Diffeomorphic Point Matching with Applications in Biomedical Image Registration

Hongyu Guo[1] and Anand Rangarajan[2]

[1]Department of Computing Sciences
Texas A&M University - Corpus Christi
Corpus Christi, TX, U.S.A.
hguo@sci.tamucc.edu

[2]Department of Computer and Information Science and Engineering
University of Florida
Gainesville, FL, U.S.A.
anand@cise.ufl.edu

Abstract

The need for 3D diffeomorphic matching of unlabeled point sets arises in numerous registration and shape analysis oriented biomedical imaging applications. The correspondence problem coupled with the need for estimating a diffeomorphism in 3D space makes it a very difficult problem. In the present work, we show that a joint clustering and diffeomorphism estimation strategy is capable of simultaneously estimating correspondences and a diffeomorphism in 3D space to interpolate the two unlabeled point sets. Correspondence is established between the cluster centers. Cluster centers for the two point sets having the same labels are always in correspondence. Essentially, as the cluster centers evolve during the iterations of an incremental EM algorithm, we estimate a diffeomorphism in 3D space. The approach is completely symmetric w.r.t. inversion of the two point sets being matched. We demonstrate our approach on 3D hippocampal data.

Keywords: medical imaging, image registration, diffeomorphism, clustering, hippocampus

1 Introduction

Recent applications of medical image analysis in computational anatomy and image guided surgery impose greater demands for automated, unambiguous, non-rigid point matching. However, fully automated feature based image registration techniques are still at the early stage in the non-rigid registration domain. The unknown correspondence problem, the existence of outliers all contribute to the difficulty of the problem. Moreover, the correspondence problem is often ill-posed. We attempt to formulate a precise mathematical model to well define the point matching problem. There are two important cases that need to be distinguished. When the two point sets have equal cardinality and the correspondences are known, we have the *landmark matching problem*. When the cardinality of the two point sets are not equal and the correspondences are unknown, we have the *point shape matching problem*.

Previous work on the landmark matching problem (Joshi and Miller, 2000; Camion and Younes, 2001) ignored the unknown correspondence problem and assumed that all the landmarks are labeled and the correspondence between the two sets of landmarks is known. This is the case when all the landmarks are hand-picked by experts but not the case when the feature selections become automated. Furthermore, there is a frequent need to match point sets of unequal cardinality where a point-wise correspondence cannot be assumed. On the other hand, previous work on the correspondence problem (Chui and Rangarajan, 2001; Chui and

Rangarajan, 2000; Chui, Win, Duncan, Schultz and Rangarajan, 2003; Chui and Rangarajan, 2003) did not attempt to solve for diffeomorphisms. The deformation models used were splines, like thin-plate splines, Gaussian radial basis functions and B-splines. The principal drawback of using a spline for the spatial mapping or the deformation model is the inability to guarantee that there are no local folds or reflections in the space mapping and that a valid inverse exists. There have been investigations on symmetric invertible consistent non-rigid registrations for image based matching, but it is still not guaranteed that the mapping is free of local folds (Beg and Khan, 2007; Yang, Li, Low, Deasy and Naqa, 2008; Yeung, Tang, Shi, Pluim, Viergever, Chung and Shen, 2008; Zeng and Chen, 2008). We will discuss this in more detail in Section 2.

In this work, we are not concerned with curve and surface matching because curve or surface representation of shapes usually require prior knowledge about the topology of the shapes. Point set representation is a universal representation of shapes and is especially useful when feature grouping (into curves and the like) cannot be assumed.

Below we briefly review previous approaches to non-rigid point matching. Since the approaches to solving the correspondence problem in point matching predate non-rigid mapping parameterizations, we include some of these methods as well in our survey. Our taxonomy of methods for solving the correspondence problem is as follows. The most direct method for solving for the unknown point-to-point correspondences involves matching points one to one and discarding a fraction from each set as outliers. This is called explicit correspondence in our taxonomy. More recently, there have been many approaches that avoid directly solving for the correspondences and we classify these methods as implicit correspondence methods. We also make a fundamental distinction between methods that stay within the ambit of the original point sets (sparse methods) and methods which work in the entire 3D space (dense methods). Finally, we discuss methods which attempt to constrain the point sets by imposing topologies on them and which include the topology information in matching. We briefly classify previous approaches into four categories.

1. *Sparse Representation and Explicit Correspondence*
   One of the first approaches to explicitly solve for point correspondences as well as the spatial mapping parameters is Baird's Ph.D. thesis (Baird, 1984). Since then, there have been many approaches including the ones based on eigen-analysis (Scott and Longuet-Higgins, 1991; Shapiro and Brady, 1992), linear assignment (Chui and Rangarajan, 2003; Belongie, Malik and Puzicha, 2002) and iterated closest point, which has turned out quite popular despite being a step backward from Baird's thesis (Besl and McKay, 1992). Here we do not include approaches that explicitly solve for correspondences while imposing a topology on each point set, which we discuss later. The main drawback of this class of approaches is the need of explicitly solving for point-to-point correspondences as a prerequisite to registration. As the implicit methods show, registration needs not require correspondence.

2. *Sparse Representation and Implicit Correspondence*
   In this class of approaches, explicit correspondence variables are not used. Instead, a distance measure between the point sets is directly minimized. Approaches in this class include pairwise distances (kernel correlation) (Grimson, Lozano-Perez, Wells III, Ettinger, White and Kikinis, 1994; Mjolsness and Garrett, 1990; Tsin and Kanade, 2004), robust distances between probability distributions (Jian and Vemuri, 2006) and joint clustering and matching (Chui et al., 2003). The kernel correlation method periodically gets rediscovered and hence it is quite possible that there exist references for this approach prior to 1990. Also, kernel correlation can be derived from an image matching objective function wherein the point sets are converted into images via the use of distributions (generalized functions) (Glaunes, Trouve and Younes, 2004). Since the later is a dense

matching method, we discuss this below. The approach in (Jian and Vemuri, 2006) is similar to the approach in (Chui et al., 2003) and to the one in this paper in that density functions are first constructed on the point sets. After a density function is constructed, a robust distance between two densities is used for non-rigid registration. The robust distance between densities involves an integral over $\mathbf{R}^3$. Since the non-rigid spatial mapping is unknown, this integral cannot be evaluated in a closed form. The authors in (Jian and Vemuri, 2006) ignore this issue and approximate the integral and obtain a sparse distance measure which is strikingly similar to kernel correlation. Finally the joint clustering and matching approach in (Chui et al., 2003) is a precursor to the approach in this paper in that we simultaneously fit Gaussian mixture models to two or more point sets while maintaining correspondence between the cluster centers. The drawback of this class of approaches is the huge burden on the non-rigid deformation to match the point sets while minimally deforming the space.

3. *Dense Representation and Implicit Correspondence*
   The dense representation approaches bypass the correspondence problem by moving from the point set to $\mathbf{R}^3$ and fitting distance functions (Paragios, Rousson and Ramesh, 2003), generalized functions(Glaunes et al., 2004) or density functions to the point sets (Wang, Vemuri, Rangarajan, Schmalfuss and Eisenschenck, 2006) and subsequently performing dense matching w.r.t. the non-rigid deformation. The drawback of this class of approaches is the need to warp the ambient space in $\mathbf{R}^3$ as opposed to merely deforming the point sets. The distance measure on images (or on the density functions) reduces to kernel correlation on the point sets if one chooses to ignore the spatial variability of the non-rigid deformations (Mjolsness and Garrett, 1990). If this assumption is not made, the image-based dense matching measure cannot be reduced to a sparse distance measure on the point sets.

4. *Graph-based Representation and Explicit Correspondence*
   Since the point sets are presumably sampled from an underlying shape, the strategy behind the graph-based methods is to perform a kind of perceptual grouping on the point sets by imposing a graph topology. Subsequently, the graphs are matched to obtain a shape distance. This class of approaches is not usually used for registration but is quite popular for shape indexing. Specific approaches in this class include shock graph matching (Sebastian, Klein and Kimia, 2004; Siddiqi, Shokoufandeh, Dickinson and Zucker, 1999), modal matching(Sclaroff and Pentland, 1995) and Bayesian networks(Rangarajan, Coughlan and Yuille, 2003). The fundamental problem with this class of approaches is the difficulty of matching shapes in the presence of a topology mismatch. While adding a topology to the point set can lend it a perceptual grouping structure thereby lessening the burden on matching, differences in topology require edits to be made during matching which are a significant drawback.

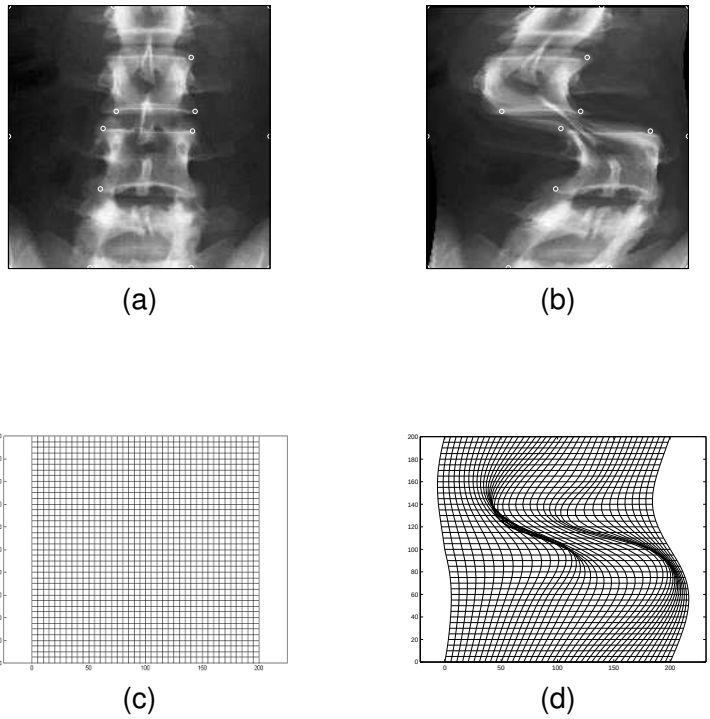
Recently there is an increasing interest in diffeomorphic image registrations. Analogies from Lagrangian mechanics (Avants, Schoenemann and Gee, 2006) and Hamiltonian mechanics (Marsland and McLachlan, 2007) have been used. Analogy from Maxwell demons in thermodynamics and statistical mechanics has also been used, where image matching is viewed as a diffusion process (Thirion, 1998; Arsigny, Commowick, Pennec and Ayache, 2006). There are pros and cons using the demons approach and Hernandez et al. give a comparison of stationary LDDMM methods and diffeomorphic demons methods (Hernandez, Olmos and Pennec, 2008). A static velocity field is used to generate and guarantee a diffeomorphic mapping (Hernandez, Bossa and Olmos, 2007). The space transformation is diffeomorphic and all those transformations form a Lie group while the velocity fields are the generating Lie algebras of the space diffeomorphic transformation groups.

Our approach is different from all these approaches in that we solve for a diffeomorphic space mapping and solve for an implicit correspondence at the same time. Our system is not Hamiltonian, but it is not a dissipating system either, unlike a diffusion system. We formulate and solve the diffeomorphic point matching problem in a sparse representation/implicit correspondence framework. In section 2, we formulate the diffeomorphism problem and the correspondence problem. We also discuss the design of an appropriate objective function. In section 3, we design an algorithm to minimize the diffeomorphic point matching objective function. In section 4, we describe and discuss the experimental results of applying our algorithm on 3D hippocampal shape point sets.

2 Diffeomorphic Point Matching and the Unknown Correspondence

2.1 The diffeomorphism

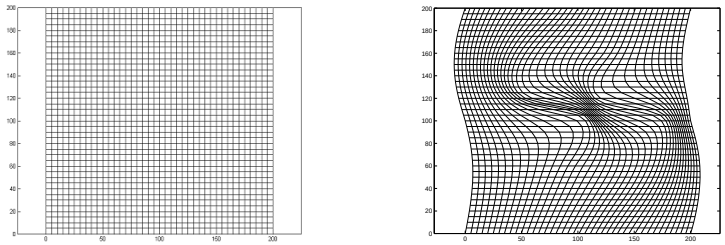
Two desirable properties of non-rigid transformations are smoothness and topology preservation. Let $\Omega_1 \subseteq \mathbf{R}^3$ and $\Omega_2 \subseteq \mathbf{R}^3$. A transformation $f : \Omega_1 \rightarrow \Omega_2$ is smooth if all partial derivatives of f, up to certain orders, exist and are continuous. A transformation $f : \Omega_1 \rightarrow \Omega_2$ preserves the topology if Ω_1 and $\mathrm{Img}(f) = \{p_2 \in \Omega_2 | \exists p_1 \in \Omega_1,\, p_2 = f(p_1)\}$ have the same topology. A smooth transformation $f : \Omega_1 \rightarrow \Omega_2$ may not preserve the topology. There are several cases when this is true. First, the smooth map f is a bijection but the inverse is not continuous. Second, the smooth map f may fail to be a bijection. That is, multiple points may be mapped to the same point and we call this the folding of space. There are two sub-cases here, one sub-case is that at some point, the tangent map of f is not an isomorphism. The other sub-case is that the tangent map of f is an isomorphism at every point but globally it is not a bijection. On the other hand, a homeomorphism may not be smooth. A transformation f that is both smooth and topology preserving is called a *diffeomorphism.* The diffeomorphism $f : \Omega_1 \rightarrow \Omega_2$ is defined as a bijection that is smooth and also has a smooth inverse. Now let us look at an example of a smooth transformation, namely, the Thin-Plate Spline (TPS)(Wahba, 1990).



(a) (b) (c) (d)

Figure 1: The folding problem of spline deformation

Figure 1 (a) shows a template image. Figure 1 (b) is the reference image obtained from some warping of the template. Some landmarks are selected and shown in the images. Figure 1 (d) demonstrates the thin-plate spline transformation in 2D space by showing the deformation of the rectangular grid shown in Figure 1 (c). We can easily see the folding of space. This is the drawback of thin-plate spline interpolation. Due to the folding of space, features in the template may be smeared in the overlapping regions. Furthermore, the transformation is not invertible. A diffeomorphic transformation is strongly desirable, which preserves the features, the topology and which is smooth. Figure 2 (b) shows the desired diffeomorphism transforming the same rectangular grid, copied to Figure 2 (a) for side by side comparison.



(a) (b)

Figure 2: Desired diffeomorphism

2.2 Joint Clustering and Diffeomorphism

There are different ways to solve for the unknown correspondence (Rangarajan, Chui, Mjolsness, Pappu, Davachi, Goldman-Rakic and Duncan, 1997; Chui and Rangarajan, 2000; Chui et al., 2003; Chui and Rangarajan, 2003). Essentially, within the framework of explicit point correspondences — as opposed to the distance function framework of implicit correspondence — we have a choice between solving for an optimal permutation and letting corresponding "labeled" points discover their optimal locations. We opt for the latter in this work because of its simplicity. The clustering in fact serves two purposes. First, it is the method to find the unknown correspondence. We initialize the two sets of cluster centers around the centroids of their data points, respectively. The cluster centers are marked with identical labels in the two sets denoting correspondence. The cluster centers evolve during the iterations of an incremental EM algorithm and they are linked by a diffeomorphism and are forced to move in lock-step with one another. Second, clustering is the modeling of the real data sets with noise and/or outliers because with two shapes represented by point samples, we cannot assume a point-wise correspondence. The correspondence is only between the two shapes and clustering is a useful way to model the shapes.

We use a Gaussian mixture model to describe the clustering of the point sets. The Gaussian mixture probability density is

$$p(x|\mathbf{r},\sigma) = \frac{1}{K}\sum_{k=1}^{K}\frac{1}{(2\pi\sigma^2)^{3/2}}\exp(-\frac{1}{2\sigma^2}||x-r_k||^2)\,, \tag{2.1}$$

where x is a point in $\mathbf{R}^3$, K is the number of clusters and $\mathbf{r}$ is the compact notation of a set of K cluster center positions $r_1, r_2, ..., r_K$, with each r_k a point in $\mathbf{R}^3$.

Given a set of N points $x_1, x_2, ..., x_N$ in $\mathbf{R}^3$ sampled from the Gaussian mixture probability density (2.1), the problem of clustering the N points into K clusters is the process of estimating the positions of K cluster centers, $r_1, r_2, ..., r_K$ that leads to the maximum log-likelihood of the observed sample

$$\log p(x|\mathbf{r}, \sigma) = \sum_{i=1}^{N} \log \sum_{k=1}^{K} \exp(-\frac{1}{2\sigma^2}||x_i - r_k||^2)\,. \tag{2.2}$$

The solutions can be found by applying the EM algorithm. As pointed out by Hathaway (Hathaway, 1986) in the mixture model context, the EM algorithm maximizing (2.2) can be viewed as an alternative maximization of the following objective

$$F(M, \mathbf{r}) = -\frac{1}{2\sigma^2} \sum_{i=1}^{N} \sum_{k=1}^{K} M_{ik}||x_i - r_k||^2 - \sum_{i=1}^{N} \sum_{k=1}^{K} M_{ik} \log M_{ik}. \tag{2.3}$$

where M is a matrix of parameters $M_{ik} \in (0, 1)$, $1 \leq i \leq N$ and $1 \leq k \leq K$. M_{ik} denotes the fuzzy membership that point x_i is in cluster k. Maximizing (2.3) is equivalent to minimizing $E(M, \mathbf{r}) = -F(M, \mathbf{r})$. $E(M, \mathbf{r})$ is the cost function or energy function.
The diffeomorphic point matching problem is described as follows: There are two sets of points $\mathbf{x} = \{x_1, x_2, ..., x_N\}$ and $\mathbf{x}' = \{x'_1, x'_2, ..., x'_{N'}\}$. We want to cluster $\mathbf{x}$ into K clusters with centers at $\mathbf{r} = \{r_1, r_2, ..., r_K\}$ and to cluster $\mathbf{x}'$ into K clusters with centers at $\mathbf{r}' = \{r'_1, r'_2, ..., r'_K\}$. We also want to find a diffeomorphism $f : \mathbf{R}^3 \to \mathbf{R}^3$ such that f has a minimal space deformation and f maps the first set of cluster centers to the second set of cluster centers: $f(r_k) = r'_k, k = 1, 2, ..., K$. Note this is equivalent to saying that the two sets of cluster centers are prelabeled and the corresponding cluster centers with the same label k are always in correspondence. The problem of finding correspondence and diffeomorphism is converted to finding the best clustering and the best diffeomorphism. In order to generate a guaranteed diffeomorphism, we introduce a continuous-time differentiable dynamical system. Such a dynamical system is defined by a one-parameter family of mappings: $\phi : \mathbf{R}^3 \times \mathbf{R} \to \mathbf{R}^3$ satisfying three conditions:

i) Identity: $\phi(x, 0) = x$,

ii) Group: $\phi(\phi(x, t_1), t_2) = \phi(x, t_1 + t_2)$, and

iii) Differentiability: $\frac{\partial}{\partial t}\phi(x, t)|_{t=0} = v(x, 0)$.

The parameter t is interpreted as time. The last property, differentiability, defines a velocity vector field. The mapping ϕ is guaranteed to be an diffeomorphism. The diffeomorphism ϕ is related to the vector field $v(x, t)$ by the transport equations

$$\frac{\partial \phi(x, t)}{\partial t} = v(\phi(x, t), t), \tag{2.4}$$

$$f(x) = \phi(x, 1) = \int_0^1 v(\phi(x, t), t)dt. \tag{2.5}$$

f is the diffeomorphism from $\mathbf{R}^3$ to $\mathbf{R}^3$ that we want for the point shape matching. We will include the regulation term of space deformation into the energy function in term of the velocity field.
Putting everything together, the joint diffeomorphic clustering energy is devised to be the following

$$E(M, M', \mathbf{r}, \mathbf{r}', v, \phi) = \sum_{i=1}^{N}\sum_{k=1}^{K} M_{ik}||x_i - r_k||^2 + 2\sigma^2 \sum_{i=1}^{N}\sum_{k=1}^{K} M_{ik}\log M_{ik} \tag{2.6}$$
$$+ \sum_{j=1}^{N'}\sum_{k=1}^{K} M'_{jk}||x'_j - r'_k||^2 + 2\sigma^2 \sum_{j=1}^{N'}\sum_{k=1}^{K} M'_{jk}\log M'_{jk}$$
$$+ \sum_{k=1}^{K} ||r'_k - \phi(r_k, 1)||^2 + 2\sigma^2\lambda \int_0^1 \int_\Omega ||Lv(x,t)||^2 dxdt.$$

In the above energy function, x_i are data points in the first point set and x'_j are the data points in the second point set to be registered. r_k are the cluster center positions in the first set and r'_k are the cluster centers in the second set. M_{ik} and M'_{jk} are the membership matrices for the first point set and the second point set respectively, as introduced in (2.3). The term $2\sigma^2\lambda \int_0^1 \int_\Omega ||Lv(x,t)||^2 dxdt$ is the regulation term for the space deformation. L is a linear differential operator on the vector field $v(x,t)$. The integration is over the deformation domain Ω. The displacement term $\sum_{k=1}^{K} ||r_k - \phi(r_k, 1)||^2$ is present in the energy cost because we adopted the inexact matching formulation. This term plays an important role here as the bridge between the clustering systems. Another reason that we prefer the deformation energy in this form is that the coupling of the two sets of clusters appear naturally through the displacement term and we don't have to introduce external coupling terms(Guo, Rangarajan, Joshi and Younes, 2004). Another advantage of this approach is that in this dynamical system described by the diffeomorphic group $\phi(x,t)$, the landmarks trace a trajectory exactly on the flow lines dictated by the field $v(x,t)$. The feedback coupling is no longer needed as in the previous approach because with this deformation energy described above, if $\phi(x,t)$ is the minimizer of this energy, then $\phi^{-1}(x,t) = \phi(x,-t)$ is the inverse mapping which also minimizes the same energy.

3 A Diffeomorphic Point Matching Algorithm

Our joint clustering and diffeomorphism estimation algorithm has two components: i) clustering and ii) diffeomorphism estimation. For the clustering part, we use the deterministic annealing approach. The clustering problem is a non-convex optimization problem. The traditional clustering techniques use descent based algorithms and they tend to get trapped in a local minimum. Rose *et al.*(Rose, Gurewitz and Fox, 1990) proposed the annealing approach using analogies to statistical physics. The clustering cost function is seen as the free energy of a Gibbs canonical distribution. The minimization of clustering cost function is seen as the simulation of a physical annealing process in which free energy is minimized. Let T be the temperature of the system and in the clustering system $T = 2\sigma^2$. Let $\beta = 1/T$ be the reciprocal temperature. Initially at $\beta = 0$, we have a single minimum for the energy, which is the global minimum and all the cluster centers are located at the same point, which is the centroid of all the data points and each data point is uniformly associated with all clusters. In the numerical implementation, we initialize all the cluster centers on a sphere of a very small radius and there are no data points within the sphere. When the temperature is lowered gradually, at a certain critical value of temperature, the clusters will split into smaller clusters and a phase transition occurs. At lower T , the free energy may have many local minima but the annealing process is able to track the global minimum.

For the diffeomorphism estimation, we expand the vector field $v(x,t)$ in terms of the kernel K of the L operator

$$v(x,t) = \sum_{k=1}^{K} \alpha_k(t) K(x, \phi_k(t)) \tag{3.1}$$

where $\phi_k(t)$ is notational shorthand for $\phi(r_k, t)$ and we also take into consideration of the affine part of the mapping when we use thin-plate spline kernel with matrix entry $K_{ij} = -\rho_{ij}$ for $\mathbf{R}^3$, where $\rho_{ij} =\| x_i - x_j \|$. After discretizing in time t, the objective in (2.6) is expressed as

$$\begin{aligned} E(M, M', \mathbf{r}, \mathbf{r}', \alpha(t), \phi(t)) &= \sum_{i=1}^{N}\sum_{k=1}^{K} M_{ik}||x_i - r_k||^2 + \sum_{j=1}^{N'}\sum_{k=1}^{K} M'_{jk}||x'_j - r'_k||^2 \\ &+ \sum_{k=1}^{K} \|r'_k - r_k - \sum_{l=1}^{K}\sum_{t=0}^{S} [P(t)d_l(t) + \alpha_l(t)K(\phi_k(t), \phi_l(t))] \|^2 \\ &+ \lambda \sum_{k=1}^{K}\sum_{l=1}^{K}\sum_{t=0}^{S} < \alpha_k(t), \alpha_l(t) > K(\phi_k(t), \phi_l(t)) \end{aligned} \tag{3.2}$$

where

$$P(t) = \begin{pmatrix} 1 & \phi_1^1(t) & \phi_1^2(t) \\ . & . & . \\ . & . & . \\ . & . & . \\ 1 & \phi_K^1(t) & \phi_K^2(t) \end{pmatrix} \tag{3.3}$$

and d is the affine parameter matrix. After we perform a QR decomposition on P, we have

$$P(t) = (Q_1(t) : Q_2(t)) \begin{pmatrix} R(t) \\ 0 \end{pmatrix}. \tag{3.4}$$

We iteratively solve for $\alpha_k(t)$ and $\phi_k(t)$ using an alternating algorithm. When $\phi_k(t)$ is held fixed, we use the following approximation to solve for $\alpha_k(t)$. The solutions are

$$\begin{aligned} d(t) &= R^{-1}(t) \left[Q_1(t)\phi(t+1) - Q_1(t)K(\phi(t))Q_2(t)\gamma(t)\right] \\ \alpha(t) &= Q_2(t)\gamma(t) \end{aligned} \tag{3.5}$$

where $K(\phi(t))$ denotes the thin-plate spline kernel *matrix* evaluated at $\phi(t) \stackrel{\text{def}}{=} \{\phi(r_k, t)|k = 1, \ldots, N\}$ and

$$\gamma(t) = (Q_2^T(t)K(\phi(t))Q_2(t) + \lambda)^{-1}Q_2^T(t)\phi(t+1). \tag{3.6}$$

When $\alpha_k(t)$ is held fixed, we use gradient descent to solve for $\phi_k(t)$:

$$\frac{\partial E}{\partial \phi_k(t)} = 2\sum_{l=1}^{N} < \alpha_k(t), \alpha_l(t) - 2W_l > \bigtriangledown_1 K(\phi_k(t), \phi_l(t)) \tag{3.7}$$

where $W_l = r'_l - r_l - \sum_{m=1}^{K} \int_0^1 \alpha_m(t)K(\phi_m(t), \phi_l(t))dt$.
The clustering of the two point sets is handled by a deterministic annealing EM algorithm which iteratively estimates the cluster memberships M and M' and the cluster centers $\mathbf{r}$ and $\mathbf{r}'$. The update of the memberships is the standard E-step of the EM algorithm(Chui et al., 2003) and is performed as shown below,

$$M_{ik} = \frac{\exp(-\beta\|x_i - r_k\|^2)}{\sum_{l=1}^{K}\exp(-\beta\|x_i - r_l\|^2)}, \forall ik \tag{3.8}$$

$$M'_{jk} = \frac{\exp(-\beta\|x'_j - r'_k\|^2)}{\sum_{l=1}^{K}\exp(-\beta\|x'_j - r'_l\|^2)}, \forall jk. \tag{3.9}$$

The cluster center update is the M-step of the EM algorithm. This step is not the typical M-step. We use a closed-form solution for the cluster centers which is an approximation. From

the clustering standpoint, we assume that the change in the diffeomorphism at each iteration is *sufficiently small so that it can be neglected*. After making this approximation, we get

$$r_k = \frac{\sum_{i=1}^{N} M_{ik} x_i + r_k - \sum_{l=1}^{K} \int_0^1 \alpha_l(t) K(\phi_l(t), \phi_k(t)) dt}{1 + \sum_{i=1}^{N} M_{ik}},$$
$$r'_k = \frac{\sum_{j=1}^{N'} M'_{jk} x'_j + \phi(r_k, 1)}{1 + \sum_{j=1}^{N'} M'_{jk}}, \forall k. \tag{3.10}$$

The overall algorithm is described below.

- **Initialization:** Initialize temperature
  $T = 0.5(\max_i \|x_i - x_c\|^2 + \max_j \|x'_j - x'_c\|^2)$ where x_c and y_c are the centroids of $\mathbf{x}$ and $\mathbf{x}'$ respectively.
- **Begin A:** While $T > T_{\text{final}}$ do
  - Step 1: Clustering
    - Update memberships according to (3.8), (3.9).
    - Update cluster centers according to (3.10).
  - Step 2: Diffeomorphism
    - Update (ϕ, v) by minimizing

      $$E_{\text{diff}}(\phi, v) = \sum_{k=1}^{K} ||r'_k - \phi(r_k, 1)||^2 + \lambda T \int_0^1 \int_\Omega ||Lv(x,t)||^2 dx dt$$

      according to (3.5) and (3.7).
  - Step 3: Annealing. $T \leftarrow \gamma T$ where $\gamma < 1$.
- **End**

4 Experiments and Results

We applied our formulation and algorithm to the 3D point data of hippocampal shapes. We first applied the algorithm to synthetic data, where we have the knowledge of ground truth and this serves as the validation of the algorithm. We then experimented with real data and evaluated the results using various measures.

4.1 Experiments on synthetic data

We select one hippocampal point set and warp it with a known diffeomorphism using the Gaussian Radial Basis Function (GRBF) kernel. We choose $\sigma = 60$ for the GRBF because with this large value of σ, we are able to generate a more global warping. Figure 3 shows the two point sets of hippocampal shapes. The set with the + markers is the original set and the set with point markers is that after GRBF warping.
We start with no added noise. We use the TPS kernel to recover the diffeomorphism via joint clustering using our algorithm. Since the reference data are synthesized, we know the ground truth and we are able to compare our result with the ground truth. After unwarping the point set with our recovered diffeomorphism, we find the squared distances between the

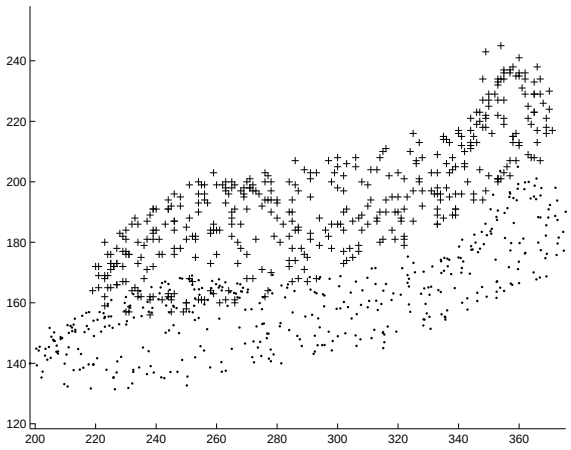


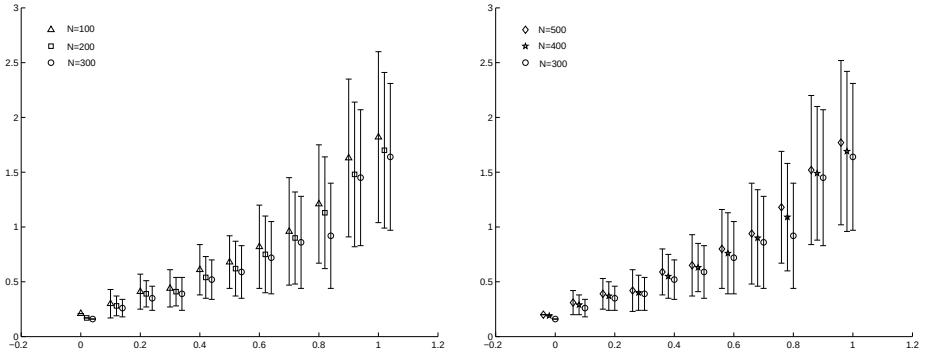
Figure 3: Synthetically warped hippocampus shape and the original hippocampus shape

corresponding data points, and find the average and then take the square root. This is the standard error for our recovered diffeomorphism. We have two free parameters, λ and T_{final}. T_{final} is determined by the limiting value of σ_T, which is in turn determined by the number of clusters. We choose a λ value such that the whole optimization process is stable in the temperature range from initial T to T_{final}. We experiment with different numbers of clusters. We list the corresponding standard errors in the first row of Table 1. It is easy to see that the standard error goes down as the number of clusters goes up from 100 to 300 and goes up again when the number of clusters increases further. This is because when we have too few clusters, the points are not well represented by the cluster centers. On the other hand, if we have too many clusters, the variance between the two shapes is too big and the deformation increases dramatically. For the standard error, there is an optimal number of clusters and in this case we find it to be 300. We need to estimate the macroscopic and microscopic dimensions of the shape in order to see how big the standard error is. We calculate the covariance matrix of the original data set. We find their eigenvalues to be $(48.1, 11.8, 4.1)$. This gives us an estimate of the macroscopic dimensions to be about 100, 24 and 8, namely twice the eigenvalues. We then find out the average distance between the nearest neighbors to be 2.65. This give us the microscopic dimension of the shape. As we can see from the table, our matching is very accurate.

noise	number of clusters				
	100	200	300	400	500
0	0.21	0.17	0.16	0.19	0.20
0.1	0.30±0.13	0.28±0.09	0.26±0.08	0.29±0.09	0.31±0.11
0.2	0.41±0.16	0.39±0.12	0.35±0.11	0.37±0.13	0.39±0.14
0.3	0.44±0.17	0.41±0.13	0.39±0.15	0.40±0.16	0.42±0.19
0.4	0.61±0.23	0.54±0.19	0.52±0.18	0.55±0.20	0.59±0.21
0.5	0.68±0.24	0.62±0.25	0.59±0.24	0.63±0.22	0.65±0.28
0.6	0.82±0.38	0.75±0.35	0.72±0.33	0.76±0.37	0.80±0.36
0.7	0.96±0.49	0.90±0.42	0.86±0.42	0.90±0.44	0.94±0.46
0.8	1.21±0.54	1.13±0.51	0.92±0.48	1.09±0.49	1.18±0.51
0.9	1.63±0.72	1.48±0.66	1.45±0.62	1.49±0.61	1.52±0.68
1.0	1.82±0.78	1.70±0.71	1.64±0.67	1.69±0.73	1.77±0.75

Table 1: Matching errors on synthetic data with different noise levels

Next we add noise to the warped data and test the robustness of our algorithm to noise. After



(a) (b)

Figure 4: Matching results on synthetic data for different number of clusters

GRBF warping, we add Gaussian noise to the warped data with different variances σ. We experiment with ten trials for each noise level from 0.1 to 1.0 and for each cluster level from 100 to 500. The standard errors and the deviation are shown in Table 1. We can see the standard error increase with the increasing noise level but it approximately stays in the range of the noise. Stronger noise does not increase the matching error dramatically and this shows the algorithm is robust against noise. This is easier to see when plotted in Figure 4 with error bars. We choose to split the five levels of clusters into two graphs because it does not look clear if they were crowded in one graph. Figure 4(a) has the errors for 100, 200 and 300 clusters and Figure 4(b) has the errors for 300, 400 and 500 clusters. We can see that at the 300 cluster level, we achieve the best matching.

4.2 Experiments on real data

We apply our algorithm on different real hippocampus data sets. Figure 5 shows two real hippocampus shapes.

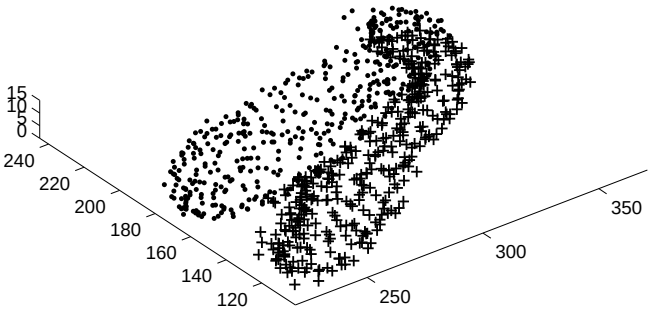


Figure 5: Real hippocampus shapes

Figure 6 demonstrates the annealing process. The *x* axis is the iteration steps. The dashed line is the scaled temperature $\sqrt{T/2}$. The solid line is the actual variance σ. During the

clustering iterations, the temperature goes down (annealing), σ also goes down, representing a better clustering solution after each iteration and σ finally converges. We observe a phase transition at temperature $T = 3.37 \times 10^3$. We also observe that there is a lower limit for σ. When the temperature gets very low, σ becomes a constant and no matter how much lower the temperature goes, the σ stays constant. This constant is determined by the number of clusters. In Figure 7(a) through (e), we show how this limit changes with the number of clusters. When the number of clusters equals or exceeds the number of data points, the limit approaches zero.

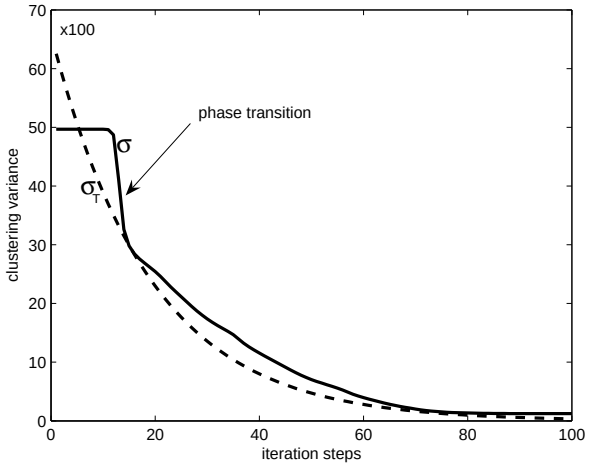
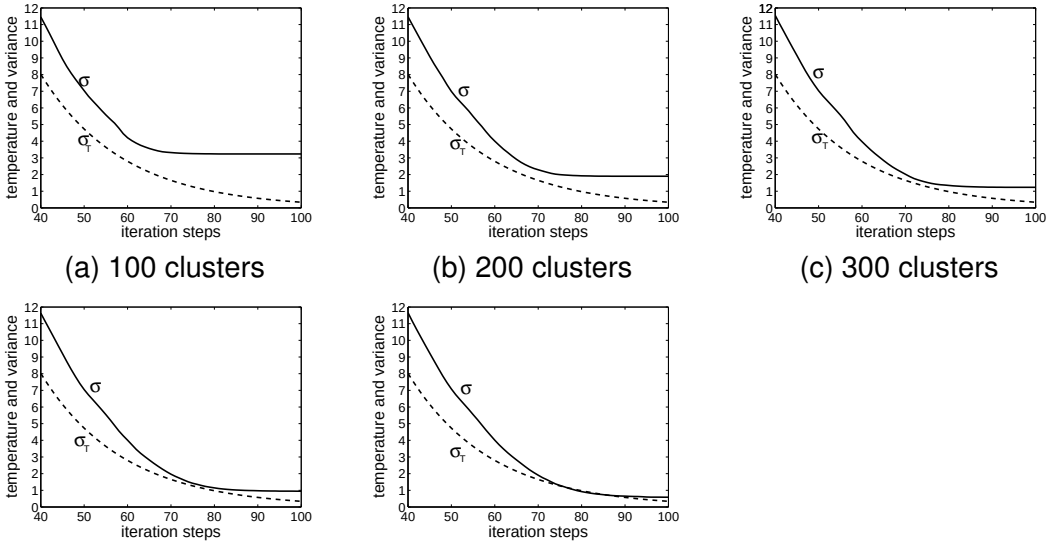


Figure 6: Deterministic annealing in the clustering process



(a) 100 clusters (b) 200 clusters (c) 300 clusters

(d) 400 clusters (e) 500 clusters

Figure 7: Limit σ vs. the number of clusters. The phase transitions are not seen here because they happen before iteration 40.

Table 2 displays the limits of σ changing with the number of clusters. Because of noise and sampling error, we should not allow this limit to go to zero. Again we observe when we have 300 clusters, we have a reasonable $\sigma = 1.2$ as we recall the average distance between the nearest neighbors is about 2.65.

number of clusters	100	200	300	400	500
σ	3.3	1.9	1.2	0.9	0.6

Table 2: Limit σ vs. number of clusters

We perform the matching for all the pairs out of ten hippocampus shapes. The matching results according to three metrics for different number of clusters are summarized in Table 3. The three metrics we use are Jensen-Shannon divergence (JS), Hausdorff distance (HD) and modified Hausdorff distance (MH). The Jensen-Shannon divergence (a special case with $\lambda = 1/2$) is defined as (Endres and Schindelin, 2003)

$$D = \sum_{i=1}^{N} (p_i \log \frac{2p_i}{p_i + q_i} + q_i \log \frac{2q_i}{p_i + q_i}), \tag{4.1}$$

where p and q are two distributions. Notice this metric is highly non-linear. When p and q are completely independent, namely in our matching case, when the two shapes are completely different, D has a maximum of $2 \log 2 = 1.39$.
The Hausdorff distance is defined as

$$H(A, B) = \max(\max_{a \in A} \min_{b \in B} \parallel a - b \parallel, \max_{b \in B} \min_{a \in A} \parallel b - a \parallel). \tag{4.2}$$

It is known that the Hausdorff distance is too sensitive to outliers. We also calculated the modified Hausdorff distance as first introduced in Dubuisson: (Dubuisson and Jain, 1994)

$$H_{\mathrm{mod}}(A, B) = \max(\frac{1}{|A|} \sum_{a \in A} \min_{b \in B} \parallel a - b \parallel, \frac{1}{|B|} \sum_{b \in B} \min_{a \in A} \parallel b - a \parallel), \tag{4.3}$$

which is the average of the minimum distances instead of the maximum of the minimum distances. It is easy to see that when the number of clusters increases, the modified Hausdorff distance decreases.

metric	clusters	trial number									
		1	2	3	4	5	6	7	8	9	10
JS	100	0.87	0.93	0.76	0.98	0.69	0.57	0.66	0.99	0.85	0.97
JS	200	0.31	0.62	0.27	0.52	0.41	0.23	0.21	0.70	0.42	0.62
JS	300	0.03	0.47	0.04	0.34	0.14	0.43	0.05	0.25	0.11	0.10
JS	400	0.13	0.05	0.16	0.09	0.18	0.78	0.14	0.19	0.68	0.18
JS	500	0.21	0.24	0.32	0.45	0.36	0.97	0.30	0.63	0.74	0.55
HD	100	7.1	9.3	7.2	8.4	9.6	9.2	6.9	8.9	9.3	7.8
HD	200	7.4	8.9	6.1	7.8	9.7	6.3	5.8	8.5	8.0	7.3
HD	300	5.7	7.2	4.9	7.2	8.0	7.1	4.4	7.0	5.9	4.7
HD	400	6.2	8.3	5.6	5.2	8.4	7.8	6.0	6.4	7.6	6.7
HD	500	7.3	8.7	6.4	6.5	8.9	8.6	7.3	8.2	9.1	8.1
MH	100	2.8	3.5	2.0	2.7	3.9	3.1	2.4	3.0	2.9	3.2
MH	200	2.0	3.1	1.7	2.4	3.6	2.8	2.2	2.6	2.7	2.8
MH	300	1.4	2.8	1.4	2.3	3.1	2.5	2.1	2.4	2.3	2.3
MH	400	1.2	2.4	1.3	1.7	2.8	2.2	1.7	2.2	2.1	1.8
MH	500	1.1	2.3	1.2	1.4	2.7	2.1	1.5	1.9	1.6	1.4

Table 3: Matching metrics of various pairs of shapes

5 Discussion

The need for a good 3D point feature matching algorithm arises in various application areas of medical image analysis. To our knowledge, this is one of the first attempts at 3D diffeomorphic point matching. We have demonstrated a 3D joint clustering and diffeomorphism algorithm and applied it to hippocampal point sets. In the process of careful validation, we investigated the role of the different numbers of clusters in the joint clustering and diffeomorphism optimization process. In the current formulation, we still have a free parameter λ, whose role has to be determined in experiments. The immediate future goal is to further address (theoretically and experimentally) the role of free parameters. The same framework can be used for atlas estimation. Finally, once we have a turnkey 3D diffeomorphic feature matching algorithm, we plan to use it for hippocampal shape classification of epilepsy patients (Vohra, Vemuri, Rangarajan, Gilmore, Roper and Leonard, 2002).

References

Arsigny, V., Commowick, O., Pennec, X. and Ayache, N. 2006. A log-euclidean framework for statistics on diffeomorphisms, *in* R. Larsen, M. Nielsen and J. Sporring (eds), *MICCAI (1)*, Vol. 4190 of *Lecture Notes in Computer Science*, Springer, pp. 924–931.

Avants, B., Schoenemann, P. and Gee, J. 2006. Lagrangian frame diffeomorphic image registration: Morphometric comparison of human and chimpanzee cortex, *Medical Image Analysis* **10**(3): 397–412.
URL: *http://dx.doi.org/10.1016/j.media.2005.03.005*

Baird, H. 1984. *Model-Based Image Matching Using Location*, MIT Press, Cambridge, MA.

Beg, M. F. and Khan, A. 2007. Symmetric data attachment terms for large deformation image registration, *IEEE Trans. Med. Imaging* **26**(9): 1179–1189.

Belongie, S., Malik, J. and Puzicha, J. 2002. Shape matching and object recognition using shape contexts, *IEEE Trans. Patt. Anal. Mach. Intell.* **24**(4): 509–522.

Besl, P. J. and McKay, N. D. 1992. A method for registration of 3-D shapes, *IEEE Trans. Patt. Anal. Mach. Intell.* **14**(2): 239–256.

Camion, V. and Younes, L. 2001. Geodesic interpolating splines, *Energy Minimization Methods for Computer Vision and Pattern Recognition*, Springer, New York, pp. 513–527.

Chui, H. and Rangarajan, A. 2000. A new feature registration framework using mixture models, *IEEE Workshop on Mathematical Methods in Biomedical Image Analysis–MMBIA 2000*, IEEE Press.

Chui, H. and Rangarajan, A. 2001. Learning an atlas from unlabeled point-sets, *IEEE Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA)*, IEEE Press, pp. 179–186.

Chui, H. and Rangarajan, A. 2003. A new point matching algorithm for non-rigid registration, *Computer Vision and Image Understanding* **89**: 114–141.

Chui, H., Win, L., Duncan, J., Schultz, R. and Rangarajan, A. 2003. A unified non-rigid feature registration method for brain mapping, *Medical Image Analysis* **7**: 112–130.

Dubuisson, M. P. and Jain, A. K. 1994. A modified Hausdorff distance for object matching, *ICPR94* pp. A:566–568.

Endres, D. M. and Schindelin, J. E. 2003. A new metric for probability distributions, *IEEE Transactions on Information Theory* **49**(7): 1858–1860.

Glaunes, J., Trouve, A. and Younes, L. 2004. Diffeomorphic matching of distributions: A new approach for unlabelled point-sets and sub-manifolds matching, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 04)*, Vol. 2, pp. 712–718.

Grimson, E., Lozano-Perez, T., Wells III, W., Ettinger, G., White, S. J. and Kikinis, R. 1994. An automatic registration method for frameless stereotaxy, image guided surgery and enhanced reality visualization, *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, IEEE Press, pp. 430–436.

Guo, H., Rangarajan, A., Joshi, S. and Younes, L. 2004. Non-rigid registration of shapes via diffeomorphic point matching, *ISBI 2004* .

Hathaway, R. 1986. Another interpretation of the EM algorithm for mixture distributions, *Statistics and Probability Letters* **4**: 53–56.

Hernandez, M., Bossa, M. and Olmos, S. 2007. Registration of anatomical images using geodesic paths of diffeomorphisms parameterized with stationary vector fields, *ICCV*, IEEE, pp. 1–8.

Hernandez, M., Olmos, S. and Pennec, X. 2008. Comparing algorithms for diffeomorphic registration: stationary lddmm and diffeomorphic demons, *Proc. MFCA Workshop of MICCAI'08*, pp. 24–35.

Jian, B. and Vemuri, B. C. 2006. A Robust Algorithm for Point Set Registration Using Mixture of Gaussians, *International Conference on Computer Vision (ICCV)*, pp. 1246–1251.

Joshi, S. and Miller, M. 2000. Landmark matching via large deformation diffeomorphisms, *IEEE Trans. Image Processing* **9**: 1357–1370.

Marsland, S. and McLachlan, R. 2007. A hamiltonian particle method for diffeomorphic image registration, *in* N. Karssemeijer and B. P. F. Lelieveldt (eds), *IPMI*, Vol. 4584 of *Lecture Notes in Computer Science*, Springer, pp. 396–407.

Mjolsness, E. and Garrett, C. 1990. Algebraic transformations of objective functions, *Neural Networks* **3**: 651–669.

Paragios, N., Rousson, M. and Ramesh, V. 2003. Distance functions for non-rigid registration, *Computer Vision and Image Understanding* **23**: 142–165.

Rangarajan, A., Chui, H., Mjolsness, E., Pappu, S., Davachi, L., Goldman-Rakic, P. and Duncan, J. 1997. A robust point matching algorithm for autoradiograph alignment, *Medical Image Analysis* **4(1)**: 379–398.

Rangarajan, A., Coughlan, J. and Yuille, A. L. 2003. A Bayesian network framework for relational shape matching, *International Conference on Computer Vision (ICCV)*, Vol. 1, pp. 671–678.

Rose, K., Gurewitz, E. and Fox, G. 1990. Statistical mechanics and phase transitions in clustering, *Physical Review Letters* **65**(8): 945–948.

Sclaroff, S. and Pentland, A. P. 1995. Modal matching for correspondence and recognition, *IEEE Trans. Patt. Anal. Mach. Intell.* **17**(6): 545–561.

Scott, G. and Longuet-Higgins, C. 1991. An algorithm for associating the features of two images, *Proc. Royal Society of London* **B244**: 21–26.

Sebastian, T. B., Klein, P. N. and Kimia, B. B. 2004. Recognition of shapes by editing their shock graphs, *IEEE Transaction on Pattern Analysis and Machine Intelligence* **26**(5): 550–571.

Shapiro, L. and Brady, J. 1992. Feature-based correspondence: an eigenvector approach, *Image and Vision Computing* **10**: 283–288.

Siddiqi, K., Shokoufandeh, A., Dickinson, S. and Zucker, S. 1999. Shock graphs and shape matching, *Intl. J. Computer Vision* **35**(1): 13–32.

Thirion, J. P. 1998. Image matching as a diffusion process: an analogy with maxwell's demons, *Medical Image Analysis* **2**(3): 243 – 260.

Tsin, Y. and Kanade, T. 2004. A correlation-based approach to robust point set registration, *European Conference on Computer Vision (ECCV)*, pp. 558–569.

Vohra, N., Vemuri, B. C., Rangarajan, A., Gilmore, R. L., Roper, S. N. and Leonard, C. M. 2002. Kernel Fisher for shape based classification in epilepsy, *Medical Image Computing and Computer-Assisted Intervention (MICCAI02)*, pp. 436–443.

Wahba, G. 1990. *Spline Models for Observational Data*, SIAM, Philadelphia, PA.

Wang, F., Vemuri, B. C., Rangarajan, A., Schmalfuss, I. M. and Eisenschenck, S. J. 2006. Simultaneous registration of multiple point-sets and atlas construction, *European Conference on Computer Vision (ECCV)*.

Yang, D., Li, H., Low, D. A., Deasy, J. O. and Naqa, I. E. 2008. A fast inverse consistent deformable image registration method based on symmetric optical flow computation, *Physics in Medicine and Biology* **53**(21): 6143–6165.

Yeung, S. K., Tang, C.-K., Shi, P., Pluim, J. P. W., Viergever, M. A., Chung, A. C. S. and Shen, H. C. 2008. Enforcing stochastic inverse consistency in non-rigid image registration and matching, *CVPR*, IEEE Computer Society, pp. 1–8.

Zeng, Q. and Chen, Y. 2008. Accurate inverse consistent non-rigid image registration and its application on automatic re-contouring, *in* I. I. Mandoiu, R. Sunderraman and A. Zelikovsky (eds), *ISBRA*, Vol. 4983 of *Lecture Notes in Computer Science*, Springer, pp. 293–304.