The Softassign Procrustes Matching Algorithm

Anand Rangarajan[1], Haili Chui[1] and Fred L. Bookstein[3]

[1] Departments of Diagnostic Radiology and Electrical Engineering
Yale University
[2] Institute of Gerontology
University of Michigan, Ann Arbor

Abstract. The problem of matching shapes parameterized as a set of points is frequently encountered in medical imaging tasks. When the point-sets are derived from landmarks, there is usually no problem of determining the correspondences or homologies between the two sets of landmarks. However, when the point sets are automatically derived from images, the difficult problem of establishing correspondence and rejecting non-homologies as outliers remains. The Procrustes method is a well-known method of shape comparison and can always be pressed into service when homologies between point-sets are known in advance. This paper presents a powerful extension of the Procrustes method to point-sets of differing point counts with correspondences unknown. The result is the softassign Procrustes matching algorithm which iteratively establishes correspondence, rejects non-homologies as outliers, determines the Procrustes rescaling and the spatial mapping between the point-sets.

1 Introduction

One of the most commonly encountered problems in medical imaging is image matching: the process of superimposing one image onto another such that an appropriate measure of goodness of match is minimized. The plethora of image matching methods span the spectrum from matching user specified *landmarks* to directly matching image intensities (see [19] for an excellent review). When we restrict the focus to feature matching methods, the problem difficulty changes considerably depending on whether we have well defined landmarks or merely unlabeled features. The essential difficulty in feature matching methods lies in the problem of establishing correspondences or homologies. When we have landmarks, the correspondence between landmarks in one image and the other are usually known in advance. All that is required is to find a suitable spatial mapping (rigid, similarity, affine, non-rigid) between the two sets of landmarks. On the other hand, when we merely have two sets of unlabeled point features, the problem of correspondence is acute [2]. In addition, when the two *point-sets* are of different point count, we are saddled with the problems of outlier (non-homology) rejection as well as the problem of discovering homologies.

The Procrustes method of shape comparison arose as a way of superimposing point-sets with known correspondence [8]. Drawing on the insight that shape can be represented as point-sets with size removed, the Procrustes method first sought to normalize the two point sets prior to superimposition. The normalization operation essentially consisted of centering and scaling each point-set such that the centroid shifted to the (common) origin and the sums of squared distances of the points was unity. The normalization operation annihilates translation and scale factors. Consequently, the search for a similarity transformation that aligns the point-sets can be restricted to a search for rotation alone. (Note that non-rigid operations such as thin-plate splines can be subsequently used to discover localized regions of dissimilarity between the point-sets [4]). The obvious question that arises at this juncture is: What happens to the Procrustes method when we do not know the correspondence between the point-sets?

The present paper answers this question and provides an algorithm for determining the correspondences or homologies and rejecting outliers or non-homologies. We see this as a powerful extension of the Procrustes method, capable of radically extending the scope of application of Procrustes methods in medical imaging. Central to our approach is the iterative combination of the search for correspondence, the spatial mapping and Procrustes rescaling. Correspondence is parameterized via a binary *match matrix* that assigns points in one set to points in the other and discards non-homologies as outliers. Parameterized in this manner, the search for correspondence is tantamount to solving *the linear assignment problem* [12]. The *softassign*, a new technique that first arose in the statistical physics and neural computation literature [14, 12] has been shown to find the optimum solution to the assignment problem. Using the softassign, our algorithm essentially alternates between solving for correspondences, the spatial mapping and Procrustes rescaling.

2 The Procrustes method revisited

The modern Procrustes literature began with Kendall's great 1984 paper on the relation between the Procrustes formalism and differential geometry in the large [11]. The recipe for calculating the standard Procrustes distance formula involves centering and scaling each form (point-set) such that the sums of squared distances of each point-set is unity and then evaluating the Euclidean distance between the point-sets after superimposing the two forms [8]. Kendall noted that this standard formula for Procrustes distance is only an approximation to the correct quantity, which is $\arccos|\sum_i z_i \bar{z}'_i|$ when 2-D shapes are written as complex vectors z_i (and z'_i) with $\sum_i z_i = 0$ and $\sum_i z_i \bar{z}_i = 1$. Kendall shows how this metric arises naturally from the construal of shapes as equivalence classes of point-sets of the plane under the operation of the similarity group (the same rotations, translations, and rescalings we have been pursuing here all along), and, furthermore, how it is the *only* such metric that satisfies certain reasonable symmetries arising from the symmetries of Euclidean distance in the original Cartesian image space. For more on the resulting smooth Riemannian manifold, which is quite a remarkable geometric object see [17].

In this Kendall formulation, which is now standard in all other applications of Procrustes methods, Procrustes distance is not determined as the sum of squares of the Euclidean distances between the two centered and scaled point-sets. While it happens to agree (approximately) with the minimum sum of squares over the operation of the similarity group, it is defined in a way independent of the selection of an element of that group. It is interesting to consider how the combinatorial context taken in this paper constrains the elegance of this more general mathematical setting.

The present paper has extended the Kendall metric, in its approximate form, to a distance function over the *union* of a sequence of these shape manifolds, one for each different count of matching points. When we change the count of outliers, we change manifolds. Other approaches to this problem have worked only in one of these manifolds at a time. For instance, [9] and [5] deal with the closely related problem of minimizing metrics within linear subspaces of the tangent space to the space of these shapes. The uniqueness of this paper lies in the combinatorial extension of the Procrustes method in order to handle the problems of determining correspondences and rejecting outliers.

3 Combining the search for correspondence and the spatial mapping

In its most basic form, the Procrustes distance requires us to center and scale each set of points so that the sum of squared distances of all points in each point-set is unity and then compute a similarity transformation between them. Then the Procrustes average shape and Procrustes residuals can be evaluated. How does this situation change when we do not have information about the homologies or point-to-point correspondences between the two point-sets? The Procrustes distance presupposes correspondence. When the homologies are unknown, we cannot even carry out the first step in the Procrustes distance calculation, namely the centering and scaling of each point-set. The reason: Ignorance of correspondence also implies ignorance of point outliers which hampers the centering and scaling of each point-set. Unfortunately, determining the correspondence beforehand is not easy if not impossible. In order to solve this chicken and egg problem, we embark on a joint formulation of the Procrustes distance and correspondence.

3.1 Theory

Denote the point-sets by $X_i \in \mathcal{R}^2,\ i = 1, 2, ..., N_1$ and $Y_j \in \mathcal{R}^2,\ j = 1, 2, ..., N_2$ respectively. N_1 and N_2 are the numbers of points in the sets X and Y respectively. When $N_1 = N_2 = N$ and the correspondences are known, the Procrustes distance between X and Y (after computing a similarity transformation between X and Y) is:

$$D_{\text{Procrustes}}(X, Y) = \sum_{i=1}^{N} \| \frac{(X_i - \mu_X)}{\sigma_X} - R(\theta) \frac{(Y_i - \mu_Y)}{\sigma_Y} \|^2. \tag{1}$$

where

$$\theta = \arctan\left(\frac{\sum_{i=1}^{N}\left[X_i^c(2)Y_i^c(1) - X_i^c(1)Y_i^c(2)\right]}{\sum_{i=1}^{N}\left[X_i^c(1)Y_i^c(1) + X_i^c(2)Y_i^c(2)\right]}\right) \tag{2}$$

and

$$X_i^c = \frac{X_i - \mu_X}{\sigma_X}, \text{ and } Y_i^c = \frac{Y_i - \mu_Y}{\sigma_Y}.$$

The numbers inside the parenthesis [$X(1)$ or $X(2)$] denote the abscissa and ordinate of the point-set. In (1), μ_X and μ_Y are the centroids of the two point-sets X and Y respectively and σ_X and σ_Y are the sums of squared distances of each point-set from the centroid.

$$\mu_X = \frac{\sum_{i=1}^{N} X_i}{N},\ \mu_Y = \frac{\sum_{i=1}^{N} Y_i}{N},\ \sigma_X^2 = \sum_{i=1}^{N} \|X_i - \mu_X\|^2 \text{ and } \sigma_Y^2 = \sum_{i=1}^{N} \|Y_i - \mu_Y\|^2. \tag{3}$$

When the correspondence and the similarity transformation relating the two point-sets are unknown, we can combine the estimation of all of these parameters in the following manner:

$$E(M,\theta,t,s) = \sum_{i=1}^{N_1}\sum_{j=1}^{N_2} M_{ij}\left(\|\frac{\sqrt{s}(X_i - \mu_X)}{\sigma_X} - t - R(\theta)\frac{(Y_j - \mu_Y)}{\sqrt{s}\sigma_Y}\|^2 - \alpha\right)$$

$$\text{subject to} \sum_{i=1}^{N_1} M_{ij} \leq 1,\ \sum_{j=1}^{N_2} M_{ij} \leq 1, \text{ and } M_{ij} \in \{0,1\}. \tag{4}$$

Equation (4) describes an optimization problem from which the transformation parameters—rotation matrix $R(\theta)$, translation t, scale s and the two centroids μ_X and μ_Y—can be obtained by minimization. (Ignore for the moment the estimation of the two "variance" parameters σ_X and σ_Y.) However, (4) also sets up an optimization problem on the point correspondences. A set of correspondence variables $\{M_{ij}\}$—a *match matrix*—has been defined such that:

$$M_{ij} = \begin{cases} 1 \text{ if point } X_i \text{ corresponds to point } Y_j \\ 0 \text{ otherwise,} \end{cases}$$

M is a correspondence match matrix which indicates when homologies have been found. If M_{ij} is one, feature "i" in slice 1 and feature "j" in slice 2 are homologies. The constraints on M_{ij} enforce one-to-one correspondence between homologies and also robustness [3]: a point feature in one image may have no corresponding homology and should be discarded as an outlier. The parameter $\alpha > 0$ controls the degree of robustness: as α is increased, less points in both sets are rejected as outliers. (Finer control over robustness is possible by having two parameters, one for each set.) While estimation of α is obviously an important problem, we do not address this issue here. This technique is closely related to the robust statistics approach based on M-estimators and influence functions [3].

Before describing our approach to solving for correspondence, it is worth examining the spatial mapping in (4). In (4), μ_X, μ_Y, σ_X, σ_Y, t, s and θ constitute the spatial mapping parameters. Since the correspondence matrix M is also unknown, Equation (3) for the centroids and the variances cannot be used. While we can differentiate (4) and solve for the centroids, our first aim is to simplify (4) itself. To this end, we begin with the following Ansatz: the centroids and variances in (4) are set using a weighted mean and variance formula:

$$\begin{aligned}
\mu_X &= \frac{\sum_{i=1}^{N_1}\sum_{j=1}^{N_2} M_{ij}X_i}{\sum_{i=1}^{N_1}\sum_{j=1}^{N_2} M_{ij}}, \\
\mu_Y &= \frac{\sum_{i=1}^{N_1}\sum_{j=1}^{N_2} M_{ij}Y_j}{\sum_{i=1}^{N_1}\sum_{j=1}^{N_2} M_{ij}}, \\
\sigma_X^2 &= \sum_{i=1}^{N_1}\sum_{j=1}^{N_2} M_{ij}\|X_i - \mu_X\|^2, \\
\sigma_Y^2 &= \sum_{i=1}^{N_1}\sum_{j=1}^{N_2} M_{ij}\|Y_j - \mu_Y\|^2
\end{aligned} \tag{5}$$

The intuition behind this Ansatz will become clearer as we proceed. Having set the four parameters—the centroids and the variances—we now estimate the translation t and the scale s parameters.

Differentiating (4) w.r.t. the translation parameter t and setting the result to zero, we get

$$t = \sum_{i=1}^{N_1}\sum_{j=1}^{N_2} M_{ij}\left[R(\theta)\frac{(Y_j - \mu_Y)}{\sqrt{s}\sigma_Y} - \frac{\sqrt{s}(X_i - \mu_X)}{\sigma_X}\right] \;=\; 0. \tag{6}$$

Equation (6) directly follows from the Ansatz for the centroids. Similarly, differentiating (4) w.r.t. the scale parameter s and setting the result to zero, we get

$$s^2 = \frac{\sum_{i=1}^{N_1}\sum_{j=1}^{N_2} M_{ij}\|X_i - \mu_X\|^2\sigma_Y^2}{\sum_{i=1}^{N_1}\sum_{j=1}^{N_2} M_{ij}\|Y_j - \mu_Y\|^2\sigma_X^2} = 1 \tag{7}$$

Equation (7) directly follows from the Ansatz for the variances. We have shown that the translation and scale parameters can be removed from the objective function since they are annihilated by the centroids and variances. From (6) and (7), we may simplify the objective function in (4):

$$\min_{M,\theta} E(M,\theta) = \sum_{i=1}^{N_1}\sum_{j=1}^{N_2} M_{ij}\left(\|\frac{(X_i - \mu_X)}{\sigma_X} - R(\theta)\frac{(Y_j - \mu_Y)}{\sigma_Y}\|^2 - \alpha\right)$$

$$\text{subject to} \sum_{i=1}^{N_1} M_{ij} \le 1, \; \sum_{j=1}^{N_2} M_{ij} \le 1, \text{ and } M_{ij} \in \{0,1\}. \tag{8}$$

The match matrix M is actually a correspondence matrix. Assume for the moment that the correspondences M_{ij} and the rotation matrix $R(\theta)$ are known. It is easy to show that the Ansatz for the centroids and the variances reduces to the Procrustes centroids and variances in (3). In addition, the objective function in (8) reduces to the Procrustes distance measure between the point-sets X and Y.

Merely showing that the Procrustes distance emerges when we set the correspondence and the rotation matrices is not enough. We need an algorithm for the simultaneous estimation of the correspondence and rotation matrices as well as the the centroids and variances since they are functions of the match matrix.

The above optimization problem in (8) contains two interlocking optimization problems—one on the rotation between the two point-sets and the other on the point-to-point feature correspondences. When the correspondences are known, a constrained *least squares* problem results from which the rotation matrix can be obtained. When the rotation is known, we obtain a *linear assignment* problem [1] for the correspondences. (The presence of outliers makes our problem slightly different from the traditional linear assignment problem, but for the most part we ignore the technical distinction between the two.) While polynomial time linear assignment algorithms exist in the literature, we'll describe a new algorithm that is better suited to interact with the constrained least squares solution for the rotation. The result is a two stage algorithm which alternates between solving for the rotation and the correspondence match matrix.

The main difficulty in jointly solving for the rotation and the correspondence lies in the seemingly disparate natures of the two problems, one continuous, the other combinatorial. Our approach is designed to overcome this problem. First, ignore the effect of outliers on the match matrix in (8). The match matrix then becomes a *permutation matrix* with binary entries and all rows and columns summing to one. Then we invoke the Birkhoff-von Neumann theorem which states that "the set of $(N \times N)$ doubly stochastic matrices is the convex hull of the set of $(N \times N)$ permutation matrices" [12]. (A doubly stochastic matrix is a square matrix with all positive entries and rows and columns summing to one.) In other words, permutation matrices (or the permutation group) are the vertices of an N dimensional doubly stochastic polytope. Consequently, the linear assignment problem can be solved by linear programming and related interior-point methods which relax the constraint that the match matrix entries be binary and merely require them to be positive. Due to the linear objective, the solution always occurs at a vertex of the doubly stochastic polytope—a permutation matrix. Following the Birkhoff–von-Neumann theorem, we relax the match matrix constraints from permutation matrix constraints to doubly stochastic matrix constraints. (When outliers are included, the above description is slightly modified). Doubly stochastic matrix constraints require positivity of each match matrix entry with rows and columns summing to one. We enforce the row and column constraints using Lagrange parameters and the positivity constraint via a *barrier function* [13]:

$$F(M, \theta, \kappa, \lambda) = \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} M_{ij} \left(\| \frac{(X_i - \mu_X)}{\sigma_X} - R(\theta) \frac{(Y_j - \mu_Y)}{\sigma_Y} \|^2 - \alpha \right) \quad (9)$$

$$+\sum_{i=1}^{N_1} \kappa_i \left(\sum_{j=1}^{N_2+1} M_{ij} - 1 \right) + \sum_{j=1}^{N_2} \lambda_j \left(\sum_{i=1}^{N_1+1} M_{ij} - 1 \right) + \frac{1}{\beta} \sum_{i=1}^{N_1+1} \sum_{j=1}^{N_2+1} M_{ij} \log M_{ij}.$$

In (9), κ and λ are two Lagrange parameters enforcing the row and column constraints respectively and $\beta > 0$ is a control parameter of an $x \log x$ *entropy* barrier function [14, 20] which enforces the positivity constraint. We have transformed the original mixed continuous-combinatorial optimization problem into a nonlinear optimization problem.

The entropy barrier function used above can be formally derived using the *saddle-point* approximation—a well known technique in statistical physics [10, 20]. The details are beyond the scope of this paper.

3.2 The Softassign Procrustes Algorithm

Moving on to the development of the algorithm, first note that all quantities in (9) are continuous valued. We now derive an algorithm which alternates between finding the spatial mapping and the correspondences.

Fixing the rotation for the moment, we can solve for the match matrix M in the free energy of (9). Minimizing (9) w.r.t. M, we get

$$M_{ij} = \exp\left[-\beta \left(\| \frac{(X_i - \mu_X)}{\sigma_X} - R(\theta) \frac{(Y_j - \mu_Y)}{\sigma_Y} \|^2 - \alpha + \kappa_i + \lambda_j \right) - 1 \right] \quad (10)$$

In (10), β is held fixed and we have yet to determine the two Lagrange parameters κ and λ. The exponentiation in (10) keeps all the match matrix entries positive and is a consequence of using the $x \log x$ barrier function. Solving for the two Lagrange parameters using a gradient ascent approach is likely to be slow and inefficient. Fortunately, the row and column constraints in (8) and (10) can be efficiently solved using a remarkable theorem due to Sinkhorn [16]: A doubly stochastic matrix is obtained from any square matrix with positive entries by the simple process of alternating row and column normalizations. (Once again, the presence of outliers leads to minor modifications to Sinkhorn's theorem.) While it may appear that Sinkhorn's theorem is unrelated to the two Lagrange parameters κ and λ, this is not the case. Solving for the Lagrange parameters (alternating between solving for κ and then λ rather than steepest ascent) is *identical* to Sinkhorn's algorithm [14]. With β held fixed, the exponentiation in (10) followed by Sinkhorn's algorithm yields a doubly stochastic matrix as expected. Slowly increasing β results in a permutation matrix and the optimal solution to the linear assignment problem [12, 14]. Since the exponentiation and Sinkhorn's algorithm at fixed β is central to our approach, we refer to this as the softassign.

Fixing the correspondence match matrix, we can solve for the rotation. The rotation matrix is updated using closed form solutions at each temperature with the match matrix held fixed. Differentiating (8) w.r.t. θ and setting the result to zero, we get

$$\theta = \arctan\left(\frac{\sum_{i=1}^{N_1}\sum_{j=1}^{N_2} M_{ij}\left[X_i^c(2)Y_j^c(1) - X_i^c(1)Y_j^c(2)\right]}{\sum_{i=1}^{N_1}\sum_{j=1}^{N_2} M_{ij}\left[X_i^c(1)Y_j^c(1) + X_i^c(2)Y_j^c(2)\right]}\right) \tag{11}$$

where

$$X_i^c = \frac{X_i - \mu_X}{\sigma_X}, \text{ and } Y_j^c = \frac{Y_j - \mu_Y}{\sigma_Y}.$$

The numbers inside the parenthesis $[X(1)$ or $X(2)]$ denote the abscissa and ordinate of the point-set. Having specified the rotation update, we turn to the estimation of the centroids and variances. Differentiating (8) w.r.t. (μ_X, μ_Y) and setting the result to zero, we obtain

$$\sum_{i=1}^{N_1}\sum_{j=1}^{N_2} \frac{M_{ij}}{\sigma_X}\left[\frac{(X_i - \mu_X)}{\sigma_X} - R(\theta)\frac{(Y_j - \mu_Y)}{\sigma_Y}\right] = 0$$
$$\sum_{i=1}^{N_1}\sum_{j=1}^{N_2} \frac{M_{ij}}{\sigma_Y}\left[\frac{(X_i - \mu_X)}{\sigma_X} - R(\theta)\frac{(Y_j - \mu_Y)}{\sigma_Y}\right] = 0 \tag{12}$$

It is straightforward to show that the Ansatz for the centroids in (5) is one solution to (12). (Showing that it is a minimum is also straightforward.) The same approach does not work for the variances, however. We think that it is difficult to construct an objective function on the centroids *and* the variances such that the Procrustes distance emerges as a minimum. Due to this inability to specify a global optimization problem on all relevant parameters, the Ansätze for the variances in (5) remain.

As the deterministic annealing parameter β is increased, the match matrix entries "harden"—they approach binary values. Outlier rejection occurs in the limit when $\beta \to \infty$ with the outliers becoming binary valued. We do not have a proof of convergence (to a local minimum) for the softassign Procrustes algorithm. The presence of the Ansätze for the centroids and the variances makes it difficult to prove convergence. Variable and constant definitions can be found in Table 1.

The Softassign Procrustes Algorithm
Initialization: θ to zero, β to β_0 and $\hat{M}_{ij}$ to $1 + \epsilon_{ij}$
Center and scale each point-set as in the Procrustes method.
Begin A: Deterministic Annealing. Do A until $\beta \geq \beta_f$
  Begin B: Do B until M converges or # of iterations $> I_0$
    Begin C: Centroids, Variances and Rotation matrix update.
      Update μ_X, μ_Y, σ_X, σ_Y using the Ansatz in (5)
      Update θ using analytical solution in (11)
    End C
    Begin D: Softassign.
      $Q_{ij} \leftarrow -\|\frac{(X_i - \mu_X)}{\sigma_X} - R(\theta)\frac{(Y_j - \mu_Y)}{\sigma_Y}\|^2 + \alpha$

$M^0_{ij} \leftarrow \exp(\beta Q_{ij})$

Begin E: Sinkhorn. Do E until $\hat{M}$ converges or # of iterations $> I_1$

Update $\hat{M}$ by normalizing the rows:

$\hat{M}^1_{ij} \leftarrow \frac{\hat{M}^0_{ij}}{\sum_{j=1}^{N_2+1} \hat{M}^0_{ij}}$

Update $\hat{M}$ by normalizing the columns:

$\hat{M}^0_{ij} \leftarrow \frac{\hat{M}^1_{ij}}{\sum_{i=1}^{N_1+1} \hat{M}^1_{ij}}$

End E

End D

End B

$\beta \leftarrow \beta_r \beta$

End A

Table 1. Variable and constant definitions for the softassign Procrustes matching algorithm

β	control parameter of the deterministic annealing method
β_0	initial value of the control parameter β
β_f	maximum value of the control parameter β
β_r	rate at which the control parameter β is increased
$\{\epsilon_{ij}\}$	small positive random variable
$\{M_{ij}\}$	match matrix variables
$\{\hat{M}_{ij}\}$	match matrix variables including the outliers
$\{Q_{ij}\}$	$\{-\Vert\frac{(X_i-\mu_X)}{\sigma_X} - R(\theta)\frac{(Y_j-\mu_Y)}{\sigma_Y}\Vert^2 + \alpha\}$
I_0	maximum # of iterations allowed at each value of the control parameter, β
I_1	maximum # of iterations allowed for Sinkhorn's method
	(back and forth row and column normalizations)

4 Results

We have applied the softassign Procrustes matching algorithm to the problem of autoradiograph alignment [15]. Primate autoradiographs of functional (metabolic) activation were obtained using the 2-DG method developed by [18]. Alignment of sequential pairwise sections is the first stage of reconstructing a 3-D high resolution functional map of the primate brain.

Slices 379-385 (left parietal and temporal cortex) are shown in Figure 1. First we run a Canny edge detector [6] on each slice. The edge detector is a single scale (Gaussian filter with width σ) edge detector incorporating hysteresis and non-maximum suppression. The edge detector outputs are shown in Figure 2. The point sets X and Y are obtained from the edge images. Since many thousands of points result, the points are first clustered in order to reduce the point count. The degree of clustering was chosen to yield approximately 200 points in each point-set, as shown in Figure 3.

Having obtained the point sets, we now execute the softassign Procrustes algorithm. Exactly as described above, the point-sets are first centered and scaled as in the original Procrustes algorithm. The parameter values were: $\beta_0 = 10\sqrt{N_1 N_2}$, $\beta_{\max} = 1000\sqrt{N_1 N_2}$, $\beta_r = 0.93$, $\alpha = 0$. The maximum number of iterations in each Sinkhorn step was 30. We allowed a maximum of two iterations at each temperature.

The results are shown in Figure 5. Initial conditions are shown in Figure 4. The x's and o's indicate the coordinate locations of the edge (point) features. Note that a subset of the x's and o's do not match. The method has discarded them as outliers. The matrix M_{ij} has all zero rows (columns) corresponding to the outliers. all zero. All experiments used exactly same parameters. Despite the much larger rotation between some slices, the softassign Procrustes algorithm had no difficulty aligning the point-sets and discarding outliers in the process. The method is not restricted to merely aligning adjacent pairwise sections. The results shown in Figure 5 and summarized in Table 2 were obtained with an eye towards averaging of forms. Incorporating correspondence and outlier rejection in Procrustes averaging is a natural extension of our approach.

Table 2. Procrustes matching parameters

Slices	Rotation	X-Centroid	Y-centroid	X-Variance	Y-Variance
379 and 380	-3.3	[-0.0011 , 0.0013]	[-0.0039,0.0093]	0.8996	0.8789
384 and 385	9.7	[-0.0022 , 0.0020]	[-0.0044 , 0.0056]	0.8556	0.8577
379 and 385	5.3	[0.0009, 0.0024]	[-0.0034,0.0063]	0.8957	0.8704
380 and 384	1.7	[-0.0022,0.0068]	[0.0005, -0.0030]	0.8838	0.8750
379 and 384	3.0	[0.0012,0.0022]	[-0.0002,0.0022]	0.8977	0.8618
380 and 385	9.7	[-0.0011, 0.0078]	[-0.0025,0.0014]	0.8624	0.8545

5 Discussion and Extensions

This paper has extended the Procrustes distance measure to a distance function over the union of point-sets with different point counts. Previous work [9], [5] deal with the related problem of minimizing metrics within linear subspaces of the tangent space to the space of these shapes. In this approach there is no combinatorial problem, only a geometric one, the solution to which emerges directly from the nesting of a projection operation inside a linear approximation of the Procrustes step. These and related algorithms have no equivalent to the notion of "outlier" that gives the softassign Procrustes algorithm its power. And this, in turn, leads to the distinction between the two classes of Procrustes parameters in the alternating algorithm here. Translation and rotation could be incorporated in the minimizing algebra of alternations, but variance could not be. This is because translation and rotation deal with the same relation between two forms that the matching function is attempting to minimize, but the scaling step is a function of only one form at a time–it has nothing to do with the actual least-squares superposition–and thus cannot be imputed to one form by any computation based in the other. The scale σ_X of a form X is *not* the minimizer of any summed squared interlandmark distance. [4] One can continuously pass from one form to a matched form by a smooth curve in one of these shape manifolds, along which arc-length is approximated by the Procrustes distance formula; but one *cannot* continuously pass from one form to a form having one landmark more or fewer. The sum-of-squares jumps *discontinuously* from 1 to $\frac{(N+1)}{N}$, and needs to be reset by a jump. That is, the Kendall shape metrics of different point counts are incommensurate. The Ansätze of this paper are all entailed in the embedding of the shape manifold in the original picture space, where the structural difference between variance and the other two parameters is not so clear. At their mathematical foundation, however, translation and rotation pertain to similarities over the same manifold; they are thus quite different from scaling, which links manifolds of different point count.

For an extension of constructions like these to the production of averaged forms, correlations between forms and their causes or consequences, and the like, one can speculate on a variety of approaches. All define the average as a form showing the least summed squared Procrustes distance to the forms of the sample–the Fréchet mean–but they would differ in the combinatorics of that average. In one approach, once the forms are matched they are modeled as combinations of isolated points and curves, and averaged in that hybrid Procrustes space by the method of [9], which matches outline arcs rather than outline samples and thus ignores point counts along curves. That is, outliers would be estimated, when appropriate, as linear interpolants of the points of curves nearest to them, and then treated by the standard (noncombinatorial) Procrustes tactics. In a second approach, the standard Procrustes averaging algorithm would be applied (fit all forms to one form, average the fitted values, and iterate), modified for use in our context of correspondence and outlier estimation: the only points that could be "averaged" are those that were never considered to be outliers in any step of the averaging [7]. (While our initial foray into averaging [7] estimated spatial mappings and correspondences during the averaging step, it did not include the crucial Procrustes centering and rescaling steps.) If this new algorithm resulted in a nonempty point-set, that would be "an average"; of course, we have no idea how many averages exist for any set of outlines in this approach. In a third possibility, one would average the matrices M to supply a weighted Procrustes average, where the weighting is by tendency to be an outlier. We have no data sets yet to which these speculations can be turned, but expect this to be a topic of an IPMI 1999 submission.

Acknowledgments

Anand Rangarajan and Haili Chui are supported by a grant from the Whitaker Foundation. Fred Bookstein is supported by NIH grants DA–09009 and GM–37251. The former grant is jointly supported by the National Institute on Drug Abuse, the National Institute of Mental Health, and the National Institute on Aging as part of the Human Brain Project.

References

1. D. P. Bertsekas and J. N. Tsitsiklis. *Parallel and Distributed Computation: Numerical Methods.* Prentice-Hall, Englewood Cliffs, NJ, 1989.
2. P. J. Besl and N. D. McKay. A method for registration of 3-d shapes. *IEEE Trans. Patt. Anal. Mach. Intell.*, 14(2):239–256, Feb. 1992.
3. M. Black and A. Rangarajan. On the Unification of Line Processes, Outlier Detection and Robust Statistics with Applications in Early Vision. *Intl. J. Computer Vision*, 19(1):57–91, 1996.
4. F. L. Bookstein. *Morphometric tools for landmark data: Geometry and biology.* Cambridge University Press, 1991.
5. F. L. Bookstein. Landmark methods for forms without landmarks: Localizing group differences in outline shape. In A. Amini, F. L. Bookstein, and D. Wilson, editors, *Proc. of the Workshop on Mathematical Methods in Biomedical Image Analysis*, pages 279–289. IEEE Computer Society Press, 1996.
6. J. Canny. A Computational Approach to Edge Detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 8(6):679–698, Nov. 1986.
7. S. Gold, A. Rangarajan, and E. Mjolsness. Learning with preknowledge: clustering with point and graph matching distance measures. *Neural Computation*, 8(4):787–804, 1996.
8. C. Goodall. Procrustes methods in the statistical analysis of shape. *J. R. Statist. Soc. B*, 53(2):285–339, 1991.
9. W. D. K. Green. The thin-plate spline and images with curving features. In K. V. Mardia, C. A. Gill, and I. L. Dryden, editors, *Proceedings in image fusion and shape variability techniques*, pages 79–97. Leeds University Press, 1996.
10. J. Hertz, A. Krogh, and R. G. Palmer. *Introduction to the Theory of Neural Computation*, volume 1 of *A Lecture Notes volume in the Santa Fe Institute Studies in the Sciences of Complexity.* Addison–Wesley, New York, 1991.
11. D. G. Kendall. Shape-manifolds, Procrustean metrics and complex projective spaces. *Bulletin of the London Mathematical Society*, 16:81–121, 1984.
12. J. J. Kosowsky and A. L. Yuille. The invisible hand algorithm: Solving the assignment problem with statistical physics. *Neural Networks*, 7(3):477–490, 1994.
13. D. Luenberger. *Linear and Nonlinear Programming.* Addison–Wesley, Reading, MA, 1984.
14. A. Rangarajan, S. Gold, and E. Mjolsness. A novel optimizing network architecture with applications. *Neural Computation*, 8(5):1041–1060, 1996.
15. A. Rangarajan, E. Mjolsness, S. Pappu, L. Davachi, P. Goldman-Rakic, and J. Duncan. A Robust Point Matching Algorithm for Autoradiograph Alignment. In K. H. Hohne and R. Kikinis, editors, *Fourth International Conference on Visualization in Biomedical Computing*, Lecture Notes in Computer Science, LNCS 1131, pages 277–286. Springer, New York, NY, 1996.

16. R. Sinkhorn. A relationship between arbitrary positive matrices and doubly stochastic matrices. *Ann. Math. Statist.*, 35:876–879, 1964.
17. C. Small. *The Statistical theory of shape*. Springer-Verlag, 1996.
18. L. Sokoloff, M. Revich, C. Kennedy, M. H. DesRosiers, C. S. Patlak, K. D. Pettigrew, O. Sakurada, and M. Shinohara. The C14-deoxyglucose method for the measurement of local cerebral glucose utilization: theory, procedure, and normal values in the conscious and anesthetized albino rat. *J. Neurochem.*, 28:897–916, 1977.
19. P. Van den Elsen. *Multimodality matching of brain images*. PhD thesis, Utrecht University, Utrecht, Netherlands, 1993.
20. A. L. Yuille and J. J. Kosowsky. Statistical physics algorithms that converge. *Neural Computation*, 6(3):341–356, May 1994.

Figures

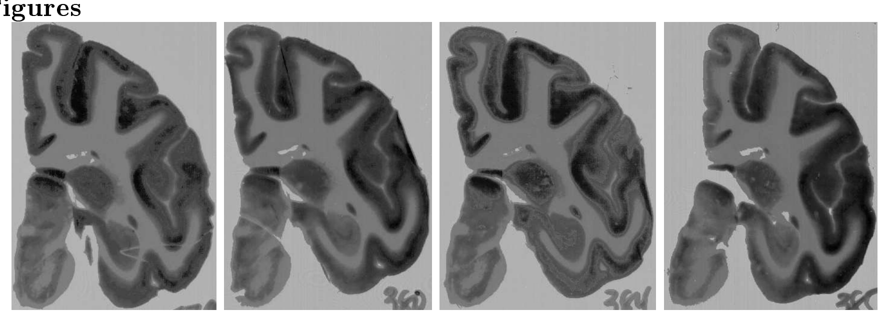

Fig. 1. Slices (a) 379 (b) 380 (c) 384 and (d) 385 of primate autoradiographs of functional (metabolic) activation. The coronal slices correspond to the left parietal and temporal cortex. After digitization, the slices were sampled down into $\simeq$ 475x350 pixels each with a resulting spatial resolution of $\simeq 100\mu$m.

Fig. 2. Canny edge images corresponding to the autoradiograph slices of Figure 1. Note the number of edges that are not in common in (a) - (d). While we have not dwelled on the instability of edge detection, it is yet another reason for our strong emphasis on robustness.

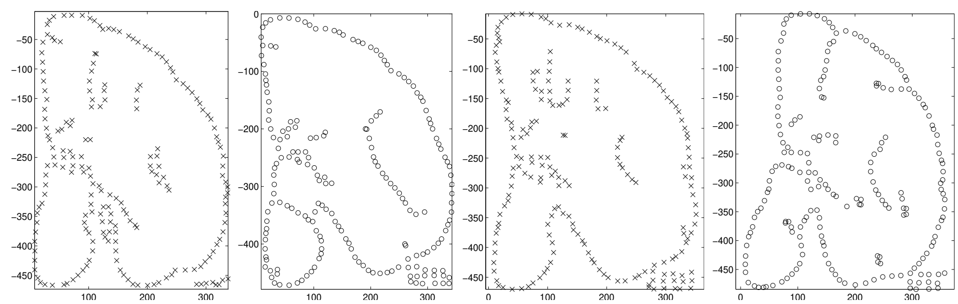

Fig. 3. Point-sets corresponding to the edge images. There were approximately 200 points in each point-set.

Fig. 4. Initial Condition

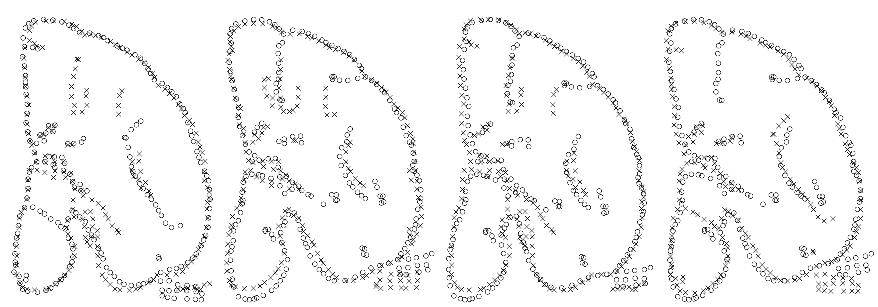

Fig. 5. Final solution found by the method. Again note the number of (mostly internal) x's and o's that do not match but remain isolated during overlay. These have been rejected as outliers and play no role towards the final stages of the algorithm.