Chapter14

Uncertainty-Driven, Point-Based Image Registration

C. Stewart

Abstract

Point-based registration is the problem of computing the transformation that best aligns two point sets, such as might be obtained using range scanners or produced by feature extraction algorithms. The Iterative Closest Points (ICP) algorithm and its variants are the most commonly used techniques for point-based registration. The ICP algorithm may be derived as the solution to a global optimization problem. A commonly-used linearization of the distance function in this optimization problem produces a useful approximation to the covariance matrix of the ICP-estimated transformation parameters. Two recent algorithms exploit this covariance matrix to improve ICP registration. One uses the covariance matrix to sample the correspondences so that the estimate is well-constrained in all directions in parameter space. A second uses the covariance matrix to guide a region-growing and model-selection technique that "grows" accurate estimates from low-order initial estimates that are only accurate in small image regions. Both show substantial improvements over standard ICP on challenging alignment problems.

14.1 Introduction

Point-based registration techniques have been used in many applications, ranging from 3d modeling and industrial inspection to medical imaging. In point-based registration, the data are geometric point sets, $\mathcal{P}$ and $\mathcal{Q}$, such as image feature locations or 3d range measurements. The points are treated as samples from curves or surfaces in $\mathbb{R}^n$, and they may have associated attributes such as intensity values or normal vectors. The goal of point-based registration is to compute the transformation, $\mathbf{M} : \mathbb{R}^n \rightarrow \mathbb{R}^n$, that best aligns the point sets. Of particular interest here are parametric transformation models of the form $\mathbf{M}(\mathbf{p}; \boldsymbol{\theta})$, where $\mathbf{p} \in \mathbb{R}^n$ is point location, and $\boldsymbol{\theta}$ is the vector of transformation mapping parameters to be

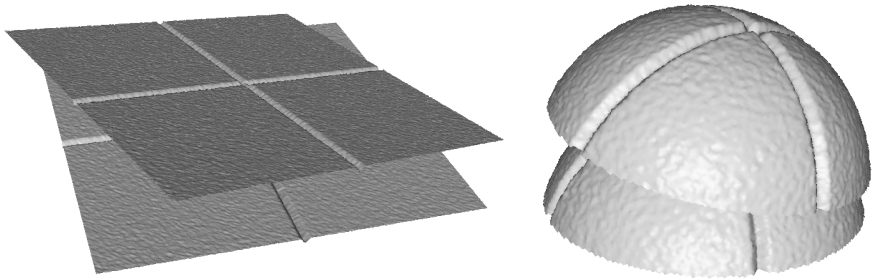

Figure 14.1. Synthetic range data sets illustrating the challenges that arise when the set of surfaces being aligned differ significantly in size. In the example on the left two planar surfaces have 1 mm deep groves cut into them. When a small amount noise is added to the data, constraints from matching points on the much larger planar surface prevent matches along the surface of the grooves from rotating the ICP alignment into place. A similar effect occurs with the alignment of two data sets from a spherical shell, shown on the right.

estimated. Similarity, affine, projective and quadratic transformations all fit into this category of parametric models.

Most approaches to point-based registration require establishing correspondence between points from $\mathcal{P}$ and $\mathcal{Q}$. If reliable correspondences are known, estimating the optimal set of transformation parameters is well-understood. On the other hand, given an accurate estimate of $\boldsymbol{\theta}$, establishing correspondence is straightforward. This poses a classic "chicken-and-egg" problem. This problem is widely addressed using the Iterative Closest Points (ICP) algorithm, discovered almost simultaneously in the early 1990's by several groups [82, 164, 185, 560, 916]. The idea of ICP is straightforward: (1) given a transformation parameter estimate, $\hat{\boldsymbol{\theta}}$, apply the transformation to a subset of $\mathcal{P}$, and for each transformed point find the closest point from $\mathcal{Q}$; (2) from these (temporary) correspondences, compute a new transformation parameter estimate $\hat{\boldsymbol{\theta}}$. These two steps are repeated until an appropriate convergence criteria is met. Important variations on ICP are discussed and analyzed in [697].

While initialization of ICP is clearly an important issue, the primary focus of this chapter is convergence. Ensuring proper convergence of ICP is challenging. Two reasons for this are illustrated in Figures 14.1 and 14.2. First, when there are significant variations in the sizes and the orientations of the surfaces to be registered, correspondence constraints from large surfaces can impede the alignment of smaller surfaces, mostly due to the effects of noise. Second, when the point sets represent complicated curve or surface patterns, such as in the vascular structure of the retina (Figure 14.2), misalignments early in the ICP process can cause mismatches that drive the algorithm to an incorrect local minimum. These mismatches often have relatively small alignment errors and therefore are not eliminated easily using robust estimation.

These two problems — one caused by a lack of balance in the constraints and one caused by incorrect correspondences — have been addressed recently

in papers from the 3d modeling literature [339], and from the medical imaging literature [765]. Underlying both is the use of uncertainty in the transformation estimate that is computed by ICP. Unlike earlier work, which studied the influence of uncertainty in point locations [276] and evaluated the uncertainty of the final ICP result [766], these two new techniques use uncertainty to guide the ICP estimation process itself. This new theme in registration could have important implications for developing more reliable and more general-purpose algorithms.

The goal of this chapter is to present this uncertainty-driven approach to registration. Section 14.2 formulates the point-based registration problem and derives both the ICP algorithm and the commonly-used normal distance form of ICP. Section 14.3 derives the transformation estimation equations and resulting approximate covariance matrix. This is used as a measure of uncertainty in the two algorithms described in Sections 14.4 and 14.5. The chapter concludes with a summary of the techniques and an outline of important questions suggested by the uncertainty-driven approach.

14.2 Objective Function, ICP and Normal Distances

Given are two point sets, $\mathcal{P}$ and $\mathcal{Q}$. These points sets are generally discrete, but they may be formed into a mesh. For expository purposes, however, they may be modeled in the continuous domain using an implicit function, e.g. $f : \mathbb{R}^n \rightarrow \mathbb{R}$, such that $\mathcal{Q} = \{\mathbf{q} \mid f(\mathbf{q}) = 0\}$. The point set registration objective function may be defined based on the proximity between transformed points from $\mathcal{P}$ and the set $\mathcal{Q}$:

$$F(\boldsymbol{\theta}; \mathcal{P}, \mathcal{Q}) = \sum_{\mathbf{p}_i \in \mathcal{P}} \min_{\mathbf{q} \in \mathcal{Q}} \|\mathbf{M}(\mathbf{p}_i; \boldsymbol{\theta}) - \mathbf{q}\|^2. \tag{14.1}$$

The goal of registration, now stated more formally, is to find the parameter estimate $\hat{\boldsymbol{\theta}}$ minimizing this objective function.

Several approaches to minimizing $F(\boldsymbol{\theta}; \mathcal{P}, \mathcal{Q})$ are possible. Here are two:

- The approach taken in the ICP algorithm alternates steps of solving the two minimization problems. The inner minimization (the matching step) in (14.1) is solved for fixed $\boldsymbol{\theta}$ to produce a correspondence set $\mathcal{C} = \{\mathbf{p}_i, \mathbf{q}_i\}$, and then the outer minimization is solved in slightly altered form by replacing the inner minimization with just the distance $\|\mathbf{M}(\mathbf{p}_i; \boldsymbol{\theta}) - \mathbf{q}_i\|^2$. If infinitesimal steps are taken in $\mathbf{q}_i$ and in $\boldsymbol{\theta}$, this converges to a local minimum of the objective function.

- $\mathcal{Q}$ is represented implicitly using a distance function in $\mathbb{R}^n$ that is 0 at locations $\mathbf{q}$ where $f(\mathbf{q}) = 0$. Example representations include Chamfer distance measures [103] and octree splines [164]. Derivatives of the objective function (14.1) may be computed based on computing derivatives of the distance function without explicitly identifying the closest point in $\mathcal{Q}$.

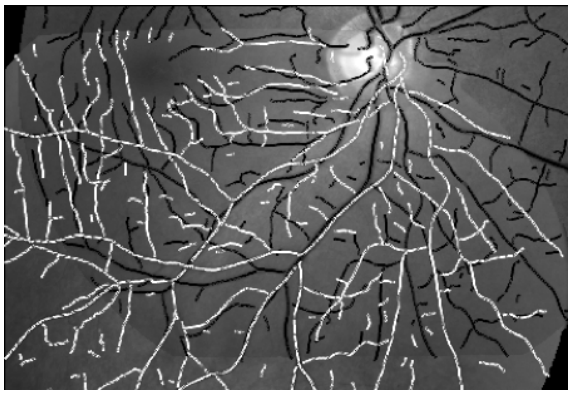

Figure 14.2. Example of misregistration of retinal images. Contours in black are blood vessel centerlines detected in one retinal image and contours in white are blood vessel centerlines detected in a second retinal image (of the same eye). The complexity of the structure of the vessels, together with a small initial misalignment, causes ICP to mismatch a significant fraction of the contours and converge to an incorrect estimate.

The focus of this chapter is on the ICP approach, which has been used widely, especially in the range image literature [697].

With the focus on ICP, the matching step must be examined in more detail. Using the implicit function definition of $\mathcal{Q}$, the minimization

$$\min_{\mathbf{q}\in\mathcal{Q}} \|\mathbf{M}(\mathbf{p}_i;\boldsymbol{\theta}) - \mathbf{q}\|^2 \tag{14.2}$$

becomes

$$\min \|\mathbf{M}(\mathbf{p}_i;\boldsymbol{\theta}) - \mathbf{q}\|^2 \text{ subject to } f(\mathbf{q}) = 0.$$

Writing this using Lagrange multipliers and introducing the simplifying notation $\mathbf{p}'_i \cong \mathbf{M}(\mathbf{p}_i;\boldsymbol{\theta})$ creates the function

$$h(\mathbf{q},\lambda) = \|\mathbf{p}'_i - \mathbf{q}\|^2 - 2\lambda f(\mathbf{q}),$$

which must be minimized simultaneously over $\mathbf{q}$ and λ. Computing partial derivatives $\partial h/\partial \mathbf{q}$ and $\partial h/\partial \lambda$ and setting the results equal to 0 yields

$$\begin{aligned} (\mathbf{p}'_i - \mathbf{q}) - \lambda \nabla f(\mathbf{q}) &= \mathbf{0} \\ f(\mathbf{q}) &= 0 \end{aligned} \tag{14.3}$$

Solving this, in turn, requires an iterative technique. Let $\mathbf{q}_i$ be the current best estimate of the closest point. After the iterations converge it will be the corresponding point for $\mathbf{p}_i$ in ICP. Linearizing f around $\mathbf{q}_i$ produces

$$f(\mathbf{q}) = (\mathbf{q} - \mathbf{q}_i)^T \boldsymbol{\eta}_i = 0 \qquad \text{and} \qquad \nabla f(\mathbf{q}_i) = \boldsymbol{\eta}_i,$$

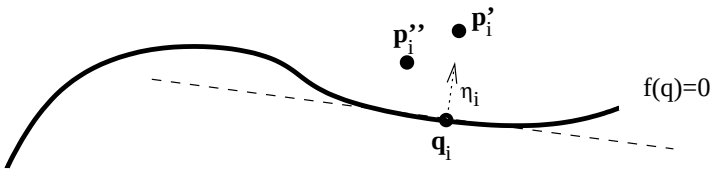

Figure 14.3. Illustrating the linearization of implicit function f that defines point set $\mathcal{Q}$. Let $\mathbf{p}'_i$ be a transformed point from $\mathcal{P}$, let $\mathbf{q}_i$ be the closest point from $\mathcal{Q}$, and let $\boldsymbol{\eta}_i$ be the local surface normal. (The linearization is pictured as the dashed line segment.) A small change in the transformation that moves $\mathbf{p}'_i$ to $\mathbf{p}''_i$ does not require recomputation of the closest point from $\mathcal{Q}$ in order to compute the (approximate) distance from $\mathbf{p}''_i$ to $\mathcal{Q}$.

where $\boldsymbol{\eta}_i$ is the normal to f at $\mathbf{q}_i$. Substituting these into (14.3) produces the system of equations

$$\begin{pmatrix} \mathbf{I} & \boldsymbol{\eta}_i \\ \boldsymbol{\eta}_i^T & 0 \end{pmatrix} \begin{pmatrix} \mathbf{q} \\ \lambda \end{pmatrix} = \begin{pmatrix} \mathbf{p}'_i \\ \boldsymbol{\eta}_i^T \mathbf{q}_i \end{pmatrix}.$$

Solving yields

$$\mathbf{q} = \mathbf{p}'_i - \boldsymbol{\eta}_i \boldsymbol{\eta}_i^T \mathbf{p}'_i + \boldsymbol{\eta}_i \boldsymbol{\eta}_i^T \mathbf{q}_i. \tag{14.4}$$

This produces an update $\mathbf{q}_i \leftarrow \mathbf{q}$. This point, however, does not satisfy $f(\mathbf{q}) = 0$, a problem that must be solved by moving along the constraint surface in direction $\mathbf{q} - \mathbf{q}_i$ rather than directly making the substitution $\mathbf{q}_i \leftarrow \mathbf{q}$. This important detail is not a concern here, however, because the current focus is on approximating the objective function.

The approximate closest point in (14.4) may be substituted back into the distance calculation equation (14.2) to yield a simplified but approximate calculation of distance. After some manipulation this yields,

$$\min_{\mathbf{q} \in \mathcal{Q}} \|\mathbf{M}(\mathbf{p}_i; \boldsymbol{\theta}) - \mathbf{q}\|^2 = a[(\mathbf{M}(\mathbf{p}_i; \boldsymbol{\theta}) - \mathbf{q}_i)^T \boldsymbol{\eta}_i]^2, \tag{14.5}$$

where $a = \boldsymbol{\eta}_i^T \boldsymbol{\eta}_i$. When $f(\mathbf{q})$ is a distance function, $a \approx 1$ because a unit step normal to the surface produces a unit change in distance. This is equivalent to assuming $\boldsymbol{\eta}_i$ is a unit vector, an assumption made throughout the remainder of this chapter. As illustrated in Figure 14.3, equation (14.5) simply reflects the fact that computing the minimum distance between a point and a linear structure does not require knowing the closest point on the linear structure; all that is needed is any point from the structure and the normal vector.

Turning back to the original problem of estimating the transformation parameters, (14.5) may be substituted into the original objective function (14.1) to obtain the approximation

$$F(\boldsymbol{\theta}; \mathcal{P}, \mathcal{Q}) = \sum_{\mathbf{p}_i \in \mathcal{P}} [(\mathbf{M}(\mathbf{p}_i; \boldsymbol{\theta}) - \mathbf{q}_i)^T \boldsymbol{\eta}_i]^2. \tag{14.6}$$

This approximation allows the calculation of the point-registration objective function without updating the correspondences.[1] It is valid as long as changes in the transformation parameters keep mapped points $\mathbf{M}(\mathbf{p}_i;\boldsymbol{\theta})$ in locations where the linearization around $\mathbf{q}_i$ is valid. This is used in deriving the covariance matrix in the next section. Equation 14.6 also leads to the "normal-distance" form of the ICP algorithm, originally proposed in [185]. The summation on the right-hand side of (14.6) is minimized for a fixed set of correspondences to estimate the next set of transformation parameters. The fact that this is a closer approximation to the true underlying objective function shows why use of normal distance constraints causes much faster and more reliable convergence of ICP [697].

14.3 Parameter Estimates and Covariance Matrices

The next step is to derive equations for estimating the transformation parameters given a fixed set of correspondences, $\mathcal{C} = \{(\mathbf{p}_i, \mathbf{q}_i)\}$. This leads directly to an approximation for the covariance matrix of the resulting estimate.

The derivation starts with a simplified form of the transformation model:

$$\mathbf{M}(\mathbf{p};\boldsymbol{\theta}) = \mathbf{p} + \mathbf{X}(\mathbf{p})\boldsymbol{\theta}. \tag{14.7}$$

A few examples will clarify this revised form. For a 3D rigid transformation using a small angle approximation (see [339], e.g.),

$$\mathbf{M}(\mathbf{p};\boldsymbol{\theta}) = \mathbf{R}\mathbf{p} + \mathbf{t} \approx \mathbf{p} + \mathbf{r} \times \mathbf{p} + \mathbf{t} = \mathbf{p} + \begin{pmatrix}\mathbf{S} & \mathbf{I}\end{pmatrix}\begin{pmatrix}\mathbf{r}\\ \mathbf{t}\end{pmatrix}.$$

Here, $\mathbf{r}$ is the vector of small angle approximations, $\mathbf{t}$ is the translation, and $\mathbf{S}$ is the skew-symmetric matrix such that $\mathbf{S}\mathbf{r} = \mathbf{r} \times \mathbf{p}$. This form is used for estimating incremental estimates of a rigid transformation. Writing an affine transformation in the form (14.7) is straightforward.[2] A 2D quadratic transformation is written

$$\mathbf{M}(\mathbf{p};\boldsymbol{\theta}) = \mathbf{p} + \begin{pmatrix}\mathbf{x}(\mathbf{p})^T & \mathbf{0}^T\\ \mathbf{0}^T & \mathbf{x}(\mathbf{p})^T\end{pmatrix}\boldsymbol{\theta}.$$

Here $\boldsymbol{\theta}$ is a 12x1 vector and if $\mathbf{p} = (u, v)^T$ then $\mathbf{x}(\mathbf{p}) = (1, u, v, u^2, uv, v^2)^T$.

Using the form of (14.7), the normal-distance ICP equation (14.6) for a fixed set of correspondences becomes

$$F(\boldsymbol{\theta};\mathcal{C}) = \sum_{(\mathbf{p}_i,\mathbf{q}_i)\in\mathcal{C}} [(\mathbf{p}_i + \mathbf{X}(\mathbf{p}_i)\boldsymbol{\theta} - \mathbf{q}_i)^T\boldsymbol{\eta}_i]^2. \tag{14.8}$$

[1] See [576] for a recent generalization to second-order approximations.

[2] Planar homographies may not be written in this form because side constraints must be imposed on the parameter vector. Different derivations of the estimation equations and covariance matrices are needed, combining the normal-distance form of (14.6) with the covariance derivations in [389, Ch. 4].

Rewriting,

$$\begin{aligned} F(\boldsymbol{\theta};\mathcal{C}) &= \sum_{(\mathbf{p}_i,\mathbf{q}_i)\in\mathcal{C}} [\boldsymbol{\eta}_i^T\mathbf{X}(\mathbf{p}_i)\boldsymbol{\theta} - \boldsymbol{\eta}_i^T(\mathbf{q}_i - \mathbf{p}_i)]^2 \\ &= (\mathbf{X}\boldsymbol{\theta} - \mathbf{y})^T(\mathbf{X}\boldsymbol{\theta} - \mathbf{y}) \end{aligned} \tag{14.9}$$

where

$$\mathbf{X} = \begin{pmatrix} \boldsymbol{\eta}_1^T\mathbf{X}(\mathbf{p}_1) \\ \vdots \\ \boldsymbol{\eta}_k^T\mathbf{X}(\mathbf{p}_k) \end{pmatrix} \qquad \text{and} \qquad \mathbf{y} = \begin{pmatrix} \boldsymbol{\eta}_1^T(\mathbf{q}_1 - \mathbf{p}_1) \\ \vdots \\ \boldsymbol{\eta}_k^T(\mathbf{q}_k - \mathbf{p}_k) \end{pmatrix}.$$

Taking the derivative with respect to $\boldsymbol{\theta}$, setting the result to $\mathbf{0}$, and solving yields the estimate,

$$\hat{\boldsymbol{\theta}} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}. \tag{14.10}$$

This has the structure of a linear regression problem. Making the simplifying assumption (discussed below) that $\mathbf{y}$ is the only random variable, the expected value of the estimate is

$$\overline{\boldsymbol{\theta}} = E[\hat{\boldsymbol{\theta}}] = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T E[\mathbf{y}].$$

Moreover, if $\mathbf{y}$ is independent and identically distributed (i.i.d.), with covariance matrix $\sigma^2\mathbf{I}$, then the covariance matrix of the parameter estimate is

$$\boldsymbol{\Sigma}_{\boldsymbol{\theta}} = E[(\hat{\boldsymbol{\theta}} - \overline{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \overline{\boldsymbol{\theta}})^T] = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1} \tag{14.11}$$

When robust weighting of the correspondences is added (see, e.g. [764]), the estimate becomes

$$\hat{\boldsymbol{\theta}} = \left(\mathbf{X}^T\mathbf{W}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{W}\mathbf{y}. \tag{14.12}$$

where $\mathbf{W}$ is a diagonal matrix of the weights of the individual constraints. The parameter estimate covariance matrix is then approximately

$$\boldsymbol{\Sigma}_{\boldsymbol{\theta}} = E[(\hat{\boldsymbol{\theta}} - \overline{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \overline{\boldsymbol{\theta}})^T] = \sigma^2(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1} \tag{14.13}$$

The approximate covariance matrix has been used in a number of algorithms, including the ones described here. Before proceeding to these, it is important to examine the assumptions and approximations underlying the foregoing derivation.

- The derivation of the covariance matrix that started from (14.6) is based on a fixed correspondence set. The prior derivation leading to (14.6) showed that (14.6) is a good approximation to the original objective function (which involves changing correspondences) when changes in the transformation are not large enough to invalidate the linearization around the points $\mathbf{q}_i$. This is true in particular as the overall algorithm — not just the estimate for a fixed set of correspondences — nears convergence.

- In deriving (14.11) from the estimate equation (14.10), the matrix $\mathbf{X}$ is assumed to depend only on deterministic quantities. For this to hold, point locations $\mathbf{p}_i$ are treated as deterministic. While this clearly underestimates the uncertainty, the effects of this should be small since the $\mathbf{p}_i$ values themselves will be much larger than errors in $\mathbf{p}_i$.

- Errors in the normal directions are assumed to be small enough that any resultant errors in projections onto the normal vectors — as in $\boldsymbol{\eta}_i^T \mathbf{X}(\mathbf{p}_i)$ and $\boldsymbol{\eta}_i^T(\mathbf{p}_i - \mathbf{q}_i)$ — are relatively insignificant. Since the errors in these projections will be proportional to the error in the orientation and since for small error angles, ϕ, $\cos\phi \approx 1$, this is reasonable, especially as the algorithm converges.

- Weight matrix $\mathbf{W}$ is also assumed to be non-random. Since each w_i depends on the error in the correspondence and therefore in the transformation itself, this is again an oversimplification.

- Finally, $\boldsymbol{\eta}_i^T(\mathbf{p}_i - \mathbf{q}_i)$ is assumed to be i.i.d. In part this says that all errors in the point positions are along the normal direction. On the negative side, this ignores errors that depend on the sensor direction [276]. On the positive side, since the point sets are treated as sets of samples from continuous manifolds, the errors in the point positions $\mathbf{q}$ tangent to the manifold keep the points (almost) on the manifold and do not change the distance measurement significantly.

Overall, it should be clear that the derived covariance matrix (a) is only a rough approximation of the true covariance matrix, (b) the approximation becomes more accuracte as the ICP estimation process nears the minimum, and (c) the primary effect of the approximation is that the magnitude of the covariance matrix is under-estimated.

14.4 Stable Sampling of ICP Constraints

This section and the next present applications of the covariance matrix estimate in ICP algorithms that address the two problems described in the introduction. This section considers the situation (Figure 14.1) where the ICP correspondences match points from the same surface in the two different data sets and are therefore in a sense "correct", but they still do not pull the estimate in the direction needed to correctly align the surfaces.

This problem is addressed in [339] by using the covariance matrix to select a subset of the correspondences that will constrain the transformation estimate as uniformly as possible in all directions. This sampling strategy is governed by a spectral decomposition of the parameter estimate covariance matrix and its

inverse:

$$\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} = \frac{1}{\sigma^2}(\mathbf{X}^T\mathbf{W}\mathbf{X}) = \sum_{j=1}^{m} \lambda_j \boldsymbol{\gamma}_j \boldsymbol{\gamma}_j^T, \qquad \boldsymbol{\Sigma}_{\boldsymbol{\theta}} = \sum_{j=1}^{m} (1/\lambda_j)\boldsymbol{\gamma}_j \boldsymbol{\gamma}_j^T. \quad (14.14)$$

The λ_j's and $\boldsymbol{\gamma}_j$'s are the eigenvalues and eigenvectors, respectively, of the inverse covariance matrix, ordered so that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m \geq 0$. The λ_j values represents the "stability" — the inverse of the variance — in direction $\boldsymbol{\gamma}_j$ in parameter space. Ideally, the stability values for each direction should be approximately equivalent. Stated another way, the condition number λ_1/λ_m should be as small as possible.

Consider the constraints from Equation (14.9) and in particular consider the projection of the constraint for correspondence i onto eigenvector j:

$$\boldsymbol{\eta}_i^T \mathbf{X}(\mathbf{p}_i)\boldsymbol{\gamma}_j. \quad (14.15)$$

The magnitude of this projection tells how much the ith point correspondence constrains the transformation in the jth direction in parameter space. Given a subset $\mathcal{C}'$ of the correspondence set $\mathcal{C}$, the value

$$s_j^2 = \sum_{(\mathbf{p}_i, \mathbf{q}_i) \in \mathcal{C}'} [\boldsymbol{\eta}_i^T \mathbf{X}(\mathbf{p}_i)\boldsymbol{\gamma}_j]^2 / |\mathcal{C}'| \quad (14.16)$$

is roughly proportional to the inverse variance of the estimate in the jth direction based on the subset. The goal of the stable sampling algorithm is to find a subset that makes these s_j^2 values as close to equal as possible, thereby constraining the estimate equally-well in all directions.

The steps involved are:

1. Compute the inverse covariance matrix and its eigenvector decomposition from a small initial set $\mathcal{C}'$ of correspondences in the region where the data sets overlap. These correspondences and the overlap region must be computed using an earlier ICP parameter estimate.

2. Compute s_j^2 for each eigenvector based on the initial set.

3. For eigenvector j with the smallest s_j^2, choose the correspondence from the overlap region that has the greatest magnitude of (14.15), add it to the correspondence set $\mathcal{C}'$, and update s_j^2 (14.16) for all eigenvectors. Note that the chosen correspondence is taken from $\mathcal{C} - \mathcal{C}'$.

4. Repeat until a sufficient number of correspondences have been selected or until the addition of a new correspondence starts to increase the approximate condition number — the ratio between the largest s_j value and the smallest. The second condition tests if the constraints available to increase the stability of the smallest eigenvalue have been exhausted.

For details of the data structures and search algorithms that make this computation efficient, see [339]. Two other important details should be mentioned here, however.

- The parameter vector $\boldsymbol{\theta}$ involves parameters of different units, including rotation angles and terms of differing orders. Numerically, the individual components of $\boldsymbol{\theta}$ are not comparable; they can differ by several orders of magnitude. To solve this, the subsets of $\mathcal{P}$ and $\mathcal{Q}$ that form the correspondences should each be centered and then normalized so that the average magnitude of corresponding points $\mathbf{p}_i$ and $\mathbf{q}_i$ are each 1 [339, 389]. All computations of the sample selection technique should be done in the centered and normalized system.
- The constraint $\boldsymbol{\eta}_i^T \mathbf{X}(\mathbf{p}_i)$ depends on a point location from $\mathcal{P}$ and a normal from $\mathcal{Q}$. This means sampling must be applied after correspondences are formed, even though many correspondences will not be used. This wasted computation may be avoided easily. Observe that after the ICP algorithm has removed the worst of the misalignments, the surface normals of the transformed points $\mathbf{p}_i$ should be roughly parallel to the normals from $\mathbf{q}_i$. Therefore, the transformed normals from $\mathbf{p}_i$ can be used in place of the normals from $\boldsymbol{\eta}_i$ in the above calculations. This means the sampling can be computed prior to establishing correspondence.

The overall computation places a third step in each iteration of ICP: (1) apply stable sampling to select a subset of the points in the overlap region, (2) establish matches (correspondences) for these points, and (3) compute the new transformation estimate using the correspondences.

Using this technique, the two problem examples shown in Figure 14.1 are each correctly aligned. For the iteration starting from the positions shown in the figure, the condition numbers dropped from 66.1 to 3.7 for the planes and from 26.9 to 4.1 for the spheres using stable sampling. The RMS alignment errors after ICP converged using stable sampling were in each case a factor of 3 lower than when a spatially-uniform sampling of point set $\mathcal{P}$ was used. See [339] for more examples.

14.5 Dual-Bootstrap ICP

The second algorithm that exploits the covariance matrix during the registration process is designed to avoid the problem of mismatches due to poor initialization. The problem occurs in particular in the registration of retinal images because of the complexity of the vascular structure and the effects of disease.

The Dual-Bootstrap algorithm described here uses points detected along the centers of blood vessel curves [146, 331] as the registration point sets $\mathcal{P}$ and $\mathcal{Q}$. Registration is initialized using matches between landmarks — branching and cross-over points of the vessels — detected in the two images. Unfortunately, images with significant pathologies sometimes have very few landmarks and even fewer that match correctly for initialization. Therefore, the approach taken is a hypothesize-and-test method, where single correspondences are generated to form initial transformation estimates that are only accurate in small image regions. The Dual-Bootstrap algorithm tests each small region and initial estimate separately

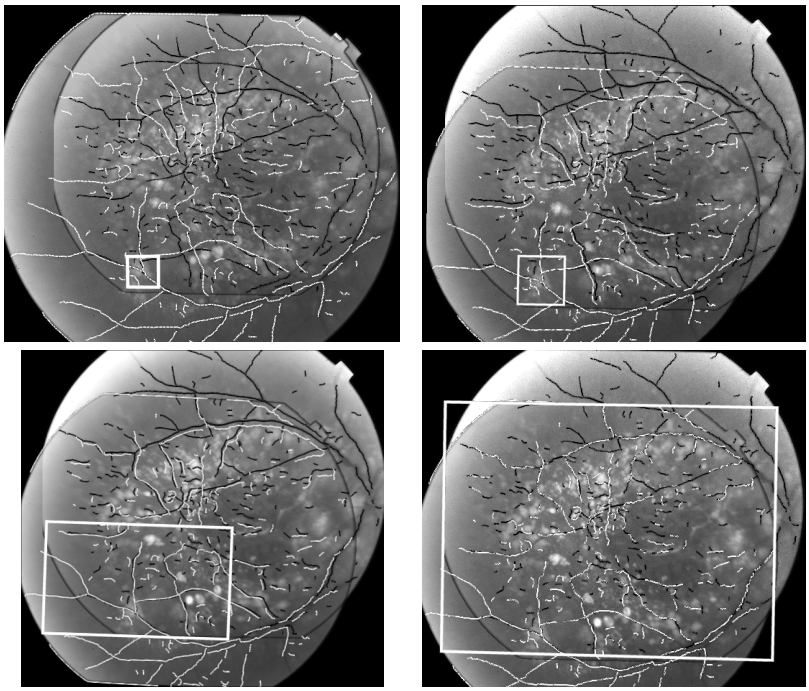

Figure 14.4. Initial (upper left), intermediate (upper right, lower left), and final (lower right) results of the Dual-Bootstrap ICP algorithm on a pair of retinal images. Vessel centerlines forming the point sets $\mathcal{P}$ and $\mathcal{Q}$ are shown using white and black contours. The rectangle drawn on top of the images shows the current region, R. The images are well-aligned within R in each iteration, and as R is expanded to cover the entire overlap region, the overall estimate converges to an accurate alignment.

by "growing" an image-wide transformation estimate. If the initial transformation is moderately accurate the Dual-Bootstrap algorithm rarely fails to produce an accurate result.

Dual-Bootstrap ICP works by iterating three steps, illustrated in Figure 14.4:

1. It applies one iteration of ICP using only points from the current region, R (the highlighted rectangle in the panels of Figure 14.4).

2. Based on the correspondences and the covariance matrix, the best transformation model is selected from among a set of possible models. Initially, when the region is small, there are only sufficient constraints for a similarity transformation. The eventual image-wide transformation is a quadratic model [147]. In between, the algorithm can select an affine transformation or a simplified version of the quadratic transformation.

3. The Dual-Bootstrap algorithm uses the uncertainty in the transformation to expand the boundary of the region, R. More stable transformations lead to faster region growth.

These steps are repeated until the entire process converges for the given initial estimate. If the final estimate covers the apparent overlap between images and is sufficiently accurate and stable, the estimate is accepted as correct. Otherwise, another starting landmark correspondence and associated region is tried. This greedy process terminates and indicates that no alignment is possible if the initial possibilities are exhausted.

The model selection and region growing steps are most relevant to the theme of this chapter, so they are discussed in more detail in the remainder of this section.

Model selection techniques [135, 799] choose the model that optimizes the trade-off between the alignment accuracy of high-order models and the stability of low-order models, with stability being measured using the covariance matrix of the parameters. The Dual-Bootstrap ICP model selection criteria is based on the expression (see [135] for a derivation):

$$\frac{d}{2}\log 2\pi - \sum_i w_i r_i^2 + \frac{1}{2}\log\det(\Sigma_{\hat{\boldsymbol{\theta}}}), \tag{14.17}$$

where d is the number of degrees of freedom in the model, $\sum_i w_i r_i^2$ is the sum of the robustly-weighted alignment errors ($r_i = (\mathbf{M}(\mathbf{p}_i;\hat{\boldsymbol{\theta}}) - \mathbf{q}_i)^T\boldsymbol{\eta}_i$), and $\det(\Sigma_{\hat{\boldsymbol{\theta}}})$ is the determinant of the parameter estimate covariance matrix. Intuitively, for higher-order models d increases, $-\sum_i w_i r_i^2$ increases (because the residuals decrease), and $\det(\Sigma_{\hat{\boldsymbol{\theta}}})$ decreases because the models are less stable. In choosing the best model, (14.17) is evaluated for a set of models using a fixed correspondence set. The model with the greatest value of (14.17) is chosen.

The growth of the region in step 3 of the Dual-Bootstrap algorithm is based on the uncertainty in the mapping of point locations on the boundary of the regions. This uncertainty is computed from the covariance of the transformation parameter estimate using fairly standard covariance propagation techniques, often called the "transfer error" [389, Ch. 4] in the computer vision literature. As before, let $\mathbf{p}' = \mathbf{M}(\mathbf{p},\hat{\boldsymbol{\theta}})$ be the mapping of point location $\mathbf{p}$. The covariance of this mapping is approximately

$$\boldsymbol{\Sigma}_{\mathbf{p}'} = \mathbf{J}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\mathbf{J}^T$$

where

$$\mathbf{J} = \frac{\partial\mathbf{M}}{\partial\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}) = \mathbf{X}(\mathbf{p}),$$

using the definition of $\mathbf{M}$ from (14.7). No uncertainty in $\mathbf{p}$ is considered because $\mathbf{p}$ is treated simply as a position in the coordinate system of set $\mathcal{P}$, not an estimated point location.

The transfer error is used to expand each of the four sides of region rectangle R (Figure 14.5). Let $\mathbf{p}_s$ be one of these points, described in a coordinate system

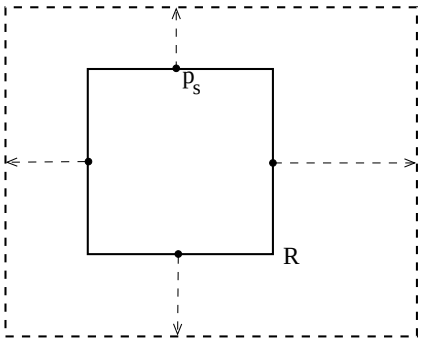

Figure 14.5. Expansion of the region R in the Dual-Bootstrap ICP algorithm. The center of each side of the region rectangle is pushed outward in inverse proportion to the transfer error variance. This means that more certainty in the transformation leads to faster growth in R. The new region is the axis-aligned rectangle formed by the four outwardly-moved points.

centered on the rectangle, and let $\mathbf{p}'_s$ be its mapping into the coordinate system of $\mathcal{Q}$. Let $\boldsymbol{\eta}_s$ be the outward normal of the side of the rectangle and let $\boldsymbol{\eta}'_s$ be the mapping of this normal into the coordinate system of $\mathcal{Q}$. The variance of $\mathbf{p}'_s$ in the outward direction is $\sigma_s^2 = {\boldsymbol{\eta}'_s}^T \boldsymbol{\Sigma}_{\mathbf{p}'_s} \boldsymbol{\eta}'_s$. Using this, the outward movement of $\mathbf{p}_s$ is:

$$\Delta \mathbf{p}_s = \beta \frac{(\mathbf{p}_s^T \boldsymbol{\eta}_s)}{\max(1, \sigma_s^2)} \tag{14.18}$$

This growth is proportional to the current distance $(\mathbf{p}_s^T \boldsymbol{\eta}_s)$ of $\mathbf{p}_s$ from the center of R, and is inversely proportional to the transfer error in the normal direction. The lower bound of 1 in the denominator prevents growth from becoming too fast. The center of each side of R is expanded outward independently using Equation 14.18, and the new region is the axis-aligned rectangle formed by the resulting four points (Figure 14.5). Parameter β controls the growth rate; the setting used in practice, $\beta = \sqrt{2} - 1$, ensures that the area of R at most doubles in each iteration.

The Dual-Bootstrap ICP algorithm has been tested on thousands of retinal image pairs, including images of unhealthy eyes in various stages of disease progression [765, 810]. Overall, when there is at least 30% overlap between images, at least one starting correspondence, and enough extracted vessels to form a stable covariance matrix, the algorithm never fails. Together, the region growth and model selection techniques work to keep the algorithm near the optimal estimate within region R. More detail about the behavior of these techniques is as follows:

- Model selection is imperfect. The algorithm tends to switch to higher-order models too early. Estimation errors in these higher-order models may lead to more mismatches, especially on the region periphery. Empirically, the implementation uses the heuristic that the quadratic model may not be used until the region has grown to 20% of the image size. A likely cause of this

problem is that the covariance matrix used underestimates the amount of uncertainty.

- Region growth, on the other hand, works extremely well. One measure of this is that halving or doubling the growth rate does not change the effectiveness of the algorithm. Removing region growth altogether, on the other hand, reduces the number of image pairs that the algorithm is able to align by 16%.

14.6 Discussion and Conclusion

This chapter has addressed the problem of point-based registration, focusing on the use of covariance-based techniques to improve the performance of the iterative closest point (ICP) algorithm in both range image and retinal image registration. The chapter started by formulating the objective function and then deriving the normal distance version of ICP. This provides a locally-accurate approximation to the overall objective function without the need for rematching. This approximation was then used to derive the equations for estimating the transformation parameters and the covariance matrix of this estimate. Several simplifying assumptions were used in deriving this matrix. These assumptions lead to an underestimate in the overall amount of uncertainty, but are a reasonable approximation as the overall ICP process nears convergence.

The chapter then summarized two algorithms in which the covariance matrix is used to modify the behavior of ICP. In the stable sampling algorithm of [339], the covariance matrix is used to guide the selection of correspondences, ensuring that all directions in parameter space are well-constrained. Geometrically, this allows the ICP algorithm to accurately align small-scale surfaces. In the Dual-Bootstrap ICP algorithm of [765], the covariance matrix is used to grow a transformation estimate and its associated region, starting from a small region surrounding a single correspondence. The covariance matrix helps avoid mismatches between vascular structures by controling the growth of the region and the selection of transformation models. Empirical results show that in both algorithms the use of the covariance matrix substantially improves the registration results.

The algorithms work well despite the approximations needed to compute the covariance matrix. The main reason for this effectiveness is that the covariance matrix plays its most important role as the algorithms near convergence. Stable sampling only has a significant effect when the dominant structures of the data are well-aligned — the small surface misalignments then appear in the eigenvectors of the smaller eigenvalues and may therefore be corrected through the sampling procedure. In the Dual-Bootstrap algorithm, the alignment is always close to convergence in region R, even when the alignment appears to be poor throughout the image. This means that the parameter estimate covariance matrix may be used to guide the growth and model selection based only on points from R.

The work described here offers a new approach to improving the performance of registration algorithms — using the uncertainty in the estimates being computed to guide further steps in the overall algorithm. This is reminiscent of recursive estimation techniques such as the Kalman filter [637], but in the new algorithms uncertainty is used more broadly, beyond the estimation equations themselves. This could point toward the development of a variety of new algorithms. Moving in this direction requires that a number of issues be addressed. On the theoretical side, a new and more accurate approximation of the covariance matrix is needed that depends on fewer assumptions. One approach might be the use of resampling methods such as the bootstrap technique from statistics [295]. On the more applied side, a second advance would be integrating uncertainty-driven methods with approaches to initialization based on keypoint matching [121]. A third advance would be incorporating uncertainty information into deformable registration, one of the most important problems in medical image analysis.

Acknowledgements

Portions of this research were supported by National Science Foundation Experimental Partnerships grant EIA-0000417, by the Center for Subsurface Sensing and Imaging Systems under the Engineering Research Centers Program of the National Science Foundation (Award Number EEC-9986821), and by the US Army INSCOM. Thanks go to Rich Radke and Charlene Tsai for comments that significantly enhanced the presentation and to Natasha Gelfand for the images in Figure 14.1.