Published in final edited form as:
SIAM J Imaging Sci. ; 6(2): . doi:10.1137/12086443X.

A PERFECT MATCH CONDITION FOR POINT-SET MATCHING PROBLEMS USING THE OPTIMAL MASS TRANSPORT APPROACH

PENGWEN CHEN[*], **CHING-LONG LIN**[†], and **I-LIANG CHERN**[‡]

[*]Mathematics, National Taiwan University (pengwen@math.ntu.edu.tw)

[†]Engineering, University of Iowa (ching@engineering.uiowa.edu)

[‡]Department of Applied Mathematics and Center of Mathematical Modeling and Scientific Computing, National Chiao Tung University, Hsin Chu, Taiwan and Mathematics, National Taiwan University (chern@math.ntu.edu.tw)

Abstract

We study the performance of optimal mass transport-based methods applied to point-set matching problems. The present study, which is based on the L2 mass transport cost, states that perfect matches always occur when the product of the point-set cardinality and the norm of the curl of the non-rigid deformation field does not exceed some constant. This analytic result is justified by a numerical study of matching two sets of pulmonary vascular tree branch points whose displacement is caused by the lung volume changes in the same human subject. The nearly perfect match performance verifies the effectiveness of this mass transport-based approach.

Keywords

Point-set matching problems; Optimal Monge-Kantorovich mass transport; Wasserstein metrics; Lung registration

1. Introduction

1.1. Literature reviews

Registration concerns matching two or more sets of image data taken at different times or from different sensors. Depending on the type of image data, registration methods can be classified into two groups: intensity-based methods and feature-based methods. Comprehensive surveys of general registration methods can be found, for instance, in [29] and [60]. Here, we are only concerned with the feature-based methods, where the features are special points in images, such as corners or salient boundary points. The associated point-set matching (registration) problem is to establish a consistent point-to-point correspondence between two point-sets. The most well-known approach is the iterative closest point (ICP) algorithm [4][58]. The advantage of ICP is the simplicity of its implementation. One limitation of ICP is its local convergence restriction (su cient overlap between the point-sets is required for initialization). Another drawback is the algorithm's sensitivity to outliers. To alleviate these challenges, a variety of robust methods have been developed[15][48]. For instance, in [15], Chui and Rangarjan proposed a robust point-matching method (TPS-RPM) that simultaneously estimates non-rigid transformations using

splines and correspondence. A type of matrix balancing technique called soft assignment is employed to establish symmetric correspondence and detect outliers.

The classic Monge-Kantorovich (MK) mass transport problem[1] [24] is to move piles of soil at a minimal cost; the solution is called the optimal mass transport. This approach has been applied to intensity-based image registration, retrieval, and morphing [52], [23], [40], [59], [20], [35], [33]. The optimal cost function is the Wasserstein distance, which measures the dissimilarity of two distributions. One advantage of the MK mass transport approach is that the global minimizer can always be reached because of the inherent convexity of the problem. However, intensity-based methods are sensitive to intensity changes, which can be caused by noise or variations in illumination[60]. Experiments demonstrate that the morphing e effect of mass transport approaches seems unavoidable and may lead to misregistration of medical images[33]. Recently, the mass transport approach has also been incorporated with kernel correlation[48] into feature-based methods[12]. When a finite kernel scale is used, the estimated correspondence is robust against distant outliers. Notice that the proposed model is an approximation of the MK problem, when the kernel scale tends to infinity.

1.2. The present study

We apply two mass transport-based methods to register two large sets (hundreds of point-pairs) of landmark data acquired from a pair of lung CT images (acquired during breath-holds) and establish their correspondence. These methods are tested on the lung CT datasets of several human subjects measured at total lung capacity (TLC) and functional residual capacity (FRC); see Fig. 1.1. Nearly perfect match results are obtained, which demonstrates the effectiveness of the mass transport methods. The results motivate us to further investigate under what conditions the optimal scheme transporting these points coincides with the ground truth correspondence.

Below are our analytic findings:

- **(a)** We establish one sufficient condition (see Prop. 2.3, Cor.2.4) for perfect matches, which states that perfect matches always occur when the product of the point-set cardinality and the norm of the curl[2] of the non-rigid deformation field does not exceed some constant.
- **(b)** We find that using the L2 mass transport cost yields a better performance than using other mass transport costs because of the existence of some correspondence invariants for the L2 mass transport cost (section 2.2).
- **(c)** In lung experiments, perfect matches are likely to occur at landmarks at lower generations (Fig. 4.2). This result supports the proposed sufficient condition when the linear elastic theory predicts the large curls occurring near the lung periphery (in Appendix B).
- **(d)** We demonstrate that the Hellinger distances-based point-set matching model[12] possesses cyclical monotonicity in the optimal transport(Remark 3.4). Experiments (section 4.3) demonstrate that (1) mass transport-based methods outperform TPS-RPM, which does not possess cyclical monotonicity; (2) the HD model is more robust against outliers than the mass transport model.

[1] A survey of theoretical studies of this problem can be found in [16] or [49].

[2] Using the Helmholtz decomposition, a vector field on R^d can be expressed as a sum of a divergence-free vector field and a curl-free vector field. Additionally, according to Brenier's polar factorization theorem [8], Theorem 3.8[49], a L2 vector-valued mapping can be expressed as a composition of the gradient of a convex function and a measure-preserving mapping.

To the best of our knowledge, this is the first study to demonstrate the outstanding matching performance of mass transport methods.

This paper is organized as follows: In sections 2 and 3, we present several theoretical studies of the mass transport-based point-matching models, including the curl-cardinality relation and the cyclical monotonicity property. In section 4, we present various matching experiments performed on lung branch points to support the theoretical analysis. Lengthy proofs, including those for the curl-cardinality relations, the curl distribution of lung deformations, the asymptotic behavior of the rigid motion estimator and the kernel scale selection in the HD model, are listed in the appendix,

2. Mass transport-based point-set matching models

In this paper, let δ be the Dirac delta function and $\|\cdot\|_2 = \|\cdot\|$ be the 2-norm. We denote the trace and the transpose of a matrix X by $T_r(X)$ and $X^\top$, respectively.

2.1. Mass transport problems and point-set matching problems

One central problem in elasticity theory concerns the determination of the deformation T on a bounded open connected subset Ω of $\mathbf{R}^3$ subjected to some applied force. The deformation T must be injective and orientation-preserving (i.e., the deformation gradient $\det(\nabla T)$ is positive) to be physically acceptable.

The primary focus of this paper is the inverse problem, which is to determine the correspondence between two unlabeled point-sets $\{x_i\}_{i=1}^n$, $\{y_i\}_{i=1}^n$ sampled from Ω and $T(\Omega)$. Here, the correspondence is described by a permutation τ on $\{1, \dots, n\}$ such that $y_i = T(x_{\tau(i)})$. To proceed, we require a stronger assumption about the transforms: T is twice continuously differentiable and has the Helmholtz decomposition $T(x) = \nabla\varphi + \nabla \times \psi$, where φ is strongly convex and $\nabla \cdot \psi = 0$. Note that φ, ψ can be determined by solving Poisson's equation (see pages 238–242 in [41]):

$$\nabla \cdot T = \nabla^2\phi, \nabla \times T = -\nabla^2\psi.$$

Suppose that the corresponding point-pairs $\{x_i, y_{\tau(i)}\}$ are nearby. Then, τ can be estimated by solving the minimization problem:

$$\min_{\tau} \sum_{i=1}^{n} \|x_i - y_{\tau(i)}\|^\alpha, \alpha \geq 1. \quad (2.1)$$

This is a combinatorial optimization problem because $n!$ possibilities must be evaluated. This difficulty can be alleviated if we instead consider a relaxed problem,

$$\min_{\mu_{i,j}} \sum_{i=1}^{n} \|x_i - y_j\|^\alpha \mu_{i,j}, \quad (2.2)$$

subject to the unit mass constraint, $\sum_{i=1}^n \mu_{i,j} = 1 = \sum_{j=1}^n \mu_{i,j}$, $\mu_{i,j} \geq 0$. This problem is known as the L^α Monge-Kantorovich mass transport problem. Here, the original permutation τ is relaxed to a correspondence matrix characterized by the doubly stochastic matrix $\{\mu_{i,j} : i, j = 1, \dots, n\}$ or the measure $\mu := \sum_{i,j=1}^n \mu_{i,j} \delta(x - x_i, y - y_j)$. More precisely, $\tau(i) = j$ if $\mu_{i,j} = 1$.

The relaxed problem described by Eq. (2.2) is a convex (in fact, linear) minimization problem, which has an optimal permutation matrix (the existence of this is guaranteed by Birkhoff's theorem) and can be solved by interior point methods[7] or primal-dual algorithms [23] (see chapter 4 in [9] also). Finally, we say that a *perfect match* occurs if the underlying correspondence between two point-sets coincides with a minimizer $\{\mu_{i,j}\}_{i,j=1}^{n}$ of Eq. (2.2).

2.2. Two special properties of the L2 cost

In this paper, the *MK method* is defined by correspondence estimation using Eq. (2.2) with $\alpha = 2$. This method has two special properties.

The first special property is the optimality condition, called cyclically monotonicity. A nonempty subset $\{(x_i, y_i)\}_{i=1}^{n}$ in $\boldsymbol{R}^d$, $d \geq 1$ is said to be cyclically monotone (p. 79 [49]) if for all $m \geq 2$ and for all subsets $\{(x_i, y_i)\}_{i=1}^{m}$,

$$\sum_{i=1}^{m} \|x_i - y_i\|^2 \leq \sum_{i=1}^{m} \|x_i - y_{i-1}\|^2, \text{with the convention } y_0 = y_m.$$

In this context, $\mu = \sum_{i=1}^{n} \delta(x - x_i, y - y_i)$ is optimal in the L2 mass transport problem when the correspondence $\{(x_i, y_i)\}_{i=1}^{n}$ is cyclically monotone. One key characterization the optimality condition is that if the support of μ satisfies cyclical monotonicity, then μ is supported in the sub-differential of a proper lower semi-continuous convex function φ (Theorem 2.27[49]), i.e., $y_i \in \partial\varphi(x_i)$, where ∂ refers to sub-gradients[37].[3]

Consider a transform $T : \boldsymbol{R}^d \to \boldsymbol{R}^d$ between two point-sets $\{x_i\}_{i=1}^{m}, \{y_i\}_{i=1}^{n}$ in $\boldsymbol{R}^d$ with $y_i = T(x_i)$. When T is the gradient of some convex function, then the correspondence can be recovered correctly by solving mass transport problems. This special class of the transforms include scalings, translations, positive definite affine transforms ($T(x) := Ax + t$, where A is positive definite) and other curl-free maps. Note that when φ is strongly convex and differentiable, the Hessian of φ is positive definite, which implies that $\nabla\varphi$ is orientation-preserving and injective.

The second special property of the L2 cost is that the correspondence is invariant under an additional transform S. Here, we discuss two correspondence invariants: (i) between $\{S(x_i)\}_{i=1}^{n}$ and $\{S(y_i)\}_{i=1}^{n}$ and (ii) between $\{S(x_i)\}_{i=1}^{n}$ and $\{y_i\}_{i=1}^{n}$.

In R^1, for one-dimensional point-sets, a *cyclically monotone* correspondence μ is monotone rearrangement (in which the spatial ordering of points is preserved); see page 75[49]. Thus, cyclical monotonicity of $\{(x_i, y_i)\}_{i=1}^{n}$ implies that $\{S(x_i), S(y_i)\}_{i=1}^{n}$, and $\{S(x_i), y_i\}_{i=1}^{n}$ are cyclical monotonic if S is an increasing function. We have both invariants (i) and (ii).

In R^d, $d > 1$, a *cyclically monotone* correspondence $\{(x_i, y_i)\}_{i=1}^{n}$ does not imply cyclical monotonicity of (i)$\{(S(x_i), S(y_i))\}_{i=1}^{n}$ or (ii) $\{(S(x_i), y_i)\}_{i=1}^{n}$. Instead, for type (i), the correspondence in L^{α} cost is invariant under rigid motions and scalings ($S(x) = Qx + t$, $S(x) = ax + t$, with $a > 0$, $t \in \mathrm{R}^d$ and Q an orthogonal matrix).[4] For invariant (ii), we will first demonstrate that the L2 cost possesses one additional (forward-backward) invariant:

[3]Regard two point-sets as finite realizations of two random variables. According to Theorem 2.32[49] or [30], for two probability measures μ, ν, where μ is absolutely continuous with respect to the Lebesgue measure, there exists a unique measurable map T such that the push-forward measure $T\#\mu = \nu$ and $T = \nabla\varphi$ for some convex function φ.

Proposition 2.1 (Forward-backward). Suppose that a matrix μ minimizes the L2 mass transport cost between $\{x_i\}_{i=1}^n$ and $\{y_i\}_{i=1}^n$. Let S(x) = Ax + t be an affine transform with a nonsingular symmetric matrix A and a vector t. Then, the matrix μ also minimizes the L2 mass transport cost between $\{S(x_i)\}$ and $\{S^{-1}(y_j)\}$:

$$\underset{\mu}{\arg\min} \sum_{i,j=1}^{n} \mu_{i,j} \|x_i - y_j\|^2 = \underset{\mu}{\arg\min} \sum_{i,j=1}^{n} \mu_{i,j} \| (Ax_i + t) - A^{-1}(y_j - t) \|^2.$$

In particular, the result holds for positive definite matrices A.

Proof. Because $\sum_{i=1}^n \mu_{i,j} = 1 = \sum_{j=1}^n \mu_{i,j}$, any minimizer $\mu_{i,j}$ for the cost $\sum_{i,j} \mu_{i,j} \| (Ax_i + t) - A^{-1}(y_j - t) \|^2 = \sum_{i,j} (-2)\, \mu_{i,j} y_j^{\intercal} x_i +$ constant terms w.r.t. $\mu_{i,j}$ also minimizes $\sum_{i,j=1}^n \mu_{i,j} \|x_i - y_j\|^2$.

Combining the first invariant with the optimal correspondence $\{(x_i, y_i)\}_{i=1}^n$ yields the optimal correspondence $\{(ax_i + t, y_i)\}_{i=1}^n$, where *a* is a positive scalar, which is invariant (ii). In practical applications, the invariant of the L2 cost eliminates the difficulty of estimating the parameters *a*, *t*.

2.3. Curl-cardinality relations

Typically, the match error of the mass transport approach is caused by two factors, nonzero curls and outliers. We first examine the effect of curl on the occurrence of mismatch. A discussion regarding outliers will be presented in the following section.

According to the Helmholtz decompositions, a three-dimensional smooth vector field can be expressed as a sum of a gradient function and a curl function. Regarding mass transport methods, point correspondence can be estimated correctly if the transform is the gradient of some convex function. However, these special transforms form a very small class, and they rarely occur in practical applications. Consider any point-set with finite cardinality *n* sampled from Ω. The following analysis demonstrates that if the curl magnitude ω_{max} (Def. 2.1) of the transform *T* is less than some upper bound *C/n*, then a perfect match can still be obtained; here, *C* is a positive constant that is related to the constant in the isoperimetric inequality[10]. Empirical experiments demonstrate that the upper bound can be significantly improved if the point-set is scattered uniformly over some region of a high-dimensional space rather than concentrated in a circle.

Definition 2.1. For each d × d matrix B, let Λ(B, k) be the k^{th} largest singular value of B. Let T be a transform on R^d, d ≥ 2. Let T_S and T_A be the symmetric and skew symmetric parts of ∇T at each x in the domain Ω, i.e., $\nabla T(x) = T_S(x) + T_A(x)$. Define the maximum curl ω_{max} of T on Ω as

[4] In general, the correspondence in the mass transport cost $\|x - y\|^\alpha$ is not invariant under affine transforms. Here is one example: Let $\{x_i\}_{i=1}^4 = \{(1,1), (0,1), (0,0), (-1,0)\}$ and $\{y_i\}_{i=1}^4 = \{(-1,1), (0,1), (0,0), (1,0)\}$. Consider two affinely transformed point-sets $\{Ax_i\}$, $\{Ay_i\}$, where *A* is a diagonal matrix with diagonal entries [*a*, 1]. For $a \in (0, 1)$ and close to zero, x_1 is matched with y_2 (consider $\{x_i\}_{i=1}^4$ and $\{y_i\}_{i=1}^4$ as $\{x_1, x_2\} \cup \{x_3, x_4\}$ and $\{y_1, y_2\} \cup \{y_3, y_4\}$). However, x_1 is matched with y_4 for *a* sufficiently greater than 1 (consider $\{x_i\}_{i=1}^4$ and $\{y_i\}_{i=1}^4$ as $\{x_1\} \cup \{x_2, x_3\} \cup \{x_4\}$ and $\{y_1\} \cup \{y_2, y_3\} \cup \{y_4\}$).

$$\omega_{max}(T;\Omega) = 2\tan^{-1}\left(\frac{\max_x \Lambda(T_A(x),1)}{\min_x \Lambda(T_S,(x),d)}\right). \quad (2.3)$$

In the case where $\Omega \subset R^3$ and $T = \nabla\varphi + \nabla\times\psi$ with φ strongly convex, we obtain

$$\Lambda(T_A,2) = 0 \text{ and } \Lambda(T_A,1) = -\Lambda(T_A,3) = 2^{-1}\|\nabla \times T\| = 2^{-1}\|\nabla^2\psi\|^2.$$

Thus, $\tan(\omega_{max}/2)$ is proportional to the maximum curl, $\max_x \|\nabla \times T(x)\|$.

To better understand ω_{max}, we first examine the upper bound on rotation angles for perfect matches.

Example 2.2. Consider one point-set consisting of n points $\{x_i \in R^2 : i = 1, \ldots, n \geq 2\}$ in a circle (centered at the origin) with radius r and with polar angle $\{\theta_i = 2i\pi/n : i = 1, \ldots, n\}$ and another point-set consisting of $\{y_i : i = 1, \ldots, n\}$ in the circle with polar angle $\theta_i + \omega$. Use the conventions $x_{n+1} = x_1$ and $x_0 = x_n$. The perfect match condition is forfeited under ω if

$$\sum_{i=1}^{n} x_i \cdot y_i = nr^2\cos\omega \leq \max\left(\sum_{i=1}^{n} x_{i+1}\cdot y_i, \sum_{i=1}^{n} x_{i-1}\cdot y_i\right),$$

which implies $\tan|\omega| \geq (1 - \cos(2\pi/n))/\sin(2\pi/n) = \tan(\pi/n)$, i.e., $|\omega| \geq \pi/n$. One can easily verify that $\Lambda(T_A; 1)/\Lambda(T_S, 2) = \tan|\omega|$ and ω_{max} in Def. 2.1 is exactly $2|\omega|$. Hence, a perfect match can be obtained if the rotation angle ω lies in $(\omega_-, \omega_+) := (-\pi/n, \pi/n)$ or $\omega_{max} \leq 2\pi/n$. Here, we term the range (ω_-, ω_+) the perfect match range. Note that this upper bound (which does not depend on the spatial size 2r) can be regarded as the angular resolution of the point-set (i.e., the average angle between points for n points $\{x_i\}_{i=1}^n$ evenly spaced in the angular range 2π).

In the following proposition and corollary, we present the *curl-cardinality* relation for rotations and general transforms. Consider a point-set $\{x_i\}_{i=1}^n$ in R^d and generate another point-set $\{y_i\}_{i=1}^n$ by rotating $\{x_i\}_{i=1}^n$ by angle ω. A rotation can be described by an orthogonal matrix R, so let $T(x) = Rx$. Perform the decomposition $\nabla T = R = R_S + R_A$, where R_S, R_A are the symmetric and skew symmetric parts of R. The magnitude of the rotation can then be measured by the ratio $\Lambda(R_A, 1)/\Lambda(R_S, d)$.

Proposition 2.3 (Curl-cardinality for rotations). Suppose that the transform T is a rotation.

- **(i)** A perfect match can be obtained for a set of n points in R^d if $\omega_{max}(T, R^d) \leq 2\pi/n$.
- **(ii)** Among all of the possible point-sets consisting of n points, the point-set that consists of the vertices of the n-sided regular polygon has the smallest perfect match range, $(\omega_-, \omega_+) := (-\pi/n, \pi/n)$.

The proof can be found in the appendix.

Corollary 2.4 (*Curl-cardinality for general deformations*). Decompose the gradient of a general transform T as a sum of the symmetric and skew symmetric parts,

$$\nabla T = T_S + T_A \text{ on } \mathbf{R}^d, d \geq 2.$$

(i) For a general transform T on $\boldsymbol{R}^{\mathrm{d}}$, the above curl-cardinality relation still holds with the constant C not necessarily equal to 2π (the difference is due to the Riemann sum approximation).

(ii) Suppose that T_S is constant, positive definite and can be factored as $T_S^{1/2} T_S^{1/2}$. The curl-cardinality relation can be improved, i.e., perfect matches occur if $\omega_{\max} \leq 2\pi/n$, where

$$\omega_{max} := 2\tan^{-1}\left(\max_x \Lambda\left(T_S^{-1/2} T_A(x) T_S^{-1/2}, 1\right)\right). \quad (2.4)$$

The proof can be found in the appendix. From the proof of the Corollary 2.4, according to the isoperimetric inequality and lemma A, one can easily verify that the upper bounds occur only when the points are exactly the vertices of the regular polygon[5].

2.4. Curl-cardinality relations

$\omega_{max} \leq C/n^{-1/d}$. At first glance, Prop. 2.3(or Cor. 2.4) might take one pessimistic about the occurrence of perfect matches. However, usually the value $\omega_+ - \omega_-$ for rotating a high-dimensional point-set is far greater than $2\pi/n$ (when the points are not concentrated in a planar circle). In fact, from the proof of the proposition, n can be reduced to a maximum length of disjoint cycles of the corresponding permutation τ. Let us examine one special case: In 2D, suppose the points are located at rectangular grid points $x_{i,j} = (i/\sqrt{n}, j/\sqrt{n})$ for $i, j = 1, \ldots, \sqrt{n} \in \mathbb{N}$. Along the rectangular boundary, there are $4\sqrt{n}$ points. The angular resolution is approximately $2\pi/4\sqrt{n} = \pi/2\sqrt{n}$. Thus, $\omega_+ - \omega_- \approx 1.57 n^{-1/2}$.[6] Similar arguments demonstrate that $\omega_+ - \omega_- \approx 1.57 n^{-1/3}$ for the cases of 3D rectangular grids.

Here, one simulation of the curl-cardinality relation in which the coefficient is greater than $\pi/2 \approx 1.57$ is presented. Consider a point-set randomly generated from the unit square $[0, 1] \times [0, 1]$ uniformly. Rotate the point-set along the z-axis by an angle ω. Each perfect match range (ω_-, ω_+) is measured under various point cardinalities, $n = 50, \ldots, 200$. The result is reported in Table 2.2 and Fig. 2.1. In the figure, the green solid line reveals a linear relationship between $n^{-1/2}$ and $\omega_+ - \omega_-$, where (approximately) $\omega_+ - \omega_- = C_2 n^{-1/2}$, with $C_2 = 2.90$. Similarly, we measure the perfect match range for 3D point-sets (randomly generated from $[0, 1] \times [0, 1] \times [0, 1]$) with different point cardinalities. The result is shown by the red dashed line,which is (approximately) $\omega_+ - \omega_- = C_3 n^{-1/3}$, with $C_3 = 2.86$.

Consequently, under the same magnitude of curls, the difficulty matching 1000 (randomly sampled) 3D points in mass transport models is similar to the difficulty of matching 100 2D points and the difficulty of matching 40 points on the four sides of a square. Hence, it is not surprising that perfect matches occur when matching 3D lung point-sets containing 1000 points. See the experimental results for details.

---

[5] In (ii), points $\{x_i\}_{i=1}^n$ are the vertices $\{y_i\}_{i=1}^n$ of the regular polygon under affine transforms, $x_i = T_S^{-1/2} y_i$

[6] To see the origin of the factor $n^{1/2}$, consider a polygon $\{x_i\}_{i=1}^m$ with constant side lengths $\|x_i - x_{i-1}\| = h$ for $i = 1, \ldots, m$ and $x_0 = x_m$. Fix *Area* (the area of the enclosed region). Since $Tr(S) = mh^2$, an upper bound for $\tan|\theta|$ is $mh^2/4Area$ (see Eq. (??)). Hence, the tightest bound for ω_{max} is given by the polygon with the minimum cardinality m.

3. Outliers in matching problems

In this section, we impose the same (small curl) assumption on transforms as in the previous section. Hence, the mass transport method yields the correct correspondence in this ideal case. We will study the outlier effect in mass transport models. In our experiment, outliers refer to the points that appear in one point-set but that are missing correspondences in the other point-set. From a mathematical viewpoint, because the points are unlabeled, it is impossible to distinguish outliers from the "inliers". Hence, perfect matches are unattainable when outliers exist. In fact, the matching result can be much worse. Sometimes, even a small number of distant outliers can cause large mismatches in the aforementioned mass transport model (see the experimental results).[7]

One possible solution is the HD model [12], in which correspondences are estimated by maximizing

$$\sum_{i,j=1}^{m,n} \sqrt{\gamma_{i,j}^{+}\gamma_{i,j}^{-}} \exp\left(-\|x_i - y_i\|^2/2\sigma^2\right), \quad (3.1)$$

with respect to nonnegative unknowns $\gamma_{i,j}^{+}$, $\gamma_{i,j}^{-}$, subject to the unit mass constraint, $\sum_{i=1}^{m}\gamma_{i,j}^{+}=1=\sum_{j=1}^{n}\gamma_{i,j}^{-}$ for all i, j. Here, the correspondence is characterized by two matrices $\gamma_{i,j}^{+}$, $\gamma_{i,j}^{-}$. This model yields the same correspondence result as the mass transport model, when σ tends to infinity (B.3 in [12]).

In this section, we will demonstrate that the optimal correspondence in the HD model is indeed cyclically monotone (see Remark 3.4). Hence, the HD model can be viewed as a re-weighted mass transport model in which the weight function is related to the spatial distance of each corresponding point-pair. When a finite kernel scale σ is used, this model is robust to distant outliers (see the experimental results presented in section 4). Indeed, for some sufficiently large σ, the correspondence determined by the majority rule(Eq. 4.2) is identical to the correspondence generated using the MK mass transport model. Thus, robustness against outliers is obtained from the use of finite kernel scales.

3.1. Duality

We will examine the optimality condition in the HD model by analyzing the duality structure between the maximization problem Eq. (3.1) and its dual problem Eq. (3.3). Let S = {(Γ^+, Γ^-) : Γ^+, Γ^- are $m \times n$ matrices with entries $\gamma_{i,j}^{+} \geq 0$, $\gamma_{i,j}^{-} \geq 0$ satisfying $\sum_{j=1}^{n}\gamma_{i,j}^{+}=\gamma_i^{+}$, $\sum_{i=1}^{m}\gamma_{ij}^{-}=\gamma_j^{-}$}. Let

$$E\left(\Gamma^+, \Gamma^-\right) = \begin{cases} \sum_{i=1,j=1}^{m,n} 2\sqrt{\gamma_{i,j}^{+}\gamma_{i.j}^{-}} K\left(x_i, y_j\right), & \text{if } \left(\Gamma^+, \Gamma^-\right) \in S \\ -\infty & \text{else.} \end{cases} \quad (3.2)$$

Let T = {(φ, ψ) : Vectors φ, ψ have positive entries φ_i; ψ_j with $\varphi_i;\psi_j \geq K(x_i, y_j)^2$}. Here, the dual problem max E is presented:

[7]For example, consider one-dimensional point-sets {x_i}, {y_i} on the x-axis with perfect matches. Add one outlier x to {x_i} and add one outlier y to {y_i}, where $x > x_i$ and $y < y_i$ for all i. Then, we obtain a 100% mismatch rate.

$$\min_{(\phi,\psi)\in T} J(\phi,\psi), \text{ where } J(\phi,\psi) := \sum_{i=1}^{m}\phi_i\gamma_i^+ + \sum_{j=1}^{n}\psi_j\gamma_j^- . \quad (3.3)$$

Indeed, these two problems are connected by the weak duality:

$$\max_{(\Gamma^+,\Gamma^-)\in S} E\left(\Gamma^+,\Gamma^-\right) \le \min_{(\phi,\psi)\in T} J(\phi,\psi). \quad (3.4)$$

We say that (Γ^+, Γ^-, φ, ψ) is a saddle point if

$$E\left(\Gamma^+,\Gamma^-\right) = J(\phi,\psi). \quad (3.5)$$

The weak and strong dualities have been studied in [12], and the strong duality can be established using the strong duality theorem (Prop. 5.2.1[3]). The following proposition summarizes the optimality conditions.

Proposition 3.1. Assume $K_{i,j} > 0$ for all i, j. Then

$$\max_{(\Gamma^+,\Gamma^-)\in S} E\left(\Gamma^+,\Gamma^-\right) = \min_{(\phi,\psi)\in T} J(\phi,\psi).$$

This result indicates the absence of the duality gap. That is, if the matrices (Γ^+, Γ^-) and the vectors (φ, ψ) are optimal, then from Eq. (3.4), the following conditions hold:

1. For all i, j, $\gamma_{i,j}^+\gamma_{i,j}^-\left(\phi_i\psi_j - K_{i,j}^2\right)=0$. Then, $\gamma_{i,j}^+=0=\gamma_{i,j}^-$ for each pair (*i*, *j*) with $\phi_i\psi_j > K_{i,j}^2$.
2. $\phi_i\gamma_{i,j}^+=\psi_j\gamma_{i,j}^-$ for all i, j.

Conversely, when these two conditions are fulfilled, (Γ^+, Γ^-, φ, ψ) is a saddle point.

From Prop. 3.1, the (*i*, *j*) entries $K_{i,j}^2$ should coincide with those in the product of the two vectors φ, ψ if $\gamma_{i,j}^+\gamma_{i,j}^- > 0$. Thus, given a matrix $\{K_{i,j} : i = 1, \ldots, m, j = 1, \ldots, n\}$ with nonzero minors[8], the correspondence matrices (Γ^+, Γ^-) are highly sparse. Here, a block coordinate descent method to compute Γ^+, Γ^- is presented.

Algorithm 3.2 (Correspondence estimation[12]).

1. Initialize matrix Γ^- with entries $\gamma_{i,j}^- = 1/n$ and matrix K with entries $K_{i,j} = \exp\left(-\|y_j - x_i\|^\alpha / 2\sigma^\alpha\right)$. Let σ be some kernel scale.
2. Repeat the iterations till they converge,

$$\gamma_{i,j}^+ \leftarrow \frac{\gamma_{i,j}^- K_{i,j}^2}{\sum_{j=1}^{n}\gamma_{i,j}^- K_{i,j}^2}, \gamma_{i,j}^- \leftarrow \frac{\gamma_{i,j}^+ K_{i,j}^2}{\sum_{i=1}^{m}\gamma_{i,j}^+ K_{i,j}^2}. \quad (3.6)$$

[8]All square sub-matrices have a nonzero determinant.

In Theorem 4.5 [13], any limit of the sequences is a *global* maximizer independent of the initial *positive* matrices Γ^+ and Γ^-. Empirically, the convergence speed is typically acceptable for point-sets with hundreds of points and small kernel scales.

3.2. Cyclical monotonicity in the HD model

Rearrange and partition each pair of discrete masses ν^+, ν^- properly, such that the matching is "bijective", i.e.,

$$\nu^+=\sum_{i=1}^{n}\gamma_i^+\delta\left(x-x_i\right),\nu^-=\sum_{i=1}^{n}\gamma_i^-\delta\left(y-y_i\right).$$

Then, the maximizer (Γ^+, Γ^-) of E can be expressed as a pair of square diagonal matrices with diagonal entries $\left\{\gamma_i^+\right\}_{i=1}^n$ and $\left\{\gamma_i^-\right\}_{i=1}^n$. The next proposition characterizes the optimal bijective matching described by $\{(\gamma_i^+,\gamma_i^-) : i = 1, \ldots, n\}$.

Proposition 3.3. The above bijective matching is optimal if and only if

$$\frac{K_{i,i}K_{j,j}}{K_{i,j}^2}\geq\sqrt{\frac{\gamma_i^+\gamma_j^-}{\gamma_j^+\gamma_i^-}}\geq\frac{K_{j,i}^2}{K_{i,i}K_{j,j}},\quad\text{for all } i,j. \tag{3.7}$$

Note that for either $y_i = y_j$ or $x_i = x_j$ we have

$$\frac{\gamma_i^-}{\gamma_j^-}=\frac{\gamma_i^+K_{i,i}^2}{\gamma_j^+K_{j,i}^2}\,or\,\frac{\gamma_i^+}{\gamma_j^+}=\frac{\gamma_i^-K_{i,i}^2}{\gamma_j^-K_{i,j}^2},\,respectively. \tag{3.8}$$

Proof. (The only-if part) Let (φ, ψ) be a minimizer of the dual problem J. Then, we have $K_{i,i}^2=\phi_i\psi_i$, $K_{j,j}^2=\phi_j\psi_j$ and $K_{i,j}^2\leq\phi_i\psi_j$, which implies

$$\frac{K_{i,i}K_{j,j}}{K_{i,j}^2}\geq\sqrt{\frac{\phi_i\phi_j\psi_i\psi_j}{\phi_i^2\psi_j^2}}=\sqrt{\frac{\psi_i}{\phi_i}\frac{\phi_j}{\psi_j}}=\sqrt{\frac{\gamma_i^+\gamma_j^-}{\gamma_i^-\gamma_j^+}},$$

where we also used $\gamma_i^+\phi_i=\gamma_i^-\psi_i$. Thus, we proved the first inequality. The second inequality is obtained by exchanging i and j.

(The if part) Suppose that Γ^+, Γ^- and $K_{i,j}$ satisfy the inequality. Let

$$\phi_i=K_{i,i}\sqrt{\gamma_i^-/\gamma_i^+},\psi_i=K_{i,i}\sqrt{\gamma_i^+/\gamma_i^-}.$$

One can then easily verify $\phi_i\gamma_i^+=\psi_i\gamma_i^-$, $\phi_i\psi_j\geq K_{i,j}^2$. From Prop. 3.1, ($\Gamma^+$, Γ^-, φ, ψ) is a saddle point. Thus, the diagonal matrices (Γ^+, Γ^-) are an optimal pair.

This proposition yields two consequences.

Remark 3.4 (Cyclical monotonicity). The inequality in Eq. (3.7) yields the following "c"-cyclical monotonicity [50], where the cost function "c" defined as $-log$ K. For any natural

number N and any subset $\{(x_1, y_1), \ldots, (x_N, y_N)\}$ of two point-sets, the following inequality holds:

$$\frac{K_{1,1}\left(\Pi_{i=2}^{N-1}K_{i,i}^2\right)K_{N,N}}{\Pi_{i=1}^{N-1}K_{i,i+1}^2} \geq \prod_{i=1}^{N-1}\sqrt{\frac{\gamma_i^+\gamma_{i+1}^-}{\gamma_i^-\gamma_{i+1}^+}} = \sqrt{\frac{\gamma_1^+\gamma_N^-}{\gamma_1^-\gamma_N^+}} \geq \frac{K_{N,1}^2}{K_{1,1}K_{N,N}},$$

i.e., we obtain the c-cyclical monotonicity

$$\sum_{i=1}^{N}\log K_{i,i} \geq \sum_{i=1}^{N}\log K_{i,i+1} \; where \; K_{N,N+1} := K_{N,1}.$$

When $K(x,y) = \exp\left(-\|x-y\|^2/\sigma^2\right)$, cyclical monotonicity implies that $\{(x_i, y_i) : i = 1, \ldots, N\}$ is included in the sub-differential of a proper lower semi-continuous convex function. From Eq. (3.8), both the upper and lower bounds of the ratio $\gamma_i^+\gamma_{i+1}^-/\gamma_i^-\gamma_{i+1}^+$ tend to *1* as $\sigma^2 \to \infty$. Hence, the pair of limit correspondences in the HD model converge to the same limit, which is exactly the correspondence matrix in the mass transport model.

Remark 3.5 (Closest point property[12]). Consider $K(x,y) = \exp\left(-\|x-y\|^\alpha/2\sigma^\alpha\right)$, $\alpha \geq 1$. Consider two point-sets $\{x_i : i = 1, \ldots, n\}$, $\{y_i : i = 1, \ldots, n\}$ with $|x_i - y_i| < min_i\{\|x_j - y_i\|, \|x_i - y_j\| : j \neq i\}$ for all i, i.e., x_i is the closest point to y_i among the first point-set and vice versa (see Fig. 3.1). $K_{i,i}K_{j,j}/K_{i,j}^2$ has a lower bound *1*. Then, the bijective matching $\{(x_i, y_i) : i = 1, \ldots, n\}$ is optimal if γ_i^+/γ_i^- is constant for each i. Hence in this case, the HD model produces the same result as the ICP method[58]. The conclusion does not depend on σ.

Remark 3.4 demonstrates that in the case of no outliers, there exists a minimum kernel scale σ_{min} such that the correspondence produced by the HD model with the kernel scale $\sigma \geq \sigma_{min}$ is the same as the correspondence μ produced by the MK method and modified by the majority rule, Eq. (4.2). To suppress the outlier effect, we should use small kernel scales if possible. According to one example in appendix D, the value of the minimum kernel scale σ_{min} is approximately proportional to the spatial size of the point-set, the largest displacement and the point cardinality (see Eq. (D.4)). In practice, because the displacement is unknown, we simply select σ to be a value between $|\Delta|/2$ and $2|\Delta|$, where $|\Delta|$ is the spatial size of point-sets.

3.3. Pre-processing for HD methods

One can use the moment method[44] to shorten the distance between the corresponding point pairs: Let $\hat{x}_i = x_i - t_x$ and $\hat{y}_i = y_i - t_y$, where $t_x := n_1^{-1}\sum_{i=1}^{n_1}x_i$, $t_y := n_2^{-1}\sum_{i=1}^{n_2}y_i$. Let $M_X := n_1^{-1}\sum_{i=1}^{n_1}\hat{x}_i\hat{x}_i^\intercal$, $M_Y := n_2^{-1}\sum_{i=1}^{n_2}\hat{y}_i\hat{y}_i^\intercal$ with singular value decompositions $M_X = U_X D_X U_X^\intercal$ and $M_Y = U_Y D_Y U_Y^\intercal$. Let

$$A = U_Y D_Y^{1/2} D_X^{-1/2} U_X^\intercal. \quad (3.9)$$

Then, $\{A\hat{x}_i\}_{i=1}^{n_1}$ and $\{\hat{y}_i\}_{i=1}^{n_2}$ have the same first two moments. Here, t_x, M_X are called the first and second (central) moments of $\{x_i\}_{i=1}^{n_1}$.

One naive approach is the following: employ the HD model on the point-sets $\{A\widehat{x}_i\}_{i=1}^{n_1}$ and $\{\widehat{y}_i\}_{i=1}^{n_2}$ instead of the original point-sets. However, in some cases, this approach could lead to poor performance because of the instability of the orthogonal matrices (in particular, the second moment has close eigenvalues[9]).

Instead, we consider the positive definite affine transform $T_1(x) = Ax + t$ as the preprocessing transform, where $A = U_X D_Y^{1/2} D_X^{-1/2} U_X^{\intercal}$, and t refers to a translation vector. Point-sets are preprocessed in the manner, $\{A(x_i - t_x)\}$, $\{y_i - t_y\}$, where $t_x, t_y \partial \mathbf{R}^d$ are chosen to match their first moments of the point-sets. We have empirically found that this preprocessing can shorten the distance between each desired point-pair and thus decrease the kernel scale requirement in the HD model. Hereafter, the *HD method* refers to the HD model applied to the preprocessed point-sets.

Two sets of experimental results are provided to demonstrate the method's performance (see Fig. 3.2 and Fig. 3.4): (i) Fig. 3.2 shows a comparison of the performance of the HD method with and without the preprocessing at matching two shapes[10] ii) Fig. 3.4 shows the matching result for the lung (TLC /FRC) branch points (shown in the left of Fig. 3.3) obtained using the mass transport-based methods (the MK and HD methods). The details of the experimental setup can be found in section 4.4. Clearly, the use of preprocessing improves both matching results.

3.4. General transforms

In applications, we often must the task of matching point-sets with large curl and large cardinality. To overcome the restriction of the curl-cardinality relation, we can incorporate other methods into the mass transport methods.

In the following, we consider the following smoothness assumption about T : Express T as the composition of two smooth transforms, $T = T_2 \bigcirc T_1$. We assume T_1 and $T \in \mathscr{N}_{\Phi(\Omega)}$, where $\mathscr{N}_\Phi$ is the reproducing kernel Hilbert space generated by a positive semi-definite kernel Φ with norm $\| \cdot \| \Phi(\Omega)$. For simplicity, let Φ be a Gaussian kernel with kernel scales σ_S and let the domain Ω be a ball.[11]

Here, we discuss two approaches for estimating T_1 (note that the transform T_2 is handled by the correspondence in the HD models).

3.4.1. HD-RBF—In section 3.3 [12], the following model is proposed to estimate the point-to-point correspondence and the transform T_1:

$$\min_{T_1} \min_{\Gamma^+,\Gamma^-} \left(-\sum_{i,j} \sqrt{\gamma_{i,j}^+ \gamma_{i,j}^-} \exp\left(-\|T_1(x_i) - y_j\|^2/\sigma_K^2 + \lambda \|T_1\|^2_{N_{\Phi(\Omega)}} \right) \right), \quad (3.10)$$

where λ is some positive parameter and the basis functions for T_1 consist of Gaussian kernels and affine transforms. We will call this the HD-RBF model (Hellinger distances-radial basis functions). Our experimental results indicate that this model performs well at matching the lung point-sets even though the model is a non-convex minimization problem. The introduction of the transform T_1 has two advantages. First, it provides one possible

[9]For instance, one point-set is a set of rectangular grid points. See the experimental results presented in Table 3.1.
[10]The point-set is provided by courtesy of Prof. Washington Mio of Florida State University.
[11]The domain Ω is required to satisfy an interior cone condition (see Definition 3.6[51]). Every ball satisfies the condition (see Lemma 3.10[51]).

method for handling large curls. In fact, when T_1 is restricted to rigid motions, the optimal rigid motion will match the first two moments of the two point-sets (see Appendix C). Second, the displacement between each desired point pair can be greatly reduced. Because the optimal selection σ depends on the largest displacement, a smaller kernel scale can be used, and we can obtain more robustness against outliers.

3.4.2. Hierarchical approaches—The second approach is a hierarchical method (coarse-fine) approach to reduce the curl of deformations. The matching challenge in mass transport models is caused by the curl of the displacement and the point cardinality. Reducing one of these factors can improve the matching performance. The hierarchical approach attempts to split the original matching problem with large cardinality and large curl into two (or more) subproblems, each of which has either small cardinality or small curl. When the curl-cardinality relation is fulfilled, perfect matches can be obtained for both subproblems.

Point-sets are selected from prominent geometric image features in images using differential operators (edge and corner operators, for instance[39]) or local descriptors (for instance, SIFT[28]). Several pairs of point-sets with different point cardinality can be generated from one pair of images. Here, we demonstrate a coarse-fine approach to compute the correspondence between (fine) point-sets X_2 and Y_2.

- Select a pair of coarse point-sets X_1, Y_1 and a pair of fine point-sets X_2, Y_2 from a pair of images. The fine point-sets have larger point cardinality than the coarse point-sets.
- Use the mass transport methods to obtain the correspondence τ from X_1, Y_1.
- Estimate the transform T_1 from the problem

$$\min |T_1|_{N_\Phi} \text{ subject to } T_1(x_i) = y_\tau(i) \text{ for } x_i \in X_1, y_i \in Y_1.$$

- Compute the correspondence via the mass transport methods on the pair of point-sets $T_1(X_2)$, Y_2. Then, the transform T_2 can be estimated in a similar manner, if necessary.

Because of the following approximation theorem, the deformation between $\{y_i\}_i$ and $\{T_1(x_i)\}_i$ usually has a smaller curl than the deformation between $\{y_i\}_i$ and $\{x_i\}_i$. According to the curl-cardinality relation, the matching performance between two updated point-sets $\{y_i\}$, $\{T_1(x_i)\}$ can be greatly improved compared with the one between the point-sets $\{y_i\}$ and $\{x_i\}$.

Theorem 3.1 (*Theorem 11.14[51]*). Let $\mathcal{N}_\Phi$ be the reproducing-kernel Hilbert space generated by a positive semi-definite kernel Φ with norm $\|\cdot\|_{\Phi(\Omega)}$. For simplicity, let Φ be a Gaussian and let the domain Ω be a ball. Let $s_{T,X}$ be the radial basis function interpolant to $T \in \mathcal{N}_{\Phi(\Omega)}$ based on T and $X = \{x_1, x_2, \ldots, x_n\}$. Then, for every $l \geq 2$, $l \in \mathbb{N}$, there exist constants h_l, $C_l < 0$ such that

$$|DT(x) - Ds_{T,X}(x)| \leq C_l h_{X,\Omega}^{l-1} |T|_{N_\Phi(\Omega)}$$

for all $x \in \Omega$, provided that $h_{X,\Omega} := sup_{x\in\Omega} min_{x_j \in X} \|x - x_j\|_2 \leq h_l$. Here, DT and $Ds_{T,X}$ refer to the first derivatives of T and $s_{T,X}$, respectively.

Hence, the curl magnitude of the difference $T - s_{T,X}$ is bounded by the norm of T and the fill distance $h_{X,\Omega}$. In the hierarchical approach, $s_{T,X}$ is the transform T_1. Suppose that the rate of

increase of the point cardinality does not exceed the rate of decrease of the curl magnitude. As a consequence of the curl-cardinality relation, a perfect match can still occur in matching the fine point-sets.

Here, one simulation result is presented. Let X be the uniform grid points in the square and Y be the deformed point-set

$$y_i = x_i + \exp\left(-\|x_i\|^2\right) \begin{pmatrix} 0.5 & 0 \\ 0 & 1 \end{pmatrix} x_i.$$

Because the deformation is not curl-free, some mismatches are expected. We generate two pairs of point-sets, (X_1, Y_1) with cardinality $n = 5^2$ and (X_2, Y_2) with cardinality $n = 10^2, \ldots, 30^2$. The transform T_1 is approximated by a linear combination of Gaussian basis functions $\exp(-\|x-y\|^2/0.1^2)$ and affine transforms. The experimental result is shown in Fig. 3.5 and Table 3.1. Note that perfect matches occur when the hierarchical approach is used. Another simulation experiment can be found in Fig. 4.2.

4. Experiments

4.1. Experimental setup

Points located at the bifurcations of the major vascular trees were manually selected and assigned correspondences from a pair of 3D CT images. First, the vascular trees from both images are automatically extracted using the segmentation software package PW2:VIDA Diagnostics (Coralville, IA, USA). The major vessel trees are then obtained by applying a morphology opening with a structure element of $3 \times 3 \times 3$ voxels, and then a connected component process is performed on the segmented vessel trees to remove small branches. A set of points located at the major vascular trees in one image is then generated by 3D skeletonization and post-analysis. Next, a semi-automatic system [32] was adopted to guide the observer in finding the corresponding points in the other image. With the semi-automatic tool, each point pair manually annotated by the observer is added to a thin-plate-spline as a warping function for the two images. The warping function is then utilized to predict its location when a new landmark point is presented to the observer. As more matching points are added to the system, the warping function becomes more accurate, and the task of the observer becomes easier, providing that there are increasingly accurate starting points.

Remark 4.1. There are several research studies related to lung registration. The intensity-based lung registration method matches the intensity patterns of the images retrieved at different lung volumes by minimizing a similarity measure [42][14][36][55][56]. The sum of the squared tissue volume difference (SSTVD) is a new similarity criterion based on preserving tissue mass of the lung at different volumes[56]. SSTVD has been demonstrated to yield an improved registration for large deformations in the lower lobes. However, such methods incur high computational costs[12], and a possible mismatch of important anatomical landmarks may occur when registration optimization falls into a local minima. In addition to intensity data, anatomical features have also been used to derive the transformation from one lung dataset to another[17][5][26][11]. To further improve the accuracy of registering intra-subject datasets across large lung volume changes, hybrid methods that utilize both anatomical landmark and intensity information have been proposed [34] [27] [54] [21]. The

[12]Computational time varies from several minutes to several hours depending on the input image size and the implementation method (e.g., sampling points, GPU, multi-threading).

correspondences between anatomical points, such as the bifurcation points of airways and/or vascular trees and vertebrae were manually designated by experts.[13]

In this work, we use two types of point-sets: (i) six pairs of CT images are used; they have subject IDs 5972, 5974, 5978, 5983, 6012 and 6019. Approximately 100–200 landmark pairs extracted from the branch points of the vessel trees are generated for each image pair. Figure 1.1 shows an example of the landmark locations; segmentations of the lungs and the vascular and airway trees are also shown for reference. (ii) One large lung point-set (subject ID H8756) consists of approximately 62500 points, which are the branch points of approximately 20 generations of airway trees in the five lung lobes generated by the volume filling technique [47] [46] (see Fig. 4.1).

Consider two sets of branch points $\{x_i\}_{i=1}^m$ and $\{y_j\}_{j=1}^n$. Because the correspondence matrices Γ^+ and Γ^- are generally different matrices, we assign the correspondence by the following majority rule,[14]

$$y_j \text{ is assigned to } x_i \text{ if } \gamma_{i,j}^+ \geq \gamma_{i,k}^+ \text{ for all } k. \quad (4.2)$$

That is, $j = arg\ \min_j \left\{\gamma_{i,j}^+ : j = 1, \ldots, n\right\}$. The matching performance of each method is evaluated using the match error, i.e., the total number of mismatched pairs from the second point-set *not including outliers*[15].

4.2. Lung point-sets without outliers

4.2.1. L^α mass transport costs—The following two experiments demonstrate that the correspondence of lung point-sets (under volume change) can be estimated using the mass transport model, specifically the L2 mass transport model (the MK method). Here, the lung point-set is extracted from subject ID H8756.

First, experiments are conducted on 6 pairs of point-sets (TLC/FRC), each of which contains 1000 points. In the first five experiments, a pair of point-sets is selected from one of the 5 lung lobes with a point ID from 1 to 1000, i.e., 1000 point-pairs at the lowest generation. The datasets of data 1, data 2, …, data 5 refer to the i^{th} lobe, where i = 1, 2, …, 5. Their spatial location is shown in Fig. 4.1. In the sixth experiment, a collection of 1000 point pairs is assembled from these five lobes (we take the first 200 point-pairs from each lobe). The correspondence is obtained by minimizing the L^α mass transport cost

$$\sum_{i,j=1}^{n} \mu_{i,j} \|x_i - y_j\|^\alpha \text{ where } \{\mu_{i,j}\} \text{ is a doubly stochastic matix.}$$

The computation is performed using the matlab function linprog, which is based on the standard interior point method[57][7].[16] Several different powers α are used, and their

[13]Some recently published work has shown that it is possible to generate large numbers of corresponding landmarks (more than 1,000) from a pair of lung CT datasets with semi-automatic tools [32] [11]. Although semi-automatic tools [32] [11] were developed to accelerate the process, the task of designating the correspondences between the points remains a labor-intensive task.

[14]If one is concerned about the symmetry, the following majority rule is suggested:

$$y_j \text{ is assigned to } x_i \text{ if } \gamma_{i,j}^+ \gamma_{i,j}^- \geq \gamma_{i,k}^+ \gamma_{i,k}^- \text{ for all } k, \quad (4.1)$$

[15]Because points have no label, we generally cannot distinguish between outliers (the points with missing correspondence) and inliers. Hence, for all experimental results, we do not consider the cases in which outliers are accidentally detected.

corresponding match errors are reported in Table 4.1. The cost with $\alpha = 2$ yields the lowest match error. Hence in the following experiments, we will focus on the L2 mass transport cost.

When $\alpha \neq 2$, the correspondence is not translationally invariant. One question arises: can the match result be improved if an extra translation t is applied to one point-set? Typically it can, but this is not for all cases. Consider the translation t that matches the mass center of two point-sets, $t = n^{-1}\left(\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i\right)$. The result is reported in Table 4.2. By comparing this table with Table 4.1, we see that the results are better than the results for the case without translations.

Next, the mass transport model is tested on all of the 62,500 point-pairs for subject H8756. Unfortunately, this is a very difficult linear programming problem because of dense constraints. The total number of (primal and dual) unknowns is at least $62500 + 62500^2$. Because of the significant computation cost, we will match a small subset of all of the point-pairs each time (we use 500 point-pairs each trial and a total of 125 matching trials).[17] In the i^{th} experiment, we match a pair of point-sets consisting of 500 point pairs with ID numbers from $500(i - 1) + 1$ to $500i$, where i ranges from 1 to 125. The number of mismatches is reported in the left side of Fig. 4.2. Note that perfect matches occur in the first 20 pairs of point-sets, as shown in the stars with an x-coordinate not exceeding 10, 000. Mismatches occur for the branch points with higher point IDs, i.e., those located at higher generations (generations greater than 12; see the right subfigure of Fig. 4.2). This result is consistent with our theoretical analysis: a large curl exists on the branch points close to the boundary (Appendix B). For a fixed point cardinality, a mismatch error is more likely to occur, when the applied deformation has a larger curl (this follows from the curl-cardinality relation).

Finally, we report results of an experiment in which a hierarchical approach is used. We estimate the transform T using a pair of coarse point-sets, which are each composed of 50 points selected from each 500–point point-set. The total number of mismatches in the above experiment is 2, 671 out of 62, 500, which is 4.3%. When a coarse point set (50 points selected from each set of 500 points) is used, the number of mismatches is reduces to 282, which is 0.45%. See the right subfigure of Fig. 4.2. (The mismatch error is zero for each pair of coarse point-sets with a cardinality of 50).

4.3. Comparisons to other methods

Matching lung point-sets is not an easy task for other methods. Here, we provide several comparisons to other methods. One naive approach is to simply assign the correspondence of each point as the closest point in the other point-set; this approach is known as the closest point method (CP). In general, the result is not symmetric, i.e., when x_1 is assigned to y_1, the assignment can change if the roles of the two point-sets are exchanged. To fix this inconsistency, the Sinkhorn matrix balancing (SB) [43] is introduced in the well-known TPS-RPM method. This method assigns a unit mass to each point, and a doubly stochastic matrix is generated to represent the matching result symmetrically. In the TPS-RPM method, one extra row and one extra column are added to address the outliers (this is called the soft assignment). Moreover, the TPS-RPM method includes the following appealing procedure: a sequence of transforms T can be estimated, when the kernel scale σ is gradually

[16] The matlab code to setup the input arguments of linprog will be available at http://www.math.ntu.edu.tw/~pengwen and other source code will be provided upon request.

[17] The experiment with 500 × 125 point-pairs took approximately 4 hours (to complete the 125 matching trials; each trial required approximately 2 minutes) using the interior point method on a server with 60 G of memory. We also conducted several experiments where we matched a point-set with a size of 1, 000. However, it converged too slowly. For matching 2,000 point-pairs, 60 GB of memory is also not sufficient to execute the interior point method.

decreased from a large value to a small positive value. The TPS-RPM method is a state-of-the-art method for medical imaging analysis (see [19] [6]). To clarify the performance, no outliers are added to the point-sets.

Table 4.3 shows a comparison of the results obtained using the closest point (CP) method, the Sinkhorn balancing (SB) method, the TPS-RPM, the MK method and the HD-RBF method ($\sigma_K = 100$, $\sigma_S = 50$). Clearly, the HD-RBF method gives the best correspondence result.[18] In comparison with the MK method, T_1 (H6019) with a nonzero curl is estimated in the HD-RBF model such that the transform from $\{T_1(x_i)\}_{i=1}^n$ to $\{y_i\}_{i=1}^n$ has a smaller curl. When the HD method was applied to the six subjects preprocessed by affine transforms given by Eq. (3.9), the same matching result was generated. Finally, note that there is little difference between the matching performance of the SB and TPS-RPM methods, even though the latter method makes some effort to estimate the transformations.[19]

To illustrate the differences between TPS-RPM and the mass transport methods (MK and HD), we select two small point-sets from the two point-sets (see Fig. 4.3(e)). The two point-sets are separately matched by the SB method, the CP method, the HD model and the MK method.[20] Each correspondence is represented by a matrix, in which the (i, j) entry is some "likelihood" of assigning the i^{th} point in the first point-set to the j^{th} point in the second point-set (see Fig. 4.3). Clearly, the mass transport methods (HD, MK) outperform the other two methods. The HD model in Fig. 4.3(c) provides an approximate solution for the mass transport problem (see Fig. 4.3(d)). Moreover, the performance of the SB method is worse than that of the CP method.

Lastly, we like to highlight that the correspondence generated using the TPS-RPM method is essentially different from the one from the mass transport method, despite their similar appearance. In TPS-RPM, the kernel matrix $\{K_{i,j} := \exp(-\|x_i - y_j\|^2/\sigma^2)\}_{i,j}$ is always dense, because the Sinkhorn matrix balancing does not change the ratio $K_{i1,j1}K_{i2,j2}/K_{i1,j2}K_{i2,j1}$. The dense stochastic matrix is irrelevant to the permutation matrix obtained in the mass transport methods (except for the case in which these points are the closest point pairs). Determination of the correct correspondence using the TPS-RPM method relies on the correct estimation of the transforms. However, even for two 1D point-sets without outliers, the doubly stochastic matrix produced by SB is generally not the permutation that places both point-sets in order. In these cases, the TPS-RPM method could fail to produce a transform T based on thin-plate splines such that the desired point-pairs form the closest point pairs. The soft assignment is forced to remove several "extra" points and label them as "outliers" to generate a permutation matrix at the end of the deterministic appealing.

4.4. Lung point-sets with outliers

4.4.1. Five lung lobes of subject H8756—In the following section, we compare the MK method with the HD method for a pair of point-sets, selected from the five lung lobes

[18]In fact, in subject ID *H*6012, the points with ID 6 and ID 8 in one point-set are incorrectly assigned to the points with ID 8 and ID 6 in the other point-set. The points are quite close to each other (they are at approximately a 1.9mm apart). Here, the results of using the HD-RBF method with different λ values are presented:

- For the cases where λ = 1 and 10, in the six subjects, the errors were 0, 0, 0, 0, 2(*H*6012), 2(*H*6019); in the five lobes, there were zero errors.
- For the case where λ = 0.02, in the six subjects, the errors were 0, 3, 8, 0, 2(*H*6012), 15(*H*6019); in the five lobes, the errors were 0, 0, 0, 0, 2, 2.

[19]The computational cost of matching a pair of two point-sets is very high. For subject H6012, the time spent was 235 seconds (200 inner iterations for SB and 20 outer iterations for the spline estimation) for the TPS-RPM; in comparison, the time required for the HD method was less than 2 seconds.

[20]The kernel scale is σ = 100 for both the SB and HD methods.

and to which randomly generated outliers have been added. The spatial size of the point-sets is approximately 200 mm, and thus we choose multiples of 100 for kernel scales.

Two point-sets *X*, *Y* with the same cardinality are assembled from two parts: (i) 200 point pairs of one lung lobe in the TLC and FRC phases and (ii) outliers with the same cardinality. Both the point-sets *X*, *Y* include outliers (with the same cardinality) generated using a normal distribution *Normal*(μ_i,σ), with means $\mu_1 = (2/n)\sum_{i=1}^{n/2} x_i$ and $\mu_2 = (2/n)\sum_{i=n/2+1}^{n} y_i$.[21] We choose multiples of the identity matrices $s^2 I$ for covariances σ, with $s^2 = 30^2, 60^2, 120^2$ and 180^2. Figure 3.3 shows the point-set (TLC) of the first lobe with a set of 25 outliers added.

We perform matching experiments with the outlier cardinality from 0 to 25 (chosen from the aforementioned 25 outliers). The HD method is applied on the point-sets pre-processed by the positive definite affine transforms. The result of each lung lobe is reported separately in the five figures (Fig. 4.4). For visual clarification, we only report the average mismatches of the four s^2 values. The y-axis represents the number of mismatches (not including outliers). The HD method, with σ_K ranging from 12.5 to 400, outperforms the MK method.

4.4.2. Six subjects—We repeat the same outlier generation rule and experimental procedure on the vessel branch points of six subjects (see the right subfigure of Fig. 3.3). The result is reported in Fig. 4.5. The HD method with σ_K = 100, 200 outperforms the MK method (σ_K = *inf*). Hence, the HD method with the proper kernel scale is more robust against the existence of outliers than the MK method.

4.5. Conclusion

In this work, we consider the lung deformation between TLC and FRC as one example. Nearly perfect matching results validate the effectiveness of the mass transport model, especially when the L2 mass transport cost is used. This result motivates our theoretical analysis. Generally, the performance of a point-set matching method depends on the assumed transforms. In this work, the deformation is assumed to be a vector field with a small curl (with respect to the point cardinality). The mechanics of perfect matches are studied from the viewpoints of the curl-cardinality relation and an L2-invariant property. The theoretical analysis and experimental results suggest that the mass transport method is a suitable tool for handling point-set registration problems subject to elastic deformations. One potential application is to cluster analysis of both airway and vessel tree branching patterns and classification of normal and abnormal lungs based on lung deformation.

Remark 4.2. The correspondence estimation problem also arises in shape analysis, e.g., [1] [38], where shapes are modeled as compact smooth Riemannian manifolds equipped with an intrinsic metric. One stable intrinsic metric is the diffusion metric, which is based on the solution of the heat equation. Gromov-Wasserstein distances are introduced for shape matching[31] [38]. These distances are also defined based on the idea of optimal mass transport and the heat kernel signature[45]. In [38], the authors propose a game-theoretic approach to recover sparse correspondences between shapes. Gromov-Wasserstein distances measure the overall distortion of the pairwise distances within point-sets, whereas Wasserstein distances measure the transporting cost between correspondences. One advantage of Gromov-Wasserstein distances is that they are invariant under rigid transforms.

[21] Choosing a nonzero mean shift yields a stronger outlier effect. Empirically, the MK method and the HD method with proper kernel scales both work quite well for the case where $\mu_1 n^{-1}\sum_{i=1}^{n} x_i$ and $\mu_2 n^{-1}\sum_{i=1}^{n} y_i$.

An interesting question (which is beyond the scope of this paper) arises: do similar conditions to ensure perfect matches (under certain classes of deformations) exist?[22]

Acknowledgments

Authors like to thank referees for their valuable comments and suggestions. C. Lin likes to thank Mathematics Division, the National Center for Theoretical Sciences (Taipei office), Taiwan, for hosting his visit during the course of this study.

P. Chen is supported by the grants NSC-99-2115-M-002-014-MY2, NSC-101-2115-M-002-007, 99R70339 and 99R30244.

C. Lin is supported by NIH Grant No. R01-HL-094315.

I. Chern is supported by NSC 99-2115-M-002-003-MY3.

Appendix A. Proofs of curl-cardinality relations

To prove these relations, we need the following lemma.

Lemma A.1. Fix a skew symmetric d × d matrix B. Let $\mathscr{A}$ be the set of all skew symmetric d × d matrices A of the form $vw^\intercal - wv^\intercal$ with v, w in $\mathbf{R}^d$. We have

$$Tr\,(BA) \leq \Lambda\,(B;1)\,Tr(2AA^\intercal)^{1/2}.$$

Also the maximum of $\mathrm{Tr}(2AA^\intercal)^{-1/2}|\mathrm{Tr}(BA)|$ is obtained if and only if the vectors v, w corresponding to the maximizer A lie in the subspace spanned by singular vectors of B corresponding to Λ(B; 1).

Proof. According to the matrix theory (page 107), a skew symmetric matrix *B* can be expressed as

$$B=\sum_{j=1}^{k}\beta_j\left(u_{2j-1}u_{2j}^\intercal - u_{2j}u_{2j-1}^\intercal\right),$$

where $\{\beta_1 \geq \beta_2 \geq \ldots \geq \beta_\kappa > 0\}$ are singular values and $\{u_j\}_{j=1}^{2k}$ are orthonormal singular vectors of *B*. Observe that for $v,w \in \mathbf{R}^d$,

$$\begin{aligned} & Tr\,(B\,(wv^\intercal - vw^\intercal)) \\ = \;& 2\sum_{i=1}^{k}\beta_i \det\begin{pmatrix} v\cdot u_{2i-1} & v\cdot u_{2i} \\ w\cdot u_{2i-1} & w\cdot u_{2i}\end{pmatrix} \leq 2\beta_1\sum_{i=1}^{k}\left|\det\begin{pmatrix} v\cdot u_{2i-1} & v\cdot u_{2i} \\ w\cdot u_{2i-1} & w\cdot u_{2i}\end{pmatrix}\right| \\ \leq \;& 2\beta_1\left(\|v\|^2\|w\|^2-(v\cdot w)^2\right)^{1/2} = \beta_1 Tr(2AA^\intercal)^{1/2}, A := vw^\intercal - wv^\intercal. \end{aligned}$$

The second and third equalities hold only when the above determinant is positive and *v*,*w* are a pair of vectors lying in the subspace spanned by u_1, u_2.[23]

[22] We thank the referees for informing us of the Gromov-Wasserstein metrics.

[23] The second inequality comes from the following fact. Let $x = [v\cdot u_1, v\cdot u_2, \ldots, v\cdot u_{2k-1}, v\cdot u_{2k}]$, $y = [w\cdot u_1, w\cdot u_2, \ldots, w\cdot u_{2k-1}, w\cdot u_{2k}]/\|w\|$ and $z = [w\cdot u_2, -w\cdot u_1; \ldots, w\cdot u_{2k}, -w\cdot u_{2k-1}]/\|w\|$. Then $x\cdot y = v\cdot w$, $y\cdot z = 0$ and $\|x\| = \|v\|$, $\|y\| = \|z\| = 1$. Since $x\cdot y$, $x\cdot z$ are the scalar projections of *x* onto orthogonal vectors *y*, *z*, then $\|x\|^2 \geq (x\cdot y)^2+(x\cdot z)^2$, i.e., $(x\cdot z\|w\|)2 \leq \|v\|^2\|w\|^2-(v\cdot w)^2$, which yields the second inequality.

A.1. Proofs of Prop. 2.3

The case $n = 2$ is obvious. Here, we focus on the case for $n \geq 3$. Let R be a rotation matrix with a minimum of $\Lambda(R_A; 1)$, which leads to the occurrence of a mismatch. Then, there exists some $m \in \mathbf{N}$, $m \leq n$ and some relabeling on $\{Rx_i\}_{i=1}^n$, such that

$$\sum_{i=1}^{m}\|Rx_i - x_i\|^2 \geq \sum_{i=1}^{m}\|Rx_i - x_{i+1}\|^2, i.e., \sum_{i=1}^{m} x_i^\intercal R\left(x_i - x_{i+1}\right) \leq 0.$$

We focus on the case of $m = n$, corresponding to the smallest $\Lambda(R_A; 1)$ for mismatches. Reformulating the inequality yields

$$\sum_{i=1}^{n} Tr\left(R\left[\left(x_i - x_{i+1}\right)\left(x_i - x_{i+1}\right)^\intercal - \left(x_{i+1}x_i^\intercal - x_i x_{i+1}^\intercal\right)\right]^\intercal\right) \leq 0. \quad \text{(A.1)}$$

Equivalently,

$$Tr\left(RS^\intercal\right) \leq Tr\left(RA^\intercal\right), i.e., Tr\left(R_S S^\intercal\right) \leq Tr\left(R_A A^\intercal\right), \quad \text{(A.2)}$$

where the following notations are used, $S := \sum_{i=1}^{n} S_i$ with $S_i := (x_i - x_{i+1})(x_i - x_{i+1})^\intercal$ and $A := \sum_{i=1}^{n} A_i = \sum_{i=1}^{n} \widehat{A}_i$ with

$$\widehat{A}_i := x_{i+1}x_i^\intercal - x_i x_{i+1}^\intercal, A_i := \left(x_{i+1} - x_1\right)\left(x_i - x_1\right)^\intercal - \left(x_i - x_1\right)\left(x_{i+1} - x_1\right)^\intercal.$$

Note that

$$\min_{R_S} Tr\left(R_S S^\intercal\right) \leq Tr\left(R_S S^\intercal\right) \leq Tr\left(R_A A^\intercal\right) \leq \max_{A} Tr\left(R_A A^\intercal\right). \quad \text{(A.3)}$$

Examine the LHS and RHS of Eq. (A.3): (i)For the LHS of Eq. (A.3), let a be the smallest eigenvalue of R_S and L be the length $L = \sum_{i=1}^{n}\|x_i - x_{i+1}\|$. Then

$$\min_{R_S} Tr\left(SR_S^\intercal\right) = \min_{R_S}\sum_{i=1}^{n}\left(x_{i+1} - x_i\right)^\intercal R_S\left(x_{i+1} - x_i\right) \geq aTr\left(S\right) \geq an^{-1}L^2,$$

where the second inequality is derived from Cauchy-Schwartz inequalities and equality holds if and only if $\| x_i - x_{i+1} \|$ is constant for each i. (ii)For the the RHS of Eq. (A.3), observe that

$$\max_{A} Tr\left(AR_A^\intercal\right) = \sum_{i=1}^{n}\max_{A_i} Tr\left(A_i R_A^\intercal\right) \leq \Lambda\left(R_A, 1\right)\max_{A_i}\sum_{i=1}^{n} Tr\left(2A_i A_i^\intercal\right)^{1/2}, \quad \text{(A.4)}$$

where equality holds only if $\{x_i - x_1\}_{i=2}^n$ lie in the subspace spanned by the singular vectors of R_A corresponding to $\Lambda(R_A, 1)$(Lemma A.1), i.e., the set $\{x_i\}_{i=1}^n$ is coplanar. Note that

$$4^{-1}\sum_{i=1}^{n} Tr(2A_i A_i^{\intercal})^{1/2} = \sum_{i=1}^{n} 2^{-1}\left[\|x_i - x_1\|^2 \|x_{i+1} - x_1\|^2 - ((x_i - x_1)\cdot(x_{i+1} - x_1))^2\right]^{1/2}$$

is the area of a polygon assembled from n triangles with vertices $\{x_i, x_{i+1}, x_1\}_{i=1}^{n}$. Hence, Eq. (A.4) implies that

$$\max_{A} Tr\left(AR_A^{\intercal}\right)/4\Lambda\left(R_A, 1\right)$$

is bounded above by the maximum area of the closed polygon with given side lengths. From [25], the maximum area occurs when the polygon is inscribed in a circle. Let *Area* be the area of the polygon.

In summary, Eq. (A.3) yields

$$an^{-1}L^2 \leq 4\Lambda\left(R_A, 1\right) Area, \text{then } \Lambda\left(R_A, 1\right)/a \geq L^2/\left(4nArea\right) \geq \pi/n,$$

where the isoperimetric inequality, $L^2/Area \geq 4\pi$ (page 33, [10]) is used. In fact, the isoperimetric inequality for polygons[25] ($L^2/Area \geq 4n\tan(\pi/n)$) shows that

$$\Lambda\left(R_A, 1\right)/a = \Lambda\left(R_A, 1\right)/\Lambda\left(R_S, d\right) \geq \tan\left(\pi/n\right). \quad \text{(A.5)}$$

The above arguments also show that the lower bound of $\Lambda(R_A, 1)/\Lambda(R_S, d)$ is reached only when $\{x_i\}_{i=1}^{n}$ are vertices of a regular polygon. On the other hand, according to Example 2.2, the lower bound can be obtained by choosing $\{x_i\}_{i=1}^{n}$ as vertices of a regular polygon, which completes the proof of the perfect match condition $\omega_{max} \leq 2\pi/n$.

A.2. Proof of Cor. 2.4

Define L and *Area* as in A.1.

Proof. [Cor. 2.4(i)] For simplicity we only consider $d = 3$ here and present the proof in the following. The occurrence of mismatches implies the existence of $n \in \boldsymbol{N}$ and some subset $\{x_i\}_{i=1}^{n}$ with convention $x_{n+1} := x_1$, such that

$$\sum_{i=1}^{n}\left(x_i - x_{i+1}\right)\cdot T\left(x_i\right) = \sum_{i=1}^{n} x_{i+1}\cdot\left(T\left(x_{i+1}\right) - T\left(x_i\right)\right) \leq 0 \text{ for some } n.$$

Then

$$\sum_{i=1}^{n}\left(x_i - x_{i+1}\right)\cdot\left(T\left(x_i\right) - T\left(x_{i+1}\right)\right) \leq -\sum_{i=1}^{n}\left(x_i - x_{i+1}\right)\cdot\left(T\left(x_i\right) + T\left(x_{i+1}\right)\right). \quad \text{(A.6)}$$

The left hand side of Eq. (A.6) has a lower bound,

$$\sum_{i=1}^{n}\left(\left(x_{i+1} - x_i\right)^{\intercal} T_S\left(\xi_i\right)\left(x_{i+1} - x_i\right)\right) \geq n^{-1}\left(\min_{x}\Lambda\left(T_S\left(x\right), 3\right)\right)\left(\sum_{i=1}^{n}\left|x_{i+1} - x_i\right|\right)^2 \quad \text{(A.7)}$$

which is an approximation of $n^{-1}(\min_x \Lambda(T_S(x), 3))L(\partial\Omega)^2$(the existence of ξ_i is ensured by the mean value theorem). On the other hand, the right hand side of Eq. (A.6) can be regarded as the Riemann sum of one line integral,

$$\begin{aligned} & \sum_{i=1}^{n} (x_i - x_{i+1}) \cdot (T(x_i) + T(x_{i+1})) \approx 2\int_{\partial\Omega} T(x) \cdot dx \\ = \; & 2\int_{\Omega} (\nabla \times T(x)) \cdot \mathbf{n} da, \end{aligned}$$

which is bounded above by $2(\max_{x\in\Omega} |\nabla \times T(x)|)Area$, where da is the area element.

If the above two approximation errors can be neglected, then together with the inequality $L(\partial\Omega)^2 \geq 4\pi Area(\Omega)$, we have

$$4\pi n^{-1} Area(\Omega) \leq n^{-1} L(\partial\Omega)^{-2} \leq \frac{2\max_{x\in\Omega} |\nabla \times T(x)|}{\min_x \Lambda(T_S(x), 3)} Area(\Omega).$$

Hence,

$$\frac{\max_{x\in\Omega} |\nabla \times T(x)|}{\min_x \Lambda(T_S(x), 3)} \geq \frac{2\pi}{n},$$

which verifies the curl-cardinality relation $\omega_{max} \leq C/n$ for perfect matches, where C might be different from 2π due to the approximation difference of the Riemann sums.

Proof. [Cor. 2.4(ii)]

Obviously, the left-hand side of Eq. (A.6) has a lower bound:

$$\sum_{i=1}^{n} a\|x_i - x_{i+1}\|^2 \geq aL^2/n, \text{where } a = \Lambda(T_S(\xi_i); 1)$$

and ξ_i is some point between x_i, x_{i+1}(the mean-value theorem). However, when T_S is constant, we can derive a tighter bound. Let $x_i = T_S^{-1/2} y_i$ for $i = 1,\ldots,n$. Then the mean value theorem indicates that the existence of ξ_i such that the left-hand side becomes

$$\sum_{i=1}^{n} (x_i - x_{i+1}) \cdot (T_S(\xi_i)(x_i - x_{i+1})) = \sum_{i=1}^{n} \|y_i - y_{i+1}\|^2.$$

The right-hand side of Eq. (A.6) can be reformulated as

$$\begin{aligned} & \sum_{i=1}^{n} (x_i + x_{i+1}) \cdot (T(x_i) - T(x_{i+1})) = \sum_{i=1}^{n} (x_i - x_1 + x_{i+1} - x_1) \cdot (T(x_i) - T(x_{i+1})) \\ = \; & \sum_{i=1}^{n} (\widehat{x}_i + \widehat{x}_{i+1}) \cdot ((T_S + T_A(\eta_i))(\widehat{x}_i - \widehat{x}_{i+1})) = \sum_{i=1}^{n} Tr\left(T_A(\eta_i)\left(\widehat{x}_i \widehat{x}_{i+1}^{\intercal} - \widehat{x}_{i+1} \widehat{x}_i^{\intercal}\right)\right), \\ = \; & \sum_{i=1}^{n} Tr\left(T_S^{-1/2} T_A(\eta_i) T_S^{-1/2}\left(\widehat{y}_i \widehat{y}_{i+1}^{\intercal} - \widehat{y}_i \widehat{y}_i^{\intercal}\right)\right), \end{aligned}$$

where n_i is some point lying on the segment between x_i and x_{i+1} (by the mean value theorem), and $\widehat{x}_i = x_i - x_1$, $\widehat{y}_i = y_i - y_1$. Divide the skew polygon with vertices $\{y_i\}_{i=1}^n$ into $n - 2$ triangles, each of which has vertices $\{y_1, y_i, y_{i+1}\}_{i=2}^{n-1}$. By Lemma A.1, the right-hand side can be regarded as

$$\sum_{i=2}^{n-1} Tr\left(T_S^{-1/2} T_A(\eta_i) T_S^{-1/2} A_i\right) \leq \sum_{i=1}^{n} \beta_1(\eta_i)\, Tr(2A_i A_i^{\intercal})^{1/2},$$

where $A_i = \widehat{y}_i \widehat{y}_{i+1}^{\intercal} - \widehat{y}_{i+1} \widehat{y}_i^{\intercal}$ and $\beta_1(\eta_i)$ is the largest singular value of the skew symmetric matrix $\left\{T_S^{-1/2} T_A(\eta_i) T_S^{-1/2}\right\}_{i=1}^{n}$. Note that the equality holds only when $\{y_i\}_{i=1}^n$ are coplanar.

Following this procedure, the same arguments in the proof of Prop. 2.3 yield

$$\tan\frac{\omega_{max}(T;\Omega)}{2} \geq \frac{L^2}{4n\,Area},$$

where $\omega_{max}(T; \Omega)$ is the largest singular value of $T_S^{-1/2} T_A(x) T_S^{-1/2}$ among all x in Ω. Hence, we have the curl-cardinality relation for perfect matches: $\omega_{\max} \leq 2\pi/n$ according to the isoperimetric inequality.

Appendix B. Large curls near the lung periphery

We will provide one explanation for the occurrence of large curls near the boundary of the domain, namely the lung periphery. First, we briefly describe the mechanics of breathing. The repeated inflation and deflation of the lungs are controlled by the respiratory muscles, the diaphragm and the intercostal muscles. During inhalation, the diaphragm and the intercostal muscles contract and create negative pressure (relative to atmospheric pressure) surrounding the lungs. The expansion decreases the pressure in the chest cavity and allows air flow in, which inflates millions of alveoli. The region around the lungs in which this negative pressure acts is called the pleural space and is filled with a very thin layer of lubricating fluid that separates the outer surface of the lungs from the inner surface of the rib cage ([2], page 4).

The lungs, which are made of spongy and elastic tissue, are commonly modeled as a *linear*, isotropic and homogeneous medium[53]. In this situation, the displacement field $u(x) := T(x) - x$ on the domain Ω (lungs) satisfies the Lamé equilibrium equations in the linear elasticity theory:

$$\mu\nabla^2 u + \nabla((\mu+\lambda)\nabla\cdot u) = f, \text{subject to serval boundary conditions on } u,$$

where $\mu > 0$ and $\lambda > 0$ are called the Lamé constants and the vector function f is the body force.

We will study the spatial distribution of the curl $\nabla \times T = \nabla \times u$. According to the classic uniqueness result of the Lamé equations, u is unique up to some rigid body displacement if traction is prescribed over the entire surface. Based on the superposition principle, the solution u can be constructed as the sum of a particular solution of the inhomogeneous

equilibrium equations and a solution of the homogeneous equilibrium equations subject to the desired boundary condition.

To proceed, we assume that the body forces are derived from a scalar potential, i.e., $f = \nabla\xi$. This assumption is valid in many cases, i.e., for the gravity forces. Consider a particular solution of the gradient form $u = \nabla\varphi$. Then

$$\nabla\left((\lambda+2\mu)\nabla^2\phi - \xi\right) = 0.$$

Hence, one particular solution can be obtained by solving Poisson's equation:

$$\nabla^2\phi = \frac{\xi}{2\mu+\lambda}.$$

Note that u is curl-free, and thus the body force does not contribute any curl on the displacement field u (However, the nonzero f does affect the boundary condition for u).

According to Helmholtz's theorem, a vector field u on a bounded domain Ω in $\mathbf{R}^3$ can be decomposed into a sum, $u = \nabla\tilde{\phi} + \nabla\times\psi$ with $\nabla\cdot\psi = 0$. Substituting the decomposition into the homogeneous Lamé equations leads to

$$\mu\nabla\times\nabla^2\psi + (\lambda+2\mu)\nabla\left(\nabla^2\tilde{\phi}\right) = 0. \quad \text{(B.1)}$$

The desired displacement field is the one satisfying the boundary condition.

Note that $\nabla\times u = -\nabla^2\psi$. By applying the divergence and curl operator on Eq. (B.1), we have $\nabla^2\nabla^2\tilde{\phi} = \nabla^2\nabla^2\psi = 0$. Thus, for each unit vector $\nu \in \mathbf{R}^3$, $\nu\cdot\nabla^2\psi$ is harmonic, which implies that the extreme values of $\nu\cdot\nabla^2\psi$ occur at the boundary $\partial\Omega$ (Theorem 2.3, page 15, [18]). Hence, regardless of the boundary condition, the maximum magnitude of the curl $\nabla\times u$ occurs at the boundary of the elastic object. This theoretical result is consistent with our lung experiments: mismatches occur at the branch points of higher generations.[24]

Appendix C. Rotations

Here we study one asymptotic property in the kernel correlation method [48] and Eq. (3.10) about the rigid motions, $T(x) = Qx + t$.

Denote two point-sets by $X := \{x_i : i = 1, \ldots, n_1\}$ and $Y := \{y_j : j = 1, \ldots, n_2\}$ in $\mathbf{R}^d$. The transformation T is estimated through

$$\min_T \{K(Y,Y) - 2K(Y,T(X)) + K(T(X),T(X))\}, \quad \text{(C.1)}$$

where σ is the *kernel scale* and

$$K(X,Y) = (n_1 n_2)^{-1}\sum_{i,j=1}^{n_1 n_2}\exp\left(-\|x_i - y_j\|^2/\sigma^2\right).$$

[24]Note that $T(x) = x+u$. Here, we ignore the variation of ω_{max} contributed from the denominator $1+\Lambda\left(\nabla^2\tilde{\phi}\right)$.

A minimizer T depends on the size σ. To avoid local optimal solutions, the kernel scale σ should not be too small. Experiments show that the number of local optimal solutions is reduced, when σ is large[48].

Observe that a minimizer (Q, t) in Eq. (C.1) maximizes the second term in Eq. (C.1),

$$\max_{Q,t} \sum_{i=1,j=1}^{n_1,n_2} \exp\left(-\|y_j - Qx_i - t\|^2/\sigma^2\right),$$

since the remaining terms are constant. Denote the first moment and the second (central) moment of X by t_X, M_X,

$$t_X = n_1^{-1}\sum_{j=1}^{n_1} x_j,\ M_X = n_1^{-1}\sum_{j=1}^{n_1}(x_j - t_X)(x_j - t_X)^\intercal. \quad \text{(C.2)}$$

Likewise, let t_Y, M_Y be the first moment and the second moment of Y.

The following proposition shows that in the asymptotic case $\sigma \to \infty$, the rigid motion "matches" both the first moments and the eigenvectors of the second moments of two point-sets. This asymptotic result also holds for the rigid motion in the HDRBF model, the first term of Eq. (3.10)[25]. This study shows the transforms T_1 with nonzero curl can be estimated via optimizing the cost function with kernel scales sufficiently large.

Proposition C.1 (*Rigid motions*). Consider two point-sets $\{y_i : i = 1, \ldots, n_2\}$ and $\{Qx_i + t : i = 1, \ldots, n_1\}$ with $\sum_{i=1}^{n_1} x_i = 0$. Let $T(x; \sigma) = Q(\sigma)x + t(\sigma)$ be a maximizer of the function in Eq. (C.1), with $Q(\sigma) \in \boldsymbol{R}^{d\times d}$ being an orthogonal matrix and $t(\sigma) \in \boldsymbol{R}^d$. Let (Q, t) be a limit of $(Q(\sigma), t(\sigma))$ as $\sigma \to \infty$.

- Then the first moments of these two point-sets are identical,
  $n_2^{-1}\sum_{j=1}^{n2}(y_j - t) = n_1^{-1}\sum_{i=1}^{n_1} Qx_i$.
- Assume that the second (central) moments M_X, M_Y both have distinct eigenvalues with SVD

$$M_X = U_X^\intercal D_X U_X,\ M_Y = U_Y^\intercal D_Y U_Y. \quad \text{(C.3)}$$

Then $Q = U_Y^\intercal U_X$, and thus the eigenvectors of $QM_X Q^\intercal$ and M_Y are paral-lelized. Note that $QM_XQ^\intercal$ is the second moment of the rotated first point-set,
$QM_XQ^\intercal = n_1^{-1}\sum_{i=1}^{n_1}(Qx_i)(Qx_i)^\intercal$.

Proof.

As σ^2 approaches infinity, $1/\sigma^2 \to 0$, the Taylor expansion shows
$\max_{Q,t}(n_1n_2)^{-1}\sum_{i=1,j=1}^{n_1,n_2}\left(1 - \|y_i - Qx_i - t\|^2/\sigma^2 + \|y_i - Qx_i - t\|^4/2\sigma^4\right) + O\left(1/\sigma^6\right)$.
Since $\sum_{i=1}^{n_1} x_i = 0$, the leading term is $\sum_{i=1,j=1}^{n_1,n_2}\|y_j - Qx_i - t\|^2 = \sum_{i,j}\left(\|y_j - t\|^2 + \|x_i\|^2\right)$,

[25]We use the initialization $\gamma_{i,j}^{+} = 1/n_1$, $\gamma_{i,j}^{-} = 1/n_1$ to estimate T_1 in the non-convex minimization problem.

As $\sigma \to \infty$, the optimal translation vector t tends to $n_2^{-1}\sum_{j=1}^{n_2} y_j$, which implies that $\sum_{j=1}^{n_2}(y_j - t) = 0$.

For the notation simplicity, we replace $(y_j - t)$ by y_j, then $\sum_{j=1}^{n_2} y_j = 0$. The asymptotic behavior of Q depends on the next leading term

$$\left(2\sigma^4 n_1 n_2\right)^{-1} \sum_{i=1,j=1}^{n_1 n_2} \|y_i - Qx_i\|^4 = \left(2\sigma^4 n_1 n_2\right)^{-1} \sum_{i=1,j=1}^{n_1 n_2} 4\left(Y_j^\intercal, Qx_i\right)^2 + \text{some contants}.$$

Hence, Q is a maximizer of the function

$$\sum_{i=1,j=1}^{n_1 n_2} (n_1 n_2)^{-1}\left(y_j^\intercal Q x_i\right)^2 = \sum_{i=1,j=1}^{n_1 n_2} (n_1 n_2)^{-1} Tr\left(Q x_i x_i^\intercal Q^\intercal y_i y_i^\intercal\right) \quad \text{(C.4)}$$

$$= Tr\left(Q M_X Q^\intercal M_Y\right) = Tr\left(D_X \overbrace{U_X Q^\intercal U_Y^\intercal} D_Y \overbrace{U_Y Q U_X^\intercal}\right) \quad \text{(C.5)}$$

which can be expressed as $Tr\left(D_X Q_1 D_Y Q_1^\intercal\right)$ with $Q_1 := U_X Q^\intercal U_Y^\intercal$. Note that $Q_1 Q_1^\intercal = I$. By the matrix theory[26],

$$\max_{Q_1}\left\{Tr\left(D_X Q_1 D_Y Q_1^\intercal\right) : Q_1 \text{ orthogonal}\right\} \quad \text{(C.6)}$$

$$\le \max_{U,V}\left\{Re\, Tr\left(D_X U D_Y V\right) : U, V \text{unitary}\right\} \le \sum_{i-1}^{d} \sigma_i\left(D_X\right)\sigma_i\left(D_Y\right), \quad \text{(C.7)}$$

with the singular values σ_i of D_X, D_Y both arranged in a decreasing order. On the other hand, clearly when Q_1 is the identity matrix, the maximal value is attained. Therefore, from the definition of Q_1, the optimal matrix Q is given by $U_Y^\intercal U_X$, which implies that $Q M_X Q^\intercal = U_Y^\intercal D_X U_Y$. Compared with $M_Y = U_Y^\intercal D_Y U_Y$, in the asymptotical case the eigenvectors of $QM_XQ^\intercal$ and M_Y are parallelized.

Appendix D. Finite kernel scales

Generally, the correspondence in the HD model with finite kernel scales is not a permutation matrix. Here is one example to illustrate the recovery of the binary correspondence by the majority rule, provided that the kernel scales σ are large enough. Besides, as the number of points increases, the requirement on σ becomes harsh.

[26] Given two $m \times n$ matrices B_1, B_2, consider the problem

$$\min_{U,V} \|B_1 - U B_2 V\|^2, \text{equivalently}, \max_{U,V} \text{Re}\,\text{tr} B_1^* U B_2 V,$$

where two "rotations" U, V are $m \times m$ and $n \times n$ unitary matrices, respectively. The maximum value is $\sum_{j=1}^{q} \sigma_j\left(B_1\right)\sigma_j\left(B_2\right)$, where the singular values $\{\sigma_j\left(B_1\right)\}_{j=1}^{q}, \{\sigma_j\left(B_2\right)\}_{j=1}^{q}$ of B_1 and B_2 are both arranged in a decreasing order(page 435, [22]).

Consider one point-set consisting of $2n+1$ points, $\{x_i=(i-1)\,a\}_{i=1}^{2n+1}$ with spatial resolution $a \in \mathbf{R}^d$ and its translated point-set $\{y_i=(i-1)\,a+b\}_{i=1}^{2n+1}$ with translation $b \in \mathbf{R}^d$ and $b \cdot a > 0$. Let $\Delta = na$, then the spacial size of the point-set is $|\Delta|$. The following computation provides a lower bound σ, approximately $|\Delta|\sqrt{2na \cdot b/|a|}$, such that $(i)\ \gamma_{i,j}^{+}=0=\gamma_{i,j}^{-}$ for all $j \neq i, i+1$, (ii) Both $\gamma_{i,i}^{+}$ and $\gamma_{i,i}^{-}$ are greater than 1/2.

Let us adopt the notations in section 3.2 and assume the bijective matching structure $\left\{\left(\gamma_i, \alpha_i^{-1}\gamma_i\right), \left(\alpha_{i+1}\left(1-\alpha_i^{-1}\gamma_i\right), 1-\alpha_i^{-1}\gamma_i\right) : i=0,1,\ldots,2n\right\} \cup \left\{\left(\gamma_{2n+1}, \alpha_{2n+1}^{-1}\gamma_{2n+1}\right)\right\}$, where $\{\alpha_i\}_{i=0}^{2n+1}$ will be determined later (see Fig. D.1). Clearly,

$$\alpha_i\left(1-\alpha_{i-1}^{-1}\gamma_{i-1}\right)+\gamma_i=1, i.e., (\alpha_i-1)+\gamma_i=\gamma_{i-1}\alpha_i/\alpha_{i-1}. \quad \text{(D.1)}$$

Claim: for $i = 0, 1, \ldots, 2n$,

$$\alpha_{i+1}/\alpha_i =: G=\left(K_{i+1,i}^{-1}K_{i,i}\right)^2=\exp\left(-(2b+a)\cdot a/\sigma^2\right).$$

For even indices i, the optimal condition in Eq. (3.8) of the partition $\alpha_i^{-1}\gamma_i, 1-\alpha_i^{-1}\gamma_i$ yields

$$\frac{\alpha_i^{-1}\gamma_i}{1-\alpha_i^{-1}\gamma_i}=\frac{K_{i,i}^2\gamma_i}{K_{i+1,i}^2\left(\alpha_{i+1}\left(1-\alpha_i^{-1}\gamma_i\right)\right)},$$

i.e., $\alpha_{i+1}\alpha_i^{-1}=K_{i,i}^2/K_{i+1,i}^2$. Similar arguments show $\alpha_{i+1}\alpha_i^{-1}=K_{i,i}^2/K_{i+1,i}^2$ for odd indices i. Hence, the claim is verified.

Hence, $\alpha_i\alpha_k^{-1}=G^{i-k}$. Note that $\gamma_0 = 1$. By eliminating $\gamma_1, \ldots, \gamma_{\kappa-1}$, the recursive relation (Eq. D.1) leads to

$$\gamma_k+k\alpha_k=\alpha_k\sum_{i=0}^{\kappa}\alpha_i^{-1}=\left(G^k+G^{k-1}+\ldots+1\right). \quad \text{(D.2)}$$

By the symmetry, $\alpha_i=\alpha_{2n-i}^{-1}$ and then $\alpha_n = 1$. Thus,

$$\gamma_n=\left(G^n+G^{n-1}+\ldots+1\right)-n=(1-G)^{-1}\left(1-G^{n+1}\right)-n. \quad \text{(D.3)}$$

From Eq. (D.2), $\alpha_\kappa = G^{\kappa-n}$ and the expression for G^{n+1} in Eq. (D.3),

$$\gamma_k-\gamma_{k+1}=G^{k-n}\left((k+1)\,G-k-G^{n+1}\right)=G^{k-n}\,(1-G)\,(\gamma_n+n-k-1),$$

which implies that $\gamma_\kappa > \gamma_{\kappa+1}$ for $\kappa < n$, and γn is the minimum. Hence, the success of the majority rule relies on $\gamma n > 1/2$. By convexity, $(1-G)^{-1}(1-G^{n+1}) \geq n+1-n(n+1)(1-G)/2$. Then $\gamma n \geq 1-n(n+1)(1-G)/2$. Along with Eq. (D.3), we obtain one sufficient condition, $1 > n(n+1)(1-G)$. Let $\Delta := na$. By the linear expansion of G, the minimal kernel scale is approximately a multiple of the spatial size of the point-set,

$$\sigma \geq \sqrt{n\,(n+1)\,(2b+a)\cdot a} = \sqrt{(1+1/n)\,(2nb\cdot\Delta/|\Delta|+1)}|\Delta|. \quad \text{(D.4)}$$

This result explains our selection on σ.

REFERENCES

[1]. Albarelli, A.; Rodola, E.; Torsello, A. A game-theoretic approach to fine surface registration without initial motion estimation. Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference; Jun. 2010 p. 430-437.

[2]. Jason, HT. Lung mechanics : an inverse modeling approach. Krieger; 2009. Bates.

[3]. Bertsekas DP. Nonlinear Programming. Athena Scientific. 2003

[4]. Besl PJ, McKay ND. A method for registration of 3-D shapes. IEEE Trans. Pattern Anal. Mach. Intell. 1992; 14:239–256.

[5]. Margrit, Betke; Harrison, Hong; Deborah, Thomas; Chekema, Prince; Jane, P. Ko Landmark detection in the chest and registration of lung surfaces with an application to nodule registration. Medical Image Analysis. 2003; 7:265–281. [PubMed: 12946468]

[6]. Bondar, Luiza; Hoogeman, Mischa S.; Vásquez Osorio, Eliana M.; Heijmen, Ben J. M. A symmetric nonrigid registration method to handle large organ deformations in cervical cancer patients. Medical Physics. 2010; 37:3760–3772. [PubMed: 20831084]

[7]. Boyd, Stephen; Vandenberghe, Lieven. Convex Optimization. Cambridge University Press; 2004.

[8]. Brenier Y. Polar factorization and monotone rearrangement of vector-values functions. Comm. Pure Appl. Math. 1991; XLIV:375–417.

[9]. Burkard, Rainer; Dell'Amico, Mauro; Martello, Silvano. Assignment Problems. Society for Industrial and Applied Mathematics; Philadelphia, PA, USA: 2009.

[10]. Carmo, MP. Differential geometry of curves and surfaces. Prentice-Hall; 1976.

[11]. Castillo, Richard; Castillo, Edward; Guerra, Rudy; Johnson, Valen E.; McPhail, Travis; Garg, Amit K.; Guerrero, Thomas. A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets. Physics in Medicine and Biology. 2009; 54:1849–1870. [PubMed: 19265208]

[12]. Chen P. A novel kernel correlation model with the correspondence estimation. JMIV. 2011; 39:100–120.

[13]. Chen P, Gui C. Alpha divergence based mass transport models for image matching problems. Inverse problem and Imaging. 2011; 5(3):551–590.

[14]. Christensen GE, Song JE, Lu W, Naqa IE, Low DA. Tracking lung tissue motion and expansion/compression with inverse consistent image registration and spirometry. Med. Phys. 2007; 34:2155–2163. [PubMed: 17654918]

[15]. Chui H, Rangarajan A. A new algorithm for non-rigid point matching. CVIU. 2003; 89:114–141.

[16]. Evans, LC. Partial Differential equations methods and Monge-Kantorovich mass transfer. Yau, ST., editor. International Press; Boston: 1997.

[17]. Fan, Li; Chen, Chang W.; Reinhardt, Joseph M.; Hoffman, Eric A. Evaluation and application of 3D lung warping and registration model using HRCT images. Proc. SPIE. 2001; 4321:234–43.

[18]. Gilbarg, David; Trudinger, Neil S. Elliptic partial differential equations of second order. Springer-Verlag, Berlin; New York: 1983.

[19]. Groher M, Zikic D, Navab N. Deformable 2D–3D registration of vascular structures in a one view scenario. Medical Imaging, IEEE Transactions on. 2009; 28:847–860.

[20]. Haker S, Zhu L, Tannenbaum A, Angenent S. Optimal mass transport for registration and warping. International Journal of computer vision. 2004; 60:225–240.

[21]. Han, Xiao. Feature-constrained nonlinear registration of lung CT images. MICCAI 2010 Grand Challenges in Medical Image Analysis: Evaluation of Methods for Pulmonary Image Registration. 2010

[22]. Horn, RA.; Johnson, CR. Matrix Analysis. Cambridge University Press; 1985.

[23]. Kaijser T. Computing the Kantorovich distance for images. J. Math. Imaging and Vision. 1998; 9:173–191.

[24]. Kantorovich LV. On the transfer of masses. Dokl. Akad. Nauk. SSSR. 1942; 37:227–229.

[25]. Karlsson, M.; Berndtsson, B. Preprint (Department of Mathematical Sciences, Chalmers University of Technology and Göteborg University). Chalmers tekniska högskola; 2005. The maximum polygon area and its relation to the isoperimetric inequality.

[26]. Kaus MR, Brock KK, Pekar V, Dawson LA, Nichol AM, Jaffray DA. Assessment of a model-based deformable image registration approach for radiation therapy planning. International Journal of Radiation Oncology Biology Physics. 2007; 68:572–580.

[27]. Li B, Christensen GE, Hoffman EA, Mclennan G, Reinhardt JM. Pulmonary CT image registration and warping for tracking tissue deformation during the respiratory cycle through 3D consistent image registration. Medical Physics. 2008; 35:5575–5583. [PubMed: 19175115]

[28]. Lowe, David G. Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vision. 2004; 60:91–110.

[29]. Antoine Maintz JB, Viergever Max A. A survey of medical image registration. Medical Image Analysis. 1998; 2:1–36. [PubMed: 10638851]

[30]. McCann RJ. Existence and uniqueness of monotone measure-preseving maps. Duke Math. J. 1995; 80:309–323.

[31]. Mémoli, Facundo. Spectral Gromov-Wasserstein distances for shape matching. Workshop on Non-Rigid Shape Analysis and Deformable Image Alignment (ICCV workshop, NORDIA'09); Oct. 2009

[32]. Murphy K, van Ginneken B, Pluim J, Klein S, Staring M. Semi-automatic reference standard construction for quantitative evaluation of lung CT registration. Proc. MICCAI. 2008; 2008; 5242:1006–1013.

[33]. Museyko O, Stiglmayr M, Klamroth K, Leugering G. On the application of the Monge-Kantorovich problem to image registration. SIAM J. Imaging Sciences. 2009; 2:1068–1097.

[34]. Paquin D, Levy D, Xing L. Hybrid multiscale landmark and deformable image registration. Math. Biosci. Eng. 2007; 4:711–713. [PubMed: 17924721]

[35]. Rabin J, Delon J, Gousseau Y. A statistical approach to the matching of local features. SIAM J. Imaging sciences. 2009; 2:931–958.

[36]. Reinhardt JM, Ding K, Cao K, Christensen GE, Hoffman EA, Bodas SV. Registration-based estimates of local lung tissue expansion compared to xenon CT measures of specific ventilation. Med. Image Anal. 2008; 12:752–763. [PubMed: 18501665]

[37]. Rockafellar, RT. Convex analysis. Princeton University Press; 1970.

[38]. Rodolé, Emanuele; Bronstein, Alex M.; Albarelli, Andrea; Bergamasco, Filippo; Torsello, Andrea. A game-theoretic approach to deformable shape matching. The XXV IEEE Conference on Computer Vision and Pattern Recognition (CVPR2012). 2012

[39]. Rohr, K. Computational imaging and vision. Kluwer Academic Publishers; 2001. Landmark-based image analysis: using geometric and intensity models.

[40]. Rubner Y, Tomasi C, Guibas LJ. The earth mover's distance as a metric for image retrieval. Int. J. Comput. Vis. 2000; 40:99–121.

[41]. Saada, AS. Pergamon unified engineering series. Pergamon Press; 1974. Elasticity: theory and applications.

[42]. Sarrut D, Boldea V, Miguet S, Ginestet C. Simulation of four-dimensional CT images from deformable registration between inhale and exhale breath-hold CT scans. Med. Phys. 2006; 33:605–617. [PubMed: 16878564]

[43]. Sinkhorn R. A relationship between arbitrary positive matrices and doubly stochastic matrices. Ann. Math. Statist. 1964; 35:876–879.

[44]. Sprinzak, Josef; Werman, Michael. Affine point matching. Pattern Recognition Letters. 1994; 15:337–339.

[45]. Sun, Jian; Ovsjanikov, Maks; Guibas, Leonidas. A concise and provably informative multi-scale signature based on heat diffusion. Proceedings of the Symposium on Geometry Processing, SGP '09; Aire-la-Ville, Switzerland, Switzerland: Eurographics Association; 2009. p. 1383-1392.

[46]. Tawhai MH, Hunter PJ, Tschirren J, Reinhardt JM, McLennan G, Hoffman EA. CT-based geometry analysis and finite element models of the human and ovine bronchial tree. J. Appl. Physiol. 2004; 97:2310–2321. [PubMed: 15322064]

[47]. Howatson Tawhai M, Pullan AJ, Hunter PJ. Generation of an anatomically based three-dimensional model of the conducting airways. Annals of Biomedical Engineering. 2000; 28:793–802. [PubMed: 11016416]

[48]. Tsin Y, Kanade T. A correlation-based approach to robust point set registration. ECCV. 2004; 3:558–569.

[49]. Villani C. Topics in Optimal Transportation. Graduate Studies in Mathematics, AMS. 2003

[50]. Villani, C. Optimal transport: Old and New. Springer Verlag (Grundlehren der mathematischen Wissenschaften); 2008.

[51]. Wendland, Holger. Scattered Data Approximation. Cambridge University Press; 2005.

[52]. Werman M, Peleg S, Rosenfeld A. A distance metric for multi-dimensional histograms. Comp. Vis. Graphics Image Proc. 1985; 32:328–336.

[53]. Werner, René; Ehrhardt, Jan; Schmidt, Rainer; Handels, Heinz. Patient-specific finite element modeling of respiratory lung motion using 4D CT image data. Medical Physics. 2009; 36:1500–1511. [PubMed: 19544766]

[54]. Yin Y, Hoffman EA, Ding K, Reinhardt JM, Lin CL. A cubic B-spline-based hybrid registration of lung CT images for a dynamic airway geometric model with large deformation. Physics in Medicine and Biology. 2011; 56:203–218. [PubMed: 21149947]

[55]. Yin Y, Hoffman EA, Lin CL. Local tissue-weight-based nonrigid registration of lung images with application to regional ventilation. SPIE Medical Imaging. 2009; 7262:72620C.

[56]. Yin Y, Hoffman EA, Lin CL. Mass preserving nonrigid registration of CT lung images using cubic B-spline. Med. Phys. 2009; 36:4213–4222. [PubMed: 19810495]

[57]. Zhang, Yin. Solving large-scale linear programs by interior-point methods under the matlab environment. Optimization Methods and Software. 1998; 10:1–31.

[58]. Zhang Z. Iterative point matching for registration of free-form curves and surfaces. International Journal of Computer Vision. 1994; 13:119–152.

[59]. Zhu L, Yang Y, Haker S, Tannenbaum A. An image morphing technique based on optimal mass preserving mapping. IEEE Image Processing. 2007; 16:1481–1495.

[60]. Zitova B, Flusser J. Image registration methods: a survey. Image and Vis. Compu. 2003; 21:977–1000.

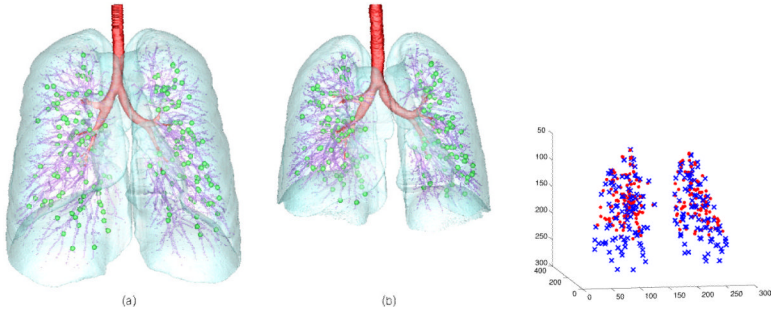

Fig. 1.1.

3D CT images of lungs(Left): (a) TLC (total lung capacity) and (b) FRC (functional residual capacity). Branch points are marked by green dots. The lungs, the airway tree, and the vessel tree are marked by cyan, red, and purple, respectively. Right: Two point-sets selected from CT images. The unit of the coordinates is mm. The direction z is along the lung height, i.e., a small z corresponds to the apex and a large z corresponds to the base. TLC is shown by markers × and FRC is shown by markers •.

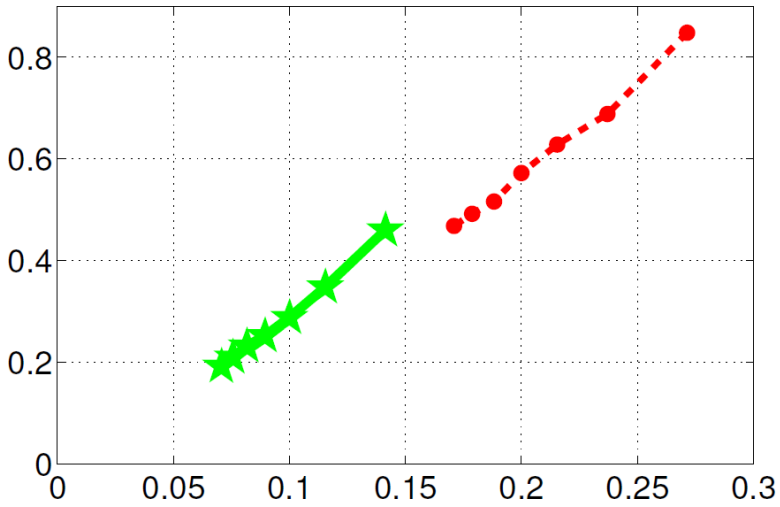

Fig. 2.1.
Robustness of the extra curl for randomly sampled point-sets. The graph is generated using Table 2.2. The y-axis is $\omega_+ - \omega_-$, and the x-axis shows $n^{-1/2}$(the green solid line) and $n^{-1/3}$(the red dashed line) for 2D and 3D, respectively.

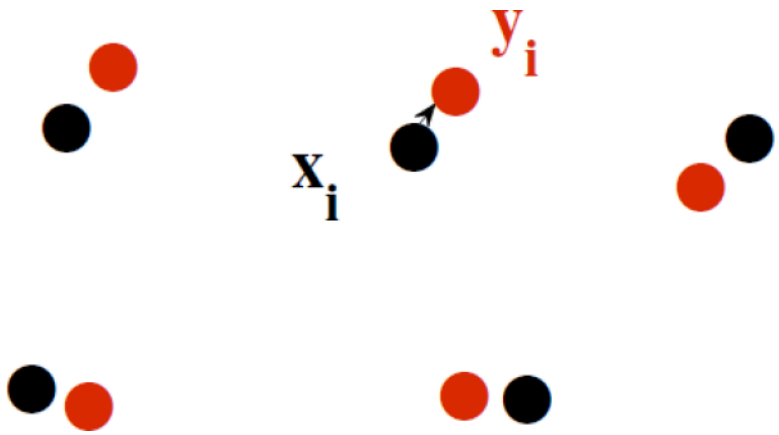

Fig. 3.1.
Illustration of closest point pairs. x_i, y_i are a pair of closest points.

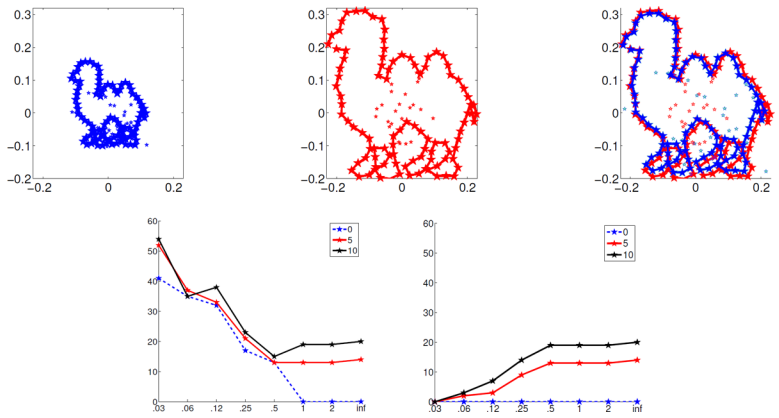

Fig. 3.2.

Top row: Point-set X, point-set Y and the matching between $T_1(X)$, Y, where an affine transform T_1 is estimated from the first two moments. Second row: the left subfigure shows the matching performance between X, Y; the right subfigure shows the matching performance between $T_1(X)$, Y.

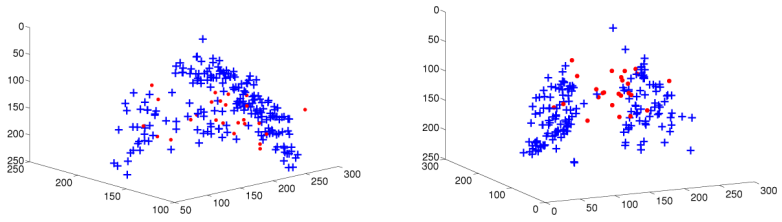

Fig. 3.3.
Left: 25 outliers (red •) are added to the TLC point-set of the first lobe (shown in blue +). Right: 25 outliers (red •) are added to the TLC point-set with subject ID H6972 (shown in blue +).

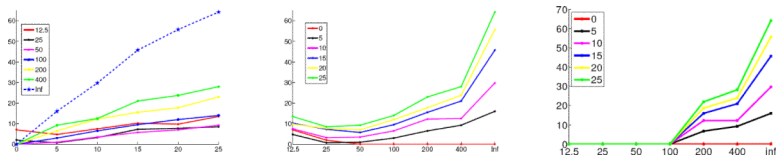

Fig. 3.4.

Mismatches of the mass transport-based methods applied to lobe 1 under the existence of outliers. Left: The x-axis is the cardinality of outliers. The results from the MK method and the HD model are depicted as a blue dashed line and solid lines. Here, no positive definite affine transform is used. Middle: Another viewpoint of the left subfigure. The x-axis is the kernel scaleσ. σ = inf refers to the MK method. Right: The mismatches of the MK and HD methods are presented. A positive definite a ne transform is used in the HD method to match the moments of the two point-sets.

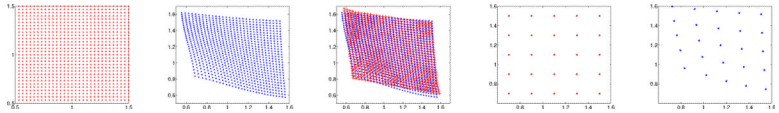

Fig. 3.5.

Matching results for hierarchical approaches. The left-1 and left-2 subfigures show the finer point-sets X_2 and Y_2, and the middle subfigure shows the point-sets $T_1(X_2)$ and Y_2, where the transform T_1 is estimated from the coarser point-sets X_1 and Y_1 in the left-4 and left-5 subfigures.

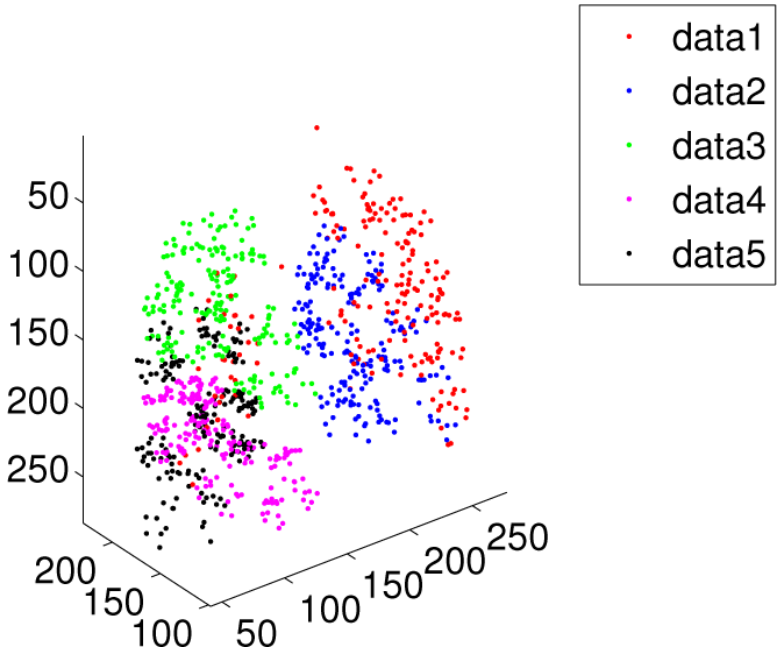

Fig. 4.1.
A 3D view of five lung lobes marked with five different colors.

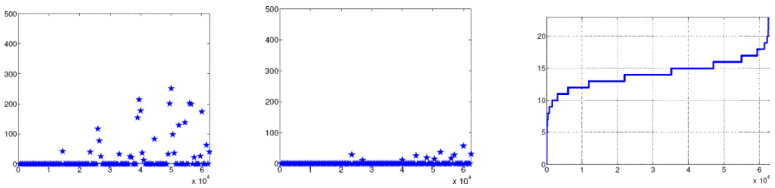

Fig. 4.2.

Mismatch errors for matching point-sets consisting of *500* point pairs consecutively selected from the 62,500 point-pairs. The left and middle subfigures show the mismatches without (left) and with (middle) transforms that were estimated using a pair of coarse point-sets with a cardinality of *50* points. The x-axis shows the last point ID for each point-set. For instance, x = *10, 000* refers to the point-set with IDs from *99, 501* to *10, 000*. Right: The y-axis shows the corresponding generation information for each point pair. For example, the first 10^4 branch points have generation IDs less than *13*.

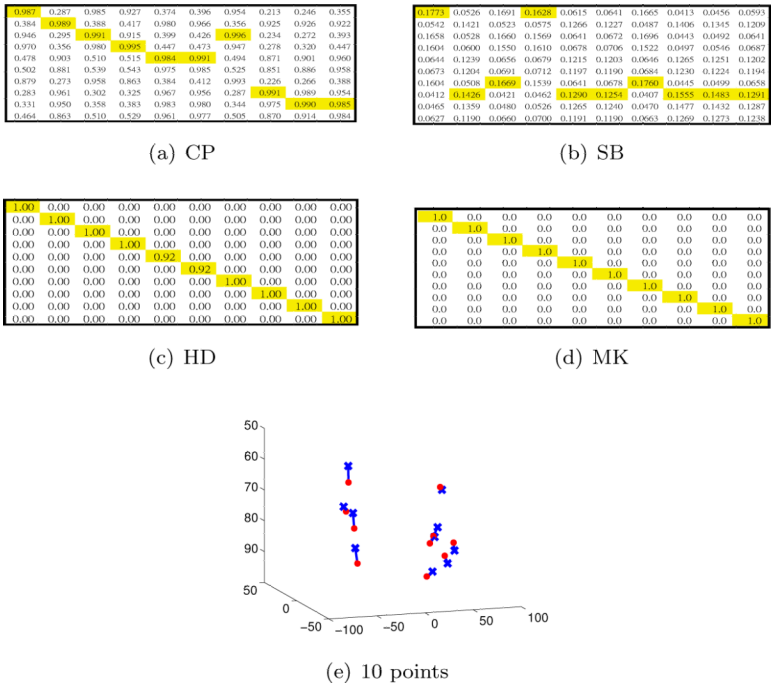

Fig. 4.3.

Matching ten points. The ground truth correspondence is the identity matrix. (a–d) show the correspondence of each method. Each diagonal entry is expected to be the maximal entry of each column, which is marked in yellow. The match errors are *3*(a), *7*(b), *0*(c) and *0*(d), respectively. (a) The matrix $K_{i,j} = \exp(-\|x_i - y_j\|^2/100^2)$. (b) The doubly stochastic matrix $\{\mu_{i,j} : i, j = 1, \ldots, 10\}$ after performing Sinkhorn matrix balancing on $K_{i,j}$. (c) $\left\{\gamma_{i,j}^{+}\gamma_{i,j}^{-}\right\}_{i,j=1}^{10}$ by the HD model. (d) $\{\mu_{i,j}\}_{i,j=1}^{10}$ is the minimizer of the MK method. (e) Illustration of two point-sets with 10 points each. The lines show the ground truth correspondence.

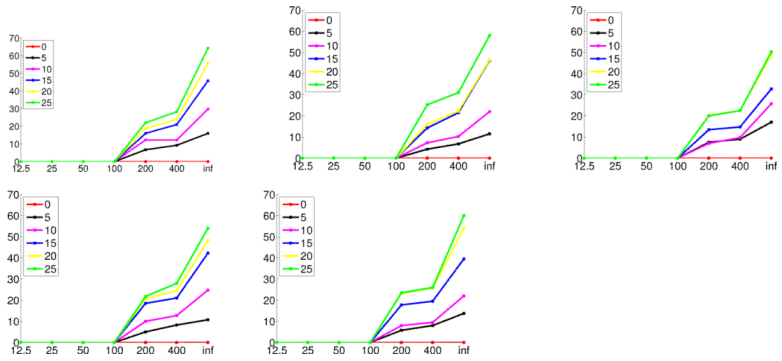

Fig. 4.4.

Mismatches for the MK (σ_K = inf) and HD methods applied to data from 5 lobes to which outliers have been added. The x-axis shows the kernel scale σ_K. The first row shows the results from the 1st, 2nd and 3rd lobes (left to right). The second row shows the results from the 4th and 5th lobes.

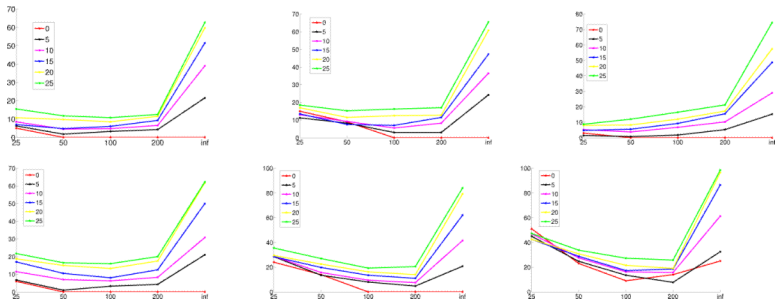

Fig. 4.5.

Mismatches for the MK and HD methods applied to six subjects to which outliers have been added. The first row (left to right) shows the results of the following subjects: ID H5972 (171 point-pairs), H5974 (188 point-pairs) and H5978 (181 point-pairs). The second row shows the results of the following subjects: ID H5983 (128 point-pairs), H6012 (254 point-pairs) and H6019 (243 point-pairs).

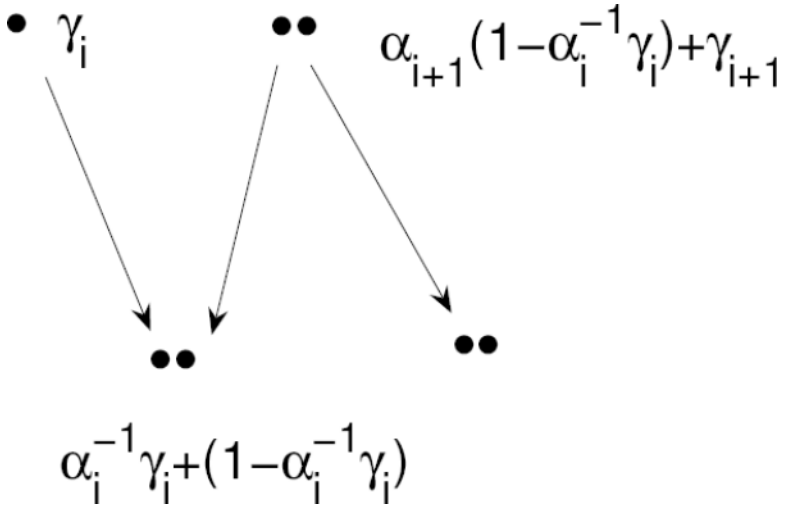

Fig. D.1.

Illustration of the bijective structure. Unit masses are divided into $\alpha_i^{-1}\gamma_i, 1-\alpha_i^{-1}\gamma_i$ and $\alpha_{i+1}\left(1-\alpha_i^{-1}\gamma_i\right)$, γ_{i+1} respectively.

Table 2.1

A summary of mathematical symbols

T, S	general deformations (vector-valued functions)
T_A, T_S	symmetric and skew symmetric parts of the ∇T matrices
φ, ψ(except in sec. 3)	functions in the Helmholtz decomposition of T
φ, ψ(in sec. 3)	dual variables in the HD model
Q, R	rotations (orthogonal matrices)
A	affine transforms (nonsingular matrices)
t	translations (vectors)
σ	kernel scales
K, $K_{i,j}$	the Gassian kernel function and its entries in matrix form
X, Y, x_i, y_j	point-sets and points in the point-set
μ, $\mu_{i,j}$	correspondence matrices in mass transport and their entries
$\Gamma^+, \Gamma^-, \gamma^+_{i,j}, \gamma^-_{i,j}$	correspondence matrices in the HD model and their entries
τ	permutation maps
Λ(B; k)	the kth largest singular value of the matrix B
ω_{max}	the magnitude of the "curl" of T defined in Eqs. (2.3) and (2.4)

Table 2.2

$(\omega_+ - \omega_-) \times 100$ based on 10 random samples

n	50	75	100	125	150	175	200
2D	46 ± 4.4	34.8 ± 1.6	28.8 ± 1.6	25.2 ± 1.88	23.3 ± 1.88	21 ± 0.89	19.2 ± 1.4
3D	84.8 ± 6.9	68.8 ± 4.4	62.8 ± 3.4	57.2 ± 3.9	51.6 ± 2.6	49.2 ± 2.4	46.8 ± 1.6

Table 3.1

Matching results for hierarchical approaches, as shown in the third column. The transform $T1$ is estimated by matching the coarser point-set with cardinality 5^2. The second column shows the mismatches between X_2 and Y_2 that occur when using the HD model directly. The 4th and 5th columns show a comparison of the results obtained using the HD-RBF model with different λ values. The poor performance of the HD-RBF model is partially caused by the identical eigenvalues for the second moment of X_2.

n	X_2, Y_2	Hierarchy $T_1(X_2), Y_2$	HD-RBF λ = 1	HDRBF λ = 100
10^2	0	0	0	0
15^2	0	0	0	0
20^2	140	0	144	0
25^2	390	0	395	370
30^2	688	0	693	657

Table 4.1

Mismatches of the mass transport model to match 1000 points under different α values. Index 1 to 5 refers to the point-sets from the 1st lobe to the 5th lobe. Index 6 refers to a pair of point-sets assembled by five sets of 200 point-pairs from each lobe.

Index	α = 0.9	α = 1	α = 1.01	α = 1.05	α = 1.2	α = 1.5	α = 2	α = 3	α = 4
1	526	196	121	29	0	0	0	0	7
2	900	508	404	209	53	0	0	0	18
3	414	19	0	0	0	0	0	0	0
4	910	490	349	146	0	0	0	78	144
5	867	362	266	148	0	0	0	0	60
6	416	13	5	0	0	0	0	0	36

Table 4.2

Mismatches of the mass transport model with their mass centers matched.

Index	$\alpha = 1$	$\alpha = 1.01$	$\alpha = 2$	$\alpha = 3$	$\alpha = 4$
1	8	2	0	0	0
2	88	70	0	38	40
3	2	0	0	0	0
4	86	79	0	0	28
5	166	154	0	0	0
6	4	2	0	0	0

Table 4.3

Mismatches of various methods. The parameter $\lambda = 1$ is used in both the TPS-RPM and the HD-RBF methods. In the TPS-RPM, we use the deterministic appealing rule, $\sigma = 100 \times 0.9^k$, $k = 0, 1, 2, \ldots, .$

Subject ID Cardinality of point-pairs	H5972 171	H5974 188	H5978 181	H5983 128	H6012 254	H6019 243
CP	127	147	130	128	199	189
SB ($\sigma = 100$)	167	179	178	126	250	237
TPS-RPM	160	179	177	119	248	241
MK	0	0	0	0	0	25
HD-RBF ($\sigma = 100$)	0	0	0	0	2	2