Discovery and visualization of structural biomarkers from MRI using transport-based morphometry

Shinjini Kundu,

Medical Scientist Training Program, University of Pittsburgh, 526 Scaife Hall, 3550 Terrace Street, Pittsburgh, PA 15261, USA

Soheil Kolouri,

ISSL Group, HRL Laboratories, Malibu, CA 90265, USA

Kirk I. Erickson,

Brain Aging & Cognitive Health Lab, Department of Psychology, University of Pittsburgh, 3601 Sennot Square, Pittsburgh, PA 15260, USA

Arthur F. Kramer,

Beckman Institute, University of Illinois, 405 North Mathews Ave, Urbana, IL 61801, USA

Edward McAuley, and

Exercise Psychology Laboratory, Department of Kinesiology and Community Health, Louise Freer Hall, 906 S Goodwin Avenue, Urbana, IL 61801, USA

Gustavo K. Rohde

Biomedical Engineering, Electrical and Computer Engineering, Box 800759, Room 1115, 415 Lane Road (MR5 Building), University of Virginia, Charlottesville, VA 22908, USA

Abstract

Disease in the brain is often associated with subtle, spatially diffuse, or complex tissue changes that may lie beneath the level of gross visual inspection, even on magnetic resonance imaging (MRI). Unfortunately, current computer-assisted approaches that examine pre-specified features, whether anatomically-defined (i.e. thalamic volume, cortical thickness) or based on pixelwise comparison (i.e. deformation-based methods), are prone to missing a vast array of physical changes that are not well-encapsulated by these metrics. In this paper, we have developed a technique for automated pattern analysis that can fully determine the relationship between brain structure and observable phenotype without requiring any *a priori* features. Our technique, called transport-based morphometry (TBM), is an image transformation that maps brain images loss-lessly to a domain where they become much more separable. The new approach is validated on structural brain images of healthy older adult subjects where even linear models for discrimination, regression, and blind source separation enable TBM to independently discover the characteristic changes of aging and highlight potential mechanisms by which aerobic fitness may mediate brain health later in life. TBM is a generative approach that can provide visualization of physically meaningful shifts in tissue distribution through inverse transformation. The proposed

---

Correspondence to: Shinjini Kundu.

framework is a powerful technique that can potentially elucidate genotype-structural-behavioral associations in myriad diseases.

Keywords

magnetic resonance imaging; computer-aided detection; aging; transport-based morphometry

1. Introduction

Recent advances in magnetic resonance imaging (MRI) technology have enabled high-resolution imaging across many new modalities. Tissue properties can be now be measured at an unprecedented level of precision and detail. These developments hold promise to illuminate structural changes underlying diseases commonly considered medical mysteries. Unfortunately, the changes can often be subtle, spatially diffuse, and complex, escaping detection by visual inspection. For example, Figure 1 demonstrates how the common morphologic pattern that differentiates individuals who are most aerobically fit from those who are least fit defies identification by gross inspection alone. However, significant differences in several quantitative parameters are reported to exist from prior studies on these images [1, 2, 3]. Thus, there is a growing role for computer-aided techniques to aid in vision and detection of morphologic patterns from MRI. Computer-aided techniques are needed to answer the following questions: are there morphologic differences that differentiate these groups? If so, what are they?

Unfortunately, traditional techniques for MRI analysis have difficulty with analysis in the image domain as well, as they require features to be pre-specified and are prone to missing a multitude of physical changes that are not adequately assessed by these finite feature sets. For example, popular biomedical image analysis softwares such as WND-CHRM [4] or FreeSurfer [5] extract a number of pre-specified numerical descriptors, such thickness, volume, texture statistics, etc., from the images and test whether these quantities are statistically different between image sets using a trial and error approach. In fact, WND-CHRM extracts nearly 3000 generic features from the images for testing. However, not only is testing descriptors a tedious process, numerical descriptors such as SIFT, Gabor features, or histogram statistics often do not have direct biological meaning. Another major limitation of these approaches is that the analysis does not incorporate anatomic prior information, missing an opportunity to compare variations in terms of known anatomy. Deformation-based methods, which include deformation-based morphometry (DBM) [6], tensor-based morphometry [7], and voxel-based morphometry (VBM) [8], also have limitations. Deformation-based methods rely on a nonrigid registration to align images before comparing them pixel-wise. In practice, perfect structural and functional alignment cannot be ensured, and subtle changes in pixel alignment can vastly change the results obtained and undermine accuracy [9]. Another limitation in these techniques is that the deformation fields are not unique. Hence, the results of tensor-based morphometry and deformation-based morphometry, which compare pixelwise across the determinant of Jacobian or deformation fields respectively, will vary depending on the particular field generated by the algorithm. While these methods incorporate anatomical information, projecting images onto individual

pixels assumes that the changes are localized into clusters and misses spatially diffuse changes. Furthermore, deformation fields model changes in local volume expansion/ contraction in order to match gross shapes. However, as Figure 2 illustrates, deformation fields cannot fully match images with zero error because they cannot quantify differences in tissue topology or texture. Figure 3 illustrates several neurologic diseases for which the main variation is in tissue texture rather than brain volume contraction/expansion - multiple sclerosis lesions and brain tumors. The authors of deformation-based techniques state that these registration-based methods are a way to index into pixels based on amount of gray matter per unit volume [10], but cannot offer insight into the physical meaning of these changes [11] as these methods are not *generative*. If a technique is not generative, then an observable datapoint, or a brain MR image cannot be generated given a feature set such as a density map. However, a generative method would afford the ability to visualize the shifts in morphologic profile as dynamic changes across a series of brain MRIs to illuminate structural mechanisms.

In this paper, we describe transport-based morphometry (TBM) [13, 14, 15, 16, 17], which has the potential to enable fully automated MRI analysis without loss of information. Rather than analyzing images in the image domain where they may not be easily separable, we first transform them to a domain that enhances separability. Prior work demonstrates that when we transform 1D and 2D signals in other applications using TBM [13, 14, 15, 16, 17, 18, 19], complex and nonlinear morphology in the image domain can be described by linear classification and regression models in the transform domain. Furthermore, the key advance of TBM is that it is generative and enables direct visualization of the interface between signal classes through inverse TBM transformation [19]. The TBM technique computes the distance needed to morph one image with respect to a common template using the mathematics of optimal mass transport (OMT). Optimal mass transport has a long mathematical history, starting in the 1700s with Monge and most recently due to its myriad applications in signal and data analysis [19]. As Figure 2 illustrates, unlike deformation-based approaches, OMT can match both shape and texture variations simultaneously; thus, information is not missed. However, TBM has never been applied for MRI-based detection and pattern analysis as current formulations and solutions to TBM are designed for smaller signals [13, 14, 15, 16, 17, 18].

In this paper, we demonstrate a TBM framework that is suitable for analysis of radiology data, the majority of which comprises three-dimensional data. We hypothesize that transforming MRI data using the new TBM approach can facilitate both discovery as well as visualization of discriminating differences in a manner similar to 1D and 2D signal analysis previously reported [13, 14, 15, 16, 17, 18, 19]. Ultimately, the goals of discovering objective clinical markers and understanding structure-function relationships would be facilitated by a technique that could assess structural changes underlying clinical phenotype in a fully automated manner without information loss and visualize the shifts in tissue distribution as a series of radiology images as part of a unified framework.

The specific contributions of this work are:

- Novel, robust formulation and solution of transport-based morphometry (TBM) enabling its first application to radiology data and computational validation

- A new image transform to facilitate pattern analysis on MRI data, with equations for analysis and synthesis as well as description of how the TBM pipeline can be used for discrimination, regression, and unsupervised learning

- Demonstration of TBM on real world neuroimage analysis problems showing the advantage of a generative technique in identification of morphologic changes as well as visualization compared to current morphometry techniques

The remainder of this paper is organized as follows. In Section 2, we summarize key theoretical results and present equations for forward and inverse TBM transformation. In Section 3, we present our TBM solver suitable for MRI data. Section 4 describes a framework for regression, discrimination, and blind signal separation tasks in the transform domain. In Section 5, experimental methodology is presented. Section 6 presents results showing robustness of the proposed solver for 3D data and the ability of TBM to accurately assess dependent brain morphology changes with age in Section 6, whereas the traditional approaches based on diffeomorphic anatomic registration through exponentiated lie algebra algorithm (DARTEL) fails to detect these changes [12]. Finally, in Section 7, we present discussion of the results and in Section 8, we conclude this paper. The Appendix material details the derivation and experimental validation for our solver in Section 3.

2. Optimal Mass Transport for Signal Transformation

This section summarizes key theorems related to optimal mass transport, equations for signal transformation using TBM, and OMT minimization.

2.1. Overview of optimal transport theory

Let Ω be a measurable space. Let μ and σ be probability measures defined on Ω, with corresponding positive probability densities I_1 and I_0, respectively. A mass preserving transform f that pushes σ to μ, or $f_\# \sigma = \mu$, satisfies the following,

$$\int_A d\sigma(x) = \int_{f(A)} d\mu(x), \quad \forall A \subset \Omega. \quad (1)$$

Figure 4 illustrates μ and σ, as well as the map $f_\# \sigma = \mu$. Such a mass preserving (MP) mapping f is in general not unique; in fact, infinitely many MP mappings may exist that satisfy Equation 1. However, we are interested in finding the MP mapping that is optimal in the sense of *mass transport*, which we will define further in Equation 2. Optimal mass transport theory developed two major formulations: one in the continuous domain utilizing a transport map called the Monge formulation, and one able to work with discrete masses such as dirac delta called the Kantorovich formulation. These are further described in [19]. In this paper, we consider digital signals as being sampled from a continuous domain and employ the Monge formulation of the problem.

Let MP be the set of all such mass preserving mappings, $MP := \{f : \Omega \rightarrow \Omega / f_{\#}\sigma = \mu\}$. The optimal MP mapping in the mass transport sense can be written according to Monge's formulation, which minimizes the following cost function,

$$\min_{f \in MP} \int_{\Omega} c(x, f(x))d\sigma(x) \quad (2)$$

Here, $c : \Omega \times \Omega \rightarrow \mathbb{R}^+$ is the cost functional. The functional c measures mass transportation cost and is often chosen to be the L^p-norm for which Equation (2) becomes the L^p-Wasserstein distance. The L^2-Wasserstein distance, $c(x, y) = |x - y|^2$, in particular has attracted rich attention in the image analysis, computer vision, and machine learning communities. For $c(x, y) = |x - y|^2$, Brénier [22] showed that there exists a unique optimal transportation map $f \in MP$ for which,

$$\int_{\Omega} |x - f(x)|^2 d\sigma(x) \leq \int_{\Omega} |x - g(x)|^2 d\sigma(x), \forall g \in MP \quad (3)$$

when (i) $\Omega = \mathbb{R}^n$ and the probability measures have finite second-order moments (i.e. their densities vanish in the limit),

$$\int_{\Omega} |x|^2 d\mu(x) < \infty \text{ and } \int_{\Omega} |x|^2 d\sigma(x) < \infty,$$

and (ii) when σ is absolutely continuous with respect to Lebesgue measure.

For certain measures (e.g. when σ is not absolutely continuous) the Monge formulation of the optimal transport problem is ill-posed in the sense that there is no transport map that rearranges σ into μ. In such scenarios the Kantorovich formulation of the problem is preferred.

Moreover, Brénier showed through polar factorization [22] that the transport map f must be the gradient of a convex function $\phi : \Omega \rightarrow \mathbb{R}$, $f = \nabla \phi$. The preceding property implies that when Ω is a convex and connected subset of $\mathbb{R}^n$, the optimal transport map is *curl free*.

2.2. Linear optimal transport analysis framework

By considering magnetic resonance images to be smooth density functions, the similarity in *spatial distribution* of two tissues can be quantified based on the L^2-Wasserstein distance. The L^2-Wasserstein distance defines a metric between images by identifying a unique spatial transformation for each brain image.

Any MRI modality that generates scalar intensity maps (i.e. T1-weighted, T2-weighted, FLAIR, fractional anisotropy, etc.) is amenable to analysis by the TBM framework. In this work, we analyze T1-weighted images, where treating images as densities enables

comparison between images where the absolute intensity value may not be physically meaningful.

Consider a set of magnetic resonance images $I_1, \ldots, I_K : \Omega \rightarrow \mathbb{R}^+$, corresponding to experimental subjects $1, \ldots, K$, where $\Omega = [0, 1]^3$, the images are first intensity normalized to produce densities such that

$$\int_\Omega I_m(x)dx = 1 . \quad (4)$$

where $m \in 1, \ldots, K$. A common reference image I_0, is chosen and the optimal transport mappings are calculated from the reference image to each subject's MRI, I_m. Let $f_m : \Omega \rightarrow \Omega$ be a mass preserving mapping from I_0 to I_m. Then, the analysis equation [17] that transforms images to their corresponding representation in transform domain can be written based on

$$f_m^*(x) = arg \min_{f_m \in \text{MP}} \int_\Omega |f_m(x) - x|^2 I_0(x)dx, \quad (5)$$

$$\text{s.t.} \quad \det(Df_m(x))I_m(f_m(x)) = I_0(x) \ \text{for} \ \forall x \in \Omega$$

Here, Df_m is the Jacobian of the mapping f_m and MP is the family of all mass preserving mappings from I_0 to I_1. The existence of a unique solution f_m^* to above optimization was shown by Brénier [22].

The transport maps $f_m^*(x)$ are vector fields that define the direction and amount of mass transport needed to morph $I_m(x)$ into $I_0(x)$. OMT defines a nonlinear distance metric, as Figure 5 shows, where the arcs, or geodesics, on the manifold between two images I_0 and I_m correspond to nonlinear OMT distances and are represented by $f_m^*(x)$. The metric space defined by the OMT-based distance metric is a Riemannian manifold, which is equipped with an inner product. Thus, projecting the manifold locally at I_0 to the tangent space maps the geodesics f_m^* to linearized versions in the tangent space, called the linearized optimal transport (LOT) metric.

Then, it can be shown that $\hat{I}_m(x) = (f_m^*(x) - x)\sqrt{I_0(x)}$ provides a natural isometric linear embedding for image I_m with respect to the LOT [13, 17]. This linear embedding is generative, thus, any arbitrary point in the LOT space can be directly inverted and visualized in the image domain [13, 17, 18, 19] as a new image according to the synthesis equation [13, 17]

$$I(x) = det(Df^{-1}(x))I_0(f^{-1}(x)) \qquad (6)$$

where $f^{-1}(x)$ is the inverse mapping of $f(x)$

Unlike current approaches based on non-rigid registration, LOT defines a *unique* spatial transformation for each image, with equations for analysis and synthesis. Uniqueness makes LOT an invertible transformation. In contrast, current registration-based correspondences are not unique, mathematically speaking. As the authors of these techniques state, unless the generative model is invertible, the parameters studied have no physical meaning [11]. LOT is a powerful technique for image analysis because it enables generative modeling in which discovery of structural biomarkers and direct visualization of morphologic shifts are unified within a single framework. In the transport domain, simple Euclidean operations on the LOT-transformed embeddings, such as linear classification and linear regression correspond to nonlinear operations on the OMT manifold [17]. Thus complex, spatially diffuse, nonlinear morphologic shifts in the image domain can be captured by simple Euclidean operations in the LOT domain. Most importantly, physical shifts in tissue morphology can be directly visualized through inverse LOT transformation.

3. Proposed approach

We have developed a method for solving optimal transport that enables the transport-based morphometry technique to be extended to large 3D volumetric images such as MRI. The authors offer this OMT approach as a viable option for carrying out TBM transformation for 3D volumetric images. While numerical OMT is a vast field, a detailed review or evaluation of OMT algorithms in general is beyond the scope of this paper.

3.1. Variational formulation of the problem

In order to find the optimal transport map, we reformulate the minimization in (2) by relaxing the MP constraint. Assuming that Ω is a convex and connected subset of $\mathbb{R}^n$, as it is the case for most image analysis problems (i.e. $\Omega = [0, 1]^n$), and assuming that the probability measures μ and σ are atomless and absolutely continuous, we can write the differential counterpart of Equation (1) as,

$$det(Df(x))I_1(f(x)) = I_0(x), \quad \forall f \in MP \quad (7)$$

where D is the Jacobian matrix, and $det(.)$ denotes the determinant operator. The minimization in (2) for $c(x, y) = |x-y|^2$ can first be relaxed into the following optimization problem,

$$\operatorname{argmin}_f \; \frac{1}{2} \int_\Omega |x - f(x)|^2 I_0(x) dx \quad (8)$$

$$s.t. \; \left\| det(Df)I_1(f) - I_0 \right\|^2 \le \varepsilon$$

for some small $\varepsilon > 0$. Next, we use the result from Brénier's theorem which states that the optimal transport map is a curl free mass preserving map. Therefore we propose to modify the optimization problem in Equation (8) by regularizing the objective function with the curl of the mapping f,

$$\text{argmin}_f \; \frac{1}{2}\int_\Omega \left| x - f(x) \right|^2 I_0(x)dx + \frac{\gamma}{2}\int_\Omega |\nabla \times f(x)|^2 dx \quad (9)$$

$$s.t. \; \left\| det(Df)I_1(f) - I_0 \right\|^2 \le \varepsilon$$

where γ is the regularization coefficient and $\nabla \times (.)$ is the curl operator. Therefore, the map f is sought that minimizes the total mass transport $\int |x - f(x)|^2 I_0(x)dx$ subject to the mass-preserving constraint $\|det(Df)I_1(f) - I_0\|^2 < \varepsilon$. We note that modifying (8) to penalize the objective with the curl $\|\nabla \times f\|^2$ does not change the optimal solution, but solving (9) in practice helps guide the solution toward the curl free map. We can relax the optimization problem above further and write it as a regularized (or penalized) unconstrained optimization problem,

$$\text{argmin}_f \; \frac{1}{2}\int_\Omega |x - f(x)|^2 I_0(x)dx + \frac{\gamma}{2}\int_\Omega |\nabla \times f(x)|^2 dx + \quad (10)$$

$$\frac{\lambda}{2}\int_\Omega (det(Df(x))I_1(f(x)) - I_0(x))^2 dx$$

Hence the formulation above contains terms explicitly signifying properties of MP mapping $\|det(Df)I_1(f) - I_0\|^2$ and a curl-free mapping $\|\nabla \times f\|^2$. The last term implicitly penalizes mappings that are not diffeomorphic when $det(Df(x))$ crosses zero.

The optimization problem in Equation (10) is not a convex problem. We use a multiscale variational optimization technique to help guide the solution toward the global optimum. We will see in the results section that the multiscale scheme is able to achieve solutions comparable to that obtained using convex methods when they apply. Section 3.3 describes the multiscale variational solver we devise for the optimization in (10).

3.2. Euler-Lagrange equations

The objective function in (10) can be written as,

$$M(f) = \int_\Omega \mathscr{L}(x, f(x), Df(x))dx. \quad (11)$$

The Euler-Lagrange equations for the transport field f then are of the form,

$$\frac{dM}{df^i} = \frac{\partial \mathscr{L}}{\partial f^i} - \sum_{k=1}^{n} \frac{d}{dx^k}\left(\frac{\partial \mathscr{L}}{\partial f^i_{x_k}}\right), \quad i = 1, \ldots, n \quad (12)$$

where the superscripts denote the coordinate index for the vectors, and the subscripts denote partial derivatives, $f^i_{x_k} = \frac{\partial f^i}{\partial x^k}$. Writing the Euler-Lagrange equations for the objective function in (10) leads to,

$$\frac{dM}{df} = (f - id)I_0 + \lambda(det(Df)\nabla I_1(f) - \nabla \cdot (adj(Df)I_1(f)))I_{error} + \gamma(\nabla \times \nabla \times f) \quad (13)$$

where $id(x) = x$ is the identity function, $adj(.)$ denotes the adjugate operator, $\nabla \cdot (.)$ is the divergence operator, and $I_{error} = det(Df)I_1(f) - I_0$. The derivation for the equation above is presented in the Appendix. Equation (13) is a key result from our formulation. The complexity for computing each gradient descent update step here is $\mathcal{O}(NlogN)$, where N is the number of pixels or voxels in the image. The computational complexity is determined by the cost of computing gradients $\mathcal{O}(N)$ in (13), but is dominated by the cost of cubic interpolation in computing the $det(Df)I_1(f)$ term.

3.3. Multiscale accelerated gradient descent

We can guide the solution toward the globally optimal solution by a multiscale scheme as depicted in Figure 6. Nesterov's accelerated gradient descent method [23] is used at each scale to find the corresponding optimal transport map from Equation (10). The optimal transport map is then interpolated and used as the initial point for the accelerated gradient descent method in the next scale (finer scale).

The accelerated gradient descent update for k'th iteration ($k > 1$) at each scale is as follows,

$$\begin{cases} g^k = f^{(k+1)} + \frac{k-2}{k+1}(f^{(k-1)} - f^{(k-2)}) \\ f^k = g^k - \alpha_k \frac{dM(g^{(k)})}{df} \end{cases} \quad (14)$$

where α_k is the gradient descent step size, and is automatically chosen at each gradient descent update such that the maximum displacement is fixed. The update at k = 1 is the usual gradient descent update.

Here, we have presented a viable approach for computing optimal transport minimization for MRI datasets, enabling a transport-based morphometry approach with radiology images.

4. Modeling shape and appearance of the brain

As previous work has demonstrated that transforming signals to the transport domain using OMT may increase their separability [13, 14, 15, 16, 17, 18, 19], we describe a framework for regression, discrimination, and blind signal separation in the transport space.

The data matrix $X \in \mathbb{R}^{d \times K}$ stores the vectorized transport maps x_m corresponding to each subject $m \in 1, \ldots, K$ where d is the number of elements in the vectorized transport map and K is the number of subjects. Figure 7 is the system diagram that illustrates the LOT transformation pipeline. In practice, the analysis is performed on the dimensionality-reduced data matrix that can provide a linear embedding for transport maps which, for the input data, can be used to fully represent each transport map. The reason for performing the computations in a reduced-dimension subspace is to enable standard regression techniques to be applied, as applying them in full high dimensional space ($\sim 10^8$) would be computationally expensive for most software packages.

4.1. Regression and correlation analysis with a clinical variable

The influence of an independent clinical variable $v \in \mathbb{R}^{K \times 1}$ on brain tissue distribution can be investigated by computing the direction in the transport domain w_{corr} such that the linear correlation with age is maximized according to (15) [15]. Here, X represents the reduced-dimension data matrix.

$$w_{corr} = arg\max_{w} \frac{w^T X v}{\sqrt{w^T w}} = \frac{Xv}{\sqrt{v^T X^T X v}} \quad (15)$$

Here, the direction $w = \bar{x} + v w_{corr}$ is a vector field that represents the direction and magnitude by which tissue is re-distributed due to v and ν represents the increment or decrement to sample along the maximally correlated direction. Pearson's correlation coefficient is computed on centered v and X.

The images corresponding to the computed direction w can be visualized through inverse TBM transformation by Equation (6) and illustrate the morphology that is associated with outcome v.

4.2. Discriminant analysis to differentiate groups of subjects

Another class of problems facilitated by the TBM technique is that of discriminating classes based on MRI appearance, such as the one posed in Figure 1. For these problems, penalized linear discriminant analysis (PLDA) [24] performed in the transport domain can find the direction in transport space that maximally separates C classes. The PLDA direction is given by (16)

$$w_{PLDA} = arg \max_{\|w\| = 1} \frac{w^T S_T w}{w^T (S_W + \alpha I) w} \quad (16)$$

where $S_T = \frac{1}{M} \sum_m (x_m - \bar{x})(x_m - \bar{x})^T$. Here, $\bar{x} = \frac{1}{M} \sum_{m=1}^{M} x_m$

The within-class scatter matrix is $S_W = \sum_C \sum_{n \in C} (x_n - \bar{x}_c)(x_n - \bar{x}_c)^T$. The parameter α controls the tradeoff between the traditional linear discriminant analysis (LDA) direction and one that lies in the principal component analysis (PCA) subspace. The parameter α can be chosen by plotting the stability of the subspace as a function of α.

Sampling along and inverting the direction w_{PLDA} yields images showing the typical morphology of a class and how it changes as one progresses from one class to another.

4.3. Visualizing principal phenotypic variations in the brain

Given the covariance matrix S_T defined in Section 4.2, the principal components are given by the eigenvectors of S_T. The eigenvectors represent the directions in the transport space that capture the main modes of variability in the dataset [15].

The factorization in Equation (17) gives both the principal components and eigenvalues, where the diagonal components of Σ represent the variance for each principal component.

$$S_T = U \sum U^T \quad (17)$$

For high dimensional data, the covariance matrix can be implicitly represented using the approach in [25]. Each principal component can be inverted and visualized to yield the principal phenotypic variations that comprise the images in the dataset.

5. Computational experiments

Here we describe image acquisition, preprocessing, morphometry analysis, and statistical learning steps. The code was prototyped in MATLAB (MathWorks, Natick, MA) using built-in libraries.

5.1. Datasets

5.1.1. MRI pattern analysis using transport-based morphometry—The ability of transport-based morphometry to aid in regression, discrimination, and signal separation tasks was assessed on images of 135 healthy subjects, ranging in age from 58 to 81 years (mean age 66.6 years, standard deviation 5.9 years). Both male and female subjects are included. T1-weighted brain images were collected using a 3D Magnetization Prepared Rapid Gradient Echo Imaging (MPRAGE) protocol with 144 contiguous slices. Images were acquired on a 3 T Siemens Allegra scanner with repetition time = 1,800 ms, echo time =

3.87 ms, field of view (FOV) 256 mm, and acquisition matrix 192×192 mm, flip angle = 8 [26]. These images provide an expanded dataset of older subjects on which age-related brain morphology can be investigated.

5.2. Multiscale variational optimal transport

5.2.1. Image preprocessing—Images were skull-stripped and affinely registered to the MNI template using Statistical Parametric Mapping (SPM) software version 12 [27]. The merits and demerits of the existing brain tissue segmentation methods are discussed in [28], which offer insight into the comparative performance of methods based on clustering, thresholding, convolutional neural networks (CNNs), and Markov Chain Monte Carlo (MCMCs).

Images were normalized so that the sum of intensities was equal in both images (equal *mass*). By normalizing to a large positive number, 10^6, numerical precision errors resulting from computations with small numbers are avoided. We also add a small constant 0.1 to the normalized images and renormalize so that they are strictly positive [29] to ensure that the OMT problem is well-posed.

The template image was chosen to be the Euclidean average of the sample images. In prior work utilizing linear optimal transportation for pattern analysis, both the Euclidean average [13], and Frechet mean [17] have been used to approximate the 2-Wasserstein distance between images. In previous work, substituting a smooth template with its sharper nearest neighbor did not increase or decrease discrimination accuracy significantly [17].

5.3. Experiment 1: Modeling the effects of aging on brain tissue distribution

Regression analysis was performed to assess the relationship between brain tissue distribution and age using the approach outlined Section 4.1. The common reference image I_0 was computed by the Euclidean average of all the subjects. Statistical significance of the computed direction is assessed using permutation testing with T = 1000 tests.

The reults of regression analysis in the transport space were compared to those obtained using deformation-based analysis. The DARTEL [12] toolbox in SPM12 [27] was used to compute deformation fields. DARTEL is commonly used to perform standard VBM and DBM analysis. Images were skull-stripped, segmented, and affine registered to the MNI template similar to the OMT procedure before non-rigid registration by performed by DARTEL. The most correlated direction was computed on the tissue density maps of DARTEL-registered images using Equation (15) for VBM and using the deformation fields for DBM analysis. Modulated versions are used to compensate for the effects of spatial normalization [30].

The TBM analysis is also performed on segmented gray matter and white matter tissue maps separately to enable comparison to VBM.

5.4. Experiment 2: Assessing the effects of aerobic fitness on brain health

Discriminant analysis between high aerobic fitness vs. low aerobic fitness groups is performed using the PLDA approach in transport space described in Section 4.2. Aerobic

fitness is measured by vO_2 L/min. The individuals were grouped into low-fit and high-fit groups based on those who had a vO_2 L/min greater than one standard deviation above the mean (high-fit: n = 22) and lower than one standard deviation below the mean (low-fit: n = 16).

Discriminant analysis using TBM was compared with that performed using deformation fields instead of transport maps generated by DBM. In VBM analysis, voxel-wise comparison on modulated tissue maps was performed to seek the voxel clusters that are significantly different between the two classes using two-sample t-test, uncorrected for multiple comparisons or cluster thresholding.

5.5. Experiment 3: Visualizing principal phenotypic variations

Unsupervised learning using PCA was performed to visualize the top three principal phenotypic variations in the transport space using the approach described in Section 4.3.

6. Results

6.1. Modeling normal variability in the brain

Figure 8 shows the fraction of variance captured by principal components of the image domain (raw voxel values after affine registration) compared to modulated DARTEL registration and optimal mass transport solved using the described approach. Fewer components are needed to represent more of the variance in the transport space than either for image domain or DARTEL pixelwise comparison. Therefore, the information about variability in the dataset appears to be better captured by examining tissue distribution using OMT rather than comparing tissue intensities individually, before or after nonrigid registration. The intuition for how OMT better capture variability in the dataset was previously illustrated by Figure 5.

6.2. Modeling the effects of aging on brain tissue distribution through TBM regression

Aging is clinically known to be associated with tissue atrophy and disproportionate loss of tissue from frontotemporal regions [31]. In this section, TBM is compared with DBM and VBM in the ability to independently discover and model these changes.

6.2.1. Assessing global changes—The direction maximally correlated with age computed in the transport space using TBM is found statistically significant (Pearson's r = 0.4605, p < 0.001). Figure 9a shows the data when it is projected onto the maximally correlated direction, with each datapoint representing a subject's image.

The most correlated direction shown in Figure 9a can be inverted to visualize the dynamic changes in morphology underlying the aging process. Figure 9b shows images generated by TBM inverse transformation (images are colorized to aid visual interpretation). We see that the changes captured by the TBM regression framework are well-corroborated by known changes in the clinical literature [31]. Specifically, the changes shown here are enlarging ventricles, especially in slices 75 and 66. There is global tissue thinning, and enlargement of the occipital horns of the lateral ventricles in slice 58. Normal anatomic landmarks

characteristic of the brain are also clearly visible in Figure 9b, such as the internal capsule in slice 66 and thalamus in slice 58.

We also compare the results of our TBM regression analysis with that using deformation fields generated by DARTEL, commonly used in DBM analysis. The relationship between the deformation fields and age is found to be statistically significant (Pearson's r = 0.2918, p = 0.0240), as Figure 10a shows, suggesting that there are significant shape changes with age that are captured by a DBM approach. Compared to the visualizations generated by TBM, we see that the those yielded by the deformation-based approach depict global shape changes but not texture changes. In all slices generated by deformation-based analysis, normal tissue landmarks are distorted. Especially, at the gray-white interface, there are is a ring-like texture that does not represent normal brain anatomy. Examining the associated images generated using the deformation-based approach, we see that while volume expansion of the ventricles is correctly identified, expected changes in tissue distribution are not well-captured using deformations alone. For example, in slice 75, the expected tissue thinning is not well-depicted in the frontal areas. In slice 66, there is an area of bilateral focal hyperintensity near the ventricles. This represents the distorted internal capsule that is correctly represented in slice 66 of Figure 9b. Other normal landmarks such as thalamus and putamen are notably absent in slice 58 as well as the occipital horns of the lateral ventricles. Thus, while DBM can capture global shape changes, such as enlargement of ventricles, texture information is not well modeled and many normal landmarks and are distorted.

6.2.2. Assessing gray matter and white matter changes—Transport-based morphometry was also applied to explore the effect of age on gray matter and white matter maps separately. Figure 11b shows the effect of age on gray matter distribution. The relationship is statistically significant with Pearson's r = 0.4271 and p<0.001. There is thinning of the gray matter tissue when progressing from a 53 year old brain to a 79 year old brain, most markedly in the temporal lobe as can be seen in slice 75. Atrophy can be seen in all the slices by enlargement of the spaces.

The relationship between white matter distribution and age similarly shows atrophy and enlargement of the ventricles in Figure 12. The relationship is statistically significant with Pearson's r = 0.4026 and p = 0.005. In addition, there appears to be disproportionate loss of white matter tissue from the frontal and temporal regions, which is best illustrated in slices 75 and 66 in Figure 12b.

In contrast, regression analysis performed on the modulated density maps registered by DARTEL, used for VBM, was unable to find a significant relationship between age and either gray matter morphology (Pearson's r = 0.6787, p = 0.1980) or white matter morphology (Pearson's r = 0.4965, p = 0.3870).

Figures 13 and 14 show the images generated by attempting to fit a regression model on individual pixel values on a fixed grid. We see that in both cases, progressing from age 53 to age 79, the intensity at voxels in the cortical gray matter is shown to decrease. However, no gross differences in shape are depicted. Similarly, examining the white matter images generated by regression on VBM maps (Figure 14), the intensity in the frontal white matter

appears to grossly decrease, especially in slices 75 and 58. However, neither of these relationships were statistically significant, nor do they adequately depict atrophy and loss of tissue from frontotemporal regions.

Overall, the known effects of aging on the brain are best assessed and depicted by the transport-based morphometry technique. A DBM approach does not adequately model tissue texture and VBM can identify intensity changes at fixed voxel locations, but these do not appear to be statistically significant when it comes to modeling the effect of aging on the brain tissue.

6.3. Assessing the effects of aerobic fitness on brain health through TBM discrimination

The effects of aerobic fitness on the brain are assessed by separating high-fit individuals from low-fit individuals using the PLDA approach for discriminant analysis, comparing the ability of TBM, DBM, and VBM to discover and visualize the interface between the two groups.

Clear separation in the training subspace is an expected result whether raw pixels, deformation fields or transport maps are used, as Figures 15a and 16a show, but when visualizing the interface between classes, TBM demonstrates clear advantages in physical interpretability. Visualizing the interface between the high-fitness and low-fitness groups using TBM in Figure 15b, we see that brains corresponding to low-fit individuals appear to demonstrate changes in tissue distribution that are similar to those due to advancing age in Figure 9b. Similarly, those individuals belonging to the high-fit group have brain morphology that appears to be resemble those of younger subjects in an older adult population as the ventricles appear smaller and tissue architecture in the frontotemporal regions are better preserved. Thus, it appears that fitness preserves areas of the brain that are affected in normal aging.

Comparing the results to that obtained utilizing the deformation fields generated by DARTEL that are used in DBM analysis, the DBM visualizations show distortion of tissue topology. Figure 16b appears to depict enlargement of ventricles with low fitness, but normal anatomic landmarks are not easily visualized, including the interface between gray and white matter. Additionally, texture variations are not well-assessed.

The analysis is performed on gray matter and white matter maps separately as well in order to compare the performance of TBM with that of VBM. Figure 18 and 17 show the results when transport-based morphometry is performed on white matter maps and gray matter maps individually.

The interface between the groups is visualized using TBM, which shows loss of temporal lobe gray matter with low fitness in Figure 17b. White matter changes visualized by TBM shows loss of frontotemporal white matter with low fitness and enlarging ventricles in Figure 18b. The pattern of changes in brain tissue distribution seen is similar to that seen in aging.

For VBM analysis, the voxels were compared individually to identify those which had significant differences in intensity across the tissue density maps. The heat maps illustrating voxelwise differences are identified in Figure 19. The clusters here are uncorrected for

multiple comparisons, but significance was selected at level $p < 0.01$. The clusters identified by VBM appear to be spatially distributed across the entire brain. There are changes identified both in occipital and frontal regions of gray matter, as well as in the periventricular white matter affecting frontal regions predominantly. Interestingly, these are some of the same regions identified to be affected by fitness in Figure 15b. However, the global shifts in tissue profile such as atrophy are not captured or well-indexed by a voxelwise analysis, which is better suited for localizing to clusters.

Therefore, while VBM is better suited to localize changes to specific clusters or anatomic regions, and DBM does not adequately assess tissue topology, transport-based morphometry is able to more fully assess the role of fitness on brain health on tissue distribution later in life to generate direct visualizations of the interface between the two classes.

6.4. Visualizing principal phenotypic variations using TBM for unsupervised learning

Finally, TBM can be used to visualize the top three PCA directions generated in transport space using TBM to gain a sense of the principal modes of variation in the dataset, which show variations in brain size, level of tissue atrophy, and prominence of midbrain structures shown by Figure 20.

7. Discussion

We demonstrate a fully automated technique for MRI analysis that facilitates discovery of structural shifts associated with observable phenotypes called transport-based morphometry (TBM). The results confirm our hypothesis that designing a TBM framework suitable for analysis of radiology images can facilitate tasks of regression, discrimination, and signal separation in the transform domain. Our approach is able to assess and visualize aging-related morphologic changes in a fully automated manner. The changes discovered independently by TBM match those that are well-accepted clinically. Additionally, our technique is able to investigate morphological differences between high-fitness and low-fitness groups and yield physically meaningful visualization of the interface between the two classes, suggesting a new mechanism by which fitness may mediate brain health later in life. Finally, transport-based morphometry is shown to enable signal separation by allowing visualization of biologically interpretable principal phenotypic variations.

Traditional methods for assessing structural correlates in neuroimages, such as those that utilize numerical descriptors or pixelwise comparison are able to test only a subset of the information available. Deformation-based morphometry computes local volume expansion/contraction in terms of deformation fields [11], but cannot quantify differences in texture. We see that DBM is able to identify volume expansions, but the deformation fields lose information about tissue topology, distorting the texture of normal landmarks in the image. Voxel-based morphometry identifies voxels on a fixed grid, but is primarily designed to characterize regionally specific changes [11]. However, VBM has difficulty in assessing nonlinear or spatially diffuse changes such as atrophy or tissue thinning, as has been previously reported [11]. Thus, results obtained using deformation-based analysis are influenced by limitations of the method, which confound biological insights. Furthermore, in registration-based methods, it is a well-described challenge that transformation parameters

that result in feature similarity may not result in a correspondence that is physically meaningful [32]. The authors of VBM and DBM state that unless the generative model is invertible, the parameters generated for analysis have no physical meaning [11].

Transport-based morphometry is an invertible transformation that allows generative modeling of the shifts in morphologic distribution underlying discrimination and regression models. TBM analyzes tissue spatial distribution in the transform domain, where we see that even linear regression and discrimination techniques in the transform domain are sufficient to assess and visualize changes in tissue distribution that are nonlinear, spatially diffuse, and affect various regions of the brain in unequal ways. There are several reasons why transforming images to transport space enhances a range of pattern analysis tasks. First, optimal mass transport provides a metric by which to compare nonlinear signals through morphing rather than registration whereby distances between images in the image domain can be modeled in terms of geodesics on a Riemannian manifold in transport space [17, 13, 19]. By projecting these geodesics locally to the tangent space, linearized versions of these metrics are available and as we see in this paper, Euclidean models in the transport space can capture a range of nonlinear morphologic changes. Second, the optimal mass preserving mapping has *a unique minimizer* with a bijective relationship with the source image with respect to a template. Therefore, in addition to enabling complex relationships to be more easily modeled, transport-based morphometry is a generative technique. An observable datapoint can be generated from any arbitrary point in the manifold. Because TBM is generative, a linear model can be directly visualized as a series of physically interpretable images through inverse TBM transformation [19]. In contrast, VBM and DBM are not generative methods. We see indeed that compared to DBM and VBM, transport-based morphometry provides enhanced insight into the morphologic shifts in the data. Using the TBM technique, we are able to adequately assess both global atrophy and local frontotemporal thinning with aging. In contrast, DBM was able to depict only local volume contraction and VBM was able to localize to individual voxels undergoing density changes. Furthermore, compared to VBM and DBM, transport-based morphometry coupled with discriminant analysis revealed a possible mechanism by which aerobic fitness mediates brain health later in life. By analyzing tissue spatial distribution using OMT, TBM can capture important phenomena that is not considered by existing techniques. Finally, our formulation and TBM solver are fully general to any image modality and encompasses a wide range of problems in regression, discrimination, and unsupervised learning. Thus, our approach opens the door to numerous research and clinical advances.

There are several limitations of this work. First, our approach for optimal transport minimization is non-convex. Although the approach does not guarantee theoretically that global minima will be achieved, the experimental results demonstrate that the multiscale scheme guides the minimization to the global minima and the results are comparable to those using convex formulations in 2D. We pose it as a future problem to couple the TBM framework presented in this paper with solvers that can overcome limitations with large 3D images and at the same time are convex. Another limitation is that analyzing the spatial distribution of voxels requires a normalization of images. Thus, the TBM transform does not directly consider whether there are statistically significant differences in the sum of voxel

intensities. However, the latter limitation is easily remedied, as the sum of voxel intensities can be included as a feature when statistical analyses are performed in the feature domain.

8. Conclusion

In conclusion, we presented a novel image transformation framework for MRI data to losslessly facilitate discovery of trends as well as yield biologically interpretable visualization of the morphologic changes associated with a variety of clinical outcomes. We demonstrate that our fully automated approach facilitates regression, discrimination, and blind signal separation with significant advancement over currently used techniques. Our approach is able to independently discover aging-related changes that are well-corroborated clinically and provide new insight into the effects of fitness on the brain, unlike traditional methods. The results validate that our approach can be used as a statistical learning tool in diseases for which gene-structure-behavior relationships are not well-known.

Acknowledgments

Appendix A: Derivation of Euler-Lagrange Equation

Here we present the derivation of the Euler-Lagrange equation in (13). Starting from the objective function in Eq. (10) we have

$$M(f) = \frac{1}{2}\left\|det(Df)I_1(f) - I_0\right\|^2 + \frac{\lambda}{2}\|\nabla \times f\|^2 \qquad (18)$$

$$= \underbrace{\frac{1}{2}\int_\Omega (det(Df(x))I_1(f(x)) - I_0(x))^2 dx}_{M_1(f)} + \underbrace{\frac{\lambda}{2}\int_\Omega |\nabla \times f(x)|^2 dx}_{M_2(f)}$$

where the first term, $M_1(f)$, enforces f to be mass preserving while the second term, $M_2(f)$, enforces f to be curl free. Starting with the first term we can write the Euler-Lagrange equation as,

$$\frac{dM_1}{df^i} = \frac{\partial \mathscr{L}_1}{\partial f^i} - \sum_{k=1}^{n} \frac{d}{dx^k}\left(\frac{\partial \mathscr{L}_1}{\partial f^i_{x^k}}\right), \quad i = 1, \ldots, n \quad (19)$$

where we have,

$$\frac{\partial \mathscr{L}_1}{\partial f^i} = det(Df)\frac{\partial I_1(f)}{\partial f^i}I_{error} \quad (20)$$

Let C be the cofactor matrix of Df. Then $det(Df)$ can be written as the sum of the cofactors of any columns or rows of Df,

$$det(Df) = \sum_{i=1}^{n} f^i_{x^j}C_{i,j}, \quad \forall j \in \left\{1, ..., n\right\} \quad (21)$$

$$= \sum_{j=1}^{n} f^i_{x^j}C_{i,j}, \quad \forall i \in \left\{1, ..., n\right\}$$

Using the cofactor matrix, C, we can write,

$$\frac{\partial \mathscr{L}_1}{\partial f^i_{x^k}} = C_{i,k}I_1(f)I_{error} \quad (22)$$

And from Equations (20) and (22) we have,

$$\frac{dM_1}{df^i} = I_{error}(det(Df)\frac{\partial I_1(f)}{\partial f^i} - \sum_{k=1}^{n}\frac{d}{dx^k}C_{i,k}I_1(f)) \quad (23)$$

and writing the vector form of the above equation for all i and using $C^T = adj(Df)$ we can write,

$$\frac{dM_1}{df} = I_{error}(det(Df)\nabla I_1(f) - \nabla \cdot (adj(Df)I_1(f))). \quad (24)$$

For the second term, we have

$$\frac{\partial \mathscr{L}_1}{\partial f^i} = 0 \quad (25)$$

Furthermore, assuming that $n = 2, 3$ and using the Levi-Civita symbol we can write the norm squared of the curl of f as follows,

$$|\nabla \times f|^2 = \sum_{p=1}^{n}(\sum_{l=1}^{n}\sum_{m=1}^{n}\varepsilon^{plm}f^m_{x^l})^2 \quad (26)$$

which leads to,

$$\frac{\partial \mathscr{L}_1}{\partial f^i_{x_k}} = \lambda \sum_{p=1}^{n} \varepsilon^{pki} (\sum_{l=1}^{n} \sum_{m=1}^{n} \varepsilon^{plm} f^m_{x_l}). \quad (27)$$

Therefore we have,

$$\frac{dM}{df^i} = -\lambda \sum_{k=1}^{n} \frac{d}{dx^k} (\sum_{p=1}^{n} \varepsilon^{pki} (\sum_{l=1}^{n} \sum_{m=1}^{n} \varepsilon^{plm} f^m_{x_l})) \quad (28)$$

$$= -\lambda \sum_{k=1}^{n} \sum_{p=1}^{n} \varepsilon^{pki} \frac{d}{dx^k} (\sum_{l=1}^{n} \sum_{m=1}^{n} \varepsilon^{plm} f^m_{x_l})$$

$$= \lambda \sum_{k=1}^{n} \sum_{p=1}^{n} \varepsilon^{ikp} \frac{d}{dx^k} (\sum_{l=1}^{n} \sum_{m=1}^{n} \varepsilon^{plm} f^m_{x_l})$$

$$= \lambda (\nabla \times \nabla \times f)^i$$

Finally, combining Equations (24) and (28) will lead to,

$$\frac{dM}{df} = I_{error} (det(Df) \nabla I_1(f) - \nabla \cdot (adj(D_f) I_1(f))) + \lambda \nabla \times \nabla \times f \quad (29)$$

Appendix B: Validating OMT registration on MRI Data

8.1. Optimal transport minimization

The equations for analysis and synthesis can be solved in closed form only for 1D signals [33], but for higher-dimensional signals, they must be solved using optimization techniques. Many solvers have been described in the OMT literature, although special challenges including drift, artifact, and computational time/complexity arise in numerical OMT of large image sets that may exceed millions of voxels.

For example, Haker *et al.*[34] solve for an initial MP map f_0 (not unique or optimal) through the Knothe-Rosenblatt rearrangement [33, 35] and then progressively update the initial map using composition (with another MP map *s* that satisfies $s_\# \sigma = \sigma$) so that it becomes curl free to signify optimality. As pointed out by Haber *et al.*[36] and Rehman *et al.*[37], however, there exist two main shortcomings to the preceding numerical approaches. First, a robust method is needed to obtain an initial MP mapping, and the obtained initial map is often far from the optimal transport map. Second, and much more importantly, such methods update the transformation in a space which is tangential to the linearized MP constraint. Hence, for any finite step update used in the optimization, $f_0(s^k)$, the mapping deviates or *drifts* from the set of mass preserving mappings *MP*. While the level of *drift* may or may not be acceptable in practice for a 2D solution, the drifting is amplified for 3D images as

demonstrated in Section 6. The *drift* phenomenon necessitates solution by alternative methods for 3D images.

Convergent methods have also been proposed based on a fluid dynamics formulation of the problem [38, 39] or based on the solution of the Monge-Ampère equation [40], but these formulations come at the cost of an additional virtual time dimension, which is computationally expensive. Computational cost also becomes a challenge in approaches utilizing a system of linear equations that arise from the finite-difference implementation of the linearized Monge-Ampere equation [41, 42].

Another family of solvers [43, 44, 29] are based on Kantorovich's formulation of the problem. In short, Kantorovich's formulation searches for the optimal transport plan π defined on $\Omega \times \Omega$ with marginals μ and σ that minimizes the following,

$$\min_{\pi \in \prod(\mu, \sigma)} \int_{\Omega \times \Omega} c(x, y) d\pi(x, y) \quad (30)$$

Here $\Pi(\mu, \sigma)$ is the set of all transport plans with marginals μ and σ. Chartrand *et al* [29] solves the dual problem to the Kantorovich formulation. Chartrand *et al* obtain the optimal transport map through a gradient descent solution. The obtained transport map as pointed out in [29] and shown in this paper see Figure 22 comes with the trade off of undesired *artifacts*, especially when the images are not smooth. Thus, additional work is needed to overcome challenges related to quality of MP match with the Chartrand *et al.*[29] approach.

In summary, in order to test the hypothesis that transport-based morphometry can both extract discriminant information and produce visualization of differences on MRI data, an OMT solution is required that can overcome computational challenges for large 3D data.

We validated our OMT approach on healthy, adult brain images obtained from the IXI dataset, Biomedical Image Analysis Group at the Imperial College in London [45].

10 images were selected at random from Guy's Hospital in UK. Subjects were male and ranged from 41 to 86 years of age at the time of imaging (mean age 57.8 years, standard deviation 15.7 years). The images were T1-weighted images, obtained using a Philips Medical Systems Intera 1.5 T scanner, with the following imaging parameters. Repetition time = 9.813 ms, echo time = 4.603 ms, number of phase encoding steps = 192, echo train length = 0, reconstruction diameter = 240, flip angle = 8 °. The images are $128 \times 128 \times 128$ matrix, with 1 mm^3 resolution.

8.1.1. Optimization parameters—The step size for accelerated gradient descent is chosen such that the maximum displacement per update is 0.01 of a pixel, the same for our method and comparison methods. At every step of gradient descent, there is a check to maintain that the mapping is diffeomorphic and the step size is reduced as necessary in order to ensure a diffeomorphic mapping. The parameters were obtained experimentally: $\lambda = 100$, $\gamma = 6.5 \times 10^4$ when the MSE reaches 25% of the initial MSE to steer the solution towards a curl-free MP mapping, number of scales = 3. The multiscale approach was also implemented

for the two other OMT methods in this paper, although the results did not significantly change when the scheme was used.

The termination criteria for all methods implemented is when MSE of the morphed source image relative to the template image reaches 0.55%. When the drifting phenomena in the Haker method [46] produces MSE 100× the initial MSE, we terminate the code. We report the mean L_2 norm of the curl per voxel and MSE relative to the template.

We use a numerical discretization scheme in which values are placed at pixel or voxel centers. A consistent second-order finite difference approximation was used for all differential operators, utilizing the DGradient toolbox for MATLAB [47].

8.2. Experiment: Validating OMT registration on MRI

We compare the approach of Haker *et al* [46] and Chartrand *et al* [29] to the approach described in this paper. The computational complexity of the gradient descent update step for all three methods implemented is $\mathcal{O}(NlogN)$. All methods implemented utilized the same preprocessed images.

All unique pairs of images were registered to each other for the 10 images, resulting in 45 total registration problems. The statistics reported in this paper are based on the registrations performed in turn with our method, that of Chartrand *et al*, and that of Haker *et al*.

The following three experiments investigate TBM for MRI pattern analysis enabled by our OMT approach.

8.3. Comparing MP registration methods

We report results for both 2D and 3D MP registration using optimal transport. The 2D image dataset was derived from the 3D dataset by extracting the same axial slice from the middle of every 3D brain image. The solver and accelerated gradient descent update equations are the same whether working with 2D or 3D images. We compare our OMT approach to current methods in the literature based on both the Monge formulation ([46]) and Kantorovich formulation ([29]) to demonstrate that our method is robust to the challenges of other OMT approaches.

8.3.1. 2D optimal mass transport—Table 2 displays the mapping statistics for the set of 2D images. The mass transported is lowest for the Chartrand *et al.*'s method and compares favorably to the mass transport achieved using our method. The MSE for Chartrand *et al* indicates that this method also produces the poorest MP mapping of the three methods. The MSE achieved is 3–4 times that achieved by the other two methods, demonstrating that this method is prone to artifacts.

Our method achieves the lowest MSE in addition to mass transport distance. As reviewed in Section 2, the optimal MP mapping is the MP mapping that achieves minimum mass transport. The curl, a measure of optimality of the MP mapping, is 0 by design for the method of Chartrand *et al.*as expected (see Section 2). The mean curl is the highest for our method compared to the other two methods, but is still small in an absolute sense (10^{-4}).

Our method produces the best results in terms of mass-preservation and mass transport. In Figure 21, we see the optimal transport fields and their corresponding morphings for several 2D brain images. Visually, the transport fields and quality of morphings are similar for all three methods.

The proposed method was prototyped in MATLAB using built-in functions. The average runtime for 256×256 brain images with 3 scales was 19.36 ± 7.91 seconds.

8.3.2. 3D optimal mass transport—While in 2D all three methods seem to produce visually similar OMT mappings, we see in 3D that several phenomena become evident. Examining Figure 23, we compare the plots of curl and relative MSE over gradient descent iterations for several brain images. We see that the magnitude of the curl (on the order of 10^6) is large for the Haker *et al.*'s method. The curl for the Chartrand *et al.*'s method remains at 0 by design. Our method produces curls for all images that tend toward zero with iterations of gradient descent.

We see that relative MSE with the Haker *et al.*'s approach increases significantly until we terminate the code when the MSE reaches 100x its initial value. Hence, starting after around 100 iterations of gradient descent, the phenomenon of *drift* with the Haker *et al.*'s approach becomes evident. The relative MSE of the Chartrand *et al.*'s approach decreases, but remains large in magnitude (5–10%) at termination. The large MSE results in visual artifacts in the quality of the MP match, which we can see in Figure 22. In contrast, for our method, all images are able to achieve the 0.55% termination criterion.

Table 2 corroborates the plots in Figure 23. Our method produces the lowest relative MSE (best MP mapping), and all brain images are able to achieve the termination criterion of 0.55%. Furthermore, our curl at termination is 8 orders of magnitude lower than that obtained using the Haker method.

In terms of mass transported, the Chartrand *et al.*'s method produces the lowest transport distance, although the MSE of the MP mapping is about 5–10x higher than that achieved using our method. We can also see artifacts visually in the mappings produced by the Chartrand *et al.*'s method compared to our method (Figure 22).

In Figure 22, we compare axial, sagittal and coronal slices mapped using our method and that of Chartrand *et al.*'s (The method of Haker *et al.* failed to produce a viable solution, which is why it is not shown.) We see that mappings produced by our method result in visually similar images to the target image I_0, whereas those produced by the Chartrand *et al.*'s method contain several artifacts.

Overall, our method outperforms both comparison methods for 3D images. Our method achieves the lowest MP mapping, while at the same time achieving small curl and mass transported.

The median runtime was under 20 minutes per brain in MATLAB using built-in libraries on a general purpose computer. There is significant opportunity for improvement with an implementation in native C.

Thus, we see that our approach is able to overcome traditional limitations of drift, artifact, and impractical computational complexity. Our approach enables the goal of pattern analysis on MRI using a transport-based morphometry approach.

Biographies

Shinjini Kundu received her B.S. and M.S. degrees in electrical engineering from Stanford University and her Ph.D. in biomedical engineering from Carnegie Mellon University. She is part of the Medical Scientist Training Program (MSTP) at the University of Pittsburgh, where she is currently pursuing her M.D. degree. Her research interests are in pattern recognition for biomedical images, magnetic resonance imaging, and computer-aided detection for radiology.

Soheil Kolouri received his B.S. degree in electrical engineering from Sharif University of Technology, Tehran, Iran, in 2010, and his M.S. degree in electrical engineering in 2012 from Colorado State University, Fort Collins, Colorado. He received his doctorate degree in biomedical engineering from Carnegie Mellon University in 2015, were his research was focused on applications of the optimal transport in image modeling, computer vision, and pattern recognition. His thesis, titled, Transport-based pattern recognition and image modeling, won the best thesis award from the Biomedical Engineering Department at Carnegie Mellon University.

Kirk Erickson received his B.S. degree in Psychology and Philosophy in 1999 from Marquette University. In 2005 he received his Ph.D. from the University of Illinois at Urbana-Champaign and was a post-doc at the Beckman Institute for Advanced Science and Technology at the University of Illinois until 2008. He is currently an Associate Professor of Psychology and the Center for the Neural Basis of Cognition at the University of Pittsburgh.

Arthur Kramer is the Director of the Beckman Institute for Advanced Science & Technology and the Swanlund Chair and Professor of Psychology and Neuroscience at the University of Illinois. A major focus of his labs recent research is the understanding and enhancement of cognitive and neural plasticity across the lifespan. Professor Kramer is a fellow of the American Psychological Association, American Psychological Society, and a recipient of a NIH Ten Year MERIT Award. Professor Kramers research has been featured in a long list of print, radio and electronic media including the New York Times, Wall Street Journal, Washington Post, Chicago Tribune, CBS Evening News, Today Show, NPR and Saturday Night Live.

Edward McAuley is a Shahid and Ann Carlson Khan Endowed Professor of Applied Health Sciences at the University of Illinois at Urbana-Champaign. He holds appointments in the Departments of Kinesiology and Community Health, Psychology, Internal Medicine, and the Beckman Institute for Advanced Science and Technology. He the director of the Exercise Psychology Laboratory at Illinois and has published over 380 articles and chapters. He has served as the Chair of the Psychosocial Risk and Disease Prevention study section of the National Institutes of Health. He is an elected fellow of the Society of Behavioral Medicine and the Gerontological Society of America. His research agenda has focused primarily on

physical activity, aging, and well-being in healthy adults and breast cancer survivors and the role played by exercise training in neurocognitive function, brain health, and psychological well-being.

Gustavo K. Rohde earned B.S. degrees in physics and mathematics in 1999, and the M.S. degree in electrical engineering in 2001 from Vanderbilt University. He received a doctorate in applied mathematics and scientific computation in 2005 from the University of Maryland. He is currently an associate professor of Biomedical Engineering and Electrical and Computer Engineering at University of Virginia.

References

1. Erickson KI, Voss MW, Prakash RS, Basak C, Szabo A, Chaddock L, Kim JS, Heo S, Alves H, White SM, et al. Proceedings of the National Academy of Sciences. 2011; 108:3017–3022.

2. Colcombe SJ, Erickson KI, Raz N, Webb AG, Cohen NJ, McAuley E, Kramer AF. The Journals of Gerontology Series A: Biological Sciences and Medical Sciences. 2003; 58:M176–M180.

3. Colcombe SJ, Erickson KI, Scalf PE, Kim JS, Prakash R, McAuley E, Elavsky S, Marquez DX, Hu L, Kramer AF. The Journals of Gerontology Series A: Biological Sciences and Medical Sciences. 2006; 61:1166–1170.

4. Shamir L, Orlov N, Eckley DM, Macura T, Johnston J, Goldberg IG. Source code for biology and medicine. 2008; 3:1. [PubMed: 18241331]

5. Fischl B, Dale AM. Proceedings of the National Academy of Sciences. 2000; 97:11050–11055.

6. Ashburner J, Hutton C, Frackowiak R, Johnsrude I, Price C, Friston K, et al. Human brain mapping. 1998; 6:348–357. [PubMed: 9788071]

7. Ashburner J, Good C, Friston KJ. NeuroImage. 2000; 11:S465.

8. Ashburner J, Friston KJ. Neuroimage. 2000; 11:805–821. [PubMed: 10860804]

9. Bookstein FL. Neuroimage. 2001; 14:1454–1462. [PubMed: 11707101]

10. Salmond C, Ashburner J, Vargha-Khadem F, Connelly A, Gadian D, Friston K. Neuroimage. 2002; 17:1027–1030. [PubMed: 12377176]

11. Friston K, Ashburner J. Neuroimage. 2004; 23:21–24. [PubMed: 15325348]

12. Ashburner J. Neuroimage. 2007; 38:95–113. [PubMed: 17761438]

13. Wang W, Slepcev D, Basu S, Ozolek JA, Rohde GK. International journal of computer vision. 2013; 101:254–269. [PubMed: 23729991]

14. Wang W, Ozolek JA, Slepcev D, Lee AB, Chen C, Rohde GK. Medical Imaging, IEEE Transactions on. 2011; 30:621–631.

15. Basu S, Kolouri S, Rohde GK. Proceedings of the National Academy of Sciences. 2014; 111:3448–3453.

16. Park SR, Kolouri S, Kundu S, Rohde GK. Applied and Computational Harmonic Analysis. 2017

17. Kolouri S, Tosun AB, Ozolek JA, Rohde GK. Pattern recognition. 2016; 51:453–462. [PubMed: 26858466]

18. Kolouri S, Park SR, Rohde GK. IEEE transactions on image processing. 2016; 25:920–934. [PubMed: 26685245]

19. Kolouri S, Park SR, Thorpe M, Slepcev D, Rohde GK. IEEE Signal Processing Magazine. 2017; 34:43–59.

20. Jeffery DR. Infections in medicine. 2002; 19:73–79.

21. Carlson, N. Fluorescence guided resection. 2011. https://engineering.dartmouth.edu/brainidb

22. Brenier Y. Communications on pure and applied mathematics. 1991; 44:375–417.

23. Nesterov Y, et al. Gradient methods for minimizing composite objective function. Technical Report, UCL. 2007

24. Wang W, Mo Y, Ozolek JA, Rohde GK. Pattern recognition letters. 2011; 32:2128–2135. [PubMed: 22140290]

25. Cootes TF, Taylor CJ, Cooper DH, Graham J. Computer vision and image understanding. 1995; 61:38–59.

26. Erickson KI, Prakash RS, Voss MW, Chaddock L, Hu L, Morris KS, White SM, Wójcicki TR, McAuley E, Kramer AF. Hippocampus. 2009; 19:1030–1039. [PubMed: 19123237]

27. W. T. C. for Neuroimaging. Spm12. 2016. URL: http://www.fil.ion.ucl.ac.uk/spm/software/spm12/

28. Dora L, Agrawal S, Panda R, Abraham A. IEEE Reviews in Biomedical Engineering. 2017

29. Chartrand R, Vixie K, Wohlberg B, Bollt E. Applied Mathematical Sciences. 2009; 3:1071–1080.

30. Mechelli A, Price CJ, Friston KJ, Ashburner J. Current medical imaging reviews. 2005; 1:105–113.

31. Brody, H. The regulatory role of the nervous system in aging. Karger Publishers; 1970. p. 9-21.

32. Crum WR, Hartkens T, Hill D. The British journal of radiology. 2004; 77:S140–S153. [PubMed: 15677356]

33. Villani, C. Optimal transport: old and new. Vol. 338. Springer Science & Business Media; 2008.

34. Haker S, Zhu L, Tannenbaum A, Angenent S. International Journal of Computer Vision. 2004; 60:225–240.

35. Bonnotte N. SIAM Journal on Mathematical Analysis. 2013; 45:64–87.

36. Haber E, Rehman T, Tannenbaum A. SIAM Journal on Scientific Computing. 2010; 32:197–211. [PubMed: 21278828]

37. ur Rehman T, Haber E, Pryor G, Melonakos J, Tannenbaum A. Medical image analysis. 2009; 13:931–940. [PubMed: 19135403]

38. Benamou JD, Brenier Y. Numerische Mathematik. 2000; 84:375–393.

39. Maas J, Rumpf M, Schönlieb C, Simon S. arXiv preprint arXiv:1504.01988. 2015

40. Benamou JD, Froese BD, Oberman AM. Journal of Computational Physics. 2014; 260:107–126.

41. Saumier LP, Agueh M, Khouider B. IMA Journal of Applied Mathematics. 2015; 80:135–157.

42. Oberman AM, Ruan Y. arXiv preprint arXiv:1509.03668. 2015

43. Solomon J, de Goes F, Studios PA, Peyré G, Cuturi M, Butscher A, Nguyen A, Du T, Guibas L. ACM Transactions on Graphics (Proc SIGGRAPH 2015), to appear. 2015

44. Cuturi M. Advances in Neural Information Processing Systems. :2292–2300.

45. Imperial College London. http://brain-development.org/ixi-dataset/, accessed 3-21-16

46. Haker S, Zhu L, Tannenbaum A, Angenent S. International Journal of Computer Vision. 2004; 60:225–240.

47. Mathworks File Exchange, Dgradient. http://www.mathworks.com/matlabcentral/fileexchange/29887-dgradient.2016
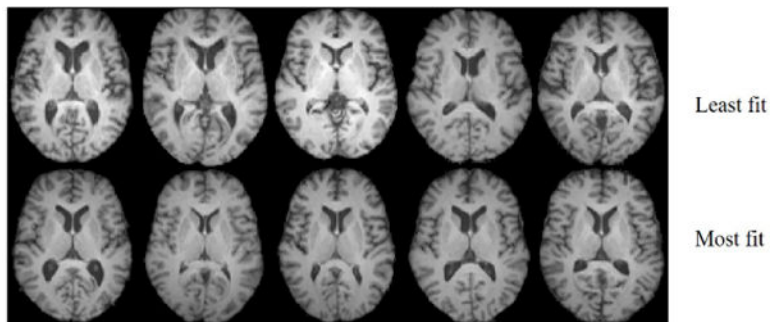
Figure 1.
MR images belonging to 10 older adults in their 6*th* or 7*th* decades of life. The corresponding axial slice is depicted for each subject for comparison. The dataset from which the images are drawn is described further in 5.1.1. The images correspond to subjects who are $+\sigma$ above the population mean aerobic fitness (most fit) or $-\sigma$ below the mean aerobic fitness (least fit) as assessed by vO_2 L/min. Prior groups have reported differences in quantitative tissue density [2], brain volume [3], as well as hippocampal volume [1] as a function of fitness on a subset of images from the same dataset. The goal is to determine whether there is a common morphologic feature that separates these groups, and if so, visualize it in a physically interpretable manner.
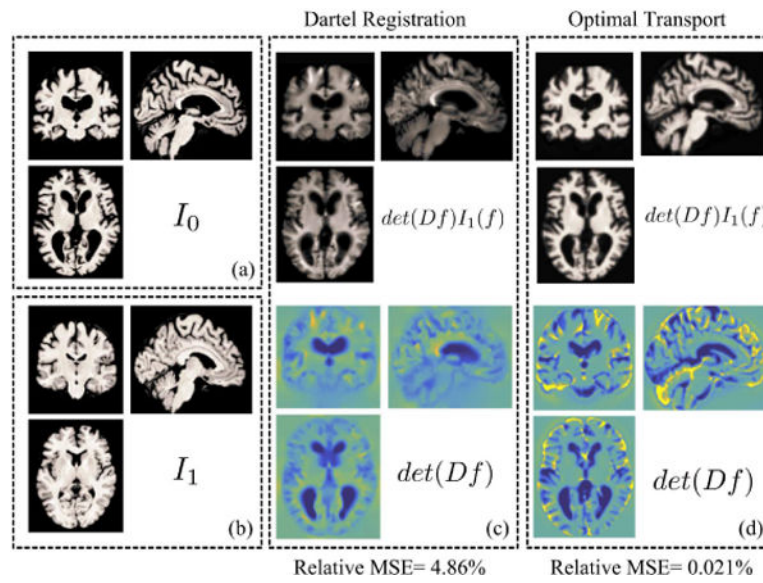
Figure 2.
Compared to deformation fields computed using DARTEL [12], transport maps computed using optimal mass transport (OMT) captures both shape and texture differences between I_0 and I_1 and match images perfectly, up to an interpolation error. The three boxes in the top row should all look the same. However, deformation fields lose texture variation information, thus resulting in high MSE when attempting to match source and target images.
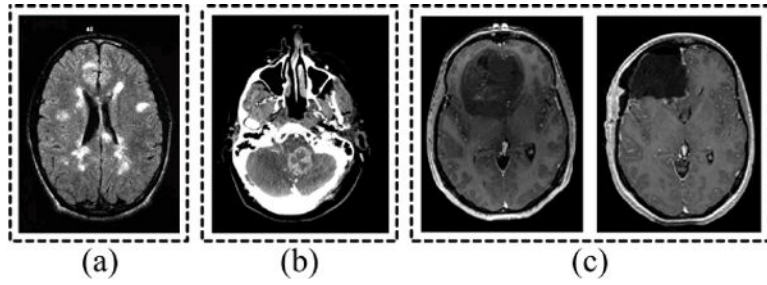
Figure 3.
Neurologic conditions where pathology affects biophysical properties of tissue manifesting as texture variation. (a) Multiple sclerosis (source: [20]), (b) metastatic breast cancer tumor (source: Dept. of Radiology, University of Pittsburgh Medical Center), (c) GBM (left) and debulking procedure (right) (source: [21])

$$d\sigma(x) = I_0(x)dx$$
$$d\mu(x) = I_1(x)dx$$
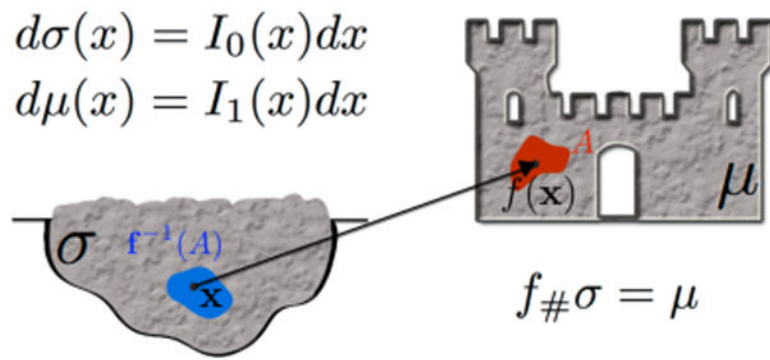
Figure 4.
Given a pile of dirt and a castle of the same total mass, the coupling is sought between units of mass in the pile of dirt and units of mass in the castle that is optimal because it minimizes the transportation cost
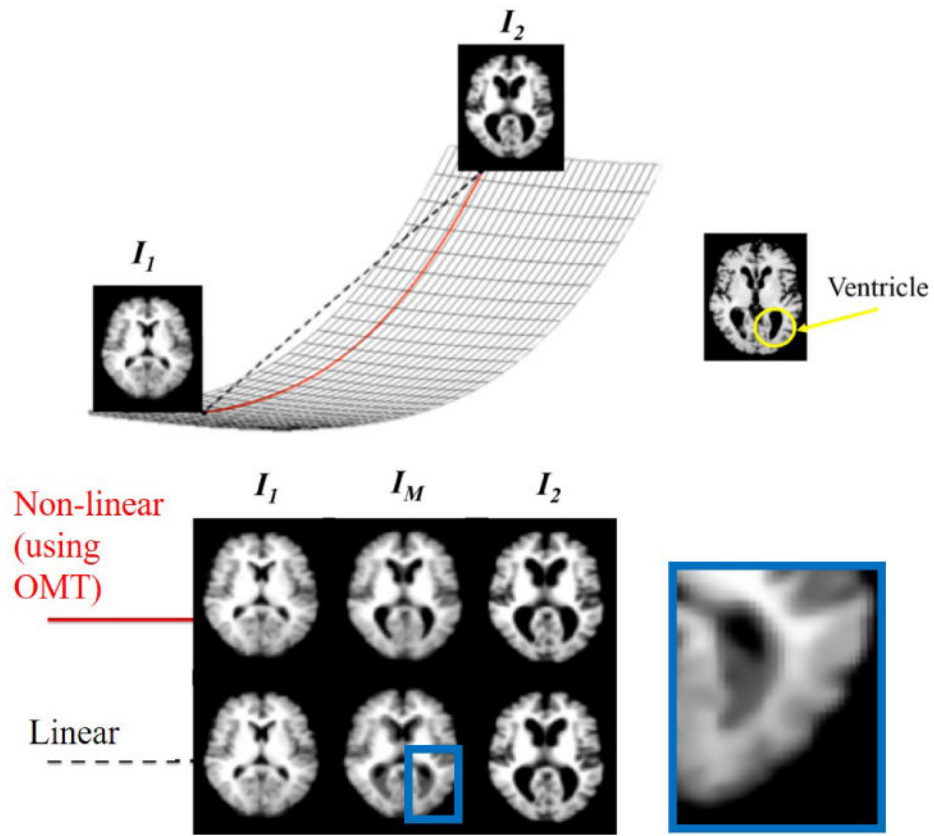
Figure 5.
Compared to the simple Euclidean distance (black), the OMT distance (red) between images defines a nonlinear distance metric between a pair of images, represented by the red arc on the manifold. Compared to simple Euclidean interpolation to estimate the middle image I_M, which leads to artifacts in the ventricles and other artifacts, the nonlinear OMT-based distance appears to better capture the natural structure of the brain.
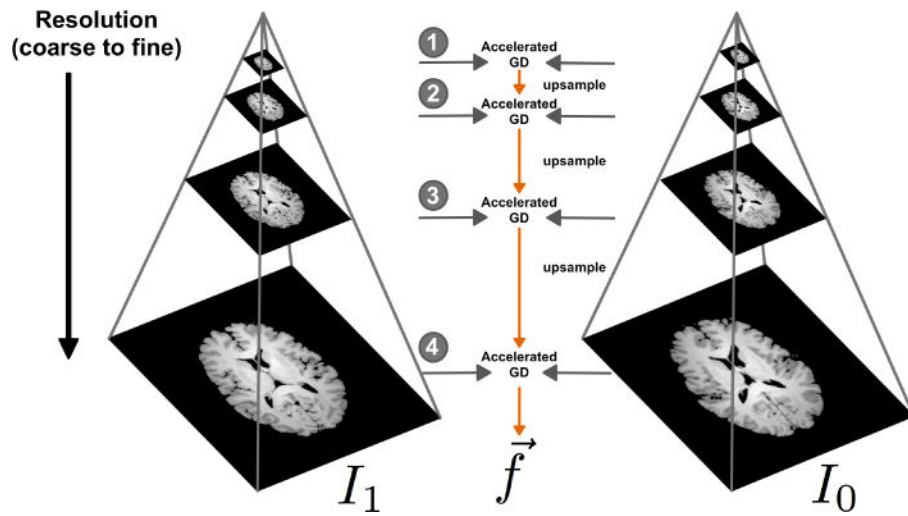
Figure 6.
The schematic of the multiscale approach devised in this paper. The solution to the accelerated gradient descent is first calculated at a coarse level and then refined as the optimization proceeds.
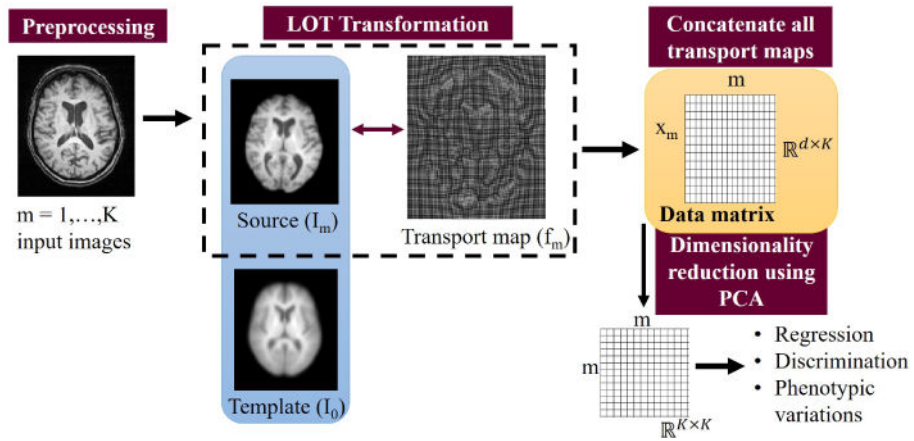
Figure 7.
System diagram. Images are first skull-stripped, intensity normalized and affine registered. A series of transport maps is computed using LOT transformation, and subsequent pattern analysis is performed in the transport space. Inverse LOT transformation provides visualization of the regression, discrimination, or principal component directions in the transport space for physical interpretation.
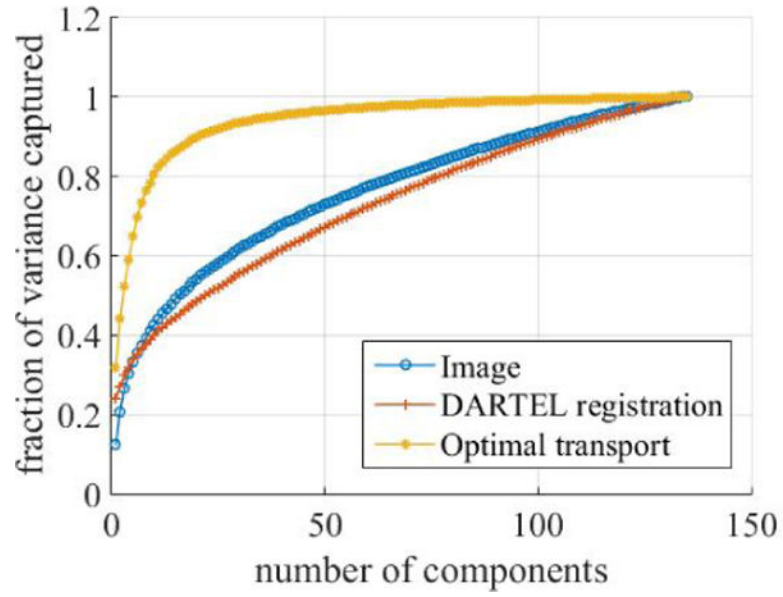
Figure 8.
Compared to the models utilizing pixel-wise comparison (Eulerian and DARTEL registration), the model based on OMT is able to capture more of the variability in the dataset with fewer principal components.
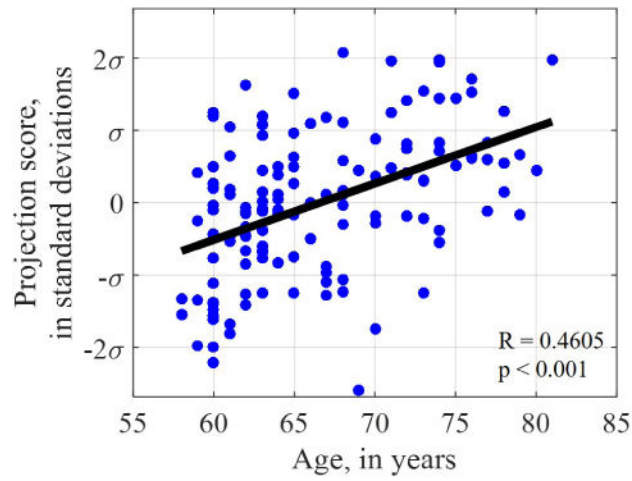
(a)
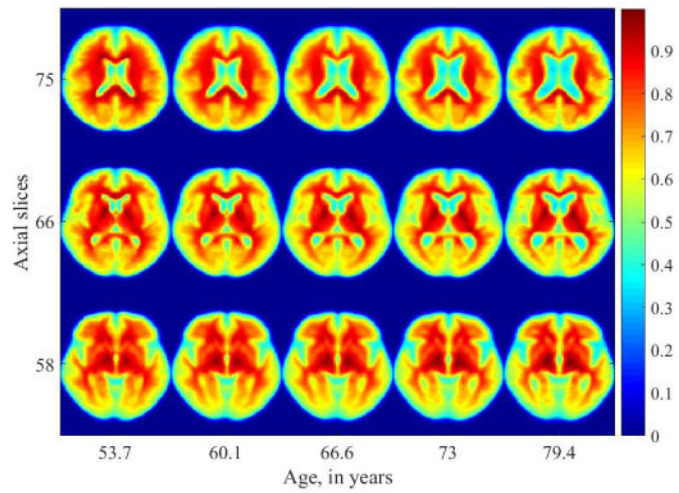
(b)

Figure 9.
(a) Projection of data onto the direction that maximizes linear correlation with age, (b) Visualization of changes in tissue distribution that are statistically dependent on age. The vertical axis shows various axial slices from a 3D dataset from rostral (towards head) to caudal (towards toe). The horizontal axis shows the effect of increasing age from left to right on that axial slice. We see enlarging ventricles, and global atrophy of both gray matter and white matter with increasing age.
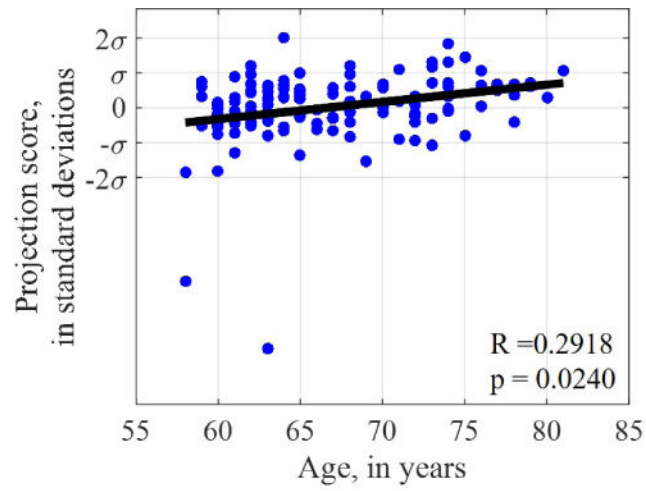
(a)
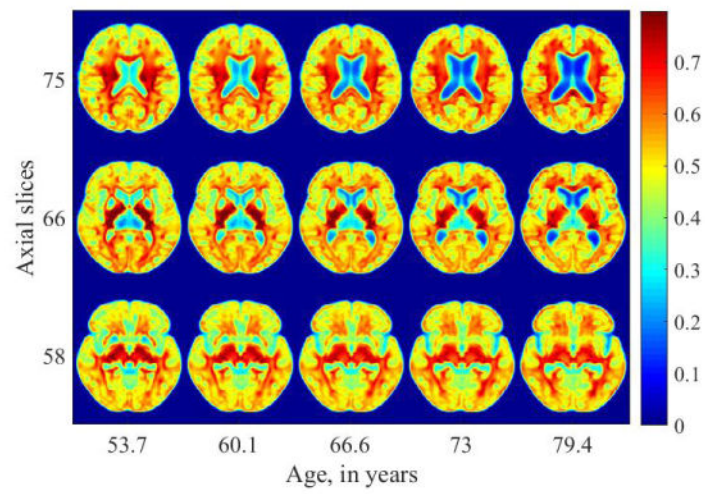
(b)

Figure 10.
Visualization of aging-related changes captured by DARTEL deformation fields used in deformation-based morphometry. Normal tissue texture is not well-modeled using a deformation-based approach.
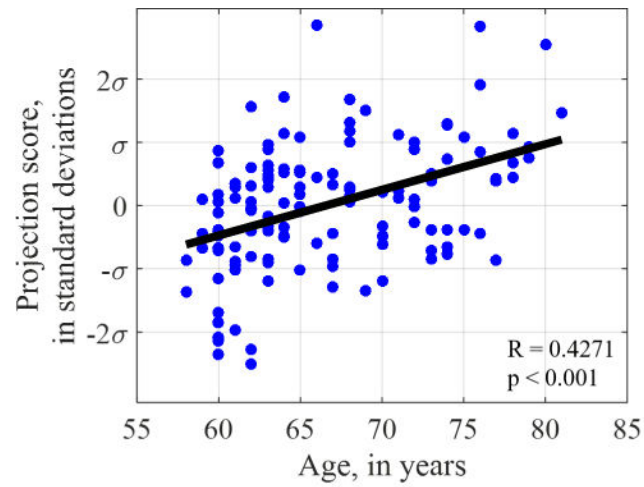
(a)
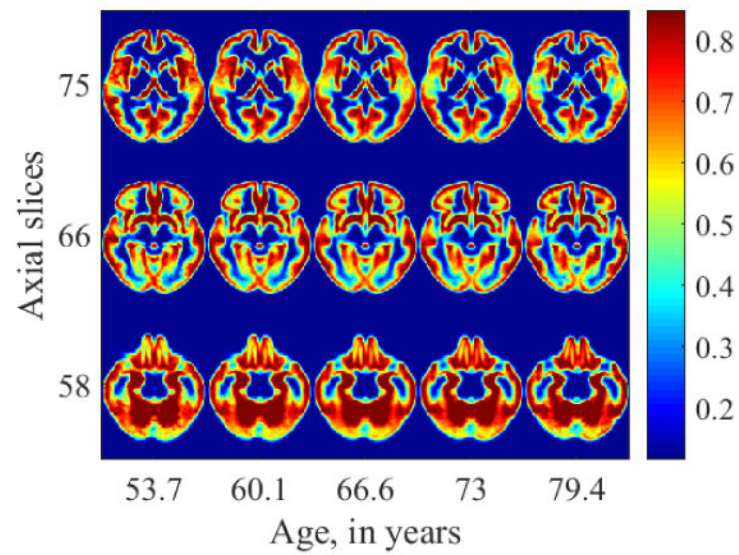
(b)

Figure 11.
Visualization of aging-related changes found by transport-based morphometry on gray matter channels. There is atrophy and loss of GM from temporal lobes.
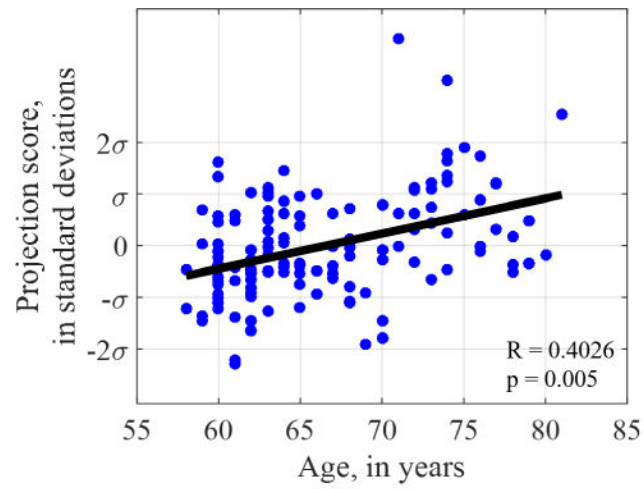
(a)
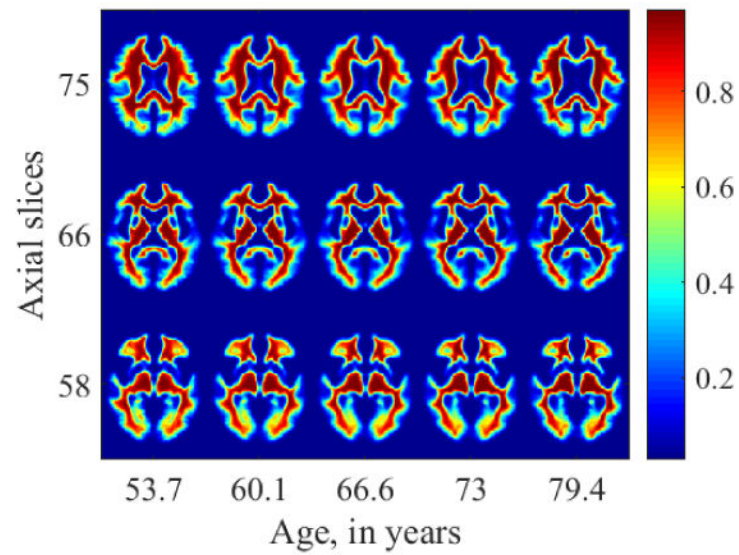
(b)

Figure 12.
Visualization of aging-related changes found by transport-based morphometry on white matter channels. TBM depicts overall atrophy and loss of white matter disproportionately from the frontal lobes.
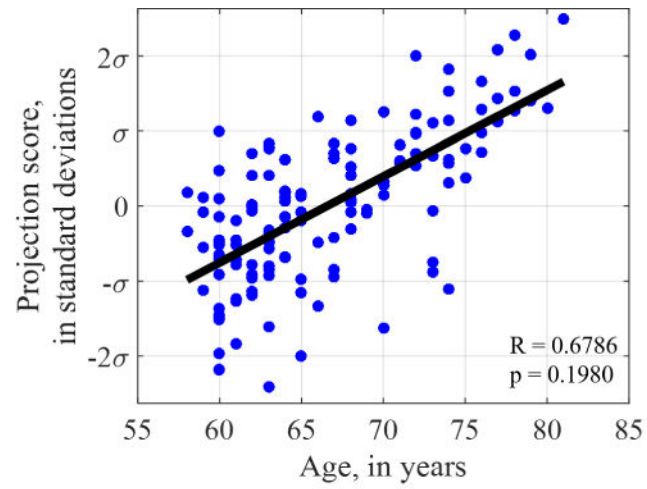
(a)
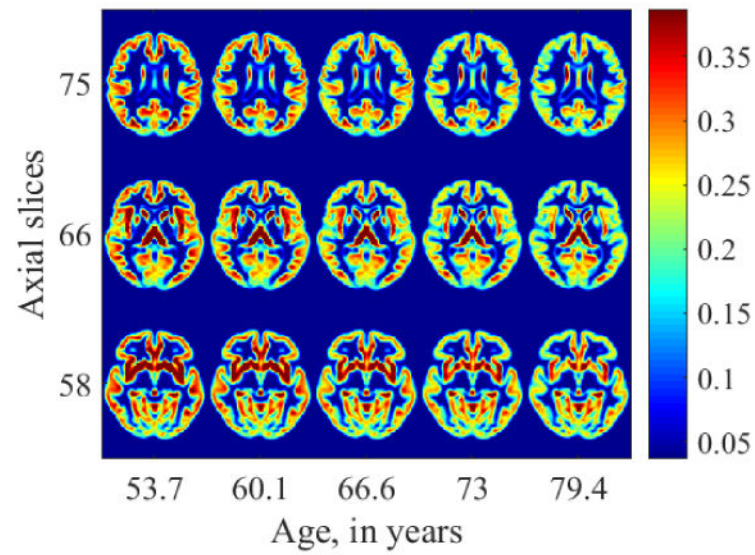
(b)

Figure 13.
Visualization of aging-related changes found by voxel-based morphometry on gray matter maps.
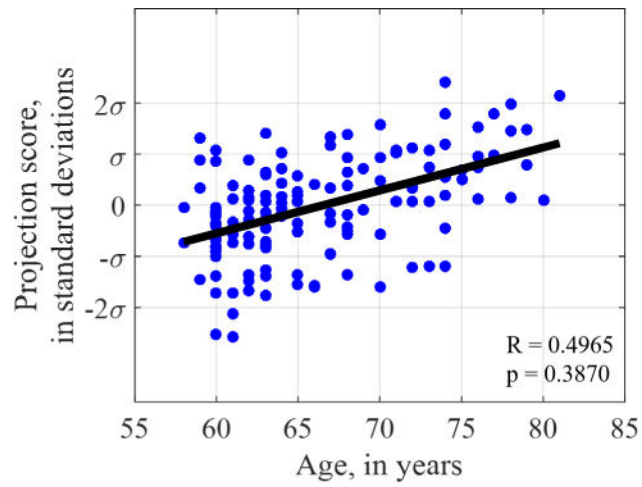
(a)
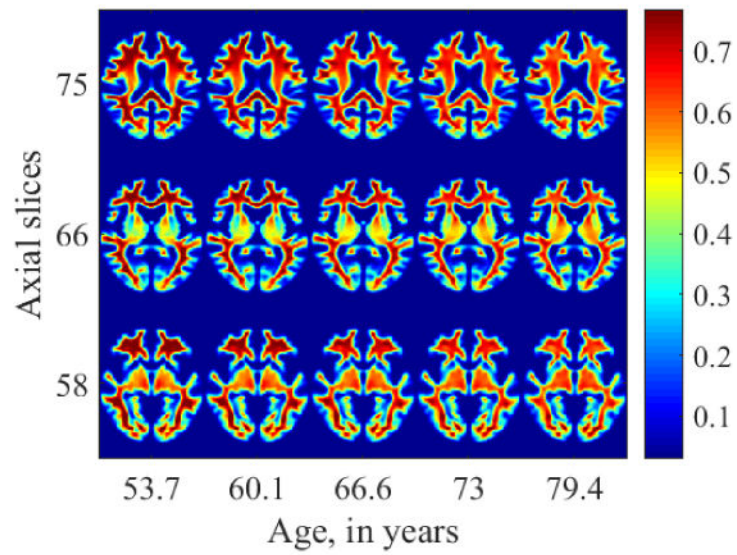
(b)

Figure 14.
Visualization of aging-related changes found by voxel-based morphometry on white matter maps.
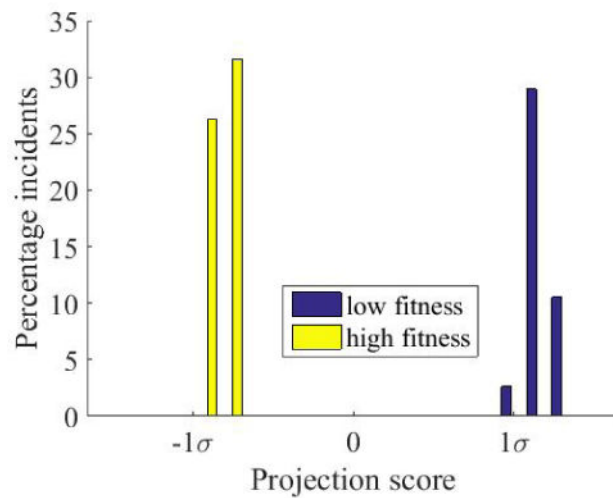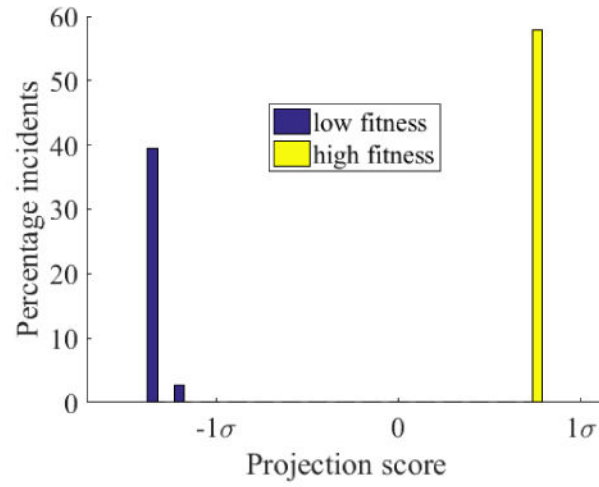
(a)
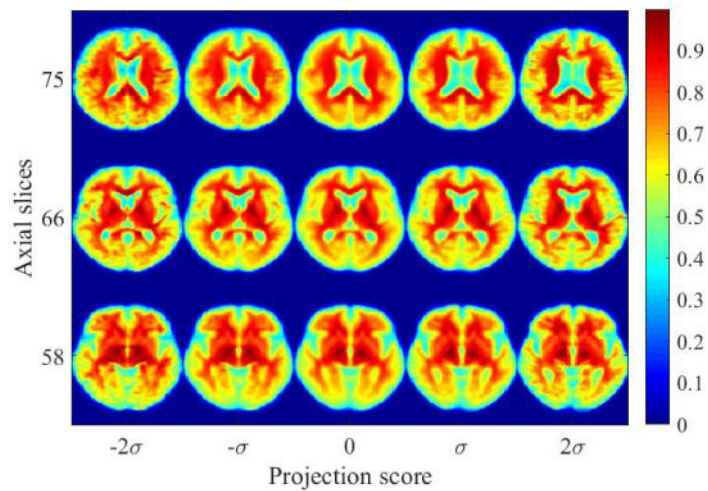
(b)

Figure 15.
(a) High-fit and low-fit individuals can be perfectly separated based on their transport maps given by TBM when projected onto the most discriminant direction computed by PLDA, (b) images illustrating the differences between high-fit and low-fit individuals generated based on transport maps.
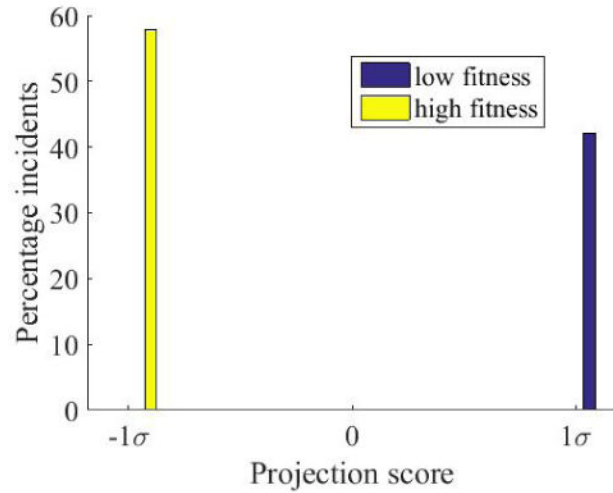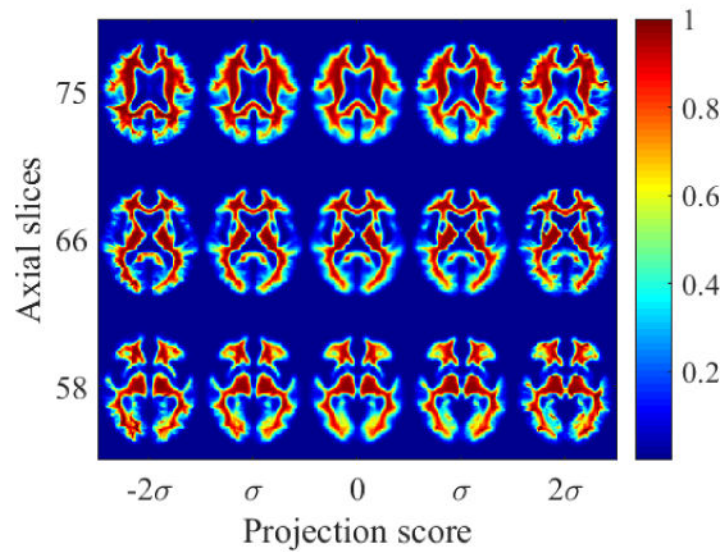
(a)
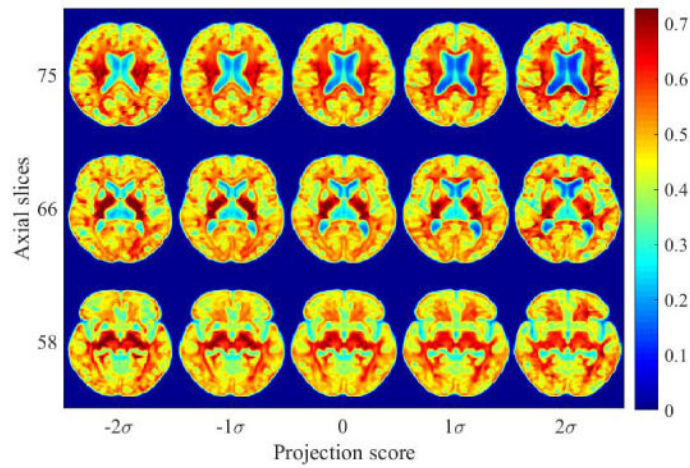
(b)

Figure 16.
(a) High-fit and low-fit individuals can be perfectly separated based on the deformation fields given by DARTEL when projected onto the most discriminant direction computed by PLDA, (b) modulated images illustrating the differences between high-fit and low-fit individuals generated based on deformation fields.
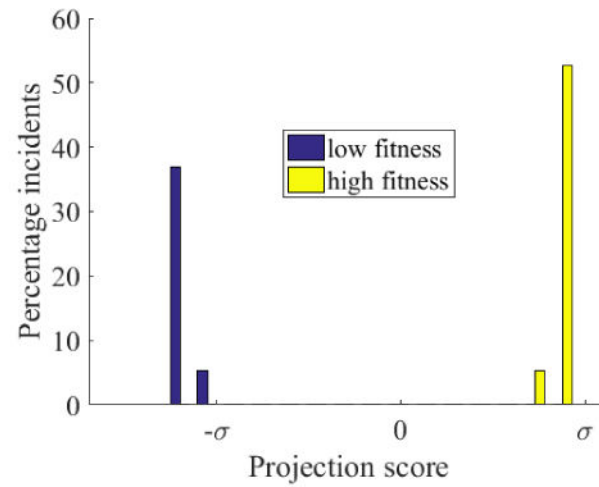
(a)
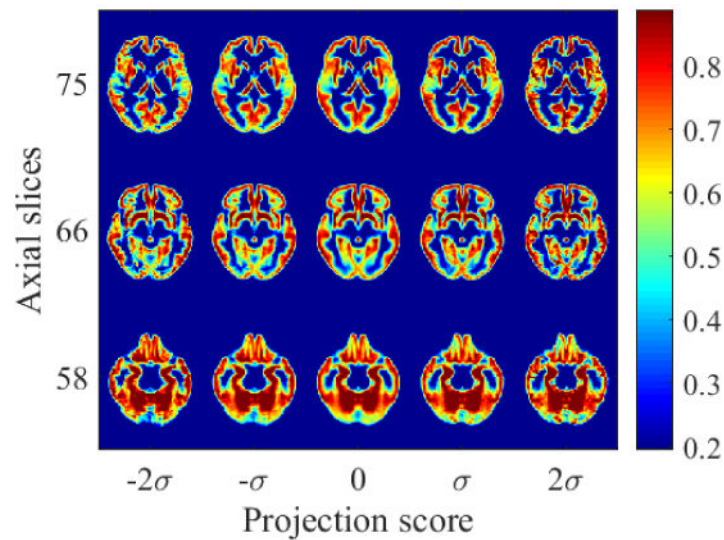
(b)

Figure 17.
(a) High-fit and low-fit individuals can be perfectly separated based on transport maps of gray matter when projected onto the most discriminant direction computed by PLDA, (b) modulated images generated by TBM depicting the gray matter differences between high-fit and low-fit individuals showing loss of tissue from temporal lobe in slice 75 with low fitness.

(a)

(b)

Figure 18.
(a) High-fit and low-fit individuals can be perfectly separated based on transport maps of white matter when projected onto the most discriminant direction computed by PLDA, (b) modulated images generated by TBM depicting the white matter differences between high-fit and low-fit individuals show that fitness appear to protect frontotemporal white matter architecture.
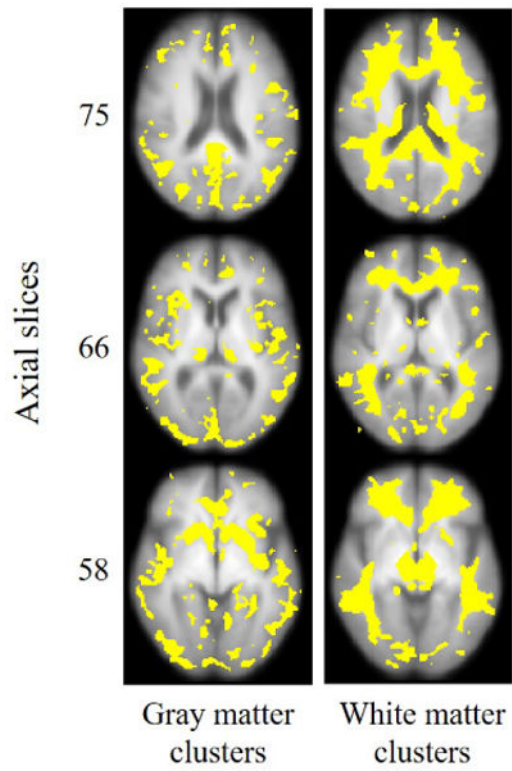
Figure 19.
Heat maps showing the voxels on modulated gray matter and white matter densities whose intensity levels are significantly different between high-fitness and low-fitness groups (p<0.01).
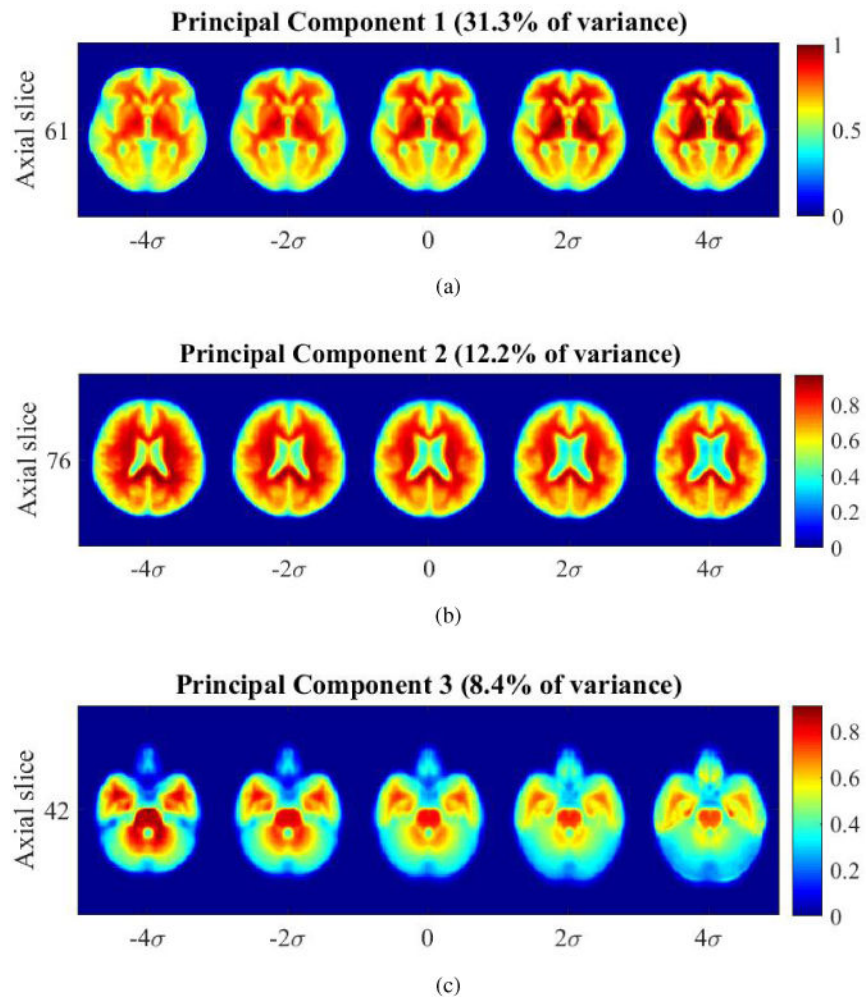
Figure 20.
Visualizing top principal components in transport space. (a) PC1: variability in brain size,
(b) PC2: variability in brain tissue atrophy and size of ventricles, (c) PC3: variability in
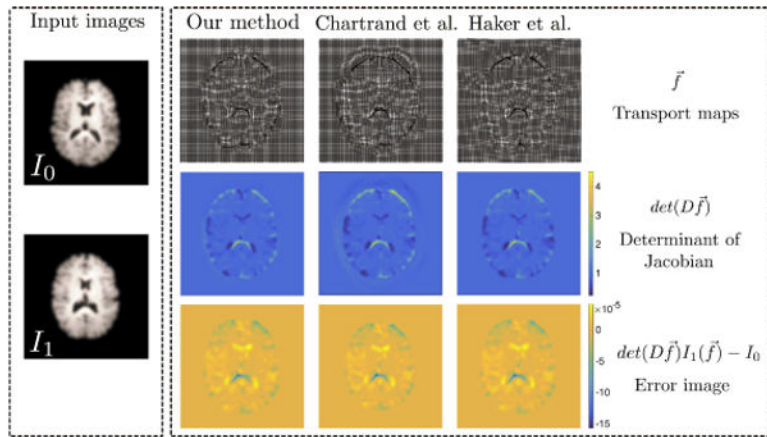prominence of midbrain and brainstem structures

Figure 21.

The source image I_1, target image I_0, and their calculated optimal transport map $\vec{f}$, corresponding determinant of Jacobian matrices, and the error image for the Haker method, the Chartrand method and our method. All are comparable for 2D OMT
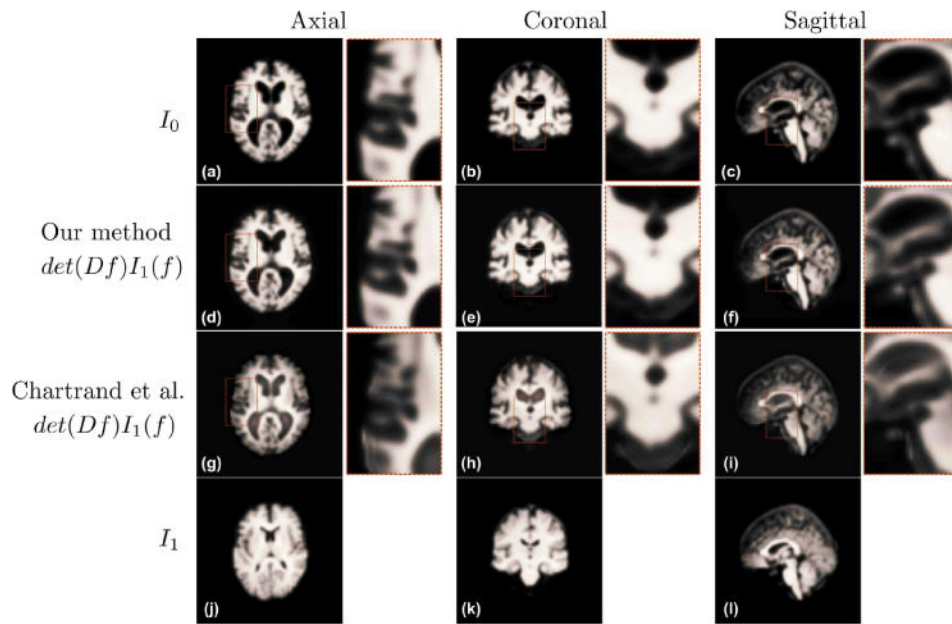
Figure 22.
The target image (a–c), the morphed image in axial, coronal and sagittal cuts using our method (d–f) and the method presented by Chartrand *et al.*[29] (g–i), and the source image (j–l)
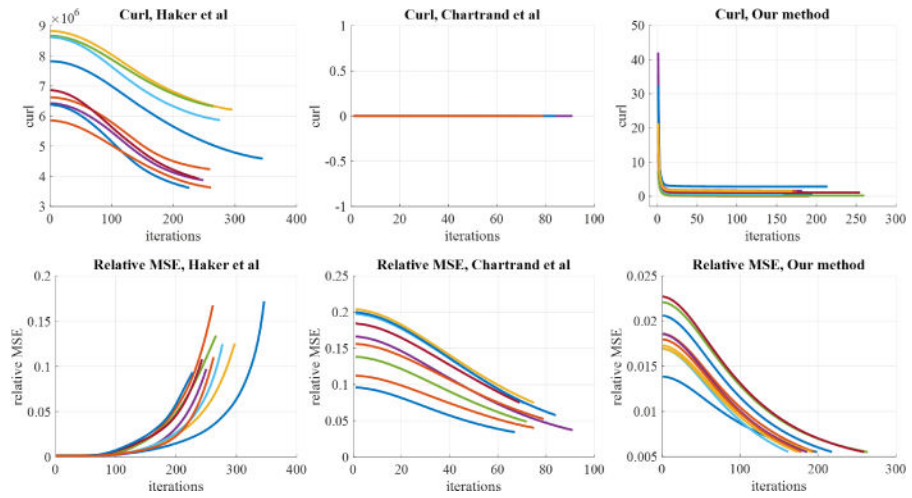
Figure 23.
We see the plots for MSE, curl and mass transported for all three methods. The plots for our method are shown only for the last scale of the GP, using an initial point already close to the final point.

Table 1

Comparing methods of solving OMT in 2D

2D	OT Mapping Statistics		
Method	*Relative MSE*	*Mean curl*	*Mass transported*
Our method	$0.23 \pm 0.056\%$	$(8.7 \pm 5.3) \times 10^{-4}$	1.63 ± 0.57
Chartrand *et al.*	$1.8 \pm 2.9\%$	0	1.56 ± 0.46
Haker *et al.*	$0.45 \pm 0.59\%$	$(7.0 \pm 0.16) \times 10^{-6}$	2.37 ± 0.79

Table 2

Comparing methods of solving OMT in 3D

3D	OT Mapping Statistics		
Method	*Relative MSE*	*Mean curl*	*Mass transported*
Our method	0.55 ± 0.0011%	0.37 ± 0.61	1.3 ± 0.50
Chartrand *et al.*	3.0 ± 1.8%	0	0.07 ± 0.04
Haker *et al.*	9.9 ± 2.5%	$(4.5 \pm 1.6) \times 10^6$	12.5 ± 4.8