Optimal Mass Transport and Image Registration

Steven Haker
Department of Radiology, Surgical Planning Laboratory
Brigham and Women's Hospital, Boston, MA 02115
Email: haker@bwh.harvard.edu
Allen Tannenbaum
Departments of Electrical and Computer and Biomedical Engineering
777 Atlantic Drive,
Georgia Institute of Technology
Atlanta, GA 30332-0250
Email: tannenba@ece.gatech.edu

Abstract

Image registration is the process of establishing a common geometric reference frame between two or more data sets from the same or different imaging modalities possibly taken at different times. In the context of medical imaging and in particular image guided therapy, the registration problem consists of finding automated methods that align multiple data sets with each other and with the patient. In this paper we propose a method of elastic registration based on the Monge–Kantorovich problem of optimal mass transport.

1 Introduction

In this paper, we propose a method for image warping and elastic registration based on the classical problem of optimal mass transport. The mass transport problem was first formulated by Gaspar Monge in 1781, and concerned finding the optimal way, in the sense of minimal transportation cost, of moving a pile of soil from one site to another. This problem was given a modern formulation in the work of Kantorovich [12], and so is now known as the *Monge–Kantorovich problem.*

This type of problem has appeared in econometrics, fluid dynamics, automatic control, transportation, statistical physics, shape optimization, expert systems, and meteorology [16]. It also naturally fits into certain problems in computer vision [6]. In particular, for the general tracking problem, a robust and reliable object and shape recognition system is of major importance. A key way to carry this out is via *template matching*, which is the matching of some object to another within a given catalogue of objects. Typically, the match will not be exact and hence some criterion is necessary to measure the "goodness of fit." For a description of various matching procedures, see [11] and the references therein. The matching criterion can also be considered a *shape metric* for measuring the similarity between two objects.

The registration problem is one of the great challenges that must be addressed in order to make image-guided surgery a practical reality. Registration is the process of establishing a common geometric reference frame between two or more data sets obtained by possibly different imaging modalities. In the context of medical imaging, this is an essential technique for improving preoperative and intraoperative information for diagnosis and image-guided therapy. Registration has a huge literature devoted to it with numerous approaches ranging from statistical to computational fluid dynamics to various types of warping methodologies. See [17] for a number of recent papers on the subject as well as an extensive set of references.

Indeed, multimodal registration methods play a central role in image-guided therapy systems. First, they allow for the fusing of information from each imaging modality, providing better and more accurate information than can be obtained from each image viewed separately. An example is the fusion of functional imaging with anatomical information from MRI for better localization of damaged brain areas. Second, they allow quantitative comparison of images taken at different times, from which information about evolution over time can be inferred. An example is the monitoring of tumor growth in image sequences. Third, when registering preoperative and intraoperative images, they provide a larger field of view and higher image quality than that available with the intraoperative images alone. An example is the fusion of video images obtained by the laparoscope's video camera with MRI data. Fourth, they allow for the updating of a preoperative image or model using intraoperative tracking data.

0-7695-1278-X/01 $10.00 © 2001 IEEE

Multimodal registration proceeds in several steps. First, each image or data set to be matched should be individually calibrated, corrected for imaging distortions and artifacts, and cleared of noise. Next, a measure of similarity between the data sets must be established, so that one can quantify how close an image is from another after transformations are applied. Such a measure may include the similarity between pixel intensity values, as well as the proximity of predefined image features such as implanted fiducials, anatomical landmarks, surface contours, and ridge lines. Next, the transformation that maximizes the similarity between the transformed images is found. Often this transformation is given as the solution of an optimization problem where the transformations to be considered are constrained to be of a predetermined class. Finally, once an optimal transformation is obtained, it is used to fuse the image data sets.

The method we propose in this paper is designed for elastic registration, and is based on an optimization problem built around the L^2 Kantorovich–Wasserstein distance taken as the similarity measure. The constraint that we will put on the transformations considered is that they obey a mass preservation property. Thus, we will be matching *mass densities* in this method, which may be thought of as weighted areas in 2D or weighted volumes in 3D. We will assume that a rigid (non-elastic) registration process has already been applied before applying our scheme.

This type of mass preservation problem occurs naturally in many areas. For example, when registering the proton density based imagery provided by MR. It also occurs in functional MR, where one may want to compare the degree of activity in various features deforming over time, and obtain a corresponding elastic registration map. A special case of this problem occurs in any application where volume or area preserving mappings are considered.

We will give a precise formulation of the problem below (see Section 2), and then develop an algorithm based in part on the work of [7, 14]. The key idea is to find the optimal mapping via an equivalent problem involving certain factorizations (called "polar") of mass preserving mappings. It will turn out that this may be done via a natural gradient descent technique. The details are given in Section 3. We will illustrate our results on some synthetic densities and on real imagery in Section 4.

2 Formulation of the Problem

We now give a modern formulation of the Monge–Kantorovich problem. Let Ω_0 and Ω_1 be two subdomains of $\mathbf{R}^d$, with smooth boundaries, each with a positive density function, μ_0 and μ_1, respectively. We assume

$$\int_{\Omega_0} \mu_0 = \int_{\Omega_1} \mu_1$$

so that the same total mass is associated with Ω_0 and Ω_1. We consider diffeomorphisms $\tilde{u}$ from (Ω_0, μ_0) to (Ω_1, μ_1) which map one density to the other in the sense that

$$\mu_0 = |D\tilde{u}|\, \mu_1 \circ \tilde{u}, \tag{1}$$

which we will call the *mass preservation* (MP) property, and write $\tilde{u} \in MP$. Equation (1) is called the *Jacobian equation.* Here $|D\tilde{u}|$ denotes the determinant of the Jacobian map $D\tilde{u}$. In particular, Equation (1) implies, for example, that if a small region in Ω_0 is mapped to a larger region in Ω_1, then there must be a corresponding decrease in density in order for the mass to be preserved. A mapping $\tilde{u}$ that satisfies this property may thus be thought of as defining a redistribution of a mass of material from one distribution μ_0 to another distribution μ_1.

There may be many such mappings, and we want to pick out an optimal one in some sense. Accordingly, we define the L^p Kantorovich–Wasserstein metric as follows:

$$d_p(\mu_0, \mu_1)^p := \inf_{\tilde{u} \in MP} \int \|\tilde{u}(x) - x\|^p \mu_0(x)\, dx. \tag{2}$$

An *optimal MP map*, when it exists, is one which minimizes this integral. This functional is seen to place a penalty on the distance the map $\tilde{u}$ moves each bit of material, weighted by the material's mass.

The case $p = 2$ has been extensively studied and will the the one proposed in this paper for registration. The L^2 Monge–Kantorovich problem has been studied in statistics, functional analysis, and the atmospheric sciences; see [5, 3] and the references therein. A fundamental theoretical result [13, 4, 8], is that there is a unique optimal $\tilde{u} \in MP$ transporting μ_0 to μ_1, and that this $\tilde{u}$ is characterized as the gradient of a convex function w, i.e., $\tilde{u} = \nabla w$. Note that from Equation (1), we have that w satisfies the *Monge–Ampère* equation

$$|Hw|\, \mu_1 \circ (\nabla w) = \mu_0,$$

where $|Hw|$ denotes the determinant of the Hessian Hw of w.

Hence, the Kantorovich–Wasserstein metric defines the distance between two mass densities, by computing the cheapest way to transport the mass from one domain to the other with respect to the functional given in (2), the optimal transport map in the $p = 2$ case being the gradient of a certain function. The novelty of this result is that like the Riemann mapping theorem in the plane, the procedure singles out a particular map with preferred geometry.

3 Algorithms for Computing The Transport Map

There have been a number of algorithms considered for computing an optimal transport map. For example, methods have been proposed based on linear programming [16], and on Lagrangian mechanics closely related to ideas from the study of fluid dynamics [3]. An interesting geometric method has been formulated by Cullen and Purser [5].

In this section, we will employ a natural solution based on the equivalent problem of *polar factorization*; see [4, 7, 14] and the references therein. We will work with the general case of subdomains in $\mathbf{R}^d$, and point out some simplifications that are possible for the $\mathbf{R}^2$ case.

As above, let $\Omega_0, \Omega_1 \subset \mathbf{R}^d$ be subdomains with smooth boundaries, with corresponding positive density functions μ_0 and μ_1 satisfying $\int_{\Omega_0} \mu_0 = \int_{\Omega_1} \mu_1$. Let $u : (\Omega_0, \mu_0) \to (\Omega_1, \mu_1)$ be an initial mapping with the mass preserving (MP) property. Then according to the generalized results of [4, 7], one can write

$$u = (\nabla w) \circ s, \tag{3}$$

where w is a convex function and s is an MP mapping $s : (\Omega_0, \mu_0) \to (\Omega_0, \mu_0)$. This is the *polar factorization* of u with respect to μ_0. In [7], just the case of area preservation is considered, i.e., μ_0 is assumed constant, but the general case goes through as well.

Our goal is to find the polar factorization of the MP mapping u, according to the following strategy. We consider the family of MP mappings of the form $\tilde{u} = u \circ s^{-1}$ as s varies over MP mappings from (Ω_0, μ_0) to itself. If we consider $\tilde{u}$ as a vector field, we can always find a function w and another vector field χ, with $\mathrm{div}(\chi) = 0$, such that

$$\tilde{u} = \nabla w + \chi,$$

i.e., we can decompose $\tilde{u}$ into the sum of a curl-free and divergence-free vector field. Thus, what we try to do is find a mapping s which will yield a $\tilde{u}$ without any curl, that is, such that $\tilde{u} = \nabla w$. Once such an s is found, we will have $u = \tilde{u} \circ s = (\nabla w) \circ s$ and so we will have found the polar factorization (3) of our given function u.

Now, here is the key point. As we discussed above, the unique optimal solution of the L^2 Monge–Kantorovich problem has the form $\tilde{u} = \nabla w$, and so the problem of finding the polar factorization of u and finding the optimal Monge–Kantorovich mapping $\tilde{u}$ are equivalent. In essence, to solve the Monge–Kantorovich problem we create a "rearrangement" of an initial vector field u using a map s, so that the resulting vector field $\tilde{u} = u \circ s^{-1}$ has no curl. We can now give the technical details.

1 Finding an Initial Mapping

We will now propose an explicit algorithm to solve the Monge–Kantorovich problem. So we want to minimize the L^2 Kantorovich–Wasserstein distance functional over MP functions from (Ω_0, μ_0) to (Ω_1, μ_1). We will try to do this by finding an initial MP mapping u and then minimizing over $\tilde{u} = u \circ s^{-1}$ by varying s over MP mappings from Ω_0 to Ω_0, starting with s equal to the identity map. Our first task is to find and initial MP mapping u. This can be done for general domains using a method of Moser [15], or for simpler domains using the following algorithm. For simplicity, we work in $\mathbf{R}^2$ and assume $\Omega_0 = \Omega_1 = [0,1]^2$, the generalization to higher dimensions being straightforward. We define a function $a = a(x)$ by the equation

$$\int_0^{a(x)} \int_0^1 \mu_1(\eta, y)\, dy\, d\eta = \int_0^x \int_0^1 \mu_0(\eta, y)\, dy\, d\eta \tag{4}$$

which gives by differentiation with respect to x

$$a'(x) \int_0^1 \mu_1(a(x), y)\, dy = \int_0^1 \mu_0(x, y)\, dy. \tag{5}$$

We may now define a function $b = b(x, y)$ by the equation

$$a'(x) \int_0^{b(x,y)} \mu_1(a(x), \rho)\, d\rho = \int_0^y \mu_0(x, \rho)\, d\rho, \tag{6}$$

and set $u(x, y) = (a(x), b(x, y))$. Since $a_y = 0$, $|Du| = a_x b_y$, and differentiating (6) with respect to y we find

$$\begin{aligned} a'(x)\, b_y(x,y)\, \mu_1(a(x), b(x,y)) &= \mu_0(x,y) \\ |Du|\, \mu_1 \circ u &= \mu_0, \end{aligned}$$

which is the MP property we need. This process can be interpreted as the solution of a one-dimensional Monge–Kantorovich problem in the x direction followed by the solution of a family of one-dimensional Monge–Kantorovich problems in the y direction.

2 Removing the Curl

Once an initial MP u is found, we need to apply the process which will remove its curl. We begin with the following elementary and intuitive property of MP mappings, the proof of which is a simple calculation.

Lemma: The composition of two mass preserving (MP) mappings is an MP mapping. The inverse of an MP mapping is an MP mapping.

Thus, since u is an MP mapping, we have that $\tilde{u} = u \circ s^{-1}$ is an MP mapping if and only if s is, that is, if and only if

$$\mu_0 = |Ds|\, \mu_0 \circ s.$$

In particular, when μ_0 is constant, this equation requires that s be area or volume preserving.

Next, rather than working with s directly, we solve the polar factorization problem via gradient descent. Accordingly, we will assume that s is a function of time, and then determine what s_t should be to decrease the L^2 Monge–Kantorovich functional. This will give us an evolution equation for s and in turn an equation for $\tilde{u}_t$ as well, the latter being the most important for implementation. By differentiating $\tilde{u} \circ s = u$ with respect to time, we get

$$\tilde{u}_t = -D\tilde{u}\,\tilde{s}_t$$

where we've abused notation to define $\tilde{s}_t := s_t \circ s^{-1}$. We need to make sure that s maintains its MP property. Differentiating $\mu_0 = |Ds|\,\mu_0 \circ s$ with respect to time, we derive

$$\mathrm{div}(\mu_0\,\tilde{s}_t) = 0,$$

from which we see that $\tilde{s}_t$, s_t and $\tilde{u}_t$ should have the following forms:

$$\tilde{s}_t = \frac{1}{\mu_0}\zeta, \qquad (7)$$

$$s_t = \left(\frac{1}{\mu_0}\zeta\right) \circ s, \qquad (8)$$

$$\tilde{u}_t = -\frac{1}{\mu_0}D\tilde{u}\,\zeta, \qquad (9)$$

for some vector field ζ on Ω_0, with $\mathrm{div}(\zeta) = 0$ and $\langle \zeta, \vec{n}\rangle = 0$ on $\partial\Omega_0$, $\vec{n}$ being the normal to the boundary of Ω_0. This last condition ensures that s remains a mapping from Ω_0 to itself, by preventing the flow of s, given by $s_t = \left(\frac{1}{\mu_0}\zeta\right) \circ s$, from crossing the boundary of Ω_0. This also means that the range of $\tilde{u} = u \circ s^{-1}$ is always $u(\Omega_0) = \Omega_1$.

Consider now the problem of minimizing the Monge–Kantorovich functional:

$$M = \int ||\tilde{u} - x||^2 \mu_0 \qquad (10)$$

$$= \int ||\tilde{u}||^2\mu_0 - 2\int \langle \tilde{u}, x\rangle\,\mu_0 + \int ||x||^2\mu_0 \qquad (11)$$

The last term is obviously independent of time. Interestingly, so is the first,

$$\int ||\tilde{u}||^2\mu_0 = \int ||u \circ s^{-1}||^2\mu_0$$

$$= \int ||u \circ s^{-1}||^2\,|Ds^{-1}|\,\mu_0 \circ s^{-1}$$

$$= \int ||u||^2\mu_0$$

where $\mu_0 = |Ds^{-1}|\,\mu_0 \circ s^{-1}$ since s^{-1} is an MP map.

Turning now to the middle term, we do a similar trick,

$$\int \langle \tilde{u}, x\rangle\,\mu_0 = \int \langle u \circ s^{-1}, s \circ s^{-1}\rangle\,\mu_0$$

$$= \int \langle u \circ s^{-1}, s \circ s^{-1}\rangle\,|Ds^{-1}|\,\mu_0 \circ s^{-1}$$

$$= \int \langle u, s\rangle\,\mu_0,$$

and taking $s_t = \left(\frac{1}{\mu_0}\zeta\right) \circ s$, we compute

$$-\frac{1}{2}M_t = \int \langle \tilde{u}, \zeta\rangle\,.$$

Now decomposing $\tilde{u}$ as $\tilde{u} = \nabla w + \chi$, we have

$$-\frac{1}{2}M_t = \int \langle \chi, \zeta\rangle\,, \qquad (12)$$

where we've used the divergence theorem, $\mathrm{div}(\zeta) = 0$, and $\langle \zeta, \vec{n}\rangle = 0$ on $\partial\Omega_0$. Thus, in order to decrease M, we can take $\zeta = \chi$ with corresponding formulas (7)-(9) for s_t, $\tilde{s}_t$, and $\tilde{u}_t$, provided that we have $\mathrm{div}(\chi) = 0$ and $\langle \chi, \vec{n}\rangle = 0$ on $\partial\Omega_0$. Thus it remains to show that we can decompose $\tilde{u}$ as $\tilde{u} = \nabla w + \chi$ for such a χ.

Gradient Descent: $\mathbf{R}^d$:

We let w be a solution of the Neumann-type boundary problem

$$\mathrm{div}(\tilde{u}) = \Delta w \qquad (13)$$

$$\langle \nabla w, \vec{n}\rangle = \langle \tilde{u}, \vec{n}\rangle \text{ on } \partial\Omega_0, \qquad (14)$$

and set $\chi = \tilde{u} - \nabla w$. It is then easily seen that χ satisfies the necessary requirements.

Thus, by (9), we have the following evolution equation for $\tilde{u}$:

$$\tilde{u}_t = -\frac{1}{\mu_0}D\tilde{u}\,\left(\tilde{u} - \nabla\Delta^{-1}\mathrm{div}(\tilde{u})\right). \qquad (15)$$

This is a first order non-local scheme for $\tilde{u}_t$ if we count Δ^{-1} as minus 2 derivatives. Note that this flow is consistent with respect to the Monge–Kantorovich theory in the following sense. If $\tilde{u}$ is optimal, then it is given as $\tilde{u} = \nabla w$, in which case $\tilde{u} - \nabla\Delta^{-1}\mathrm{div}(\tilde{u}) = \nabla w - \nabla\Delta^{-1}\mathrm{div}(\nabla w) = 0$ so that by (15), $\tilde{u}_t = 0$.

Gradient Descent: $\mathbf{R}^2$:

The situation is somewhat simpler in the $\mathbf{R}^2$ case, due to the fact that a divergence free vector field χ can in general be written as $\chi = \nabla^{\perp}h$ for some scalar function h, where $\perp$ represents rotation by 90 deg, so that $\nabla^{\perp}h = (-h_y, h_x)$. In this case, (12) becomes

$$-\frac{1}{2}M_t = \int \langle \nabla^{\perp}f, \nabla^{\perp}h\rangle = \int \langle \nabla f, \nabla h\rangle \qquad (16)$$

where the decomposition of $\tilde{u}$ is $\tilde{u} = \nabla w + \nabla^{\perp}f$, and we can take $h = f$. The function f can be found by solving the Dirichlet-type boundary problem

$$-\,\mathrm{div}(\tilde{u}^{\perp}) = \Delta f, \qquad (17)$$

$$f = 0 \text{ on } \partial\Omega_0, \qquad (18)$$

which gives us the evolution equation

$$\tilde{u}_t = \frac{1}{\mu_0} D\tilde{u}\, \nabla^{\perp}\Delta^{-1}\mathrm{div}(\tilde{u}^{\perp}). \tag{19}$$

We may also derive a second order *local* evolution equation for $\tilde{u}$ by using the divergence theorem with (16) to get

$$\tilde{u}_t = -\frac{1}{\mu_0} D\tilde{u}\, \nabla^{\perp}\mathrm{div}(\tilde{u}^{\perp}), \tag{20}$$

where appropriate handling of the evolution at the boundary, as described in Section 4, is required.

3 Defining the Warping Map

Typically in elastic registration, one wants to see an explicit warping which smoothly deforms one image into the other. This can easily be done using the solution of the Monge–Kantorovich problem. Thus, we assume now that we have applied our gradient descent process as described above and that it has converged to the Monge–Kantorovich mapping $\tilde{u}_{MK}$.

Following the work of Benamou and Brenier, [3], (see also [8]), we consider the following related problem:

$$\inf \int\int_0^1 \mu(t,x)\|v(t,x)\|^2\, dt\, dx$$

over all time varying densities μ and velocity fields v satisfying

$$\begin{aligned} \frac{\partial \mu}{\partial t} + \mathrm{div}(\mu v) &= 0, & (21)\\ \mu(0,\cdot) &= \mu_0, \quad \mu(1,\cdot) = \mu_1. & (22) \end{aligned}$$

It is shown in [3] that this infimum is attained for some μ_{min} and v_{min}, and that it is equal to the L^2 Kantorovich–Wasserstein distance between μ_0 and μ_1. Further, the flow $X = X(x,t)$ corresponding to the minimizing velocity field v_{min} via

$$X(x,0) = x, \quad X_t = v_{min} \circ X$$

is given simply as

$$X(x,t) = x + t\ (\tilde{u}_{MK}(x) - x). \tag{23}$$

Note that when $t = 0$, X is the identity map and when $t = 1$, it is the solution $\tilde{u}_{MK}$ to the Monge–Kantorovich problem. This analysis provides appropriate justification for using (23) to *define* our continuous warping map X between the densities μ_0 and μ_1.

4 Implementation and Examples

We tested the method by optimally mapping μ_0, a 256 x 256 "Lena" image, onto μ_1, a "Tiffany" image. In Figure 1, we have the original Lena image. It has been histogrammed, and its values have been scaled from 0.25 to 4.0. In Figure 2, we show the Tiffany image histogrammed and scaled in the same manner as Lena. The initial mapping u is shown in Figure 3. The shading in this example differs from the previous one in that we have simply taken the Lena pixel values μ_0 and transported them via the mapping u. Here, the Jacobian of u ranges between 1/16 and 16. Next, in Figure 4, we show the Monge–Kantorovich optimal mapping $\tilde{u}_{MK}$, again found by applying the non-local flow (19). The Monge–Kantorovich mapping can be seen to be much more regular, and indicates the potential of the method. The processing of 1000 iterations took about 12 minutes.

Next, we use the warp function (23) to find a continuous transformation of Lena to Tiffany. This is illustrated in Figures 5 through 9, which correspond to t = 0.1, 0.3, 0.5, 0.7, 0.9, and 1.0. Here, we have shaded the target points by $|DX^{-1}|\ \mu_0 \circ X^{-1}$, which can be seen by (1) to vary smoothly between Lena at time 0.0 and Tiffany at time 1.0

Finally, in Figures 10 through 13 we show a brain deformation sequence. Here, the first and last images were given, and the intermediate two were found using our process. This type of elastic brain deformation occurs during surgery, after the skull is opened.

We have also successfully implemented the second order local flow (20), with similar results. In this case, we require that a periodic boundary condition be enforced, specifically that $\tilde{u} - x$ be periodic on the square image domain. We also used an upwinding scheme when calculating $D\tilde{u}$. While it may seem that this local flow should provide a faster method than the non-local flow (19), in practice this does not seem to be the case. Even though the non-local method requires that the Laplacian be inverted during each iteration, the problem has been set up to allow the use of fast numerical solvers which use FFT-type methods and operate on rectangular grids. We have used the Matlab solver here, which uses sine transforms followed by the solution of a tridiagonal system. Moreover, we have found that the functional is decreased substantially more during each iteration of the non-local method, using the maximum temporal step size allowed for stability in each case.

In general, the target domain Ω_1 need not be rectangular when using the non-local method. However, we note that if the periodic boundary condition described above is used, then the Laplacian in (19) can be inverted using the FFT alone, without the need to solve a subsequent matrix system. For the Lena to Tiffany warp, this reduced the processing time by 1/3.

5 Conclusions

In this paper, we presented a natural method for image registration based on the classical problem of optimal mass transportation. We showed that for an L^2 version of the problem, one could derive easily-implementable gradient descent equations to carry out the method. Our approach allows us to have non-uniform densities on both the domain and range.

Although applied here to the Monge–Kantorovich problem, the method used to enforce the mass preservation constraint is general and has other applications. In particular, the concept a harmonic mapping, defined as a minimizer of the Dirichlet integral, can be combined with a mass preservation constraint to obtain a new approach to mass-preserving diffeomorphisms [1]. A minimizer, when it exists, is called an MP map of *minimal distortion*. Non-local and local gradient descent methods for computing such a map of minimal distortion can be derived in a manner very similar to that described above for the Monge–Kantorovich functional [1]. These methods have applications to brain surface flattening and virtual colonoscopy as described in [2, 10].

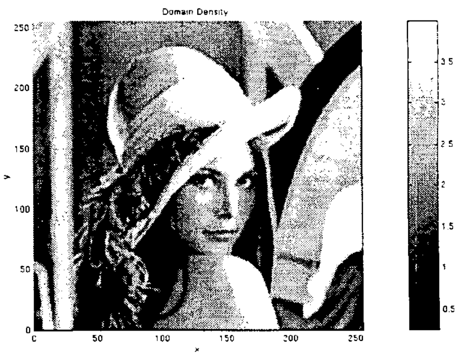

Figure 1: Original Lena Image

References

[1] S. Angenent, S. Haker, A. Tannenbaum, and R. Kikinis, "On area preserving maps of minimal distortion," in *System Theory: Modeling, Analysis, and Control*, edited by T. Djaferis and I. Schick, Kluwer, Holland, 1999, pages 275-287.

[2] S. Angenent, S. Haker, A. Tannenbaum, and R. Kikinis, "Laplace-Beltrami operator and brain surface flattening," *IEEE Trans. on Medical Imaging* **18** (1999), pp. 700-711.

[3] J.-D. Benamou and Y. Brenier, "A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem," *Numerische Mathematik* **84** (2000), pp. 375-393.

[4] Y. Brenier, "Polar factorization and monotone rearrangement of vector-valued functions," *Com. Pure Appl. Math.* **64** (1991), pp. 375-417.

[5] M. Cullen and R. Purser, "An extended Lagrangian theory of semigeostrophic frontogenesis," *J. Atmos. Sci.* **41** (1984), pp. 1477-1497.

[6] D. Fry, *Shape Recognition Using Metrics on the Space of Shapes*, Ph.D. Thesis, Harvard University, 1993.

[7] W. Gangbo, "An elementary proof of the polar factorization of vector-valued functions," *Arch. Rational Mechanics Anal.* **128** (1994), pp. 381-399.

[8] W. Gangbo and R. McCann, "The geometry of optimal transportation," *Acta Math.* **177** (1996), pp. 113-161.

[9] W. Gangbo and R. McCann, "Shape recognition via Wasserstein distance," Technical Report, School of Mathematics, Georgia Institute of Technology, 1999.

[10] S. Haker, S. Angenent, A. Tannenbaum, and R. Kikinis, "Nondistorting flattening maps and the 3D visualization of colon CT images," *IEEE Trans. of Medical Imaging*, July 2000.

[11] R. Haralick and L. Shapiro. *Computer and Robot Vision*, Addison-Wesley, New York, 1992.

[12] L. V. Kantorovich, "On a problem of Monge," *Uspekhi Mat. Nauk.* **3** (1948), pp. 225-226.

[13] M. Knott and C. Smith, "On the optimal mapping of distributions," *J. Optim. Theory* **43** (1984), pp. 39-49.

[14] R. McCann, "Polar factorization of maps on Riemannian manifolds," preprint 2000. Available at http://www.math.toronto.edu/mccann.

[15] J. Moser, "On the volume elements on a manifold," *Trans. Amer. Math. Soc.* **120** (1965), pp. 286-294.

[16] S. Rachev and L. Rüschendorf, *Mass Transportation Problems*, Volumes I and II, Probability and Its Applications, Springer, New York, 1998.

[17] A. Toga, *Brain Warping*, Academic Press, San Diego, 1999.

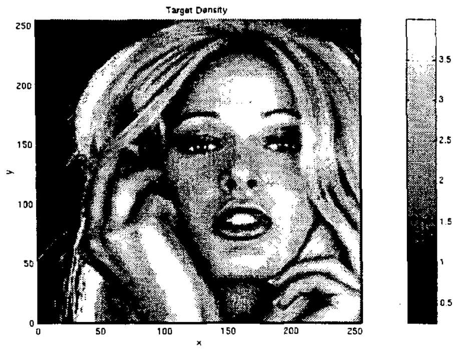

Figure 2: Target Tiffany Image

Figure 5: Lena to Tiffany Warp: $t = 0.1$

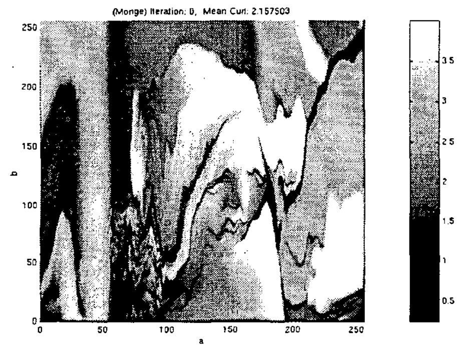

Figure 3: Initial Mass–Preserving Mapping

Figure 6: Lena to Tiffany Warp: $t = 0.5$

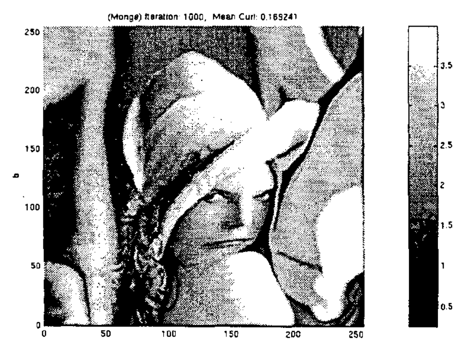

Figure 4: Final Monge–Kantorovich Mapping

Figure 7: Lena to Tiffany Warp: $t = 0.7$

Figure 8: Lena to Tiffany Warp: $t = 0.9$

Figure 9: Lena to Tiffany Warp: $t = 1.0$

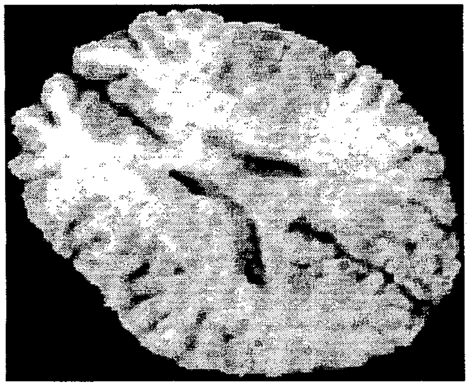

Figure 10: Brain Warping: $t = 0.00$

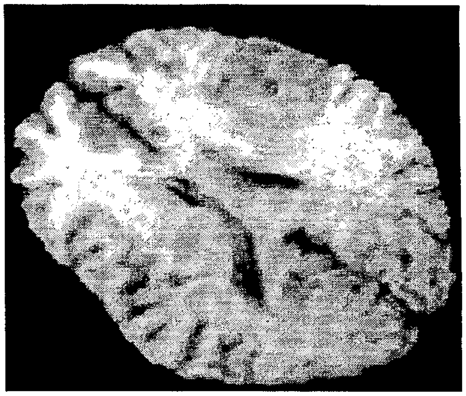

Figure 11: Brain Warping: $t = 0.33$

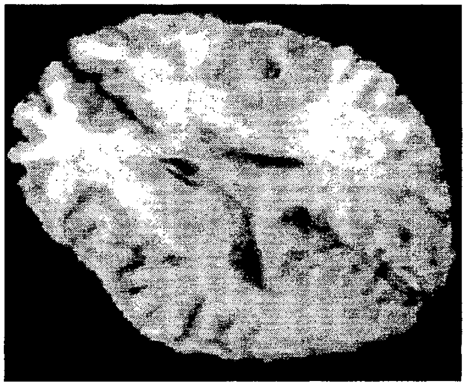

Figure 12: Brain Warping: $t = 0.66$

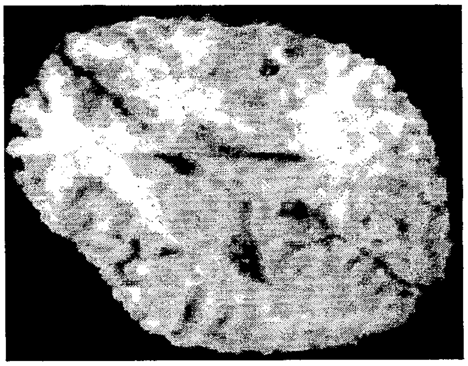

Figure 13: Brain Warping: $t = 1.00$