Membrane Nonrigid Image Registration

Geoffrey Oxholm and Ko Nishino

Department of Computer Science
Drexel University
Philadelphia, PA

Abstract. We introduce a novel nonrigid 2D image registration method
that establishes dense and accurate correspondences across images with-
out the need of any manual intervention. Our key insight is to model
the image as a membrane, i.e., a thin 3D surface, and to constrain its
deformation based on its geometric properties. To do so, we derive a
novel Bayesian formulation. We impose priors on the moving membrane
which act to preserve its shape as it deforms to meet the target. We derive
these as curvature weighted first and second order derivatives that corre-
spond to the changes in stretching and bending potential energies of the
membrane and estimate the registration as the maximum a posteriori.
Experimental results on real data demonstrate the effectiveness of our
method, in particular, its robustness to local minima and its ability to
establish accurate correspondences across the entire image. The results
clearly show that our method overcomes the shortcomings of previous
intensity-based and feature-based approaches with conventional uniform
smoothing or diffeomorphic constraints that suffer from large errors in
textureless regions and in areas in-between specified features.

1 Introduction

The goal of nonrigid image registration is to align a template image to a refer-
ence image by locally deforming the template image. Modeling nonlinear, local
deformations has important applications in many computer vision problems in-
cluding image stabilization [1], subject tracking [2,3], and medical imaging [4],
to name a few.

There are two primary approaches to nonrigid image registration: intensity-
based and feature-based. Intensity-based approaches [5,6,7] attempt to minimize
the intensity differences across the entire image. Such methods produce dense
correspondences but suffer from ambiguities arising from similar intensity re-
gions. Feature-based methods [8,9,10] compute deformations that align a sparse
set of specifically selected features. These points are then used in conjunction
with a parametric model to interpolate the recovered deformations across the
rest of the image. In addition to the separate challenge of detecting and matching
good features (which often relies on manual intervention), the overall quality of
the registration directly relies on the interpolation method. Consequently, accu-
racy inherently decays rapidly as the distance from the feature points increases.

In this paper, we introduce an automatic nonrigid 2D image registration method that establishes dense and accurate correspondences across the entire image without the need to provide feature correspondences a priori. Our key idea is to model the image as a 2D membrane embedded in a 3D spatial-intensity space. We then formulate nonrigid image registration as the process of aligning two membranes by deforming one to the other while preserving its local geometric structures. In particular, we model the elastic and bending potential energies of the membrane. By penalizing their changes, the local structures of the template membrane are preserved as it deforms to meet the reference membrane.

We derive a probabilistic formulation of this membrane nonrigid image registration. We model each template image point as a Gaussian and seek the maximum a posteriori estimate of the template image as a mixture of Gaussians given the reference image. Our main contributions are a newly derived likelihood and priors that reflect physically-motivated constraints on the membrane geometry:

Novel likelihood: We construct a Gaussian at each pixel of the template image scaled by the membrane's original curvature at that point. This naturally encodes the significance of the underlying image structure, which in turn encourages features to align with corresponding features.

Bending energy: We model the inherent flexibility of a membrane by penalizing local surface deformations in proportion to the membrane's original curvature. This corresponds to minimizing the change in potential bending energy which translates into a novel curvature-weighted second order derivative prior.

Stretching energy: We model the inherent elasticity of a membrane by penalizing surface stretching and compression. This corresponds to minimizing the change in potential elastic energy across the membrane which translates into a novel first order derivative prior.

Intuitively, this formulation leads to surface regions with prominent local structures (features of the membrane) to be preserved and aligned with each other while more smooth regions are allowed to deform more flexibly. By preserving the shape of the membrane features, their appearance in the image being modeled remain true to their underlying geometry.

We demonstrate the accuracy and effectiveness of our method on 2D slices of real brain MRIs and images of faces with different expressions. In particular, we show that in addition to a significant decrease in overall intensity error, our method establishes accurate correspondences of prominent image structures automatically. This has strong implications in various applications since local image structures usually correspond to meaningful geometric structures of the imaged scene or object, and accurately aligning such structures is of great importance.

2   Related Work

Nonrigid image registration has been a popular area of research. Here we focus on methods that specifically address the shortcomings of both intensity-based and feature-based methods. We refer the reader to surveys of the rich literature [11, 12, 4] for more thorough context.

Fischer and Modersitzki [13] combine the two approaches on a sliding scale. They initially register a set of manually established features, then incrementally shift towards a uniform intensity-based metric. Our curvature-scaled objective function has a similar effect in that it encourages the rapid registration of feature rich areas. It does so, however, without requiring predefined features or by giving priority to the registration of any subregion.

Fischer and Modersitzki [14] also introduced a "curvature-based" normalization term that encourages locally smooth deformations by penalizing sharp changes in the displacement field. Although we also describe our bending energy constraint term as "curvature-based," the two approaches are fundamentally different. Whereas their normalization term is a second-order derivative of the 2D displacement field, we impose an energy minimization prior on the membrane, i.e., the image modeled as a 3D spatial-intensity surface. This added dimension allows us to impose geometrically-induced constraints on the image deformation.

Intensity-based methods assume that corresponding regions in the imaged scene maintain the same intensity pattern in both images. Previous authors [15, 16] have noted that this assumption can lead to violations of the basic physical properties of the subject which are present despite changes in illumination. To address this they use mass or volumetric constraints specific to their given applications. More general methods like Thirion's Demons method [5] and the recent diffeomorphic extension of this work by Vercauteren et al. [7] smooth the 2D deformation field thereby preventing large feature displacements from tearing or folding the deformation field. Although smooth deformation fields are found, ambiguities arising from similar intensity patterns of non-corresponding regions result in undesirable non-local artifacts. In our physically motivated model, we avoid such local minima by preserving the shape of the image membrane thereby maintaining local structures as they move across the image. To address folding we introduce a novel prior which allows pixels to come quite close to each other without overlapping. This allows us to model the common physical occurrence of creasing which is impossible under the various smoothing models.

Recently, probabilistic formulations of nonrigid image registration have gained further attention. Jian and Vemuri [17] use a Gaussian mixture model to register two point sets by placing a Gaussian at each point. Our work is most closely related to the extension of this approach by Myronenko et al. [6] that formulates image registration as a Gaussian mixture estimation with Gaussians centered on each pixel. By placing quadratic priors on each Gaussian, they preserve the distance of each pixel to its neighbors thereby avoiding tearing and folding of the deformation field. This results in a locally smooth deformation with well-minimized intensity distance on synthetic deformations. Unavoidably, however, these priors perform less well on real-world data which exhibit more complex transformations that cannot be modeled with assumptions of smoothness. Since accurate correspondences are of primary concern in many applications, we show that the minimization of an intensity distance is an insufficient objective function without shape preserving constraints.

3   Bayesian Membrane Registration

We model the image as a 2D membrane in a 3D space. In order to ensure this membrane approximates the actual imaged surface, the intensities are normalized and the height of each pixel is set proportionately to the log of the normalized intensity.[1] As noted by Koenderink and van Doorn [18], by using the log-intensity we ensure a geometrically invariant intensity encoding which eliminates any effect intensity magnitude may otherwise have on our geometric constraints while simultaneously achieving a degree of symmetry between the Cartesian pixel coordinates and heights of the points on the membrane.
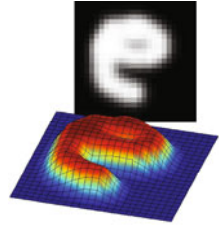
More precisely, we view the image coordinates $\widehat{\mathbf{x}} = (\widehat{\mathbf{x}}_u, \widehat{\mathbf{x}}_v)$ and scaled logarithm of the normalized intensity of each point $I(\widehat{\mathbf{x}})$ together as points $\mathbf{x} = (\widehat{\mathbf{x}}_u, \widehat{\mathbf{x}}_v, I(\widehat{\mathbf{x}}))$ of a 2D membrane in a 3D space. In many cases we may assume that this membrane reflects the geometry of the imaged object. For instance, a Lambertian surface would have its normals roughly encoded in its shading and the intensity in medical images reflects the density of the subject.

Similar to Myronenko et al. [6], we formulate nonrigid image registration as a MAP estimation of a product of Gaussian mixture densities. The posterior, representing the probability of the template image $\mathbf{Y}$ given the reference image $\mathbf{X}$ and parameters θ, is formulated as

$$p(\mathbf{Y}|\mathbf{X}, \theta) \propto p(\mathbf{X}|\mathbf{Y}, \theta)p(\mathbf{Y}|\theta) \ , \tag{1}$$

where we assume uniform normalization $p(\mathbf{X})$. We have five parameters, $\theta = (\mathbf{Y}^0, \sigma_0, \beta_\mathrm{e}, \beta_\mathrm{b}, \beta_\mathrm{f})$, which we describe below. Here $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_N)^\mathrm{T}$ is an $N \times 3$ matrix containing the points in the reference membrane $\mathbf{x} = (\widehat{\mathbf{x}}_u, \widehat{\mathbf{x}}_v, I(\widehat{\mathbf{x}}))$ and $\mathbf{Y} = (\mathbf{y}_1, \ldots, \mathbf{y}_M)^\mathrm{T}$ is an $M \times 3$ matrix containing the final locations of the registered template membrane's points $\mathbf{y} = (\widehat{\mathbf{y}}_u, \widehat{\mathbf{y}}_v, I(\widehat{\mathbf{y}}))$. We denote the original, undeformed template membrane as $\mathbf{Y}^0 = (\mathbf{y}_1^0, \ldots, \mathbf{y}_M^0)^\mathrm{T}$. N and M are the number of pixels in the images (which need not be equal in size). We model the likelihood as a product of N independent Gaussian mixture densities $p(\mathbf{X}|\mathbf{Y}, \theta) = \prod_n p(\mathbf{x}_n|\mathbf{Y})$. Building on this formulation, we introduce a curvature-based scaling to each point as we discuss next.

Our key contributions lie in the three priors on the template membrane $\mathbf{Y}$,

$$p(\mathbf{Y}|\theta) \propto \exp\left(-\beta_\mathrm{e}\mathcal{E}(\mathbf{Y}) - \beta_\mathrm{b}\mathcal{B}(\mathbf{Y}) - \beta_\mathrm{f}\mathcal{F}(\mathbf{Y})\right) \ . \tag{2}$$

The first, $\mathcal{E}(\cdot)$, quantifies the amount of change in elastic potential energy in the membrane. The second, $\mathcal{B}(\cdot)$, quantifies the change in bending potential energy in the membrane. Finally, $\mathcal{F}(\cdot)$ quantifies the amount of folding, or overlap, in the membrane. Each function is weighted by a parameter $\beta_{\{\mathrm{e,b,f}\}} \in \theta$. We will now describe each component of the posterior in more detail.

---

[1] Results are consistent so long as the images are normalized consistently. Since this formulation is geometrically invariant, the scale only effects the convergence rate.
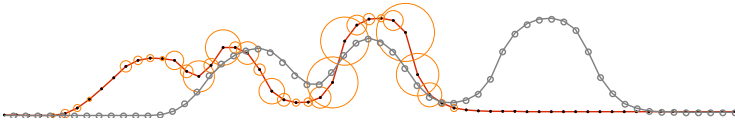
Fig. 1. Shown in this 1D example, a Gaussian is established for each pixel of the template image (solid) with standard deviation (circles) proportional to the curvature at that point. This allows prominent local structures that usually have high curvature to travel further and align with corresponding structures of the reference image (dotted) while preserving their shapes as modeled by their elastic and bending energies.

3.1 Gaussian Mixture Likelihood

The mixture density $p(\mathbf{x}_n)$ for a pixel of the reference image $\mathbf{x}_n$ is expressed probabilistically as a Gaussian mixture where each point of the template membrane is expressed as its own Gaussian distribution

$$p(\mathbf{x}_n|\mathbf{Y}) = \sum_{m=1}^{M} \frac{1}{M}\mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m) \ . \tag{3}$$

Observing that regions with prominent local structures (features) are more indicative of the membrane's overall shape, we allow points in these regions a larger range of motion by scaling the Gaussian centered at each point by the membrane's original curvature at that point. Using the squared mean curvature $H^2(\mathbf{y}_m^0)$ we model this with a per-point mean and standard deviation of

$$\boldsymbol{\mu}_m = \mathbf{y}_m, \qquad \boldsymbol{\Sigma}_m = (H^2(\mathbf{y}_m^0)\sigma_0)^2\mathbf{I}_3 \ , \tag{4}$$

where $\mathbf{I}_3$ is the identity matrix as each image dimension is statistically independent. These feature rich areas maintain their shape due to increased rigidity constraints (discussed in the next section). Intuitively, this leads to feature-rich surface regions to be preserved and aligned with each other, guiding the registration of the rest of the membrane.

   We can then express the likelihood across the entire image as an unweighted product of these Gaussian mixture densities

$$p(\mathbf{X}|\mathbf{Y}, \sigma_0) \propto \prod_{n=1}^{N} \sum_{m=1}^{M} \exp\left[-\frac{1}{2}\left\|\frac{\mathbf{x}_n - \mathbf{y}_m}{H^2(\mathbf{y}_m^0)\sigma_0}\right\|^2\right] \ . \tag{5}$$

In other words, for a given scale parameter σ_0, the final pixel locations $\mathbf{Y}$ that maximize this likelihood represent the deformation that maps the points of the template membrane to regions of the reference membrane.

   In Fig. 1 we show a simple one-dimensional, $(x, I(x))$, example where the initial template $\mathbf{Y}^0$ is shown in red, and the reference $\mathbf{X}$ is shown in gray. The relative standard deviations of the Gaussians are shown as orange circles. As shown in Fig. 2, this increase in the search space for key regions of the curve is necessary to avoid local minima and preserve the geometry of membrane features.
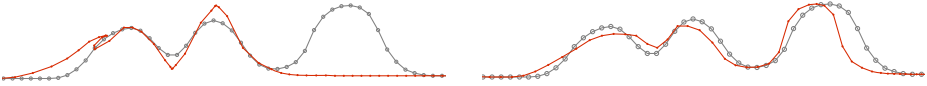
Fig. 2. Two results after registering the curves of Fig. 1 are shown (solid) relative to the target curve (dotted). Standard smoothing priors [6] (*left*) can cause local minima to be found. Here an entire peak is unregistered while two peaks have collapsed into one. Imposing our physically-based constraints (*right*) ensures that the structure of the entire curve is maintained during deformation resulting in a more accurate registration.

The shape of the deformed membrane must now be considered. Without constraints on the deformation, the pixel locations can be permuted at will to maximize the likelihood. To ensure an accurate deformation, we introduce physically-motivated priors that operate on the local geometry of the membrane.

3.2   Shape-Preserving Priors

The membrane model of an image allows us to incorporate physically-based constraints that preserve the local intensity structures of the image as it deforms. In particular, we model the elastic and bending potential energies of the membrane and impose geometric constraints that minimize the changes in these energies.

Elastic Energy. The elastic energy of a deformation captures the change in elastic potential as the membrane deforms. We define this energy as the sum of the elastic energy across all points $\mathcal{E}(\mathbf{Y}) = \sum_m \mathcal{E}(\mathbf{y}_m)$. We define the elastic energy at a point as the change in elastic potential energy at that point $\mathbf{y}$ relative to the potential at that point in the original membrane $\mathbf{y}^0$. We evaluate the potential of a point on a membrane using Hooke's law $E = \frac{1}{2}kx^2$. By assuming the elastic constant (k in Hooke's law) is uniform across the membrane we let $\beta_{\mathrm{e}} = (k/2)^2$ which is then used to weight the entire energy term. The relative displacement (x in Hooke's law) at each point naturally corresponds to the total change in distance to the point's neighbors $\mathrm{ne}(\mathbf{y})$.

By squaring the difference in potential of the relaxed and deformed membranes, we naturally quantify the amount of elastic energy at each point as

$$\mathcal{E}(\mathbf{y}) = \sum_{\mathbf{y}_i \in \mathrm{ne}(\mathbf{y})} (\left\|\mathbf{y}_i - \mathbf{y}\right\|^2 - \left\|\mathbf{y}_i^0 - \mathbf{y}^0\right\|^2)^2. \tag{6}$$

Note that because the intensity of a pixel does not change this reduces to

$$\mathcal{E}(\mathbf{y}) = \sum_{\mathbf{y}_i \in \mathrm{ne}(\mathbf{y})} (\left\|\widehat{\mathbf{y}}_i - \widehat{\mathbf{y}}\right\|^2 - 1)^2. \tag{7}$$

This prior differs considerably from stretching or elastic constraints of past work. Specifically, the first-order smoothing terms used in past work impose smoothing on the 2D deformation field itself, necessarily resulting in overly smooth local deformations.

Bending Energy. We also model the bending potential energy of the membrane and derive an energy term which quantifies the change in this potential as the membrane deforms. We define the total bending energy as the sum across all points, $\mathcal{B}(\mathbf{Y}) = \sum_m \mathcal{B}(\mathbf{y}_m)$. Our bending potential function is based on the Willmore energy $\int_S \frac{1}{2}H^2 - KdA$, where H is the mean curvature function and K is the Gaussian curvature function. By the Gauss-Bonnet theorem K is a topological invariant, and so remains constant during the deformation. Since we are concerned with the change in this energy, this term cancels out. We extend the Willmore energy to include the inherent rigidity of structural features by considering the potential of each point separately.

Whereas homogeneous membranes have uniform elasticity, the flexibility of a membrane varies with the curvature of the undeformed surface [19]. This translates to weighting the bending energy with a per-point rigidity coefficient equal to the squared mean curvature of the undeformed membrane at that point $H^2(\mathbf{Y}^0)$. This term also provides robustness to noise since a corrupt pixel will yield a high curvature value at that point. Since mean curvature is computationally expensive, we use the Laplacian $\Delta(\cdot)$ as an approximation for $H^2(\cdot)$ when computing the change in energy [20]. We define the bending energy as the weighted squared change in bending potential

$$\mathcal{B}(\mathbf{y}) = H^2(\mathbf{y}^0)\left(\Delta(\mathbf{y}) - \Delta(\mathbf{y}^0)\right)^2. \tag{8}$$

At a given point, the Laplacian of a surface is expressed using the (log) intensity heights $I(\cdot)$ of the point $\mathbf{y} = (\widehat{\mathbf{y}}_u, \widehat{\mathbf{y}}_v, I(\widehat{\mathbf{y}}))$ and its negative direction and positive direction neighbors $\mathbf{y}_-$ and $\mathbf{y}_+$ respectively

$$\Delta(\mathbf{y}) = \left(\frac{h_-(I(\widehat{\mathbf{y}}_+) - I(\widehat{\mathbf{y}})) - h_+(I(\widehat{\mathbf{y}}) - I(\widehat{\mathbf{y}}_-))}{h_+ h_- h_\pm}\right)^2, \tag{9}$$

where the distance to the positive direction neighbor h_+ the negative direction neighbor h_- and the distance between midpoints $h_\pm$ are used

$$h_+ = \|\widehat{\mathbf{y}}_+ - \widehat{\mathbf{y}}\| \;, \quad h_- = \|\widehat{\mathbf{y}} - \widehat{\mathbf{y}}_-\| \;, \quad h_\pm = \|[(\widehat{\mathbf{y}}_+ + \widehat{\mathbf{y}}) - (\widehat{\mathbf{y}} + \widehat{\mathbf{y}}_-)]/2\| \;. \tag{10}$$

As a time saving approximation we assume $h_+ = h_- = h_\pm$. We also note that the numerator is equal for $\Delta(\mathbf{y})$ and $\Delta(\mathbf{y}^0)$ since the intensities of the pixels do not change. Further, we note that $h_\pm^0$ is constant which allows us to reduce the horizontal bending penalization of Eq. 8 to

$$\mathcal{B}(\mathbf{y}) \propto H^2(\mathbf{y}^0)\left(h_\pm^{-2} - 1\right)^2. \tag{11}$$

For 2D images we consider horizontal, vertical, and two diagonal bending energies by formulating $\mathcal{B}_\rightarrow$, $\mathcal{B}_\downarrow$, $\mathcal{B}_\nearrow$, and $\mathcal{B}_\searrow$ analogously and take the sum

$$\mathcal{B}(\mathbf{y}) = \mathcal{B}_\rightarrow(\mathbf{y}) + \mathcal{B}_\downarrow(\mathbf{y}) + \mathcal{B}_\nearrow(\mathbf{y}) + \mathcal{B}_\searrow(\mathbf{y}) \;. \tag{12}$$

Folding Prior. During registration, regions of the deforming template membrane will expand and compress to meet the corresponding reference regions. As
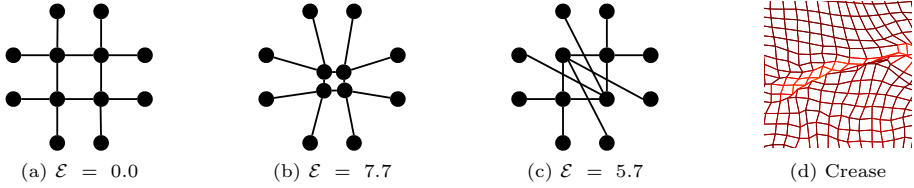
(a) $\mathcal{E} = 0.0$ (b) $\mathcal{E} = 7.7$ (c) $\mathcal{E} = 5.7$ (d) Crease

Fig. 3. The elastic penalization for areas of compression is minimized when neighboring surface patches fold over one another in featureless regions. We address this with an explicit prior on folding which allows for creases to form but eliminates folding.

shown in Fig. 3, since our elastic energy constraint encourages uniform spacing and our bending energy constraint applies primarily to feature rich areas, folding can occur. Although the bending prior discourages this in textured areas, it is not sufficient in relatively featureless regions.

Conventional methods decrease this by imposing second order derivative penalizations on the 2D deformation field [6, 14] or by specifically modeling diffeomorphic registrations [7]. Problems arise, however, in regions that change in size dramatically. As real-world objects inevitably experience such large deformations, a more accurate model should allow sharp boundaries in the deformation field as neighboring regions converge and creases form.

We allow such sharp boundaries to form with an explicit model of folding that allows pixels to come quite close to each other without penalty while strongly penalizing folding. We model this with a sigmoid function on each of the four neighboring directions of a point $\mathbf{y} = (\widehat{\mathbf{y}}_u, \widehat{\mathbf{y}}_v, I(\widehat{\mathbf{y}}))$. The folding energy of a deformation is then the sum across the deformation of each of these four values

$$\mathcal{F}(\mathbf{Y}) = \sum_{m=1}^{M} \left(\mathcal{F}_{\rightarrow}(\mathbf{y}_m) + \mathcal{F}_{\leftarrow}(\mathbf{y}_m) + \mathcal{F}_{\uparrow}(\mathbf{y}_m) + \mathcal{F}_{\downarrow}(\mathbf{y}_m) \right) . \quad (13)$$

For example, the right neighbor function is given by

$$\mathcal{F}_{\rightarrow}(\mathbf{y}) = \left(1 + \exp\{c(\widehat{\mathbf{y}}_u^+ - \widehat{\mathbf{y}}_u + t)\} \right)^{-1} , \quad (14)$$

where $\widehat{\mathbf{y}}^+$ is the right neighbor of $\widehat{\mathbf{y}}$. We establish the other three functions similarly. In this formulation a sufficiently high value for c and low value for t effectively make this a step function that penalizes the folding of neighboring pixels while allowing pixels to form sharp boundaries without penalty.

3.3   MAP Estimation

Having formulated the likelihood and prior constraints, we may estimate the maximum a posteriori using energy minimization. Specifically, the log posterior

$$\log p(\mathbf{Y}|\mathbf{X}, \theta) = \sum_{n=1}^{N} \log \sum_{m=1}^{M} e^{-\frac{1}{2} \left\| \frac{\mathbf{x}_n - \mathbf{y}_m}{H(\mathbf{y}_m^0)\sigma_0} \right\|^2} - \beta_{\mathrm{e}} \mathcal{E}(\mathbf{Y}) - \beta_{\mathrm{b}} \mathcal{B}(\mathbf{Y}) - \beta_{\mathrm{f}} \mathcal{F}(\mathbf{Y}) + C \quad (15)$$

(a) Template   (b) Reverse deformations at low resolution

(c) Reference   (d) Full resolution registrations

(e) Difference   (f) Full resolution intensity difference

(g) Detail and true landmarks   (h) Detail of intensity difference and final landmark locations
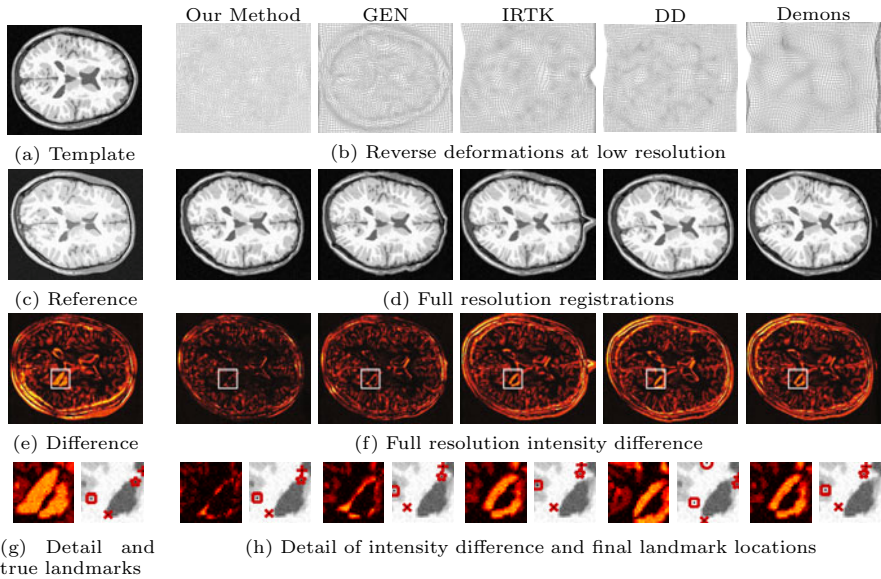
Fig. 4. Template and reference images (from BrainWeb [22]) are scaled down and registration is performed with various methods. The resulting reverse deformation grid (b) is applied to the original template image (a). These registrations (d) are subtracted from the reference image (c). The error is then visualized (f) and compared with the difference of the original template and reference images (e). Increased brightness corresponds to larger error. Detailed inspections (h) of a region requiring a large transformation (g) show that our method results in the least error both in terms of the intensity difference and in the alignment accuracy of features. This example is labeled "Brain1" in Fig. 8.

can be maximized using simulated annealing over the scale parameter σ_0 [6]. We vary σ_0 between σ_{max} and σ_{min} which depend only on the size of the images and are set automatically. The solution for each iteration is found with an interior trust region method [21]. In practice our rigidity constraints have proven robust to large values for σ_{max}. Typically $\sigma_{min} = 0.5$, $\sigma_{max} = 6$, and 6 annealing iterations are needed to converge for 100×100 images. We set $t = 1$ and $c = 5$ in our folding prior $\mathcal{F}$ to allow faster convergence of each annealing iteration. With this smooth penalization, however, resulting registrations occasionally have some amount of folding of the registration. To address this, after each annealing iteration points that have folded over each other are merged together. Unfortunately, each iteration is still quite computationally expensive, requiring as much as forty-five minutes in our current unoptimized implementation. We envision a significant speed-up with an approximate linearization of the objective function.

4   Experimental Results

We evaluate the accuracy of our model on human facial expressions and 2D slices of real brain MRIs. We compare the results with four characteristic automatic

methods. Rueckert et al. [23] introduced an intensity-based approach that uses b-splines to smooth the deformation field which they released as part of their Image Registration Toolkit (IRTK). Thiron's well known Demons method [5], uses gradient information from the reference image to determine the amount of force the deforming points must exert. This work was later extended by Vercauteren et al. [7] to specifically model diffeomorphisims in a model termed Diffeomorphic Demons (DD). Finally, we compare our work to the generalized elastic net (GEN) model of Myronenko et al. [6] that uses a Gaussian mixture formulation similar to ours but with conventional smoothing priors. When possible, we use publicly available implementations of these algorithms with default parameters.

Past work use synthetic deformations to compare their results to ground-truth deformations. Synthetic deformations, however, are generated without regard for the physical structure of the image subject and therefore provide little information about real-world accuracy. Instead, we observe that an ideal registration should conform to the structural properties of the imaged subject. A deformation field embodying this characteristic should therefore maintain accuracy even when applied to a higher resolution image. At this increased resolution we may then compare the intensity error as well as the locations of manually labeled feature points to test sub-pixel accuracy. What may be termed "under-fitting" or "over-fitting" occurs when a deformation field appears well-suited at one resolution, but reveals significant inaccuracy at higher resolutions.

In Fig. 4 we compare the results of our method on a subject from the Brain-Web database [22] with the other methods. A lower resolution version of the template image (4a) is registered to an equally down-sampled reference image (4c). The resulting inverse deformation fields (4b) show where each pixel in the resulting registration originated. The resulting high resolution registrations (4d), formed using a bilinearly interpolated inverse deformation field, are then compared with the reference image (4c) and the absolute difference is visualized as a heat map (4f) in which the brightness of the pixel increases as the error increases. Our approach produces a significantly improved registration, as evident by the greater amount of black (4f). Closer inspection (4h) shows the feature alignment accuracy of our method as evident by the close proximity of the feature points to key anatomical landmarks (4g). Here we also see that the error for this region is less than the interpolation scale (which is 3 in this case), revealing the degree of sub-pixel accuracy of our method. This example is labeled "Brain1" in Fig. 8. Note that we achieve a minimum 29% decrease in overall intensity error, while achieving a 29% decrease in feature alignment error.

Fig. 5 details the importance of shape preservation. When using a smoothing model, a landmark that was originally at the tip of a long feature looses this distinction and becomes embedded in a mass. With our method, the local geometry of the feature is preserved and a more accurate registration is achieved.

In Fig. 6 we qualitatively compare the registrations a neutral face (6a) to a smiling face (6b) from the Japanese Female Facial Expressions (JAFFE) database [24]. In analyzing the deformation fields we find that the key challenge in expression registration is the sudden appearance of dark regions that were not
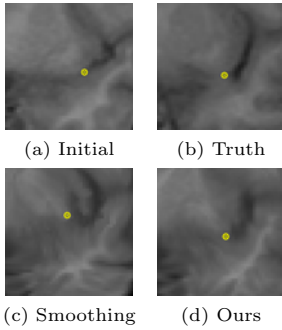
(a) Initial   (b) Truth

(c) Smoothing   (d) Ours

Fig. 5. Our shape preserving priors (d) ensure that meaningful correspondences are made as compared to smoothing models [5] (c)

(a) Template   (b) Reference   (c) Our method

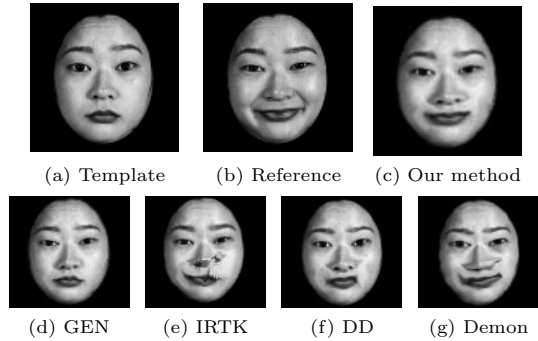(d) GEN   (e) IRTK   (f) DD   (g) Demon

Fig. 6. Faces (from JAFFE [24]) present a particular challenge due to dramatic local deformations and intensity variations in corresponding regions like the creases of a smile. Our method outperforms past work by preserving the shape of the features as they deform. This example is "Face1" in the graphs of Fig. 8.

previously present. In this example the formation of a smile introduces dramatic changes in brightness in the cheeks as creases appear. This causes gross deformations to result in the other, less-structured models. Our method, on the other hand, achieves a much more accurate registration across the entire face. Although it is not possible to recreate the creases without changing the intensity of the pixels, the shape of the lips and raise in cheeks are captured well.

Three more datasets are shown in Fig. 7. The first example shows the registration of a neutral face to a frown. Note how as the upper lip compresses, a crease forms in the chin. The co-appearance of these dark regions above and below the original lip creates a challenging ambiguity. IRTK tries to split this dark region between the lip and chin whereas the Diffeomorphic Demons method shifts the mouth down to meet the chin crease. Our method achieves an accurate result by maintaining the membrane geometry of the whole mouth as it stretches and curves down. The second example shows the registration of a neutral face to a sad face. This subtle expression demonstrates the key criticism of intensity-based methods – despite the reasonable appearance of these results each method fails to accurately align key landmarks as evidenced by the bright sections of error surrounding the facial features exhibited by every method except ours. Note also how the lower lip has been dramatically compressed in various methods to meet the highlight that shows up in the reference image. The final example shows two horizontal MRI slices of the same subject. Note how the top left portion of the ventricle has become almost completely occluded in the reference image. As is shown in the detail of Fig. 3d, our method correctly models this as a crease whereas the other models have no way to register a feature that has disappeared.

In Fig. 8 we quantitatively evaluate our results and compare them with these methods. For each dataset cluster the yellow column (labeled "None") shows the value of the error measures when no algorithm is used indicating the relative
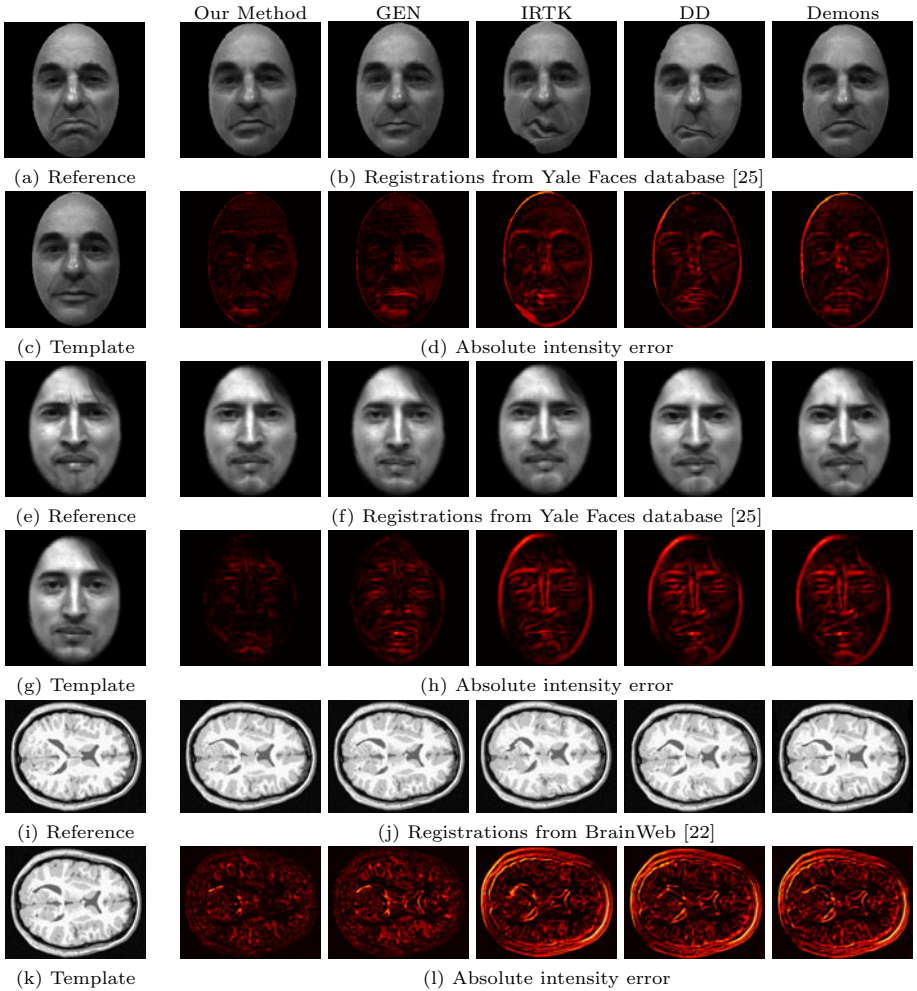
Fig. 7. Facial expressions and 2D MRI slices are registered using our method, GEN [6], IRTK [23], Diffeomorphic Demons (DD) [7], and Demons [5] methods respectively. These datasets are labeled Face3, Face4, and Brain4, in the graphs of Fig. 8

magnitude of the measures. In Fig. 8a we compare the mean squared error in intensity $[0, 1]$ of the registration. The results show that our method results in an average of 32% less error than the next best method. In Fig 8b we compare the mean error of final landmark locations (in pixels) from the manually labeled ground-truth locations. For each image pair we annotate between 12 and 27 primary feature correspondences. Here our method achieves an average of 53% less error in feature alignment than the next best method with values consistently below the interpolation scale.
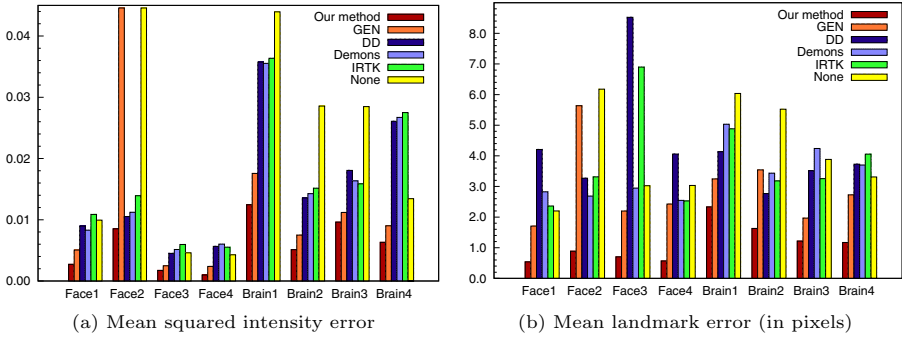
(a) Mean squared intensity error

(b) Mean landmark error (in pixels)

Fig. 8. In both graphs the yellow (right-most) bar of each grouping indicates the amount of error if no registration is performed. Our method consistently outperforms every other benchmark method.

5   Conclusion

Our method demonstrates considerable accuracy that results from our key assumption – that the image as a membrane in 3D spatial-intensity space approximates the actual surface of the subject and preserving its geometric shape reflects the true image deformation more accurately. Experimental results have shown that in many cases the assumption is valid and geometrically induced constraints increase accuracy dramatically. In particular, our method achieves higher accuracy in both the overall deformation and resulting feature correspondences. The resulting registrations exhibit a robustness to the common pitfalls of intensity-based registration techniques while maintaining particularly high accuracy for feature points automatically. This has strong implications in various applications where the accuracy of correspondences is particularly important.

References

1. Irani, M., Rousso, B., Peleg, S.: Recovery of Ego-Motion Using Image Stabilization. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 454–460 (1994)
2. Dedeoglu, G., Kanade, T., Baker, S.: The Asymmetry of Image Registration and its Application to Face Tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence 29, 807–823 (2007)
3. Weese, J., Penney, G., Desmedt, P., Buzug, T.M., Hill, D., Hawkes, D.: Voxel-Based 2-D/3-D Registration of Fluoroscopy Images and CT Scans for Image-Guided Surgery. IEEE Trans. on Info. Technology in Biomedicine 1, 284–293 (1997)
4. Maintz, J., Viergever, M.: A Survey of Medical Image Registration. ACM Computing Surveys 2, 1–36 (1998)
5. Thirion, J.P.: Non-rigid Matching Using Demons. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 245–251 (1996)

6. Myronenko, A., Song, X., Carreira-Perpinán, M.A.: Free-Form Nonrigid Image Registration Using Generalized Elastic Nets. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8 (2007)
7. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Diffeomorphic Demons: Efficient Non-parametric Image Registration. NeuroImage 45, S61–S72 (2009)
8. Wang, K., He, Y., Qin, H.: Incorporating Rigid Structures in Non-rigid Registration Using Triangular B-Splines. Variational, Geometric, and Level Set Methods in Computer Vision 3752, 235 (2005)
9. Kohlrausch, J., Rohr, K., Stiehl, H.: A New Class of Elastic Body Splines for Nonrigid Registration of Medical Images. Journal of Mathematical Imaging and Vision 23, 280 (2005)
10. Davis, M.H., Khotanzad, A., Flamig, D.P., Harms, S.E.: Elastic Body Splines: A Physics Based Approach to Coordinate Transformation in Medical Image Matching. In: IEEE Symposium on Computer-Based Medical System, vol. 8, p. 81 (1995)
11. Zitova, B., Flusser, J.: Image Registration Methods: A Survey. Image and Vision Computing 21, 977–1000 (2003)
12. Brown, L.G.: A Survey of Image Registration Techniques. ACM Computing Surveys 24, 325–376 (1992)
13. Fischer, B., Modersitzki, J.: Combination of Automatic Non-rigid and Landmark Based Registration: The Best of Both Worlds. Society of Photo-Optical Instrumentation Engineers 5032, 1037–1048 (2003)
14. Fischer, B., Modersitzki, J.: Curvature Based Image Registration. Journal of Mathematical Imaging and Vision 18, 81–85 (2003)
15. Haker, S., Tannenbaum, A., Kikinis, R.: Mass Preserving Mappings and Image Registration. In: Niessen, W.J., Viergever, M.A. (eds.) MICCAI 2001. LNCS, vol. 2208, pp. 120–127. Springer, Heidelberg (2001)
16. Haber, E., Modersitzki, J.: Volume preserving image registration. In: Barillot, C., Haynor, D.R., Hellier, P. (eds.) MICCAI 2004. LNCS, vol. 3216, pp. 591–598. Springer, Heidelberg (2004)
17. Jian, B., Vemuri, B.C.: A Robust Algorithm for Point Set Registration Using Mixture of Gaussians. In: IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, p. 1246 (2005)
18. Koenderink, J., van Doorn, A.: Image Processing Done Right. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002. LNCS, vol. 2350, pp. 158–172. Springer, Heidelberg (2002)
19. Grinspun, E., Hirani, A., Desbrun, M., Schröder, P.: Discrete Shells. In: ACM Special Interest Group on Graphics and Interactive Techniques, p. 67 (2003)
20. Wardetzky, M., Bergou, M., Harmon, D., Zorin, D., Grinspun, E.: Discrete Quadratic Curvature Energies. Comp. Aided Geom. Design 24, 499–518 (2007)
21. Coleman, T.F., Li, Y.: An Interior Trust Region Approach for Nonlinear Minimization Subject to Bounds. SIAM Journal on Optimization 6, 418–445 (1996)
22. Cocosco, C., Kollokian, V., Kwan, R., Pike, G.B., Evans, A.C.: Brainweb: Online Interface to a 3D MRI Simulated Brain Database. Functional Mapping of the Human Brain 5, 425 (1997)
23. Rueckert, D., Sonoda, L., Hayes, C., Hill, D., Leach, M., Hawkes, D.: Nonrigid Registration Using Free-Form Deformations: Application to Breast MR Images. IEEE Transactions on Medical Imaging 18, 712–721 (1999)
24. Lyons, M., Akamatsu, S., Kamachi, M., Gyoba, J.: Coding Facial Expressions With Gabor Wavelets. In: Face and Gesture Recognition, p. 200 (1998)
25. Yale: Face Database (1997), http://cvc.yale.edu/projects/yalefaces/yalefaces.html