Free-Form Nonrigid Image Registration Using Generalized Elastic Nets

Andriy Myronenko, Xubo Song and Miguel Á. Carreira-Perpiñán
Dept. CSEE, OGI, Oregon Health & Science University
{myron, xubosong, miguel}@csee.ogi.edu

Abstract

We introduce a novel probabilistic approach for nonparametric nonrigid image registration using generalized elastic nets, a model previously used for topographic maps. The idea of the algorithm is to adapt an elastic net (a constrained Gaussian mixture) in the spatial-intensity space of one image to fit the second image. The resulting net directly represents the correspondence between image pixels in a probabilistic way and recovers the underlying image deformation. We regularize the net with a differential prior and develop an efficient optimization algorithm using linear conjugate gradients. The nonparametric formulation allows for complex transformations having local deformation. The method is generally applicable to registering point sets of arbitrary features. The accuracy and effectiveness of the method are demonstrated on different medical image and point set registration examples with locally nonlinear underlying deformations.

1. Introduction

Image registration is an important problem in image analysis. It is a process of determining a geometric transformation that relates the contents of two images (the reference image and the source image) in a meaningful way and establishes the correspondence between them. Applications of image registration include combining images of the same subject from different modalities, aligning temporal sequences of images to compensate for motion of the subject between scans, image guidance during interventions and aligning images from multiple subjects in cohort studies [10, 13]. Nonrigid image registration is the most interesting and challenging work in registration today. Many nonrigid registration techniques have been proposed in the last 20 years [6, 13, 14]. Nearly all techniques are based on minimizing an objective function containing two terms: the first term measures the distance between the reference image and the registered (i.e., transformed) source image for a given transformation, and the second is a regularization term that encourages certain types of transformations (e.g. physically realizable ones). The registration is obtained by minimizing this objective over a suitable space of transformations.

Parametric methods describe the space of transformations in terms of a finite number of parameters. The most important parametric transformations use basis functions, in particular splines. Spline-based registration algorithms use control points in the source and target image and a spline function to define correspondences away from these points. In the popular thin-plate spline (TPS) [2], each control point influences the transformation globally, and fitting the TPS costs $\mathcal{O}(N^3)$ where N is the number of pixels. This limits the ability to model complex and localized deformations, unless many parameters are used at a high computational cost [6]. B-splines are only defined in the vicinity of each control point, so perturbing the position of one control point only affects the transformation near that point. However, B-splines sometimes require special care to prevent folding of the transformation [6, 16, 17]. Distances commonly used with parametric methods are the L_2 distance, the correlation and the mutual information. Parametric methods include landmark-based methods, where corresponding landmarks are determined in both images and then the transformation is fitted to interpolate the landmarks. While suitable for some applications, finding good landmarks is a crucial task that is often done manually by an expert.

Nonparametric methods do not assume a parametric form for the transformation. Nonparametric registration has mainly focused on variational approaches based on minimizing the L_2 distance. Here, one applies calculus of variations to the objective functional to derive the Euler-Lagrange equation (a system of PDEs for the transformation), which is solved numerically by discretizing the spatial domain and iteratively solving systems of equations. Depending on the regularization used, one obtains elastic, fluid, diffusion and curvature schemes [14]. Unlike for parametric methods, for variational methods the transformation is known only at the grid points and must be interpolated at other points. The spatial structure in the linear systems resulting from the discretization of the Euler-Lagrange equation allows the use of fast numerical schemes. The

1-4244-1180-7/07/$25.00 ©2007 IEEE

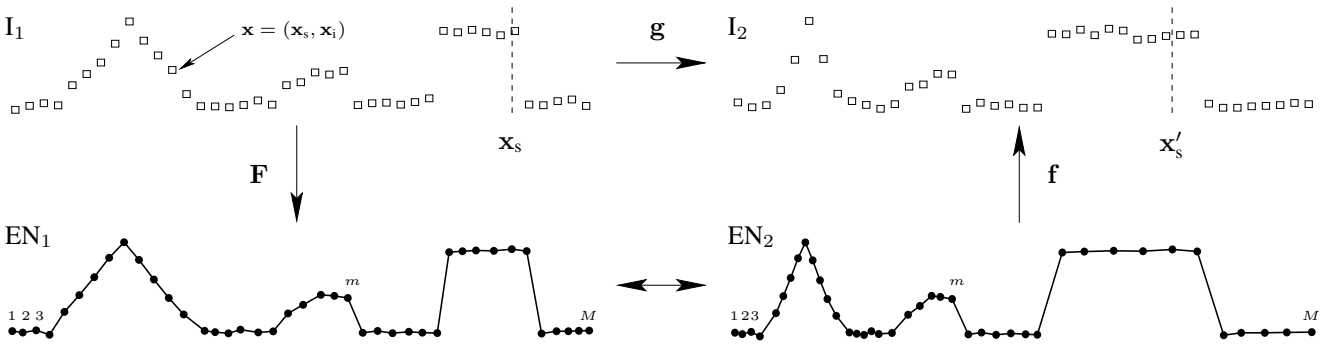

Figure 1. Illustration of the alignment method (for 1D images, for simplicity). I_1 represents a 1D intensity image in spatial (x-axis) and intensity (y-axis) space (each pixel is marked as a small □). I_2 represents the same image with local, nonlinear spatial distortion and intensity noise. EN_1 is an elastic net fitted to I_1 (in spatial-intensity space, with centroids marked •) and EN_2 is adapted from EN_1 to fit I_2. Since the centroids in EN_1 and EN_2 correspond one-to-one ($1 \leftrightarrow 1$, etc.), and the elastic net allows to define mappings between image points and centroids (see section 2), we can map any spatial location $\mathbf{x}_s$ in I_1 to a spatial location $\mathbf{x}'_s$ in I_2 through the elastic nets, thus aligning I_1 to I_2.

methods run in $\mathcal{O}(N \log N)$ (elastic, fluid and curvature) and $\mathcal{O}(N)$ (diffusion) per iteration, where N is the number of pixels.

Finally, a third class of methods, often called point-set registration methods, consider the registration of two sets of feature points (see e.g. [1, 5, 11, 18, 19]). These feature vectors may be the pixel feature vectors (consisting of the pixel spatial location and their intensity, color, etc.), or they may be a smaller set of other features extracted (manually or automatically) from the image, such as edges or landmarks. These methods bring into play not only the transformation (which is typically assumed parametric) but also the correspondences between pixels (or points) in both images. Our method is most closely related to point-set registration methods. We defer a detailed comparison with them to section 4.

In this paper we introduce a probabilistic, nonparametric registration method that can deal with nonrigid deformation of arbitrary complexity. The approach is nonparametric in that we do not assume any parametric transformation. It is based on the elastic net (EN), which is a net of connected points which jointly and smoothly move in a high-dimensional space to model a data set. An energy function can be defined to trade off accuracy of the net fitting the data (fitness term) vs net continuity (tension term). The elastic net was originally introduced as a continuous optimization method for the traveling salesman problem [7, 8] and has also been successfully applied to modeling maps of primary visual cortex. However it has had a limited use in computer vision. A generalization of elastic nets to arbitrary quadratic tension terms was investigated in [3]. Here we adapt the generalized elastic net to represent image deformations. The intuition is to position a net according to the first image and then deform it to align with the second image. The deformation produced by the elastic net directly represents the transformation between the images. This is illustrated more specifically in Fig. 1. We consider an image as a noisy 2D manifold in the spatial-intensity space, i.e., each pixel is represented by a point $\mathbf{x} = (\mathbf{x}_s, \mathbf{x}_i) \in \mathbb{R}^3$ of spatial location $\mathbf{x}_s \in \mathbb{R}^2$ and intensity $\mathbf{x}_i \in \mathbb{R}$. We model this manifold in a probabilistic way with an elastic net EN_1, which allows to map any image point onto the net, and vice versa. We then adapt EN_1 for a given image I_1 to a new image I_2 in the spatial-intensity space; again this allows to map a net point onto image space and vice versa. The alignment mapping which maps a spatial location in I_1 to another spatial location in I_2 is obtained through the deformed elastic net. Thus, the idea is to characterize the image manifold and deform it (adapt it) to the reference image manifold, extracting in the process the transformation. We describe the generalized elastic net and its adaptation to image registration in section 2, give experimental results in section 3 and discuss them in sections 4–5.

2. Image registration with generalized elastic nets

Generalized elastic nets (GEN) The *elastic net* is a Gaussian mixture with a quadratic prior on its centroids [3, 7, 8]. The centroids implicitly represent a nonlinear, low-dimensional manifold that probabilistically models a high-dimensional data set $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_N)$ (expressed as a $D \times N$ matrix). Specifically, given a collection of M D-dimensional centroids $\mathbf{Y} = (\mathbf{y}_1, \ldots, \mathbf{y}_M)$ (expressed as a $D \times M$ matrix) and a scale parameter $\sigma \in \mathbb{R}^+$, consider a Gaussian-mixture density $p(\mathbf{x}) = \sum_{m=1}^{M} \frac{1}{M} p(\mathbf{x}|m)$ with $\mathbf{x}|m \sim \mathcal{N}(\mathbf{y}_m, \sigma^2 \mathbf{I}_D)$. A smoothing or neighborhood-preserving prior on the centroids is defined as $p(\mathbf{Y}; \beta) \propto \exp\left(-\frac{\beta}{2} \sum_m \|\mathbf{y}_{m+1} - \mathbf{y}_m\|^2\right)$ where β is a regularization

hyperparameter. Without the prior, the centroids could be permuted at will with no change in the model, since the variable m is just an index. The elastic net minimizes the energy function

$$E(\mathbf{Y},\sigma) = -\sum_{n=1}^{N} \log \sum_{m=1}^{M} e^{-\frac{1}{2}\left\|\frac{\mathbf{x}_n-\mathbf{y}_m}{\sigma}\right\|^2} + \frac{\beta}{2}\sum_m \|\mathbf{y}_{m+1}-\mathbf{y}_m\|^2 \quad (1)$$

which is derived from the log posterior $\log p(\mathbf{Y}|\mathbf{X},\sigma)$ of the full model (i.e., maximum-a-posteriori estimation). We call the first term the *fitness term*, arising from the Gaussian mixture $p(\mathbf{X}|\mathbf{Y},\sigma)$, and the second term the *tension term*, arising from the prior $p(\mathbf{Y})$. The elastic net was generalized in [3, 4] to accommodate general quadratic priors. The prior can be used to convey the topological (dimension and boundary conditions) and geometric (e.g. curvature) structure of a manifold implicitly defined by the centroids. The *generalized elastic net (GEN)* minimizes the energy function

$$E(\mathbf{Y},\sigma) = -\sum_{n=1}^{N} \log \sum_{m=1}^{M} e^{-\frac{1}{2}\left\|\frac{\mathbf{x}_n-\mathbf{y}_m}{\sigma}\right\|^2} + \frac{\beta}{2}\operatorname{tr}\left(\mathbf{Y}^T\mathbf{Y}\mathbf{S}\right). \quad (2)$$

Quadratic priors are considered of the form $\mathbf{S} = \mathbf{D}^T\mathbf{D}$, so that $\operatorname{tr}\left(\mathbf{Y}^T\mathbf{Y}\mathbf{S}\right) = \left\|\mathbf{D}\mathbf{Y}^T\right\|^2$ in terms of the Frobenius norm. The matrix $\mathbf{D}$ represents a discretized differential operator. For example (for a 1D net for simplicity, and using forward differences [3]), a first-order derivative results in a sum of squared lengths $\left\|\mathbf{D}\mathbf{Y}^T\right\|^2 = \sum_m \|\mathbf{y}_{m+1}-\mathbf{y}_m\|^2$ and approximates a penalty $\int \|\nabla \mathbf{y}\|^2$ over a continuous net $\mathbf{y}$ (with an infinite number of centroids). This corresponds to a matrix $\mathbf{D}$ where each row is a shifted version of $(-1\ 1\ 0\ 0\ldots 0)$, and it was the tension term used in the original elastic net (eq.(1)), penalizing stretching of the net. A second-order derivative results in $\sum_m \|\mathbf{y}_{m+2}-2\mathbf{y}_{m+1}+\mathbf{y}_m\|^2$, etc. By choosing $\mathbf{S}$ as an appropriate combination of differential operators we can impose a desired type of smoothness on the GEN (see [4] for a discussion of the effect of different derivatives on the maps of primary visual cortex). The resulting $\mathbf{S}$ has a sparse, banded structure. We consider open boundary conditions at the image boundaries. Fig. 1 schematically shows a 1D elastic net. For 2D nets, appropriate 2D finite differences are used (e.g. having 4 neighbors for first-order derivatives) and $\mathbf{S}$ has a block-banded structure.

Adaptation of the GEN It is possible to derive an EM algorithm to estimate $\mathbf{Y}$ and σ jointly, but the GEN is usually trained with a deterministic annealing algorithm, which serves as a coarse to fine scale strategy. This minimizes E over $\mathbf{Y}$ for fixed σ, starting with a large σ and tracking the minimum to a small value of σ. For constant σ, [3] used a fixed-point iteration to find stationary points of E:

$$\frac{\partial E}{\partial \mathbf{Y}} = -\frac{1}{\sigma^2}(\mathbf{X}\mathbf{W}-\mathbf{Y}\mathbf{G}) + \beta\mathbf{Y}\mathbf{S} = \mathbf{0} \implies \mathbf{Y}\mathbf{A} = \mathbf{X}\mathbf{W} \quad (3)$$

with weight matrix $\mathbf{W} = (w_{nm})$ and invertible diagonal matrix $\mathbf{G} = \operatorname{diag}(g_m)$

$$w_{nm} = \frac{e^{-\frac{1}{2}\left\|\frac{\mathbf{x}_n-\mathbf{y}_m}{\sigma}\right\|^2}}{\sum_{m'=1}^{M} e^{-\frac{1}{2}\left\|\frac{\mathbf{x}_n-\mathbf{y}_{m'}}{\sigma}\right\|^2}} \qquad g_m = \sum_{n=1}^{N} w_{nm}$$
$$\mathbf{A} = \mathbf{G} + \sigma^2\beta\mathbf{S}. \quad (4)$$

The weight w_{nm} is the responsibility $p(m|\mathbf{x}_n)$ of centroid y_m for generating point $\mathbf{x}_n$, g_m is the total responsibility of centroid $\mathbf{y}_m$, and the matrix $\mathbf{X}\mathbf{W}$ is a list of average centroids. We solve for $\mathbf{Y}$ in the system of eq. (3) and iterate, since $\mathbf{W}$ and $\mathbf{G}$ depend on $\mathbf{Y}$. In [3], the system (3) was solved using Cholesky factorisation. While this is robust and efficient (since it takes advantage of the sparsity structure of $\mathbf{S}$), here we use a faster method based on linear conjugate gradients (CG) [15]. Linear CG solves an $M\times M$ positive definite linear system in at most M steps, each costing $\mathcal{O}(M^2)$ for full $\mathbf{A}$ and $\mathcal{O}(k_1 M)$ for sparse $\mathbf{A}$ (with k_1 nonzeros per row), and has two important advantages: (1) we can initialize the linear CG from the previous $\mathbf{Y}$ value (which will be close to the solution) rather than solving each system anew, as Cholesky does; (2) we can run only a few linear CG steps and obtain an approximate but good enough solution rather than an exact, costly one. This considerably accelerates the overall annealed algorithm without sacrificing accuracy. We obtain further acceleration by truncating the Gaussian kernel so w_{mn} are nonzero only for the nearest neighbors (obtained for free from the pixel grid). For low-dimensional problems ($D \leq 3$) we could also use the fast Gauss transform [9] to compute $\mathbf{X}\mathbf{W}$ in $\mathcal{O}(D(N+M))$.

The computational cost of our method is then as follows. Building the $\mathbf{X}\mathbf{W}$ matrix takes $\mathcal{O}(DNM)$ if $\mathbf{W}$ is full and $\mathcal{O}(Dk_2M)$ if the weights are nonzero only for the k_2 nearest neighbors (on average). Each step of CG costs $\mathcal{O}(Dk_1M)$ if $\mathbf{A}$ has k_1 nonzeros per row. Thus the total cost is $\mathcal{O}(DM(k_1+k_2)k_3)$ if we run k_3 CG steps altogether (along the annealing schedule).

3. Experimental results

Registration We now show how the elastic net framework can be adapted to image registration. First, we represent two images I_1 and I_2 in the spatial-intensity space (thus, the feature vectors have $D = 3$ dimensions). Then we construct an elastic net with as many centroids as pixels in image I_1. This net $\mathbf{Y}$ is initialized with each centroid representing the spatial-intensity value of one pixel in

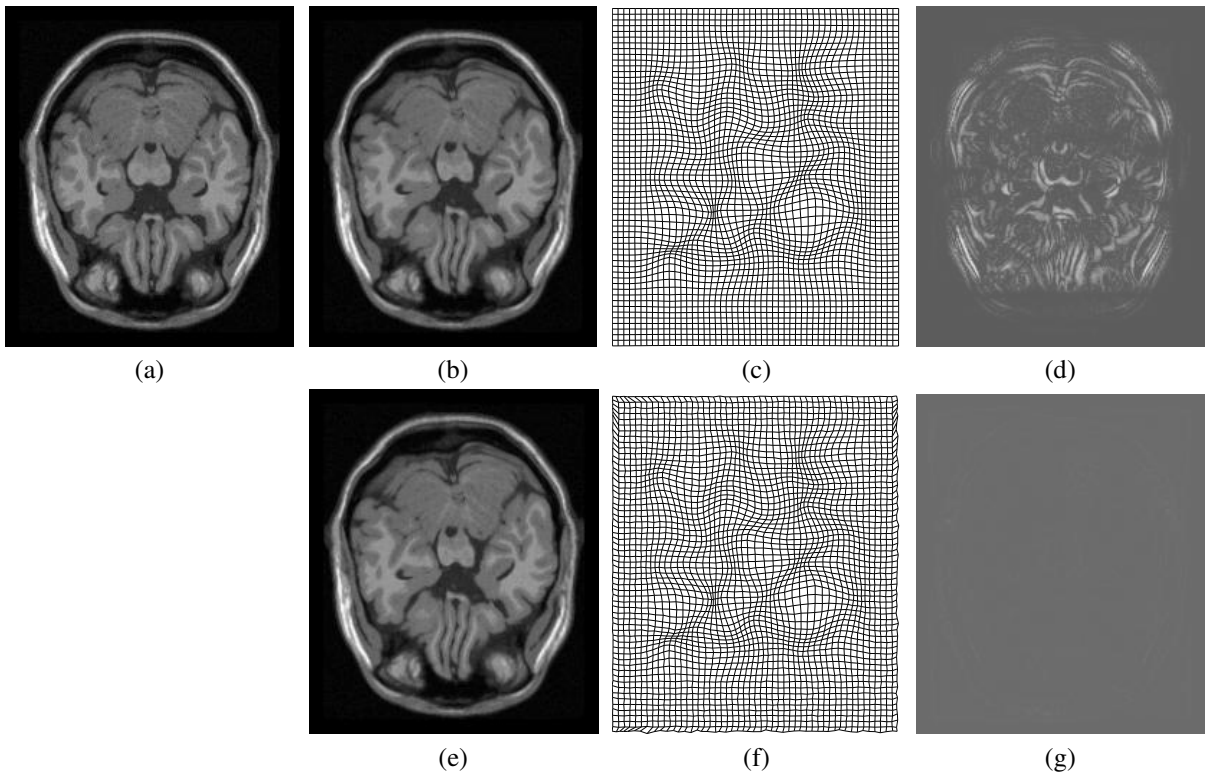

Figure 2. GEN registration with known deformation. (a) Original MR slice; (b) original image deformed using TPS; (c) transformation represented by TPS control points; (d) absolute value of intensity difference between original and deformed images; (e) registration result of the algorithm (original image is registered onto the deformed one); (f) transformation found by the algorithm (almost equal to the true one); (g) absolute value of intensity difference between registered and original deformed images (almost zero).

I_1 (i.e., $\mathbf{Y} = \mathbf{X}_1$). The net is adapted by adjusting the centroids to fit data $\mathbf{X}_2$, representing image I_2 in spatial-intensity space, by minimizing the energy (3). The reason for using the same number of centroids as there are pixels in I_1 is that the final centroid locations, when E is minimized, directly show the displacement of each pixel in I_1 when it is deformed into I_2. As a result, no interpolation is needed—unlike e.g. most variational methods, which need to interpolate the registered image after each iteration in order to compare it with the reference image. In general, we can choose to have more or fewer centroids than pixels. In this case the displacement of a pixel $\mathbf{x}$ in I_1 can be found by interpolation using the probabilities $p(m|\mathbf{x})$ and $p(\mathbf{x}|m)$ provided by the GEN.

We assume that the deformation between two images is only spatial, not in intensity. This translates to constraining the intensity components in the centroid vectors to be constant. In other words, the free parameters for centroid $\mathbf{y}_m = (\mathbf{y}_{ms}, \mathbf{y}_{mi})$ are $\mathbf{y}_{ms}$ only, and the optimization updates only apply to $\mathbf{y}_{ms}$. Doing so is important to produce only spatial deformation for I_1 when fitting it to I_2. In general, intensity variations across images can be accommodated by updating the complete $\mathbf{y}_m = (\mathbf{y}_{ms}, \mathbf{y}_{mi})$. We use the following penalty matrix: $\mathbf{S} = \beta_1 \mathbf{D}_1^T \mathbf{D}_1 + \beta_2 \mathbf{D}_2^T \mathbf{D}_2$, where $\mathbf{D}_1$ and $\mathbf{D}_2$ are first- and second-order derivatives, and their relative strengths are controlled by problem-dependent hyperparameters β_1 and β_2. We need the first and second derivatives to prevent tearing and folding, which are not physically reasonable in the registration of medical images [6].

In all experiments, the image intensities are first rescaled (to the range $[0, 15]$) to allow the use of a single σ for all dimensions, and the images are coarsely aligned using cross-correlation to eliminate translation (normalizing for rotation or affine transformation was not required for our examples). The resulting data sets $\mathbf{X}_1, \mathbf{X}_2 \subset \mathbb{R}^3$ were used to adapt the elastic net. The aligned dataset $\mathbf{X}_1'$ (obtained from the spatial deformation given by the GEN and the original intensity values) was postprocessed with bicubic interpolation to produce the aligned image. The hyperparameters β_1 and β_2 were set manually for each type of image (e.g. brain MRI images); once this is done, the same values can be used without further tuning for images of the same type. We ran 10 annealing iterations from $\sigma = 3$ to $\sigma = 0.5$ pixels, and about 60 CG steps in each iteration (resulting in a residual error $\|\mathbf{YA} - \mathbf{XW}\|$ for CG of around 10^{-1} pixels; this gave results as good as a residual of 10^{-5}). We truncated

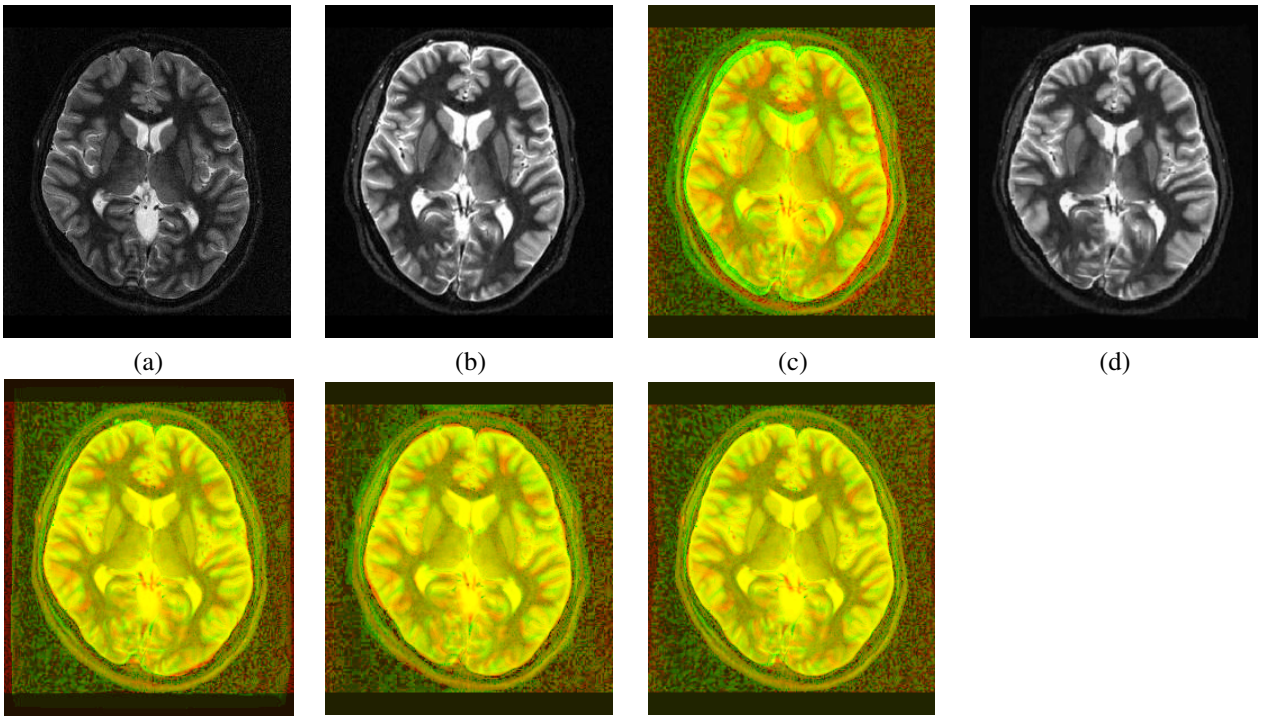

Figure 3. Registration with unknown deformation. (a) First person (reference image); (b) second person (template image); (c) composite of reference and template images; (d) registration of the template image onto the reference using GEN; (e) composite view after registration using GEN; (f) composite of the reference image and the registration result from [12]; (g) composite of the reference image and the registration result obtained with the Image Registration Toolkit [16] (please view this figure in color).

the Gaussian kernel to a width of 5σ (i.e., from 15 to 2.5 pixels).

We show results on artificial data with known nonlinear deformations and on two real-life examples. The algorithm was implemented in Matlab with subroutines in C, and tested on Pentium4 CPU 3.5GHz with 4GB RAM. The test images are 250×250 grayscale, and the registration takes about 20 minutes for each image pair. Validation of registration results on real-world images (for which no ground truth is available) is difficult, with most papers resorting to visual inspection [6, p. S145].

Table 1. Experimental results for different deformation levels.

Deform. STD	GEN: Transform RMSE (pixels)	GEN: Intensity RMSE	ITK: Intensity RMSE
1.0	0.3135	0.0044	0.0094
1.5	0.5124	0.0047	0.0104
2.0	0.9753	0.0053	0.0117
2.5	1.1152	0.0060	0.0129
3.0	1.0962	0.0059	0.0143

Brain MRI 2D images with and without known deformation A slice of MRI brain image was artificially deformed using a known transformation (Fig. 2). We define a uniform grid of control points in the original image, randomly move them and use thin-plate splines to create a locally nonlinear deformed image. Our algorithm is applied to align original image (a) onto deformed one (b). The final absolute image difference (g) is so small it is hardly visible, demonstrating the high accuracy of the method. Table 1 shows the value of root mean square error (RMSE) between true and estimated deformation as well as the intensity RMSE between original and registered images (image intensity were normalized to $[0, 1]$ interval), as a function of spatial distortion level controlled by the standard deviation (STD) of control points perturbation measured in pixels. The transformation error is at most of the order of one pixel. For comparison, we also show the intensity RMSE achieved with Image Registration Toolkit (ITK) [16]. Although ITK required heavy manual tuning, its intensity RMSE is twice more than our GEN intensity RMSE.

Figure 3 shows the images (a) and (b) from two patients, and its misalignment (c). To compare two images, we use a color composite view of the two histogram-equalized im-

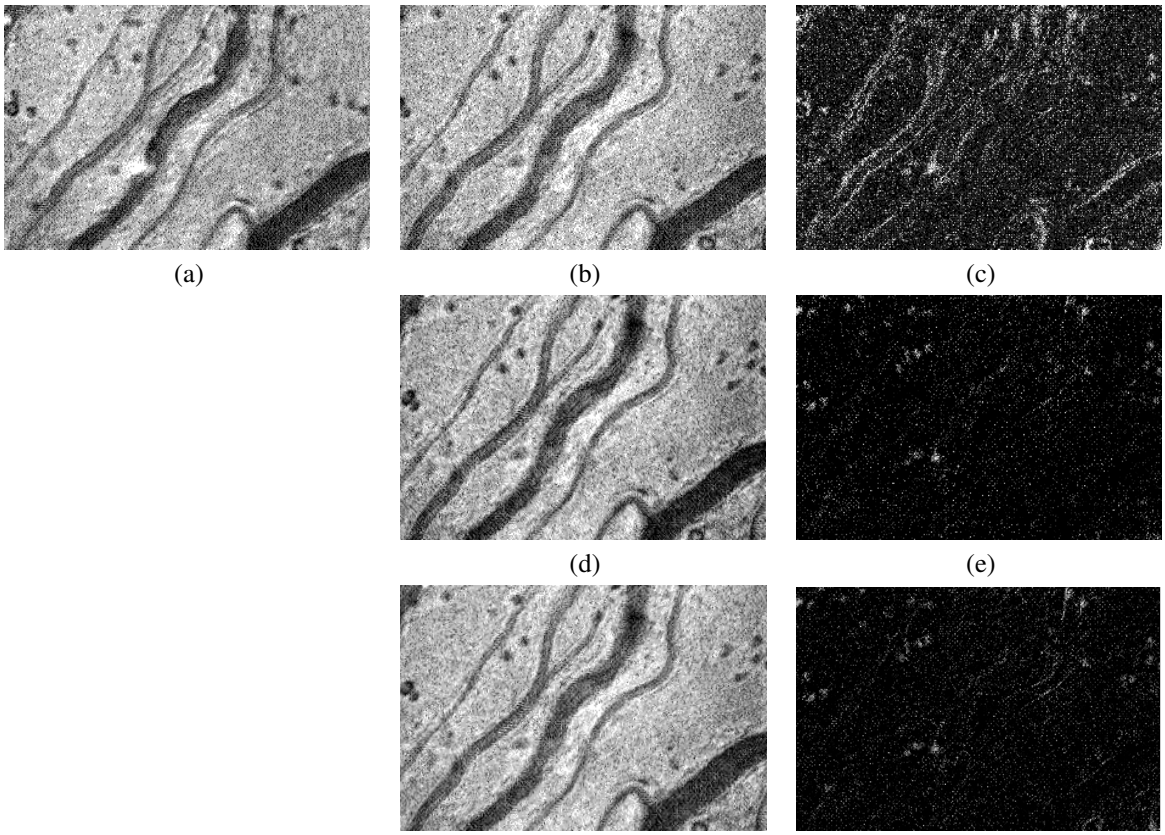

Figure 4. Microscopic video of iris: (a) frame 1; (b) frame 37; (c) absolute intensity difference between the two frames before registration; (d) registration result using GEN: image (b) is aligned with image (a); (e) absolute intensity difference between the two frames after GEN registration; (f, g) registration result using the Image Registration Toolkit [16] and its absolute intensity difference with (a).

ages with one image coded in green and the other in red, so that yellow indicates good alignment and red or green indicate misalignment. Figure 3 (d) shows the registration result of our method based on GEN ((b) is registered onto (a)), and (e) is the composite view of (d) and (a). Careful visual inspection reveals that our method (e) improves considerably over no registration (c), even though the original images have significantly different intensity ranges; note the edges of the cortex and the inner structures. We compare our method to the results from [12] (f). Overall our result is better, e.g. note the left and upper part of the cortex. We also compare our method to the registration result obtained with the state-of-the-art Image Registration Toolkit [16] (g), which achieves free form deformations using B-splines and normalized mutual information as a similarity measure. Our results are comparable in quality.

Microscopic iris images. We stabilize a video sequence of microscopic iris images through frame-by-frame registration. This is necessary to remove the severe jitter and deformation across frames in order to be able to track leukocyte motion. The deformation between frames is highly nonlinear. Our algorithm proves accurate and effective for these images (Fig. 4). Ideal registration should lead to an absolute difference image with near-zero background intensity and bright blobs corresponding to the moving leukocytes, which is exactly the case in Fig.4(e). ITK failed to work on iris images without preprocessing, even though different combinations of parameters and similarity measures were tried. We were able to get it to work only after using edge-preserving Wiener filtering to partially eliminate the distraction from cells. The deformation field obtained after the registration on filtered images were then applied to the original. The results of ITK with normalized mutual information for similarity measure are shown in Fig.4(f, g). Note that some vessel contours are still misaligned, unlike for the GEN, which also did not use any preprocessing.

Point set registration. The GEN can also be applied to point set registration, as demonstrated with the following example of corpus callosum shape registration, represented as a set of 2D points (Fig.5). Panel (a) shows the first person

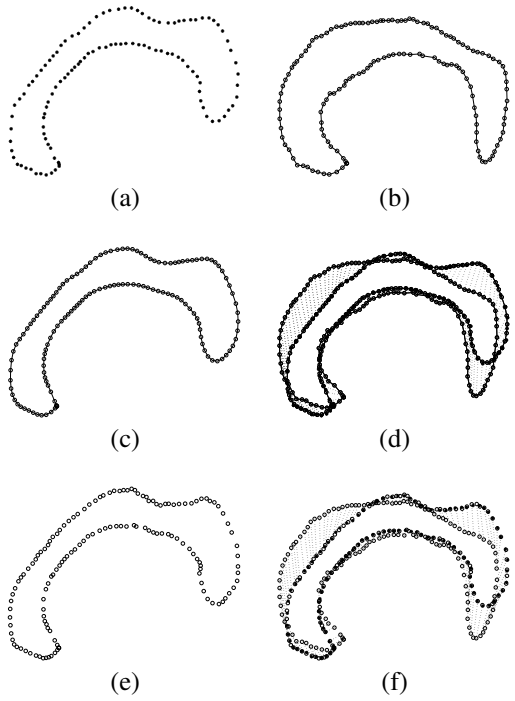

Figure 5. Corpus callosum shape registration example. (a) First point set; (b) second point set, which serves for the elastic net initialization; (c) final adapted position of the elastic net with the points in (a); (d) composite view of (a,b,c) and the displacement of elastic net from its initial position; (e,f) registration result and composite view of RPM [5], similar to those of the GEN.

point set. Panel (b) show the second person point set, which serves as initial position for the elastic net with connectivity between the adjacent points. The final adapted position of the net and its displacement from the initial position is shown in panels (c, d). For comparison we also show the result obtained using the RPM algorithm [5] (e, f), based on thin plate spline parametrization. Note that while RPM registers by estimating the transformation (by TPS), the GEN estimates directly the shape of the net, assuming the connectivity pattern is the same in both point sets (a closed contour in the Fig. 5).

4. Related work

Our motivation was to use a highly successful model of topographic maps, the generalized elastic net, to model probabilistically the image manifold and its deformation to fit a reference image. Our approach is most closely related to point-set registration methods. Chui and Rangarajan [5] use a parametric transformation $\boldsymbol{\phi}$ given by a thin-plate spline (TPS) and consider variables $w_{nm} \in [0,1]$ to model explicitly the correspondence between points $\mathbf{x}_n$ and $\mathbf{y}_m$ in the reference and source image, respectively. They minimize the objective function

$$E(\boldsymbol{\phi}, \mathbf{W}) = \sum_{n,m=1}^{N,M} w_{nm} \left\| \mathbf{x}_n - \boldsymbol{\phi}(\mathbf{y}_m) \right\|^2 + T\mathcal{R}(\boldsymbol{\phi}) + T \sum_{n,m=1}^{N,M} w_{nm} \log w_{nm} - \zeta \sum_{n,m=1}^{N,M} w_{nm} \quad (5)$$

where T is a temperature parameter, $\mathcal{R}(\boldsymbol{\phi})$ is a regularization term (second derivative of the TPS), the entropy term encourages binary correspondences, and the last term is necessary to deal with outliers (to which the first term, being a sum of squares, is very sensitive). The minimization is done by coordinate descent (alternating between minimizing over $\boldsymbol{\phi}$ for fixed $\mathbf{W}$, and vice versa) and annealed over T to avoid bad minima. The method can be seen as a generalization of the well-known iterated closest point algorithm [1, 19] by allowing soft correspondences. Using a TPS requires some heuristics in the minimization, and results in a complexity per step of $\mathcal{O}(N^3)$. The GEN differs from this in several respects. First, the GEN uses a nonparametric transformation (given by the centroids). This has several advantages: we are able to represent complex transformations and the influence of each point in the image is confined to its neighborhood, thus allowing for local deformations (which are harder to model with a global transformation such as the TPS). We can also use any regularizer we wish (e.g. higher-order derivatives or linear combinations of derivatives) by simply using the appropriate finite-difference in the matrix $\mathbf{S}$. And the computational cost of each step in the GEN is lower because of the sparse structure of $\mathbf{S}$. Second, our objective function is derived from a probabilistic model (a constrained Gaussian mixture) and thus defines a density for the point set; this introduces robustness to noise and outliers [18] without the need for an ad-hoc term in the objective function. The correspondence variables w_{nm} are implicitly obtained as posterior probabilities $p(\mathbf{y}_m|\mathbf{x}_n)$ and we need not model them nor minimize over them explicitly. (Note that, while [5] is initially formulated in terms of a Gaussian mixture, in practice eq. (5) is used.)

Our approach is also closely related to recent work [11, 18] based on modeling the point sets as density functions (kernel density estimates) and then minimizing a distance between densities over a parametric transformation: correlation in [18], L_2 distance (using a TPS as transformation) in [11]. With those distances, the resulting objective function can be obtained in closed form and optimized with respect to the transformation parameters (while the Kullback-Leibler distance does not lead to a closed form). Our work can be seen in this framework since the GEN defines a density model (Gaussian mixture), and MAP estimation is equivalent to minimizing the Kullback-Leibler distance from the reference density (kept fixed, and left as a

sum of delta functions at the points rather than as a kernel density estimate) to the source density (the GEN). However, for densities the Kullback-Leibler distance is more appropriate than the L_2 or correlation distances because the latter weigh the errors equally no matter where they happen in the density domain, while the Kullback-Leibler distance weighs them according to the reference density.

Finally, note that all these methods [5, 11, 18] were applied to point sets consisting of a small number (a few hundred) of feature vectors previously extracted from the image, while we deal with a much larger number of points ($\approx 10^5$) since we consider directly one feature vector per pixel. Naturally, our method can also be used with the smaller point sets (as in Fig. 5).

5. Discussion and conclusion

We have developed a probabilistic approach for nonparametric nonrigid image registration based on the generalized elastic net (using first- and second-order differential priors). The resulting formulation is simple and elegant, being essentially a penalized maximum likelihood problem. It produces soft correspondences naturally as posterior probabilities, without having to introduce them in an ad-hoc way, and its probabilistic basis affords some robustness to outliers and noise. The nonparametric transformation allows to model complex and localized deformations flexibly without prior knowledge about the type of transformation required, and to use sophisticated regularizers (e.g. high-order derivatives and linear combinations of them). The structured, sparse nature of the regularizer matrix allows an efficient optimization with linear conjugate gradients, faster than thin-plate splines. The method accurately registers images with nonlinear local deformations, and has robustness to image intensity distortion. When the elastic net is initialized with one centroid for each pixel in image I_1, the resulting deformed net will provide directly the displacement for each pixel without the need of image interpolation at each iteration, unlike many other registration methods. In general, the transformation can be controlled by using an arbitrary number of centroids; in this case, interpolation is necessary and is naturally given by averaging with respect to the posterior probabilities. While we have focused on intensity features, the method accommodates arbitrary features (e.g. gradient information and color components), spatial dimensions (e.g. 3D, 4D), and images of different spatial resolutions. The method is also well suited for registering consecutive frames in an image sequence, by successively adapting the net from one image to the next.

Acknowledgements

This work is partially supported by NIH grant NEI R01 EY013093, NSF grant IIS–0313350 (awarded to X. Song) and NSF CAREER award IIS–0546857 (awarded to Miguel Á. Carreira-Perpiñán).

References

[1] P. J. Besl and N. D. McKay. A method for registration of 3–D shapes. *IEEE PAMI*, 14(2):239–256, Feb. 1992.

[2] F. L. Bookstein. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE PAMI*, 11(6):567–585, June 1989.

[3] M. Á. Carreira-Perpiñán, P. Dayan, and G. J. Goodhill. Differential priors for elastic nets. In *Proc. of the 6th Int. Conf. Intelligent Data Eng. and Automated Learn. (IDEAL'05)*, volume 3578 of *LNCS*, pages 335–342, 2005.

[4] M. Á. Carreira-Perpiñán and G. J. Goodhill. Influence of lateral connections on the structure of cortical maps. *J. Neurophysiol.*, 92(5):2947–2959, Nov. 2004.

[5] H. Chui and A. Rangarajan. A new point matching algorithm for non-rigid registration. *Computer Vis.and Image Understanding*, 89(2–3):114–141, Feb.–Mar. 2003.

[6] W. R. Crum, T. Hartkens, and D. L. G. Hill. Non-rigid image registration: Theory and practice. *Brit. J. of Radiology*, 77(special issue):S140–S153, Dec. 2004.

[7] R. Durbin, R. Szeliski, and A. Yuille. An analysis of the elastic net approach to the traveling salesman problem. *Neural Computation*, 1(3):348–358, 1989.

[8] R. Durbin and D. Willshaw. An analogue approach to the traveling salesman problem using an elastic net method. *Nature*, 326(6114):689–691, Apr. 16 1987.

[9] L. Greengard and J. Strain. The fast Gauss transform. *SIAM J. Sci. Stat. Comput.*, 12(1):79–94, Jan. 1991.

[10] D. L. G. Hill, P. G. Batchelor, M. Holden, and D. J. Hawkes. Medical image registration. *Physics in Medicine and Biology*, 46(3):R1–R45, Mar. 2001.

[11] B. Jian and B. C. Vemuri. A robust algorithm for point set registration using mixture of Gaussians. In *ICCV*, pages 1246–1251, 2005.

[12] J. Kybic and M. Unser. Fast parametric elastic image registration. *IEEE Img. Proc.*, 12(11):1427–1442, 2003.

[13] J. B. A. Maintz and M. A. Viergever. A survey of medical image registration. *Med. Im. Anal.*, 2(1):1–36, Mar. 1998.

[14] J. Modersitzki. *Numerical Methods for Image Registration*. Oxford University Press, 2004.

[15] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer-Verlag, New York, 1999.

[16] D. Rueckert, L. Sonoda, C. Hayes, D. Hill, M. Leach, and D. Hawkes. Non-rigid registration using free-form deformations: Application to breast MR images. *IEEE Transactions on Medical Imaging*, 18(8):712–721, 1999.

[17] R. Szeliski and S. Lavallee. Matching 3-D anatomical surfaces with non-rigid deformations using octree-splines. *IJCV*, 18(2):171–186, 1996.

[18] Y. Tsin and T. Kanade. A correlation-based approach to robust point set registration. In *ECCV*, volume 3, pages 558–569, 2004.

[19] Z. Zhang. Iterative point matching for registration of free-form curves and surfaces. *IJCV*, 13(2):119–152, Oct. 1994.