Optimal mass transport for deformable registration and warping of 2D images

Reuben R. Shamir

School of Engineering and Computer Science, The Hebrew University, Jerusalem, Israel.

We implement a warping method which is using the properties of elastic materials to compute the deformation between two images. The proposed elastic registration method incorporates the L2 Kantorovich-Wasserstein distance, also known as the Earth Mover's Distance (EMD), as a similarity measure. The implemented paper presents an efficient partial differential equation approach for a first-order solution of this problem that is simpler than existing works suggested high-order solutions and is computationally simpler than existing works based on linear programming.

Specifically, we implement the following paper
Steven Haker, Lei Zhu, Allen Tannenbaum, and Sigurd Angenent, "Optimal mass transport for registration and warping", *International Journal of Computer Vision*, 60(3), p. 225-240, 2004.

1. Introduction

Image to image registration is the process of aligning the different coordinate systems of two or more image data sets by computing a common reference frame between them. The registration allows the integration of multiple image data sets acquired with different modalities (MRI, CT, etc.), at different times and under varying patient pose and position. Normally, the registration transformation is computed to minimize a predefined similarity measure which quantifies how close an image set is from the other one. The implemented paper [1] presents a warping method, which is using the properties of elastic materials to compute the deformation between two images. The proposed elastic registration method incorporates the L^2 Kantorovich-Wasserstein distance, also known as the Earth Mover's Distance (EMD), as a similarity measurement. This similarity measure has 5 key advantages: 1. It is parameter free; 2. it is symmetrical; 3. it requires no landmarks selection; 4. the solution is unique, and; 5. it considers changes in density that are resulted from changes in area or volume. The implemented paper presents an efficient partial differential equation approach for a first-order solution of this problem. The method is simpler than existing works suggested high-order solutions and is computationally simpler than existing works based on linear programming. Next, we briefly describe the proposed method and afterwards we describe its implementation and present experimental results.

2. Methods

2.1 Formal description of the problem

Let Ω_0 and Ω_1 be two sub-domains of R^d, with smooth boundaries, each with a positive density function, μ_0 and μ_1, respectively. The following property is assumed:

1) $$\int_{\Omega_0} \mu_0(x)\, dx = \int_{\Omega_1} \mu_1(x)\, dx$$

Mass Preservation (MP) property is defined as follows.

2) $$\mu_0(x) = |Du|\, \mu_1(u(x))$$

Where u is a diffeomorphism map from Ω_0 to Ω_1, and Du is its Jacobian.

Contact info: email: rubke@cs.huji.ac.il, URL: www.cs.huji.ac.il/~rubke

Note that, if a small region in Ω_0 is mapped to a large region in Ω_1 then the density will be decreased. In this paper, an optimal MP map $\tilde{u}$ is computed to minimize the L^2 Kantorovich-Wasserstein metric defined as follows.

3) $$d^2(\mu_0, \mu_1, u) = \int_{\Omega_0} \left\| u(x) - x \right\|^2 \mu_0(x)\, dx$$

where u is a MP map. It turns out that the optimal solution $\tilde{u}$ is unique and is characterized as a gradient of a convex function ω.

4) $$\tilde{u} = \nabla \omega$$

To avoid large differences in corresponding pixels intensities a comparison term penalizing the changing of intensity is introduced.

5) $$M \equiv C(I_0, I_1, u) + \alpha \cdot d^2(\mu_0, \mu_1, u)$$

Noteworthy, $C(I_0, I_1, u)$ may be any comparison term, like the squared error, normalized correlation or mutual information.

2.2 The proposed solution

A. Polar factorization and Rearrangement maps

Given $u : (\Omega_0, \mu_0) \to (\Omega_1, \mu_1)$, which is an initial diffeomorphic map that preserves the MP property, then the Polar Factorization of u with respect to μ_0 is defined as

6) $$u(x) = \nabla \omega\ (s(x)),$$

Where $s(x)$ is a MP mapping $s : (\Omega_0, \mu_0) \to (\Omega_0, \mu_0)$.

Since the optimal MP map that is minimizing the L^2 Kantorovich-Wasserstein metric has the form $\tilde{u}(x) = \nabla \omega(x) = u\ (s^{-1}(x))$, it is equivalent to the Polar Factorization problem.

The solution strategy is as follows. The algorithms starts with an initial MP maps u_0 and s, and define the current-iteration optimal solution as

7) $$\tilde{u}(x) = \nabla \omega(x) + \chi = u_0\ (s^{-1}(x))$$

Where χ is a vector field with $div(\chi) = 0$. Then, the algorithm is iteratively computes a mapping s such that finally we get $\tilde{u}(x) \approx \nabla \omega(x)$, a curl free vector field.

B. Finding an initial mapping

The authors suggest computing the initial mapping u_0 with an iterative algorithm that solves the optimal-mass transport in one dimension along lines parallel to the axes, iteratively - axis after axis. In the 2D case, the initial mapping u_0 is computed as a combination of the optimal mass transport map along the x-axis, denoted as $a(x)$ and the optimal mass transport map along the y-axis, denoted as $b(x, y)$, such that $u_0 = (a(x), b(x, y))$.

C. Removing the curl

The authors solve the Polar Factorization problem via gradient descent. $s(x)$ and $u(x)$ are defined as a function of time and the value of $\frac{\partial s}{\partial t}$ and $\frac{\partial u}{\partial t}$ that decreases the L^2 Kantorovich-Wasserstein metric is derived as follows.

To preserve the MP property the derivatives are of the form:

8) $$\frac{\partial s}{\partial t} = \left(\frac{\varsigma}{\mu_0}\right) \circ \left(s(x)\right) \quad \text{and} \quad \frac{\partial u}{\partial t} = -\frac{1}{\mu_0} Du \ \varsigma$$

Where ς is a vector field on Ω_0 with $div(\varsigma) = 0$ and $\langle \varsigma, \vec{n} \rangle = 0$ on $\partial\Omega_0$, $\vec{n}$ is the normal to the boundary of Ω_0.

Now, if we take the derivative of the L^2 Kantorovich-Wasserstein functional, denoted as M, we get

9) $$-\frac{1}{2}\frac{\partial M}{\partial t} = \int_{\Omega_0} \left\langle u(x), \mu_0\left(\frac{\partial s}{\partial t}(s^{-1}(x))\right) \right\rangle \, dx$$

Following equations (7) and (8), and the divergence theorem, the authors show that

10) $$-\frac{1}{2}\frac{\partial M}{\partial t} = \int_{\Omega_0} \langle \chi, \varsigma \rangle \, dx$$

Therefore, defining $\varsigma = \chi$ in formula (8) will decrease M.

Next the authors show how to decompose u as $u = \nabla\omega + \chi$.

For the general case of $\Omega_0 \subset R^d$, ω is defined as the solution for the Neumann type boundary problem

11) $$\Delta\omega = div(u)$$

12) $$\langle \nabla\omega, \vec{n} \rangle = \langle u, \vec{n} \rangle \text{ on } \partial\Omega_0$$

If we define $\chi = u - \nabla\omega$ it satisfies the requirements of ς in equation (8).

With this definition and equation (8), we get

13) $$\frac{\partial u}{\partial t} = -\frac{1}{\mu_0} Du \ (\text{u}-\nabla(\Delta^{-1}\text{div(u)}))$$

Where $\Delta^{-1}\text{div(u)}$ denotes the solution ω.

In the simpler case of $\Omega_0 \subset R^d$, χ can be written as $\chi = \nabla^{\perp} f = \left(-\frac{\partial f}{\partial y}, \frac{\partial f}{\partial x}\right)$ for some scalar function f, and u is decomposed into $u = \nabla\omega + \nabla^{\perp} f$.

f is found as the solution for the Dirichlet type boundary problem as follows.

14) $$\Delta f = -div(u^{\perp})$$

15) $$f = 0 \text{ on } \partial\Omega_0$$

With equation (8), we get

16) $$\frac{\partial u}{\partial t} = \frac{1}{\mu_0} Du \ \nabla^{\perp}(\Delta^{-1}\text{div}(\text{u}^{\perp}))$$

The second order local evolution equation for u is derived as follows.

17) $$\frac{\partial u}{\partial t} = -\frac{1}{\mu_0} Du \ \nabla^{\perp}\text{div}(\text{u}^{\perp})$$

D. Adding a comparison term

A comparison term is added to avoid undesired situations, like a mapping of a small high intensity region into a large low-intensity region.

Therefore, the minimization of the following functional is desired.

18) $$M(\alpha, u) = \int_{\Omega_0} \left(I_1\left(u(x)\right) - I_0(x)\right)^2 dx + \alpha^2 \int_{\Omega_0} \left\|u(x) - x\right\|^2 \mu_0(x) \, dx$$

The authors suggest minimizing it with the gradient descent method as before, but this time defines:

19) $$\frac{\partial u}{\partial t} = \frac{1}{\mu_0} Du\ \nabla^{\perp}(\Delta^{-1}\mathrm{div}(\mathrm{P}^{\perp}))$$ - for the non-local flow

And

20) $$\frac{\partial u}{\partial t} = -\frac{1}{\mu_0} Du\ \nabla^{\perp}\mathrm{div}(\mathrm{P}^{\perp})$$ - for the local flow

Where P is defined as:

21) $$P = \frac{1}{\mu_0^2}\left(I_1(u(x)) - I_0\right)^2 \nabla\mu_0(x) + \frac{2}{\mu_0}\left(I_1(u(x)) - I_0\right)^2 \nabla I_0(x) + 2\alpha^2 u$$

E. The proposed algorithm for computing the optimal MP map is as follows:

1. Compute the initial mapping u_0.
2. Set $current\_u = u_0$
3. Compute P as $P = current\_u$ for pure optimal mass computation or as suggested in equation (21) to incorporate a comparison term.
4. Solve the appropriate Poisson's equation for the non-local flow problems.
5. Compute the appropriate u_t value.
6. update: $current\_u = current\_u - \Delta t \cdot u_t$
7. go to (3)

The method stops when the energy is decreasing sufficiently slowly.

The time step Δt can be chosen to be less than $\min_{x,i}\left|\frac{1}{\mu_0}\left(\nabla^{\perp}(\Delta^{-1}\mathrm{div}(\mathrm{P}^{\perp}))\right)_i\right|^{-1}$ where i stands for the component of the vector.

F. Defining the warping map

The warping map is defined as follows.

22) $$X(x,t) = x + t(\tilde{u}(x) - x)$$

Where, $\tilde{u}$ is the optimal mass transport mapping, and t is a time parameter in the range of 0 to 1. The justification of this warping method is outlined in the text.

3 Implementation and experiments

3.1 Computing an initial mapping

Implementation

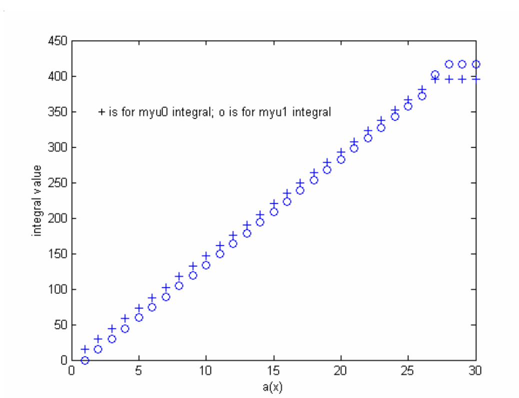
$u_0 = (a(x), b(x,y))$ is computed with a simple numerical integration technique that iteratively guarantees that $\int_0^{a(x)} \int_0^{b(x,y)} \mu_0(\beta, \lambda)\, d\beta\, d\lambda = \int_0^x \int_0^y \mu_1(x,y)\, dx\, dy$.
See 'compute_initial_mapping.m' for the actual code.

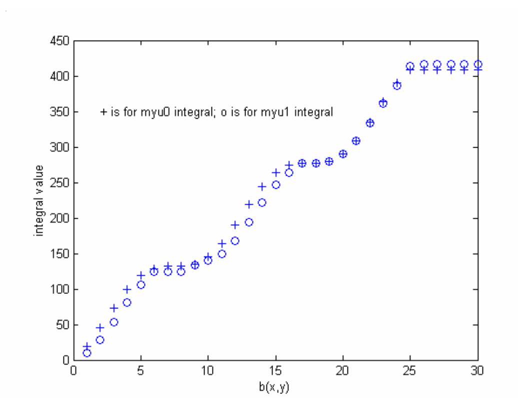
Experiments

We have performed three experiments to test the computation of u_0.
To run the experiments please use the commands: run_tests(1), run_tests(2) or run_tests(3).
The code is found at run_tests.m

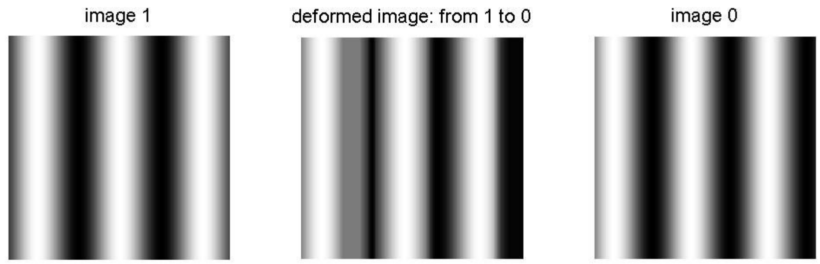
The inputs for the first three experiments are two images of horizontal or vertically sine waves with a phase difference between them normalized to the range of [0,1]. The tests compare the integrals values at different a(x) and their correlated x locations and at different b(y) and their correlated y locations. In addition we use the computed deformation map to warp one image towards the other, but using only the intensity values of the first image.
The first experiment inputs are two horizontal sine waves, the second and third experiments inputs are two vertical sine waves. The difference between the second and the third tests is the method used to equalize the total mass of the density maps. While in the second method the integrals ratio is used, in the third experiment the mass distribution is evenly equalized over the image domain (using histeq).

Results

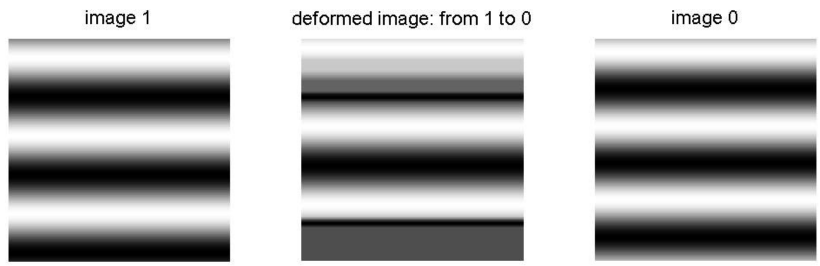




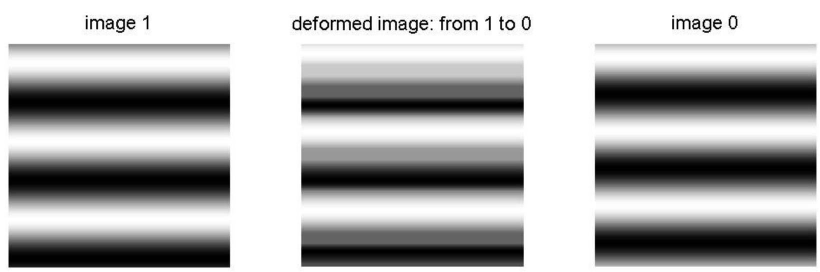
This result (run_tests(1)) shows that the initial mapping computation approximately fulfills the MP property (the integrals equation is preserved). Total mass is preserved by multiplying by the integrals ratio. Similar results were observed in run_tests(2) and run_tests(3).



Result of run_tests(1). Image 1 is deformed towards image 0 with the computed initial mapping (the intensities are of image 1 and are not mixed). we think that the undesired grey colors between the left and the middle white columns are due to the small numerical errors of the integrals approximation method. A better integral approximation method may result with more convincing results. Total mass equality is preserved by multiplying by the integrals ratio.



Result of run_tests(2). Image 1 is deformed towards image 0 with the computed initial mapping (the intensities are of image 1 and are not mixed). Total mass equality is preserved by multiplying by the integrals ratio.



Result of run_tests(3). Image 1 is deformed according image 0 with the computed initial mapping (the intensities are of image 1 and are not mixed). Total mass equality is preserved by histogram equalization method.

Note that the method used to fulfill the “same total mass” property has an effect on the deformation outcome.

3.2 Computing the optimal MP map

Implementation

The method was implemented as in the text. Next, we shortly describe the implemented modules.

<u>A Reading an image</u>

imread_to_gray.m – reads indexed or RGB images and converts them to a double precision grey level image.

B Creating the density maps
For efficiency, the images are divided into a number of blocks of a predefined size (a parameter of the method) – see compute_density_map.m.

C Computing initial mapping
As described previously this is done with compute_initial_mapping.m

D Computing the optimal mapping
The optimal map is computed at compute_optimal_mass_transport.m
The method implements the described paper. Two parameters of the methods are: 1. is it pure optimal mass transport or includes a comparison term, and 2. what method is used to fulfil "same total mass" property (multiplication by the integrals ratio or histogram equalization). After reading the images and computing an initial map, it calls to the gradient descent method (see gradient_descent.m). At each iteration u_t and the time step size are computed (see compute_ut.m).

The Jacobian elements, the gradient and the divergence are computed using standard central differences (I describe the Jacobian is a diagonal matrix), see modules: compute_gradient.m, compute_devergance.m and compute_jacobian, respectively.
We used poicalc (MATLAB function) to solve the Poisson's equation.

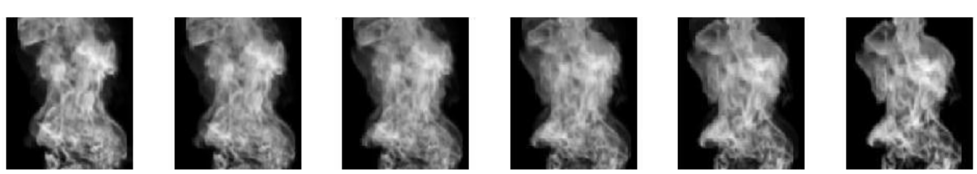
E Warping the images and presenting the results
Two warping methods are implemented to warp images based on the resulted optimal mapping between two images. The first method (transform.m) deforms an image with its own intensity values (without mixing intensity values of the two images), and the second method warps also the intensities (transform_intensity.m).

I also programmed a method to create a series of warped images (see create_image_series.m), and to superimpose a deformation map on an image (see superimpose.m).

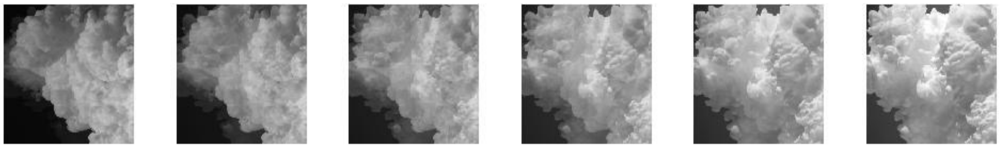
Experiments

Three pairs of cloud, flame and water images are tested. For each pair the optimal mass transport map was computed and intermediate images were created. (run: run_tests(4), run_tests(5), run_tests(6))
In the last experiment we have examined the effect of the "same total mass" property method (integral ratio vs. histogram equalization).

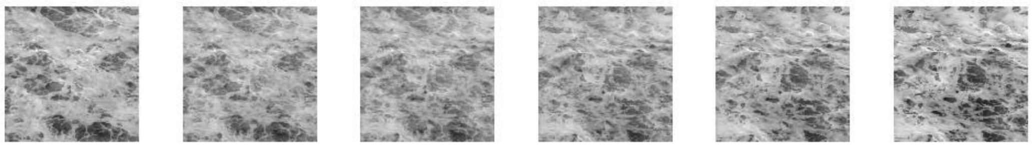
Results



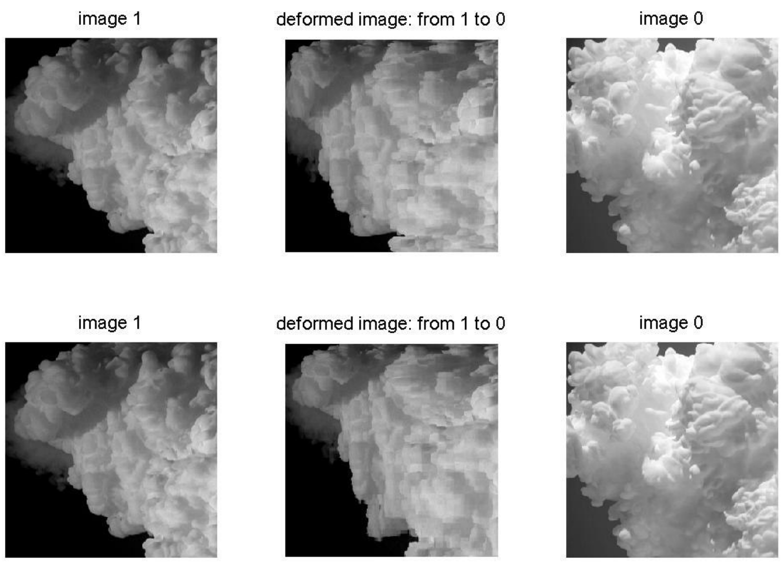
Experiment 4 (includes intensities mixture, result of run_tests(4)) .



Experiment 5 (includes intensities mixture, result of run_tests(5)).



Experiment 6 (includes intensities mixture, result of run_tests(6)).



Experiment 7
In each image series the left image is the source, the right is the target and the middle image is the source image deformed towards the target with its own intensity values (no intensity mixture is allowed between the images). When the "total mass" assumption is fulfilled by simply multiplying by the integrals ratio the result was better (upper series), than the histogram equalization method.

4 Conclusions

We have implemented a warping method which is using the properties of elastic materials to compute the deformation between two images. The proposed elastic registration method incorporates the L2 Kantorovich-Wasserstein distance, also known as the Earth Mover's Distance (EMD), as a similarity measurement. The implemented paper presents an efficient partial differential equation approach for a first-order solution of this problem that is simpler than existing works suggested high-order solutions and is computationally simpler than existing works based on linear programming. The method was verified and tested and usage examples are provided in run_tests.m.

References:

[1] Steven Haker, Lei Zhu, Allen Tannenbaum, and Sigurd Angenent, "Optimal mass transport for registration and warping", *International Journal of Computer Vision*, 60(3), p. 225-240, 2004.