C. A., CHEN, G. T. Y., SPELBRING, D. R., WEICHSELBAUM, R. R., AND CHEN, C. T. 1989. Accurate three-dimensional registration of CT, PET and/or MR images of the brain. *J. Comput. Assisted Tomogr. 13*, (Jan./Feb.), 20–26.

PRICE, K. E. 1985. Relaxation matching techniques—A comparison. *IEEE Trans. Patt. Anal. Machine Intell. 7*, 5 (Sept.), 617–623.

RANADE, S., AND ROSENFELD, A. 1980. Point pattern matching by relaxation. *Patt. Recog. 12*, 269–275.

RATIB, O., BIDAUT, L., SCHELBERT, H. R., AND PHELPS, M. E. 1988. A new technique for elastic registration of tomographic images. In *IEEE Proceedings of SPIE: Medical Imaging II*, vol. 914. IEEE, New York, pp. 452–455.

ROSENFELD, A., AND KAK, A. C. 1982. *Digital Picture Processing*. Vol. I and II. Academic Press, Orlando, Fla.

SANFELIU, A. 1990. Matching complex structures: The cyclic-tree representation scheme. In *Structural Pattern Analysis*. World Scientific, Teaneck, N.J.

SHAPIRO, L. G., AND HARALICK, R. M. 1990. Matching relational structures using discrete relaxation. In *Syntactic and Structural Pattern Recognition, Theory and Applications*. World Scientific, Teaneck, N.J.

SINGH, M., FREI, W., SHIBATA, T., HUTH, G. C., AND TELFER, N. E. 1979. A digital technique for accurate change detection in nuclear medical images—With application to myocardial perfusion studies using Thallium-201. *IEEE Trans. Nuclear Sci. NS-26*, 1 (Feb.).

SLAME, C. C., ED. 1980. *Manual of Photogrammetry*, 4th ed. American Society of Photogrammetry, Falls Church, Va.

SOLINA, F., AND BAJCSY, R. 1990. Recovery of parametric models from range images: The case for superquadrics with global deformations. *IEEE Trans. Patt. Anal. Machine Intell. 12*, 2 (Feb.), 131–147.

STOCKMAN, G. C., KOPSTEIN, S., AND BENETT, S. 1982. Matching images to models for registration and object detection via clustering. *IEEE Trans. Patt. Anal. Machine Intell. 4*, 229–241.

STYTZ, M. R., FRIEDER, G., AND FRIEDER, O. 1991. Three-dimensional medical imaging: Algorithms and computer systems. *ACM Comput. Surv. 23*, 4 (Dec.), 421–424.

SUETENS, P., FUA, P., HANSON, A. J. 1992. Computational strategies for object recognition. *ACM Comput. Surv. 24*, 1 (Mar.), 5–62.

SVEDLOW, M., MCGILLEM, C. D., AND ANUTA, P. E. 1976. Experimental examination of similarity measures and preprocessing methods used for image registration. In *The Symposium on Machine Processing of Remotely Sensed Data* (Westville, Ind., June). pp. 4A–9.

TERZOPOULOS, D., WITKIN, A., AND KASS, M. 1987. Energy constraints on deformable models: Recovering shape and non-rigid motion. In *Pro-*

*ceedings AAAI 87* (July), vol. 2. AAAI, Menlo Park, Calif., pp. 755–760.

THOMAS, I. L., BENNING, V. M., AND CHING, N. P. 1986. *Classification of Remotely Sensed Images*. Adam Hilger, Bristol, England.

TON, J., AND JAIN, A. K. 1989. Registering Landsat images by point matching. *IEEE Trans. Geosci. Remote Sensing 27*, (Sept.), 642–651.

VAN DEN ELSEN, P A., POL, E.-I. D., AND VIERGEVER, M. A. 1992. Medical image matching—A review with classification. *IEEE Eng. Med. Biol*, in press.

VAN WIE, P, AND STEIN, M. 1977 A LANDSAT digital image rectification system. *IEEE Trans. Geosci. Electr. GE-15*, 3 (July), 130–137.

VENOT, A., LEBRUCHEC, J. F., AND ROUCAYROL, J. C. 1984. A new class of similarity measures for robust image registration. *Comput. Vision Graph. Image Process. 28*, 176–184.

WIDROW, B. 1973. The rubber mask technique, Part I and II. *Patt Recog. 5*, 3, 175–211.

WOLBERG, G. 1990. *Digital Image Warping* IEEE Computer Society Press, Los Alamitos, Calif.

WOLBERG, G., AND BOULT, T. 1989. Separable image warping with spatial lookup tables. In *ACM SIGGRAPH '89* (Boston, July/Aug.). ACM, New York, pp. 369–377.

WONG, A. K. C., AND YOU, M. 1985. Entropy and distance of random graphs with application to structural pattern recognition. *IEEE Trans. Patt Anal. Machine Intell. PAMI-7*, 5 (Sept.), 599–609.

WONG, R. Y. 1977. Sensor transformations. *IEEE Trans. Syst. Man Cybernetics SMC-7*, 12 (Dec.), 836–840.

Received August 1991, final revision accepted March 1992