A Survey of Image Registration Techniques

Lisa Gottesfeld Brown

Department of Computer Science
Columbia University
New York, NY 10027

January 12, 1992

Abstract

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. Over the years, a broad range of techniques have been developed for the 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 distortions in the image and the type of registration techniques which are most suitable. Two major types of distortions are distinguished. The first type are those which are the source of misregistration, i.e., they are the cause of the misalignment between the two images. Distortions which are the source of misregistration determine the transformation class which will optimally align the two images. The transformation class in turn influences the general technique that should be taken. The second type of distortion are those which are not the source of misregistration; this type usually effects intensity values but they may also be spatial. Distortions of this type are not to be removed by registration but they make registration more difficult since an exact match is no longer possible. They effect the choice of

feature space, similarity measure and search space and 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 appropriate technique for a specific problem.

Contents

List of Tables

1 Introduction

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 82]. To register two images, a transformation must be found so that each point in one image can be mapped to a point in the second. This mapping must "optimally" align the two images where optimality depends on what needs to be matched. As an example, consider two images taken of a patient using different sensors. A CT scan (computed tomography) is able to clearly see the structures of the patient, i.e., the bones and gross anatomy. Another scan using a sensor which is sensitive to radionucleic activity such as PET (positron emission tomography) or SPECT (single photon emission computed tomography), is capable of localizing specific metabolic activity but can only indirectly sense a limited number of normal structures. Since the two images may be taken at different resolutions, from different viewpoints, and at different times, it is not possible to simply overlay the two images. However, successful registration is capable of identifying the structural sites of metabolic activities (such as tumors) that might otherwise be difficult to ascertain[Maguire 90]. In this case, registration involves finding a transformation which matches the structures found by both sensors.

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

i) Computer Vision and Pattern Recognition - for numerous different tasks such as segmentation, object recognition, shape reconstruction, motion tracking, stereomapping and character recognition,

ii) 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 chomosomes, and

iii) Remotely Sensed Data Processing - for civilian and military applications in agriculture, geology, oceanography, oil and mineral exploration and pollution, urban, forestry and target location and identification.

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 however, 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. In these tables, each class of problems is further 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 the 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 development of enumerable independent registration methodologies.

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

Table 1: Registration Problems - Part I

VIEWPOINT REGISTRATION
Class of Problems: Registration of images taken from different viewpoints *Typical Application:* Depth or shape reconstruction *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 *Example 1* *Field:* Computer Vision *Problem:* Stereomapping to recover depth or shape from disparities *Example 2* *Field:* Computer Vision *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 *Typical Applications:* Detection and monitoring of changes or growths *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 *Example 1* *Field:* Medical Image Analysis *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 *Example 2* *Field:* Remotely Sensed Data Processing *Problem:* Natural Resource Monitoring, Surveillance of Nuclear Plants, Urban Growth Monitoring

Table 2: Registration Problems - Part II

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 (e.g. affine, polynomial, or elastic) to find the optimal transformation for a particular problem. In this survey, it was found that the type of transformation used to register two images is one of the best ways to categorize the methodology and to assist in selecting techniques for particular applications. The transformation type depends on the cause of the misalignment which may or may not be all the distortions present between the two images. This will be discussed in more detail in Section 2.3. In 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 the transformation needed is global but not affine, the primary approach uses feature point mapping to define a polynomial transformation. These techniques are described in 3.3. In the last subsection 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.

An important distinction in the nomenclature that is used throughout this survey may prevent some confusion. 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 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. The important distinction that needs to be understood is between global/local *transformations* and *methods*, global/local *distortions* and global/local *computations*. Since image distortions may not need to be corrected, it is critical for understanding registration methods, that whether distortions are global or local does not depend on the type of transformation. Similarly, global/local computations refer to whether or not computations needed to determine the necessary transformation require information from the entire image or just small local regions. Again, this is distinct from the type of transformation used. Since the transformation class designates the registration approach to be taken, global and local descriptors applied to *methods* refer only to their transformation types. For example, global methods search for the optimal global transformations but may have local distortions which did not cause the misalignment. Local methods search for the optimal local transformation but are most accurate (and slower) if they require global computations since they use information from the entire image to find the best alignment.

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 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 introduced during acquisition, is detailed. In Section 3, the major registration approaches are presented as outlined above. These methods are used as examples for the last section of this survey, Section 4, which offers a framework for the broad range of possible registration techniques. Given knowledge of the kinds of distortion present, and those which need to be corrected, registration techniques select 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 or a specific class defined by the parameters of the problem. Then a feature space and similarity measure are selected which is least sensitive to irrelevant noise and 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 for the given distortions. All registration methods can be viewed as different combinations of choices for these three components: a feature space, a similarity metric and a search strategy. The feature space extracts the information in the images which will be used for matching. Then the search strategy chooses the next transformation from the transformation class which will be used to match the images. The similarity metric determines the relative merit of the match. Then the search continues based on this result until a transformation is found whose similarity measure is satisfactory. This framework for registration techniques is useful for understanding the benifits and relationships between the wide variety of existing techniques and for assisting in the selection of the appropriate technique for a specific problem.

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 2-dimensional 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 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.,

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

and g is 1D intensity or radiometric transformation.

The registration problem is the task involved in finding the optimal spatial and intensity transformations so that the images are matched with regard to the misregistration source. The intensity transformation is frequently not necessary, except, for example, in cases where there is a change in sensor type (such as optical to radar [Wong 77]) or where a simple look up table determined by sensor calibration techniques is sufficient [Bernstein 76]. After all, if the images are matched exactly, then what information can be extracted? Finding the 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 naturally implemented. If the geometric transformation can be expressed as a pair of separable functions, i.e., such that two consecutive 1-D (scanline) operations can be used to 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 89].

2.2 Transformations

The most fundamental characteristic of any image registration technique is the type of spatial transformation or mapping needed to properly overlay two images. Although many types of distortion 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 not due to differences in scene characteristics that are to be detected. The primary general transformations are affine, projective, perspective, and polynomial. These are all well-defined mappings of one image onto another. Given the intrinsic nature of imagery of nonrigid objects, it has been suggested [Maguire 89] 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 for every constant c

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

and

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

A transformation is *affine* if $T(\mathbf{x}) - T(\mathbf{0})$ is linear. Affine transformations are linear however 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. 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, θ, 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:

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

The general 2D affine transformation

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

can account for other spatial distortions as well such as skew and aspect ratio.[Van Wie 77] Without the addition of the first vector which acts as a translation, the transformation becomes linear.

The perspective transformation accounts for the distortion which occurs when a 3D scene is projected through an idealized optical image system as in Figure 1. This is a mapping from 3D to 2D. This projective distortion causes imagery to appear smaller the farther it is from the camera and more compressed the more it is inclined away from the camera. 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

$$x_i = \frac{-f x_o}{z_o - f}$$

$$y_i = \frac{-f y_o}{z_o - f}$$

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.) If the scene is composed of a flat plane tilted with respect to the image plane, a *projective* transformation is needed to map the scene plane into an image which is tilt-free and of a desired scale[Slama 80]. 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:

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

$$y_i = \frac{a_{211}x_p + a_{22}y_p + a_{23}}{a_{31}x_p + a_{32}y_p + a_{33}}.$$

Figure 1: Camera coordinates

If these transformations do not account for the distortions in the scene or not enough information is known about the camera geometry, global alignment can be determined using a polynomial transformation. This is defined in section 3.3.2. 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.

2.3 Image Distortions

An important consideration for selecting the registration method to be employed for a given problem is the *source of misregistration.* The source of misregistration is the cause of the misalignment between images, the misalignment that must be found in order to properly register the two images. The source of misregistration may be due to a change in the sensor position, viewpoint and viewing characteristics or to object movement and deformation. Other distortions, either spatial or photometric, can be present as well, which make it difficult to ascertain the source of misregistration. These distortions, which make it difficult to find the correct registration, are generally due to sensor noise or operation, changes in sensor type, and changes in scene conditions. Distortions which are the source of misregistration determine the transformation class for registration while other distortions influence the selection of the appropriate feature space, similarity metric measure and search space and strategy.

All distortion can be classified as either static/dynamic, internal/external and geometric/photometric. Static distortions do not change for each image and hence can be corrected in all images in the same procedure via calibration techniques. Internal distortions are due to the sensor. Typical internal geometric distortions in earth observation sensors [Bernstein 76] are centering, size, skew, scan nonlinearity, and radially (pin-cushion) or tangentially symmetric errors. Internal distortions 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 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 intensity distortions. The majority of internal errors and many of the photometric ones are static and thus can be removed using calibration. In this survey, the emphasis is on external geometric distortions. Intensity distortions that are not static usually arise from a change in sensor and varied lighting and atmospheric conditions. Their correction becomes important when integrating information between images and using point differences during geometric correction. Typically, the intensity histogram and other statistics about the distribution of intensities are used such as in the method developed by [Wong 77] to register radar and optical data using the Karhunen-Loeve transformation. Sometimes intensity correction is performed simultaneously with geometric correction[Herbin 89].

Since a common objective of registration is to detect a change between two images, it is important that images are matched only with regards to the misregistration source. Otherwise the change of interest will be removed at the same time. For this reason, techniques which are applied to dissimilar images often have a special need to model the misregistration source. In hierarchical search techniques described by [Hall 79], for example, matching rules are selected which are more invariant to natural or even man-made changes in scenery. [Herbin 89] considers registration as a problem of estimating the parameters of a mathematical model which describes the allowable transformations. Of the four major registration problems mentioned in Table 1 and 2, only template matching does not have as its objective to detect changes. 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 with different viewing geometries uses the disparity between images to to determine the depth of objects in the scene or their 3-dimensional shape characteristics. Lastly, registration of images acquired from different sensors integrates the different measurements to classify picture points for segmentation and object recognition. Only in standard template matching where the source of misregistration is noise (for example due to the sensor or lighting conditions) is the objective *not* to detect changes.

Not surprisingly, the more that is known about the type of distortion present in a particular system, the more effective registration can be. For example, [Van Wie 77] decomposes the error sources in Landsat multispectral imagery into those due to sensor operation, orbit and attitude 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 upon 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 reducing the load of the global corrections and improving performance.

Another class of problems in which the source of misregistration is often very usefully modeled is stereo matching and motion tracking. By exploiting camera and object model characteristics such as viewing geometry, smooth surfaces and small motions, registration 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 properties of stereo imagery. The epipolar constraint of stereopsis assures that for any point in one image, its potential matching point in the other image will lie along a line determined by the geometry of the camera viewpoints. If the surfaces in the scene are opaque, an ordering constraint is imposed along corresponding epipolar lines. 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 be possible to register the images analytically and statically. For example, if the two images differ only in their viewing geometries, and this relative difference is known, then the appropriate sequence of elementary Cartesian transformations (namely, a translation, rotation and scale change) 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. In particular, this is possible if the viewing sensor images a plane at a constant distance from the sensor at a constant scale factor, e.g., a simple optical system without optical aberrations. 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, 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 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 $P = (x_o, y_o, z_o)$ can be mapped to an image point $P' = (x_i, y_i)$ by the scale factor $f/(z_o - f)$. This can be seen from the similar triangles in the viewing geometry illustrated in Figure 1. 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 any other flat plane, given by

$$x_o cos\alpha + y_o cos\beta + z_0 = h$$

rectification can be performed by mapping (x_i, y_i) into $(fx_i/Z, fy_i/Z)$ where $Z = f - x_i cos\alpha - y_i cos\beta$ [Rosenfeld 82]. This is because the plane can be decomposed into lines each at a constant distance 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. Two pictures of the flat plane from different viewpoints can be registered by the following steps. First, the scene points (x_1, y_1, z_1) are related to their image coordinates in image 1 scaled by a factor $(z_1 - f)/f$ dependent on their depth (the z_1 coordinate) and the lens center f because of similar triangles. They must also satisfy the equation of the plane. The scene coordinates are then converted from the coordinate system with respect to the 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 90] 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. It is usually used for template matching or pattern recognition. It is a match metric, i.e., it gives a measure of the degree of similarity between an image and a template. 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:

$$C(u,v) = \frac{\sum_x \sum_y T(x,y)I(x-u,y-v)}{\left[\sum_x \sum_y I^2(x-u,y-v)\right]^{\frac{1}{2}}}$$

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 82] 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. Also, this measure is directly related to the more intuitive measure,

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

which decreases with the degree of similarity. Since the template energy $\sum_x \sum_y T^2(x,y)$ is constant, if we again normalize for the local image energy $\sum_x \sum_y I^2(x-u,y-v)$, then it is the product term or correlation, $\sum_x \sum_y T(x,y)I(x-u,y-v)$ which will effect the outcome.

A related measure, which is sometimes advantageous, is the correlation coefficient:

$$\frac{covariance(I,T)}{\sigma_I \sigma_T} = \frac{\sum_x \sum_y (T(x,y)-\mu_T)(I(x-u,y-v)-\mu_I)}{\left[\sum_x \sum_y (I(x-u,y-v)-\mu_I)^2 \sum_x \sum_y (T(x,y)-\mu_T)^2\right]^{\frac{1}{2}}}$$

where μ_T and σ_T are mean and standard deviation of the template and μ_I and σ_I are mean and standard deviation of the image. This statistical measure has the property that it measures correlation on an absolute scale which ranges from $[-1, 1]$. Under certain assumptions, the value measured by the correlation coefficient gives a linear indication of the similarity between images. This is sometimes useful in order to measure confidence in a match and to reduce the number of measurements needed when a prespecified confidence is sufficient.[Svedlow 76]

By the convolution theorem, correlation can be computed as a product of Fourier transforms. Hence, an important reason why this metric has been widely used is because it can be computed using the Fast Fourier Transform (FFT) and thus, for large images of the same size, it can be implemented efficiently. There are two major caveats however. Only the cross-correlation before normalization may be treated by FFT. Secondly, although the FFT is faster it also requires a memory capacity that grows with the log of the image area. Furthermore, both direct correlation and correlation using FFT have costs which grow at least linearly with the image area.

Template matching using correlation has many variations [Pratt 78]. If the allowable transformations include rotation or scale, for example, multiple templates can be used. As the number of templates grows, however, the computational costs quickly become unmanageable. Often smaller local features of the template which are more invariant to shape and scale, such as edges joined in a Y or a T, are used. In [Duda 73], it is suggested that a triangle be matched by first finding three separate lines and then determining if a triangle is indeed present. A better solution is offered by [Widrow 73], (elaborated upon by [Burr 81]), who introduces the *rubber template*, a template which can be locally distorted, so that information between local matches can be utilized. This is described in more detail in section 3.4

If the image is noisy, the peak of the correlation may not be clearly discernible. If the noise can be easily modeled, (or more precisely if it is additive, stationary and independent of the image and its power-spectral density is known), the image can be prefiltered and correlated simultaneously using *matched filter* techniques [Rosenfeld 82]. A similar technique uses a *statistical correlation* measure [Pratt 78] which prefilters the image and template in such a way as to maximize the peak correlation when the pair of images are optimally matched. This measure requires heavy computational costs in order to compute the eigenvalues and eigenvectors of the image covariance matrices, (unless the images can be modeled by separable Markov processes and there is no observational noise), it is usually too computationally intensive.

A far more efficient class of algorithms than traditional cross-correlation was proposed by [Barnea 72], called the sequential similarity detection algorithms (SSDAs). Two major improvements are offered. First, they suggest a similarity measure $E(u, v)$, which is computationally much simpler, based on the L_1 norm between 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 means of the template and local image window respectively. Even in the unnormalized case, however, a minimum is guaranteed for a perfect match. Correlation on the other hand, requires both normalization and the expense of multiplications.

The second improvement Barnea and Silverman introduce is a sequential search strategy. In the simplest case of translation registration, this strategy might be a sequential thresholding. For each possible window the error measure 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 error 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 upon 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. Several options for similarity measures are discussed in section 4.2.

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 to not miss good matches.

In comparison with correlation, the sequential similarity technique improves efficiency by orders of magnitude. Tests conducted by Barnea and Silverman, 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 both maximized by 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. These methods are applicable whenever low frequency noise is present.

3.2 Fourier Methods

The Fourier Transform has several properties that can be exploited for image registration. Translation, rotation, reflection, distributivity and scale, all have their counterpart in the Fourier domain. Furthermore, by using the frequency domain, it is possible to achieve excellent robustness against correlated and frequency-dependent noise. Lastly, the transform can either be efficiently implemented in hardware or using the Fast Fourier Transform. In this section, the basic methods used to register images using Fourier analysis will be described.

An elegant method to align two images which are shifted relative to one another is to use *phase correlation* [Kuglin 75]. Phase correlation relies on the translation property of the Fourier transform, sometimes referred to as the Shift Theorem. Given two images I_1 and I_2 which differ only by a displacement (d_x, d_y), i.e.,

$$I_2(x, y) = I_1(x - d_x, y - d_y)$$

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. If the exponential form of $F_i(\vec{\omega}\,) = |F_i| e^{j\phi_i(\vec{\omega}\,)}$ for $i = 1, 2$, then the phase difference is given by $e^{j(\phi_1 - \phi_2)}$. Because of the shift theorem , this phase difference is equivalent to the phase of the cross-power spectrum,

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

where $*$ is the complex conjugate. The inverse Fourier transform of the phase difference is a delta function centered at the displacement, which in this case, is the point of registration. In practise, the continuous transform must be replaced by the discrete one and the delta function becomes a unity pulse.

The method 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, this technique is particularly well-suited to images with narrow bandwidth noise. Consequently, it is an effective technique for images obtained under differing conditions of illumination since illumination functions 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 correlation measure relatively scene-independent.

On the other hand, cross-correlation is optimal if there is white noise. [Kuglin 75] suggest introducing a generalized weighting function to the phase difference before taking the inverse Fourier Transform, so that there exists 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.

Certain assumptions underlie the use of the Fourier transform which should not be overlooked. Since the images are bounded and discrete, frequency information is also bounded and discrete. By the sampling theorem, the interval between discrete samples must be small enough so the bandwidth of the signal can be reproduced or aliasing will occur. Also, since the image is bounded, or in other words, a window of the signal has been taken, a distortion due to the frequency components of the window will be introduced in the frequency-domain [Gonzalez 77]. In summary, in using the Fourier transform, it has been assumed that the images are bandlimited and accordingly Nyquist sampled, and periodic with the image size. Images are often preprocessed in order to make these assumptions more valid. For example, Gaussian smoothing can be applied to limit the bandwidth.

In an extension of the phase correlation technique, [De Castro 87] has 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 87] present the following two step process to first determined 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. Two images $I_1(x,y)$ and $I_2(x,y)$ which differ by a translation (x_d, y_d) and a rotation ϕ_0 will have Fourier transforms related by

$$F_2(\omega_x, \omega_y) = e^{-j(\omega_x x_d + \omega_y y_d)} F_1(\omega_x cos\phi_0 + \omega_y sin\phi_0, -\omega_x sin\phi_0 + \omega_y cos\phi_0).$$

By taking the phase of the cross-power spectrum as a function of the rotation angle estimate ϕ and using polar coordinates to simplify the equation we have

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

Therefore, by first determining the angle ϕ which makes the phase of the cross-power spectrum the closest approximation to a unit pulse, 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 87] applied the transform to a zero-padded image thus increasing the resolution and improving the approximation of the transform after rotation. Other interpolation techniques, for instance, nearest neighbor and bilinear interpolation, proved to be unsatisfactory. Their method is also costly because of the difficulty in testing for each ϕ. [Alliney 86] 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 transforms of the x- and y-projections are 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 87] 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 translational shift between them.) Then the power cepstrum is used to determine the translational correspondence in a similar manner to phase correlation. This has the advantage over [De Castro 87] of the computational savings gained by adding images instead of multiplying them due to the use of logarithms. The work of [De Castro 87] summarizes previous work published in Italy before 1987, but no direct comparison with [Lee 87] 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.

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. The general method 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 based on least squares regression or similar technique. Resampling of one image onto the other is performed applying the spatial mapping and an interpolation technique. In the following three sections, we will describe (1) the different types of control points and how they are matched, (2) the global mapping methods which find a single transformation from the matched control points for aligning two images and (3) the more recent work in local mapping using 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 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 image and data image is found.

Control points can either be intrinsic or extrinsic. Intrinsic control points are markers in the image which are not relevant to the data itself, are often placed specifically for registration purposes and are easily identified. They may even be placed on the sensor such as reseau marks in which case the registration is really just calibration. Fiducial chemical markers are widely used in medical imaging; these are identifiable structures placed in known positions, such as plastic "N" shaped tubing filled with $CuSO_4$ placed strategically for magnetic resonance imaging (MRI) systems [Evans 88] or stereotactic coordinate frames that identify three dimensional coordinates for positron emission tomography (PET) [Bergström 81, Bohm 83, Bohm 88, Fox 85]. 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 79].

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 knowledgeable with 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 88]. 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. These and many other features are discussed in more detail in section 4.1. 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.

After the set of features has been determined, the features in each picture must be matched. 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 each match. 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 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 involve clustering, relaxation, matching of minimum spanning trees of the two sets and matching of convex hull edges of the two sets [Goshtasby 88]. 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.

The relaxation technique described by [Ranade 80], can be used to register images under translation. In this case, the point matching and the determination of the best spatial transformation are accomplished simultaneously. Each possible match of points defines a displacement which is given a rating according to how closely other pairs would match under this displacement. The procedure is then iterated, adjusting, in parallel, the weights of each pair of points based on their ratings.

The clustering technique described by [Stockman 82] is similar in that the matching determines the spatial transformation between the two images. In this case the transformation is a rotation, scaling and translation although it could be extended to other transformations. For each possible pair of matching 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.

These schemes allow for global matching which is less sensitive to local distortions 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. Determining the point matches and the global transformation simultaneously is advantageous whenever there is little independent information for obtaining the matches first. 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. This 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 Global Methods

Global methods based on point matching use a set of matched points to generate an single optimal transformation. Given a sufficient number of points we can derive the parameters of any transformation 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 least squares regression analysis. The number of matched points must be sufficiently greater than the number of parameters of the transformation. Thus for large numbers of automatic control points, approximation makes the most sense. For intrinsic or manual control points, there are usually fewer but more accurate matches, suggesting that interpolation may be more applicable. In this case, the transformation is constrained so that the matched points are satisfied exactly. 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.

For static distortions, the form of the mapping function between the two images is known; approximation or interpolation is selected accordingly, and registration or calibration is achieved. More commonly though, the precise form of the mapping function is unknown and 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 into, and a_{ij} and b_{ij} are the constant polynomial coefficients. The order of the polynomial, m, depends on the tradeoff between accuracy and speed needed for the specific problem. For many applications, second or third order is sufficient [Nack 77, Van Wie 77]. 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.

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 76] 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 altitude deviations. [Maguire 85] fit correlation matched landmarks to a fourth order polynomial to register CT and PET images of the heart thus correcting translation, rotation, scale and skew errors. If more information is known about the transformation then a general polynomial transformation may not be needed. [Merickel 88] 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.

For a large number of control points, using the normal equations to solve the least squares approximation becomes unstable and inaccurate. This can be overcome by using orthogonal polynomials as the terms of the polynomial mapping. Orthogonal polynomials can be readily generated using the Gram-Schmidt orthogonalization process. They also 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 the error is sufficiently small [Goshtasby 88].

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.

3.3.3 Local Methods

The global point-mapping methods mentioned 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 the mapping transformation. Local methods are more powerful and can handle many distortions that global methods cannot; examples include 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 tradeoff between the power of these methods and their corresponding computational cost.

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:

$$\begin{aligned} X_i &= F_x(x_i, y_i) \\ Y_i &= F_y(x_i, y_i) \qquad\qquad i = 1, ..., N \end{aligned}$$

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 79], compared exactly these methods, testing each on several surfaces and evaluating their performance characteristics. As will be seen, the methods used in Franke's study, although not designed for this purpose, underlie much of the current work in local image registration.

Most of the methods evaluated by Franke 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 some cases, a weighted sum of the basis functions such as B-splines or Gaussian distributions is used and the weights are similarly derived from the constraints. In the simplest case, a weighted sum of neighboring points is computed where the weights may be related inversely with distance such as in linear 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. Other methods compared in Franke's study include the use of finite elements and the generalized Newton interpolant. 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 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 not 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 and sometimes global methods (which require global computations) on lower resolution data sets precede their use. Local methods which rely only on local computations are not only more efficient, but they can be locally controllable. This 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.

From the set of surface interpolation techniques discussed in the study, many registration techniques are possible. For instance, [Goshtasby 86] proposed using "optimal" triangulation of the control points to partition the image into local regions for interpolation. Triangulation decomposes the convex hull 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, he extended this method [Goshtasby 87] 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 polynomials, Goshtasby decomposed each triangle into Clough-Tocher subtriangles and assumed certain partial derivatives along the edges of the triangles were given. Similar methods, using polynomials of various orders, have been proposed by scientists in Computer Aided Geometric Design (CAGD) to fit composite surfaces to scattered data.

The piecewise cubic polynomial method can successively register images with local geometric distortion assuming the difference between images is continuous and smooth. However, where a discontinuous geometric difference exists, such as in a motion sequences where occlusion has occurred, the method would fail. Also, the Franke study concluded that methods that use triangulation can be problematic when long thin triangles occur and also 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 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 selected one based on divide and conquer with complexity $O(NlogN)$ 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.

[Ratib 88] suggests that it is sufficient for the "elastic" matching of Positron Emission Tomographic images of the heart, to first globally match the images 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 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 study, they generally do not compare well with many of the other surface interpolation techniques. However, a quadratic least squares fit at each data point in conjunction with localization of the weights was found to be one of the best methods of all.

Another registration technique proposed by Goshtasby, which is also derived from the interpolation methods discussed in Franke's study is called the local weighted mean method [Goshtasby 88]. In this 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, the square of the order of the polynomial, and the size of the image. 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 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 were used so that each polynomial's influence would be spread far enough to cover image locations without points.

The primary global portion of these calculations is the determination of the set of control points and their matches. This 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, they must do so in a single pass; there is no feedback between the point matching and the interpolation. Nor do they take advantage of several algorithmic techniques which can improve and speed up the extraction of local distortions. These are, namely, iteration, a hierarchical structure, and cooperation. 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 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 as the deformation of an elastic material. 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[Widrow 73]. 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. These were used to iteratively modify the template until the best match was found. However, it was not until more recently [Burr 81] 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 to the nearest neighbor in the complementary image is determined for each edge or feature point, both from first image and from the second. 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) 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 are 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 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.

ii) 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.

iii) Cooperation: Features in one location influence decisions at other locations.

Techniques with these characteristics 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. In their original method, developed by Broit in his Ph.D. thesis, 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. Bajcsy and Broit [Bajscy 82], applied this to 2 and 3D medical images and claim greater efficiency over Burr's method although their experiments are limited. Like Burr's method, iteration and cooperation are clearly utilized.

In her latest work with S. Kovacic, [Bajscy 89] X-ray computed tomography scans of the human brain are elastically matched with a 3D atlas. As with many local techniques, it is necessary to first globally align images 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, but this is extended in a hierarchical fashion. The same set of partial differential equations 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 locally pulling it towards the maximum of the local similarity function. 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 local similarity function is computed based on normalized correlation but which decomposes each image into its projections onto a complete system of orthonormal functions and uses only those projections relevant for matching. The system of equations are 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 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 method, iteration, cooperation and a hierarchical structure are exploited.

A very similar method was proposed by [Dengler 1986] for solving the correspondence problem in moving image sequences. To increase the speed and reliability, the external forces were computed from local binary correlations based on the sign of the Laplacian. Also, to allow discontinuities, the displacement vector field was computed from a Laplacian whose local region is limited to pixels of the same Laplacian sign. A similar hierarchical scheme was used to increase efficiency and make local neighborhoods scale invariant.

Recently, techniques similar to elastic matching have been used to recover shape and non-rigid 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 must proceed the analysis and 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 87]. 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., use the de-projection 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 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 distortions 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 the transformation needed is global but not affine, then point mapping can be used to interpolate or approximate a polynomial transformation. 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 local distortion occurs which is not the source of misregistration, then it is 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 three major components:

- feature space
- similarity metric
- search space and strategy

As described earlier, the best available knowledge of the source of misregistration determines the transformation needed. This in turn, determines the complexity and kind of method. Knowledge of other distortions (which are not the source of misregistration) can then be used to decide upon the best choices for the three major components listed above. Tables 3,4, and 5 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 three subsections, each of the components of registration is described more fully.

4.1 Feature Space

The first step in registering two images is to decide upon the feature space to use for matching. This may be the image itself, but other common fea-

Feature Spaces and Their Attributes
RAW INTENSITY - most information
EDGES - intrinsic structure, less sensitive to noise Edges [Nack 77] Contours [Medioni 84] Surfaces [Pelizzari 89]
SALIENT FEATURES - intrinsic structure, accurate positioning Points of locally maximum curvature on contour lines [Kanal 81] Centers of windows having locally maximum variances [Moravec 81] Centers of gravity of closed boundary regions [Goshtasby 86] Line intersections [Stockman 82] Fourier descriptors [Kuhl 82]
STATISTICAL FEATURES - use of all information, good for rigid transformations, assumptions concerning spatial scattering Moment invariants [Goshtasby 85] Centroid/principal axes [Rosenfeld 82]
HIGHER LEVEL FEATURES - uses relations and other higher level information, good for inexact and local matching Structural features: graphs of subpattern configurations [Mohr 90] Syntactic features: grammars composed from patterns [Bunke 90] Semantic networks: scene regions and their relations [Faugeras 81]
MATCHING AGAINST MODELS - accurate intrinsic structure, noise in one image only Anatomic atlas [Dann 89] Geographic map [Maitre 87] Object model [Terzopoulos 87]

Table 3: Feature Spaces used in Image Registration

ture spaces include: edges, contours, surfaces, other 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. The feature space is a fundamental aspect of almost all computer vision tasks and 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,

- which properties of the images will be matched, e.g., more interested in matching structures than textural properties,

- the computational cost by either reducing the search space or, on the other hand, increasing the computations necessary.

Images are usually preprocessed in an attempt to extract intrinsic structure. This reduces the effects of scene and sensor noise, forces matching to optimize structural similarity and reduces the corresponding search space. Image enhancement techniques [Gonzalez 77] can be used to emphasize structural information. For example, homomorphic filtering can be used to control the effects of illumination and enhance the effects of reflectance. Edges, because they represent much of the intrinsic structures of an image, are the most frequently used feature space. Another 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. Image statistics such as moment invariants are another popular choice although they are computationally costly (lower order moments are sometimes used first to guide the match and speed the process [Goshtasby 85],[Mahs 87]) 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. [Mitiche 83] suggests the use of *shape-specific points*, such as the centroid and the radius weighted mean, for pre-registration 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 is 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 distortions which are not relevant to the proper registration 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 either the feature space or through the similarity measure. Typical similarity measures for image or feature values 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 85] 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

Similarity Metric	*Advantages*
Normalized cross-correlation function [Rosenfeld 82]	accurate for white noise but not tolerant of local distortions, sharp peak in correlation space difficult to find
Correlation coefficient[Svedlow 76]	similar to above but has absolute measure
Statistical correlation and matched filters[Pratt 78]	if noise can be modeled
Phase-correlation [De Castro 87]	tolerant of frequency dependent noise
Sum of absolute differences of intensity [Barnea 72]	efficient computation, good for finding matches with no local distortions
Sum of absolute differences of contours [Barrow 77]	can be efficiently computed using "chamfer" matching, more robust against local distortions - not as sharply peaked
Contour/surface differences[Pelizzari 89]	for structural registration
Number of sign changes in pointwise intensity difference [Venot 89]	good for dissimilar images
Higher-level metrics: structural matching: tree and graph distances [Mohr 90], syntactic matching: automata [Bunke 90]	optimizes match based on features or relations of interest

Table 4: Similarity Metrics used in Image Registration

point mapping methods. Then the set of control point matches are 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. If grey 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. 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 84] 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 effect the classical measures according to the grey 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 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 space and search strategy. For computationally intensive features such as moment invariants, a search strategy must be designed to limit the number of features to be computed. Likewise for similarity measures such as correlation, it is important to reduce the number of measures to be computed. The greater the distortion in the image that needs to be corrected 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 distortion is present. 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 labeling, and generalized Hough transforms, linear programming, tree and graph matching, dynamic programming and heuristic search.

Search Space: The model underlying each registration technique determines the characteristics of the search space. 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 space can be significantly reduced while allowing

Search Strategy	*Advantages and Reference Examples*
Decision Sequencing	Improved efficiency for similarity optimization for rigid transformations [Barnea 72]
Relaxation Labeling	Practical approach to find global transformations when local distortions are present, exploits spatial relations between features [Hummel 83], [Price 85], [Ranade 80], [Shapiro 90]
Dynamic Programming	Good efficiency for finding local transformations when an intrinsic ordering for matching is present [Guilloux 86], [Maitre 87], [Milios 89], [Ohta 87]
Generalized Hough Transform	For shape matching of rigidly displaced contours by mapping edge space into dual "parameter" space [Ballard 81], [Davis 82]
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 84]
Hierarchical Techniques	Applicable to improve and speed up many different approaches by guiding search through progressively finer resolutions [Bajscy 89],[Bieszk 87],[Davis 82],[Paar 90]
Tree and Graph Matching	Uses tree/graph properties to minimize search, good for inexact and matching of higher level structures [Gmur 90],[Sanfeliu 90]

Table 5: Search Strategies used in Image Registration

more general transformations. In local methods, such as piecewise interpolation or elastic model-based methods, the models become more complex, 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 77] and [Kiremedjian 87] 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 88]. 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.

Another important factor in determining the appropriate model, besides the allowable transformations, is the allowable distortions. Distortion may be present which is not the source of misregistration. In particular, if the model allows only global transformations, an important issue is whether or not local geometric distortions are expected. In the latter case standard search strategies are no longer sufficient. Why would local distortions still be expected while modeling misregistration sources as global? Perhaps the best reason is that it is known that the images are globally misaligned but that differences in local geometry are of interest. An example might be in aerial photographs taken at different times.

Search Strategies: Table 5 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 82] contains several examples of computer architectures designed for registration problems in pattern processing.

For this discussion, two search strategies have been chosen to exemplify the kinds of strategies used in registration: relaxation matching and dynamic programming. 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 to find maxima, and linear equations or regression to fit polynomials. 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 89], linear programming for solving point matching with polygonal shaped point errors [Baird 84], generalized Hough transforms for shape matching [Ballard 81].

Relaxation Matching: Several researchers have investigated the use of relaxation matching as a search strategy for registration [Hummel 83], [Ranade 80]. Relaxation get its name from the iterative numerical methods which it resembles. It is usually used to find a global maximum to a similarity criteria for rigid transformations. 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 practise it is often a good short-cut over more rigorous techniques such as linear programming.

Relaxation matching techniques have been compared by [Price 85] 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 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 clearly justified. It is empirically compared on aerial photographs and thus its generality is questionable. The major contribution is the 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 efficiently 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 [Pari 90], the registration of geographic contours with maps[Maitre 87], shape matching [Milios 89], stereomapping [Ohta 87], and horizontal motion tracking [Guilloux 86]. 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.

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 practise, the cost is reduced by limiting the matches to an interval size which reflects the largest expected disparty 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

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 upon the class of transformations which will optimally map the images onto each other. The class of transformations and its complexity determine the type of method to be used. Affine transformations can be found by Fourier methods and techniques related to cross-correlation. 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 model-based approaches exploiting knowledge of the distortions. The technique can be completely specified by selecting a particular feature space, similarity metric, search space and strategy from the types of methods available for registration given the transformation class. The choices for these components of the registration method depend on the remaining distortions, spatial and photometric which obsure 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 distortions 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 and similarity measurements 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 between 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.

References

[Alliney 86] S. Alliney and C. Morandi, "Digital Image Registration using Projections," *IEEE Trans. Pattern Analyis and Machine Intelligence* **PAMI-8** , No. 2, March 1986, pp222-233.

[Bajscy 82] R. Bajscy and C. Broit, "Matching of Deformed Images," *6th Int. Conf. Pattern Recognition*, 1982, pp351-353.

[Bajscy 89] R. Bajscy and S. Kovacic, "Multiresolution Elastic Matching," *Computer Vision, Graphics, and Image Processing* **46**, 1989, pp1-21.

[Baird 84] H. S. Baird, "Model-Based Image Matching Using Location," *An ACM Distinguished Disstertation*, MIT Press, Cambridge, MA, 1984.

[Ballard 81] D. H. Ballard, "Generalizing the Hough Transform to Detect Arbitrary Shapes," *Pattern Recognition* **13**, No. 2, 1981, pp111-122.

[Barnea 72] D. I. Barnea and H. F. Silverman, "A Class of Algorithms for Fast Digital Registration," *IEEE Trans. Computers* **C-21**, 1972, pp179-186.

[Barrow 77] H. G. Barrow, J. M. Tenenbaum, R. C. Bolles and H. C. Wolf "Parametric Correspondence and Chamfer Matching: Two New Techniques for Image Matching," *Proc. Intl. Joint Conference in Artificial Intelligence*, 1977, pp659-663.

[Bernstein 71] R. Bernstein and H. Silverman "Digital Techniques for Earth Resource Image Data Processing," *Proc. Amer. Inst. Aeronautics and Astronautics 8th Annual Meeting* **21**, AIAA paper no. 71-978, October 1971.

[Bergström 81] M. Bergström, B. J. ëthius, L. Eriksson, T. Greitz, T. Ribbe and L. Widén, "Head Fixation Device for Reproducible Position Alignment in Transmission CT and Positron Emission Tomography", *Journal of Computer Assisted Tomography,* **5**, February, 1981, pp136-141.

[Bernstein 76] R. Bernstein, "Digital Image Processing of Earth Observation Sensor Data," *IBM J. Research and Development* **20**, January 1976, pp40-67.

[Bieszk 87] J. A. Bieszk and I. Fram, "Automatic Elastic Image Registration," *Proc. Computers in Cardiology,* Leuven, Belgium, September 1987, pp3-5.

[Bohm 83] C. Bohm, T. Greitz, D. Kingsley, B. M. Berggren and L. Olsson, "Adjustable Computerized Stereotaxic Brain Atlas for Transmission and Emission Tomography", *American Journal of Neuroradiology,* **4**, March, 1983, pp731-733.

[Bohm 88] C. Bohm and T. Greitz, "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*, Ed. David Ottoson and William Rosténe, Wenner-Gren International Symposium Series, **53**, 1989, pp137-140.

[Bresler 87] Y. Bresler and S. J. Merhav, "Recursive Image Registration with Application to Motion Estimation," *IEEE Trans. on Acoustics, Speech, and Signal Processing* **ASSP-35**, No. 1, Jan. 1987, pp70-85.

[Bunke 90] H. Bunke and A. Sanfeliu, editors, *Syntactic and Structural Pattern Recognition, Theory and Applications*, World Scientific Pub. Co., Teaneck, NJ, 1990.

[Burr 81] D. J. Burr, "A Dynamic Model for Image Registration," *Computer Graphics and Image Processing* **15**, 1981, pp102-112.

[Dann] R. Dann, J. Hoford, S. Kovacic, M. Reivich and R. Bajcsy, "Evaluation of Elastic Matching System for Anatomic (CT, MR) and Functional (PET) Cerebral Images," *J. of Computer Assisted Tomography,* **13**, July/August 1989, pp603-611.

[Davis 82] L. S. Davis, "Hierarchical Generalized Hough Transform and Line Segment Based Generalized Hough Transforms," *Pattern Recognition* **15**, 1982, p277-285.

[De Castro 87] E. De Castro and C. Morandi "Registration of Translated and Rotated Images using Finite Fourier Transforms," *IEEE Trans. on Pattern Analysis and Machine Intelligence* **PAMI-9** ,No. 5, September 1987, pp700-703.

[Deguchi 86] K. Deguchi, "Registration Techniques for Partially Covered Image Sequence," *IEEE Proc. 8th Intl. Conf. on Pattern Recognition*, Paris, October 1986, pp1186-9.

[Dengler 1986] J. Dengler, "Local Motion Estimation with the Dynamic Pyramid," *Eighth International Conference on Pattern Recognition*, Paris 1986, pp1289-92.

[Duda 73] R. O. Duda and P. E. Hart *Pattern Classification and Scene Analyis* John Wiley & Sons, New York, 1973.

[Evans 88] A. C. Evans, C. Beil, S. Marrett, C. J. Thompson, and A. Hakim, "Anatomical-Functional Correlation Using an Adjustable MRI-Based Region of Interest Atlas with Positron Emission Tomography," *Journal of Cerebral Blood Flow and Metabolism* **8**, 1988, pp513-530.

[Faugeras 81] O. Faugeras and K. Price, "Semantic Description of Aerial Images using Stochastic Labeling," *IEEE Trans. Pattern Analysis and Machine Intelligence,* **PAMI-3**, Nov 1981, pp638-642.

[Fox 85] P. T. Fox, J. S. Perlmutter, and M. E. Raichle, "Stereotactic Method of Anatomical Localization for Positron Emission Tomography," *Journal of Computer Assisted Tomography* **9**, 1985, pp141-153.

[Fu 82] K. S. Fu and T. Ickikawa, *Special Computer Architectures for Pattern Processing*, CRC Press, Inc., Boca Raton, FL, 1982.

[Franke 79] R. Franke, "A Critical Comparison of Some Methods for Interpolation of Scattered Data," *Naval Postgraduate School Technical Report*, NPS-53-79-003, December 1979.

[Gerlot 87] P. Gerlot and Y. Bizais, "Image Registration: A Review and a Strategy for Medical Applications," *Proc. of the 10th Intl. Conf. on Information Processing in Medical Imaging*, Utretcht, Netherlands, Ed. C. N. de Graaf and M. A. Viergever, 1987, pp81-89.

[Gmur 90] E. Gmur and H. Bunke, "3-D Object Recognition Based on Subgraph Matching in Polynomial Time," in *Structural Pattern Analysis*, World Scientific Pub. Co., Teaneck, NJ, 1990.

[Gonzalez 77] R. C. Gonzalez and P. Wintz, *Digital Image Processing*, Addison-Wesley Publishing Co., Reading, MA, 1977.

[Goshtasby 85] A. Goshtasby, "Template Matching in Rotated Images," *IEEE Trans. Pattern Analysis and Machine Intelligence* **7**, No. 3, May 1985, pp338-344.

[Goshtasby 86] A. Goshtasby, G. C. Stockman, and C. V. Page, "A Region-base Approach to Digital Image Registration with Subpixel Accuracy," *IEEE Trans. Geosci. Remote Sensing* **24**, No. 3, 1986, pp390-399.

[Goshtasby 86] A. Goshtasby, "Piecewise Linear Mapping Functions for Image Registration," *Pattern Recognition* **19**, No. 6, 1986, pp459-466.

[Goshtasby 87] A. Goshtasby, "Piecewise Cubic Mapping Functions for Image Registration," *Pattern Recognition* **20**, No. 5, 1987, pp525-533.

[Goshtasby 88] A. Goshtasby, "Image Registration by Local Approximation," *Image Vision Computing*, **6**, No. 4, November 1988, pp255-61.

[Guilloux 86] Y. Le Guilloux, "A Matching Algorithm for Horizontal Motion, Application to Tracking," *IEEE Proc. 8th Intl. Conf. on Pattern Recognition*, Paris, October 1986, pp1190-2.

[Hall 79] E. L. Hall, *Computer Image Processing and Recognition*, Academic Press, Inc., NY, 1979.

[Haralick 76] R. M. Haralick, "Automatic Remote Sensor Image Registration," *Topics in Applied Physics* **11**: *Digital Picture Analysis*, Ed. by A. Rosenfeld, Springer-Verlag, 1976, pp5-63.

[Herbin 89] M. Herbin, A. Venot, J. Y. Devaux, E. Walter, J. F. Lebruchec, L. Dubertret, and J. C. Roucayrol, "Automated Registration of Dissimilar Images: Application to Medical Imagery" *Computer Vision, Graphics and Image Processing* **47**, 1989, pp77-88.

[Hummel 83] R. Hummel and S. Zucker, "On the Foundations of Relaxation Labeling Processes," *IEEE Trans. on Pattern Analysis and Machine Intelligence* **5**, May 1983, pp267-287.

[Kanal] L. N. Kanal, B. A. Lambird, D. Lavine, and G. C. Stockman, "Digital Registration of Images from similar and dissimilar Sensors," *Proc. Intl. Conf. Cybernetics and Society*, 1981, pp347-351.

[Kiremedjian 87] G. Kiremidjian "Issues in Image Registration," *IEEE Proc. SPIE:Image Understanding and the Man-Machine Interface* **758**, 1987, pp80-87.

[Kuglin 75] C. D. Kuglin and D. C. Hines, "The Phase Correlation Image Alignment Method," *Proc. IEEE 1975 Int. Conf. Cybernetics and Society*, September 1975, pp163-165.

[Kuhl] F. P. Kuhl and C. R. Giardina, "Elliptic Fourier Features of a Closed Contour," *Computer Graphics and Image Processing,* **18**, 1982, pp 236-258.

[Lee 87] D. J. Lee, T. F. Krile and S. Mitra, "Digital Registration Techniques for Sequential Fundus Images," *IEEE Proc. SPIE: Applications of Digital Image Processing X* **829**, 1987, pp293-300.

[Mahs 87] R. Maghsoodi and B. Rezaie, "Image Registration Using a Fast Adaptive Algorithm," *IEEE Proc. of SPIE: Methods of Handling and Processing Imagery* **757**, 1987, pp58-63.

[Maguire 85] G. Q. Mauire Jr., M. E. Noz, E. M. Lee and J. H. Shimpf, "Correlation Methods for Tomographic Images using Two and Three Dimensional Techniques," Proc. of the 9th Conference of Information Processing in Medical Imaging, Washington, D.C, Ed. S. L. Bacharach, June 10-14, 1985, pp266-279.

[Maguire 90] G. Q. Mauire Jr., M. E. Noz, and H. Rusinek, "Software Tools to Standardize and Automate the Correlation of Images with and between Diagnostic Modalities," *IEEE Computer Graphics and Applications,*.

[Maguire 89] From discussions with G. Q. Maguire Jr., Professor of Computer Science, Columbia University, NY, NY.

[Maitre 87] H. Maitre and Y. Wu, "Improving Dynamic Programming to Solve Image Registration," *Pattern Recognition* **20**, No. 4, 1987, pp443-462.

[Medioni 84] G. Medioni and R. Nevatia, "Matching Images using Linear Features," *IEEE Trans. Pattern Analysis and Machine Intelligence,* **PAMI-6,** 1984, 675-685.

[Merickel 88] M. Merickel, "3D Reconstruction: The Registration Problem," *Computer Vision, Graphics and Image Processing* **42**, no.2, 1988, pp206-219.

[Milios 89] E. E. Milios, "Shape Matching Using Curvature Processes," *Computer Vision, Graphics, and Image Processing* **47,** 1989, pp203-226.

[Mitiche 83] A. Mitiche and J. K. Aggarwal, "Contour Registration by Shape-specific Points for Shape Matching," *Computer Vision, Graphics and Image Processing* **22**, 1983, pp296-408.

[Mohr 90] R. Mohr, T. Pavlidis, and A. Sanfeliu, *Structural Pattern Analysis*, World Scientific Pub. Co., Teaneck, NJ, 1990.

[Moravec 81] H. Moravec, "Rover Visual Obstacle Avoidance," *Proc. 7th Intl. Conf. Artificial Intelligence,* Vancouver, B.C., Canada, August 1981, pp785-790.

[Mort 88] M. S. Mort and M. D. Srinath "Maximum Likelihood Image Registration with Subpixel Accuracy," *IEEE Proc. of SPIE: Applications of Digital Image Processing* **974**, 1988, pp38-43.

[Nack 77] M. L. Nack, "Rectification and Registration of Digital Images and the Effect of Cloud Detection," *Proc. Machine Processing of Remotely Sensed Data*, 1977, pp12-23.

[Nahmias 94] C. Nahmias, E.S. Garnett, "Correlation between CT, NMR and PT Findings in the Brain," *NATO ASI Series* **F19**: *Pictorial Information Systems in Medicine,* Ed. K. H.Hohne, Springer-Verlag Berlin, 1986, pp507-514.

[Ohta 87] Y. Ohta, K. Takano, and K. Ikeda, "A Highspeed Stereo Matching System Based on Dynamic Programming," *Proc. of Int'l Conf. in Computer Vision*, London, 1987, pp335-342.

[Paar 90] G. Paar and W. G. Kropatsch, "Hierarchical Cooperation between Numerical and Symbolic Image Representations," in *Structural Pattern Analysis*, World Scientific Pub. Co., Teaneck, NJ, 1990.

[Pari 90] M. Parizeau and R. Plamondon, "A Comparative Analysis of Regional Correlation, Dynamic Time Warping, and Skeletal Tree Matching for Signature Verification," *IEEE Trans. in Pattern Analysis and Machine Intelligence,* **12**, No. 7, July 1990, pp710-717.

[Pratt 78] W. K. Pratt, *Digital Image Processing*, John Wiley & Sons, Inc., NY 1978.

[Pelizzari 89] C. A. Pelizarri, G. T. Y. Chen, D. R. Spelbring, R. R. Weichselbaum and C. T. Chen, "Accurate Three-Dimensional Registration of CT, PET and/or MR Images of the Brain," *J. of Computer Assisted Tomography,* **13**, January/February 1989, pp20-26.

[Price 85] K. E. Price, "Relaxation Matching Techniques–A Comparison," *IEEE Trans. on Pattern Analysis and Machine Intelligence* **7**, No. 5, September 1985, pp617-623.

[Ranade 80] S. Ranade and A. Rosenfeld, "Point Pattern Matching by Relaxation," *Pattern Recognition* **12**, pp269-275.

[Ratib 88] O. Ratib, L. Bidaut, H. R. Schelbert and M. E. Phelps, "A New Technique for Elastic Registration of Tomographic Images," *IEEE Proceedings SPIE: Medical Imaging II* **914**, 1988, pp452-455.

[Rosenfeld 82] A. Rosenfeld and A. C. Kak, *Digital Picture Processing*, Vol. I and II, Academic Press, Orlando, FL, 1982.

[Sanfeliu 90] A. Sanfeliu, "Matching Complex Structures: The Cyclic-Tree Representation Scheme," in *Structural Pattern Analysis*, World Scientific Pub. Co., Teaneck, NJ, 1990.

[Shapiro 90] L. G. Shapiro and R. M. Haralick, "Matching Relational Structures Using Discrete Relaxation," in *Syntactic and Structural Pattern Recognition, Theory and Applications*, World Scientific Pub. Co., Teaneck, NJ, 1990.

[Singh 79] M. Singh, W. Frei, T. Shibata G. C. Huth and N. E. Telfer, "A Digital Technique for Accurate Change Detection in Nuclear Medical Images - with application to Myocardial Perfusion Studies using

Thallium-201," *IEEE Trans. on Nuclear Science* **NS-26**, No. 1 February 1979.

[Slama 80] C. C. Slame, editor in chief, *Manual of Photogrammetry,*, Fourth Edition, American Society of Photogrammetry, Falls Church, VA, 1980.

[Svedlow 76] M. Svedlow, C. D. McGillem, and P. E. Anuta, "Experimental Examination of Similarity Measures and Preprocessing Methods Used for Image Registration," Symposium on Machine Processing of Remotely Sensed Data, June 1976, Purdue University, Indiana, pp4A-9.

[Stockman 82] G. C. Stockman, S. Kopstein and S. Benett, "Matching Images to Models for Registration and Object Detection via Clustering," *IEEE Trans. Pattern Analysis and Machine Intelligence* **4**, 1982, pp229-241.

[Terzopoulos 87] D. Terzopoulos, A. Witkin, and M. Kass, "Energy Constraints on Deformable Models: Recovering Shape and Non-Rigid Motion," *Proceedings, AAAI 87* **2**, July 1987, pp755-760.

[Van Wie 77] P. Van Wie and M. Stein, "A LANDSAT Digital Image Rectification System," *IEEE Trans. on Geoscience Electronics* **GE-15**, No. 3, July 1977, p130-137.

[Venot 84] A. Venot, J. F. Lebruchec, and J. C. Roucayrol, "A New Class of Similarity Measures for Robust Image Registration," *Computer Vision, Graphics, and Image Processing* **28**, 1984, pp176-184.

[Widrow 73] B. Widrow, "The Rubber Mask Technique, Part I and II," *Pattern Recognition* **5** No. 3, 1973, pp175-211.

[Wolberg 89] G. Wolberg and T. Boult, "Separable Image Warping with Spatial Lookup Tables," *ACM SIGRAPH '89 Conf. Proc., Computer Graphics* **23**, No. 3, Boston, MA, July/August 1989, pp369-377.

[Wolberg 90] G. Wolberg, *Digital Image Warping,* IEEE Computer Society Press, Los Alamitos, CA, 1990.

[Wong 85] A. K. C. Wong and M. You, "Entropy and Distance of Random Graphs with Application to Structural Pattern Recognition," *IEEE*

Trans. on Pattern Analysis and Machine Intelligence **PAMI-7**, No. 5, September 1985, pp599-609

[Wong 77] R. Y. Wong, "Sensor Transformations," *IEEE Trans. on System, Man and Cybernetics* **SMC-7**, No. 12, December 1977, pp836-840.