# REFERENCE-BASED PROBABILISTIC SEGMENTATION AS NON-RIGID REGISTRATION USING THIN PLATE SPLINES

Luca Bertelli, Pratim Ghosh, B. S. Manjunath

Electrical and Computer Engineering Department University of California, Santa Barbara 93106 Santa Barbara, CA

{lbertelli, pratim, manj}@ece.ucsb.edu

#### ABSTRACT

In this paper we demonstrate the effectiveness of reference (or atlas)-based non-rigid registration to the segmentation of medical and biological imagery. In particular we introduce a segmentation functional exploiting feature information about the reference image and we minimize it with respect to the parameters of the non-rigid transformation, akin to a regionbased maximum likelihood estimation process. The warping transformation is modeled using Thin Plate Splines, which incorporate information about the global rigid motion and the non-rigid local displacements. Extensive experimental evaluations and comparisons with other segmentation techniques on a complex biological dataset are presented. The proposed algorithm outperforms the others in both classification rate and, in particular, *localization accuracy*.

*Index Terms*— atlas-based segmentation, region based segmentation, non-rigid registration.

## 1. INTRODUCTION

Incorporating prior information in the segmentation process can be extremely important particularly in the case of biological and medical imagery, where domain knowledge often carries more information than the data itself. In particular, an *atlas* generated, for instance, by manually segmenting a given image, contains information about location and shape of the structures of interest and most importantly the *spatial relationship* between them. In addition, statistical description of the features in the different regions can be obtained.

Existing segmentation techniques, varying from feature based pixel classification to level set methods, are not suitable to incorporate all such information. Purely feature based pixel/voxel classification [1], k-means for instance, works typically one pixel at a time, independently of its spatial location. An attempt to include spatial information such as neighborhood relationships is made within the level set methods by adding geometrical constraints to the framework (see for example [2] and references therein). An elastic force is Frédéric Gibou

Mechanical Engineering Department University of California, Santa Barbara 93106 Santa Barbara, CA

gibou@engineering.ucsb.edu

added to the evolving surfaces to penalize the splitting in many parts due to noisy conditions. Attempts have been made to incorporate shape information as well ([3] and references therein), but spatial relationships between objects and structures still remain unexploited.

In contrast non-rigidly registering a known segmentation mask (the atlas) on top of the target image allows to obtain reliable segmentation results, exploiting all the above mentioned information in a principled way. Several atlas-based segmentation techniques have been proposed in the past [4, 5]. In these works the non-rigid registration is obtained in two steps, by first rigidly estimating the global motion and then by locally estimating the non-rigid displacements (via B-splines [4] or level set models [5]). In this paper we model the nonrigid registration via Thin Plate Splines adopting the method proposed in [6] to obtain a direct estimate of the transformation which maximizes a segmentation functional. This functional is defined using probability distributions computed exploiting the reference image, which underlays the atlas map. Hence, the non-rigid registration procedure can be interpreted as a region-based maximum likelihood estimation process [7].

## 2. PRIOR-BASED SEGMENTATION AS NON-RIGID REGISTRATION

## 2.1. Notation

Let us consider two images representing views of the same object, organ or tissue, namely  $I_1 : \Omega_1 \mapsto \mathbb{R}^{N_1}$  and  $I_2 : \Omega_2 \mapsto \mathbb{R}^{N_2}$  (where  $N_1, N_2$  are the dimensions of the feature space, note that  $N_1$  can be different from  $N_2$ , since the two views can be acquired with different modalities/sensors). We assume that the segmentation of the object(s) of interest in image  $I_1$  is known (*i.e.* we have an exhaustive partition of  $\Omega_1$  in N non-overlapping regions  $R_i$ ). In particular, we can define a function  $\Upsilon : \Omega_1 \mapsto \mathbb{N}$  such that  $\Upsilon(\mathbf{x}) = i$ , with  $i \in \mathbb{N}$  only if  $\mathbf{x}$  belongs to the object i or region  $R_i$ . The boundaries of the objects are hence defined by the discontinuities in the characteristic function  $\Upsilon$ .

Furthermore, let us assume that the two views are related through a warping model  $\mathbf{W}: \Omega_2 \mapsto \Omega_1 (\mathbb{R}^2 \mapsto \mathbb{R}^2)$ , parameterized in terms of the parameters vector  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)$ . Therefore, a point  $\mathbf{x}$  in one view is mapped onto the transformed view according to  $\mathbf{x}' = \mathbf{W}(\mathbf{x}, \boldsymbol{\theta})$ . Now consider the function  $\Upsilon(\mathbf{x}) : \Omega_1 \mapsto \mathbb{N}$  defined above, whose discontinuities define the object(s) boudaries in image  $I_1$ . Applying the warping to its argument, we obtain the function  $\Upsilon(\mathbf{W}(\mathbf{x}, \boldsymbol{\theta})) : \Omega_2 \mapsto \mathbb{R}$ , whose rounded version  $\overline{\Upsilon}(\mathbf{W}(\mathbf{x}, \boldsymbol{\theta})) : \Omega_2 \mapsto \mathbb{N}$  defines a segmentation in the transformed view  $I_2$ . Changing the parameters of the warping, we therefore change the segmentation of image  $I_2$ . Next we define a segmentation functional, whose minimization w.r.t. the parameter vector  $\boldsymbol{\theta}$  leads to the optimal segmentation of image  $I_2$ .

#### 2.2. Segmentation Functional Maximization

Separating an image into several regions can be considered as the partitioning which seeks "within-regions homogeneity" and at the same time "between-regions heterogeneity". In other words, we would like the pixels belonging to one region to be similar to each other according to some particular metric and dissimilar to the pixels belonging to the other regions.

Let us therefore introduce a set of functions  $f_i(\mathbf{x}) : \Omega_2 \mapsto \mathbb{R}$  with i = 1, ..., N, which give an indication of how well pixel  $\mathbf{x}$  fits in region  $R_i$ , where N is the total number of regions. The optimal segmentation can be then defined as the one which maximizes the following functional:

$$F(\boldsymbol{\theta}) = \sum_{i=1}^{N} \int_{\Omega_2} f_i(\mathbf{x}) \left( 1 - \min\{1, |\bar{\Upsilon}(\mathbf{W}(\mathbf{x}, \boldsymbol{\theta})) - i|\} \right) d\mathbf{x}$$
(1)

Maximizing this functional along the gradient direction with respect to the parameters vector  $\boldsymbol{\theta}$ , we obtain:

$$\Delta \boldsymbol{\theta} = \boldsymbol{\alpha} \nabla_{\boldsymbol{\theta}} F \tag{2}$$

which gives the update rule for the warping parameters  $\theta$  ( $\alpha$  is the stepsize and  $\nabla_{\theta}$  is the gradient w.r.t the elements of  $\theta$ ). The rounding function  $H_R$ , *i.e.* the function which maps  $H_R : \Upsilon \mapsto \overline{\Upsilon}$ , has, in the general case, a complicated expression. For the two regions case instead, it can be written as a modified version of the Heaviside or step function H(z). In particular we can write:

$$H_R(z) = H(z - \frac{3}{2}) + 1$$
(3)

which quantizes the real axis to 1 and 2 with mid-rise cutoff at  $\frac{3}{2}$ . Using  $H_R$  we can rewrite (1) yielding:

$$F(\boldsymbol{\theta}) = \int_{\Omega_2} f_2(\mathbf{x}) (H_R(\Upsilon(\mathbf{W}(\mathbf{x}, \boldsymbol{\theta}))) - 1) d\mathbf{x} \quad (4)$$
  
+ 
$$\int_{\Omega_2} f_1(\mathbf{x}) \Big( 2 - H_R(\Upsilon(\mathbf{W}(\mathbf{x}, \boldsymbol{\theta}))) \Big) d\mathbf{x}$$

Maximizing this functional along the gradient direction with respect to the parameters vector  $\theta$ , we can specialize (2) as:

$$\Delta \boldsymbol{\theta} = \boldsymbol{\alpha} \int_{\Omega_2} \left( f_2(\mathbf{x}) - f_1(\mathbf{x}) \right) \delta_R \big( \Upsilon(\mathbf{W}(\mathbf{x}, \boldsymbol{\theta})) \big)$$
  
$$\cdot \left[ \nabla \Upsilon(\mathbf{W}(\mathbf{x}, \boldsymbol{\theta})) \frac{\partial \mathbf{W}(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right]^T d\mathbf{x}$$
(5)

where, as before,  $\alpha$  is the stepsize and  $\delta_R$  is a regularized derivative of  $H_R$ .

## 2.3. Non-rigid warping using Thin Plate Splines

To describe the warping process  $\mathbf{x}' = \mathbf{W}(\mathbf{x}, \boldsymbol{\theta})$ , which nonrigidly deforms the segmentation of the first view onto the second one, we used thin plate splines (TPS) [8]. The TPS warping is a composition of an affine (*rigid*) transformation and a *non-rigid* warping:

$$\mathbf{x}' = \begin{bmatrix} A & T & W \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ 1 \\ U(||\mathbf{x} - \hat{\mathbf{x}}^{(:)}||) \end{bmatrix}$$
(6)

where  $A \in \mathbb{R}^{2\times 2}$  and  $T \in \mathbb{R}^{2\times 1}$  describe the affine/rigid transformation, while the weight matrix  $W \in \mathbb{R}^{2\times K}$  and the set of K control points  $\hat{\mathbf{x}} = (\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \dots, \hat{\mathbf{x}}_K)$  characterize the non-rigid deformation.  $U(r) = r^2 \log r^2$  is the radial basis function of the spline. Constraints on the squareintegrability of the second derivatives of the spline-based interpolation functions give the following additional relation:

$$W\begin{bmatrix} \hat{\mathbf{x}}^T & \mathbf{1} \end{bmatrix} = W P = \mathbf{0} \tag{7}$$

where  $\mathbf{1} \in \mathbb{R}^{K \times 1}$  and  $\mathbf{0} \in \mathbb{R}^{2 \times 3}$  (matrices with respectively only one and zero entries). The unknown transformation parameters in A, T and W can be uniquely determined if the transformed control points  $\hat{\mathbf{x}}'$  are available. In this case, combining (6) and (7), one could write:

$$\begin{bmatrix} \hat{\mathbf{x}}' & \mathbf{0} \end{bmatrix} = \begin{bmatrix} A & T & W \end{bmatrix} \begin{bmatrix} P^T & O \\ Z & P \end{bmatrix}$$
(8)

where  $Z_{ij} = U(||\hat{\mathbf{x}}_i - \hat{\mathbf{x}}_j||)$  and *O* is a zero matrix  $\in \mathbb{R}^{3\times 3}$ . The warping model in (6) can be therefore parameterized only in terms of the transformed control points  $\hat{\mathbf{x}}'$ , via a *linear* mapping [6] (*i.e.* the parameter vector  $\boldsymbol{\theta}$  is composed by the coordinates of the transformed control points):

$$\mathbf{x}' = \begin{bmatrix} \mathbf{\hat{x}}' & \mathbf{0} \end{bmatrix} \begin{bmatrix} P^T & O \\ Z & P \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{x} \\ 1 \\ U(||\mathbf{x} - \mathbf{\hat{x}}^{(:)}||) \end{bmatrix}$$
(9)

This linear relation makes extremely efficient the computation of the Jacobian of the warp with respect to the parameter vector (in this case the transformed control points)  $\frac{\partial \mathbf{W}(\mathbf{x},\theta)}{\partial \theta}$  in (5). In order to regularize the transformation (*i.e.* to prevent portion of the image to collapse onto sets of measure zero) we used the approach suggested in [6], consisting in replacing the matrix Z with  $Z + \lambda I$ , where I is the identity matrix. This is equivalent to adding a further constraint on the TPS smoothing. The larger is  $\lambda$ , the more rigid is the warping, since we extend the region of influence of each control point.

### 2.4. The Segmentation Functions f

The information about the segmentation of the reference image can be exploited in constructing the region indicator functions  $f_i$ . The simplest choice consists in modeling the image as piecewise constant [2]. According to this choice, each region can be represented by the average of the pixels belonging to it. The corresponding region indicator function becomes:

$$f_i(\mathbf{x}) = -||I(\mathbf{x}) - c_i||^2$$
 (10)

where  $c_i$  is the average feature vector for region *i* computed using the reference image  $I_1$ , where the segmentation in different regions is supposed to be known (we are making the assumption that  $I_1$  and  $I_2$  have been acquired with the same modality). This model, though computationally efficient, imposes an unimodality constraint on the data distribution, *i.e.* the image, which in some cases might be too restrictive. Consider for example the case in which regions are composed of two separate feature clusters. Using the model in (10)  $c_i$  can assume a intermediate value between the average of cluster one and the average of cluster two. This would produce low values of  $f_i$ , even for pixels clearly belonging to that region.

The information about the prior image can be instead exploited to estimate the probability distribution of each region [7]. Introducing N Gaussian Mixture Models (GMMs) (one for every region) and training this model using the Expectation Minimization algorithm on the reference image, we obtain N likelihood function which can then be used as region indicator functions:

$$f_i(\mathbf{x}) = p(I(\mathbf{x})|\mathcal{R}_i) \tag{11}$$

In this way we have removed the unimodality constraint on the data distribution. The non-rigid registration procedure can be therefore consider as a region-based maximum likelihood classifier.

## 3. EXPERIMENTS AND DISCUSSION

In this section we demonstrate the effectiveness of the proposed algorithm for reference-based segmentation of medical and biological images. Fig. 1 is a synthetically generated example in which a binary mask (the reference, Fig. 1(a)) is used to segment its non-rigidly deformed version (Fig. 1(c)). The example is meant to show that the recovered warping parameters (represented via the deformed grid of Fig. 1(d)) are consistent with the synthetically induced deformation.



**Fig. 1**. (a) Segmented reference (b) Initial misalignment between reference and target (c) Final segmentation of the target (d) Recovered deformation grid.



**Fig. 2.** (a) CT-scan slice of a prostate, used as segmented reference (a bone is segmented out). (b) Different CT-scan slice of the same organ, with synthetically added occlusion and noise. (c) Segmentation result of the target image.

The second example is semi-synthetic in the sense that the two images are two different slices of a CT-scan of a prostate (Fig. 2(a,c)). As shown in Fig. 2(b) the bone structure (light gray object) in the second slice is shifted and deformed w.r.t. the first slice. In addition, we introduced a synthetically generated occlusion and noise. The algorithm is still able to correctly segment the bone in the target image (Fig. 2(c)). While for this two example we have been able to successfully model the region indicator function  $f_i$  as in (10), the complex biological data-set that we describe in the next subsection required the use of the probabilistic GMM-based model in (11).

#### 3.1. Retinal Layers Segmentation

A vertical section of a retina illustrates its complex and highly patterned architecture (Fig. 3, first and third rows). Specifically it is characterized by a layered structure, where each layer has a different structure consisting of the group of cell bodies or synaptic terminals. Staining the tissue with an antibody results in collecting fluorescence images with high intensity regions corresponding to a high concentration of protein expression and the rest as dark regions. When the tissue is stained with multiple antibodies, the response to each antibody is combined to form a color image, where the color represents the protein response to each antibody.

As a result retinal layers are characterized by distinctive features (color and texture), but most importantly by their reciprocal spatial collocation (which makes these images a perfect test set for the proposed technique). Of interest to biology are measures such as number of cells in each layer, layer



**Fig. 3.** First and third rows: detected layer boundaries (in white) and ground truth (in black). Second and fourth rows: detected layer masks (with ground truth boundaries in black). First column: proposed algorithm. Second column: level set approach. Third column: ML pixel-level classification.

thickness and changes in protein distribution. Hence retinal layer segmentation is a critical step for quantitative analyses.

The dataset we used is composed by 52 images taken under 4 different experimental conditions and a segmentation ground truth (where the boundaries between layers are drawn by a human) is available for benchmarking. One image is randomly picked as the reference (and used for training the GMMs in (11)). We compared the performance of our algorithm with other two techniques. The first technique is a simple Maximum Likelihood pixel-level classification based on the the region indicator function  $f_i$  as in (11) (each pixel is assigned to the region that is more likely to belong to, without any regularization). The second technique is a probabilistic level set based technique as in [7] (or [2], substituting the Chan-Vese data term with (11)), where the regularization is granted by imposing geometric constraints (*i.e.* smoothness) on the evolving curves. As we evince from Fig. 3, the ability of the proposed method to take into account not only geometric constraints, but also more high level concept as the reciprocal position of the different layers, yields more reliable results. In particular, the splitting of layers in multiple parts and the presence of isolated blobs (which happens using the level set approach, see Fig. 3(e,k)) is avoided.

**Table 1.** Quantitative comparison: for each layer the first value in the table, the **classification rate**, is the % of pixels correctly classified and the second one, the **localization accuracy**, is the average distance of the misclassified pixels from the closest point of the layer (this penalizes pixels uncorrectly assigned to one layer and spatially far from it). Note: Layer 1 (background, dark gray) has two connected components, one on the left and one on the right of the tissue.

Method	Layer 1	Layer 2	Layer 3	Layer 4
Proposed	0.93 - 0.59	0.93 - <b>0.89</b>	0.88 - 0.56	0.83 - 1.30
Level Sets	0.89 - 7.83	<b>0.94</b> - 4.31	0.79 - 1.52	0.81 - 5.56
ML Class.	0.86 - 5.55	0.90 - 4.71	0.76 - 8.20	0.76 - 8.70

A quantitative evaluation of the performances is presented in Table 1 (see the caption for details), where we demonstrate that the proposed method clearly outperforms the other two in classification rate and especially in *localization accuracy*.

#### 4. REFERENCES

- N. Kovacevic, N. Lobaugh, M. Bronskill, A. Feinstein, and S. Black, "A robust method for extraction and automatic segmentation of brain images," *NeuroImage*, vol. 17, no. 3, pp. 1087–1100, 2002.
- [2] T F Chan and L A Vese, "Active contours without edges," *IEEE Transactions on Image Processing*, vol. 10, no. 2, pp. 266–77, February 2001.
- [3] G. Unal and G. Slabaugh, "Coupled pdes for non-rigid registration and segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2005, vol. 1, pp. pp. 168–175.
- [4] M. Chen, T. Kanade, H. Rowley, and D. Pomerleau, "Quantitative study of brain anatomy," in *IEEE Work-shop on Biomedical Image Anlysis*, 1998, pp. 84–92.
- [5] B. C. Vemuri, J. Ye, Y. Chen, and C. M. Leonard, "Image registration via level-set motion: Applications to atlasbased segmentation," *Medical Image Analysis*, vol. 7, no. 1, pp. 1–20, 2003.
- [6] J. Lim and M. H. Yang, "A direct method for modeling non-rigid motion with thin plate spline," in *IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), 2005, vol. 1, pp. 1196–1202.
- [7] N Paragios and R Deriche, "Geodesic active regions and level set methods for supervised texture segmentation," *International Journal of Computer Vision*, pp. 223–47, February 2002.
- [8] F. Bookstein, "Principal warps: Thin-plate splines and the decomposition of deformations," *IEEE Transactions* on Pattern Analysis and Machine Intelligence, vol. 11, no. 6, pp. 567–585, 1989.