Spatially Adaptive Log-Euclidean Polyaffine Registration Based on Sparse Matches

Maxime Taquet[1,2], Benoît Macq[1], and Simon K. Warfield[2]

[1] ICTEAM Institute, Université catholique de Louvain, Louvain-La-Neuve, Belgium
[2] Computational Radiology Laboratory, Harvard Medical School, USA
maxime.taquet@uclouvain.be

Abstract. Log-euclidean polyaffine transforms have recently been introduced to characterize the local affine behavior of the deformation in principal anatomical structures. The elegant mathematical framework makes them a powerful tool for image registration. However, their application is limited to large structures since they require the pre-definition of affine regions. This paper extends the polyaffine registration to adaptively fit a log-euclidean polyaffine transform that captures deformations at smaller scales. The approach is based on the sparse selection of matching points in the images and the formulation of the problem as an expectation maximization iterative closest point problem. The efficiency of the algorithm is shown through experiments on inter-subject registration of brain MRI between a healthy subject and patients with multiple sclerosis.

Keywords: Registration, Polyaffine, Log-Euclidean, Structure Tensor

1 Introduction

Medical image registration is used in a variety of applications, from atlas construction to intraoperative navigation. The goal is to find a transform T that maps an image I onto another image J. A valuable property of the transform is *diffeomorphism* which guarantees invertibility and topology preservation [9].

Different models can be chosen for T [4]. Dense field models, as returned by diffeomorphic demons [9], are very flexible. However, models with fewer parameters are typically more robust. Log-euclidean polyaffine transforms (LEPT) [2,6] are compact and diffeomorphic transforms. They are built by composing affine transforms in the log-domain, and have been successfully used to register predefined anatomical structures [6].

The nature of tissues suggests that higher order deformations could also be modeled by LEPT with more degrees of freedom. However, predefining affine regions at a lower scale would be too cumbersome or impossible.

This paper introduces a registration method that adaptively fits a LEPT based on matching points sparsely selected in the image. The main contribution is the extension of the expectation-maximization iterative closest point (EM-ICP [7]) to use LEPT. This scheme accounts for matching ambiguities due to regularities in the image and naturally integrates regularization as a prior on T.

The rest of this paper is organized as follows. Section 2 introduces the elements of the algorithm. Section 3 shows results on brain inter-subject registration. Section 4 concludes and presents some directions for future work.

2 Methods

The proposed method, summarized in Table 1, aims at optimizing the parameters of a LEPT to map image J onto I. The location of the anchors of the affine components are first defined. The parameters of the affine components are then estimated in a multi-scale approach. At each scale, corresponding points are selected and their prior matching probabilities are recorded. The optimization then alternates between updating the posterior matching probabilities and optimizing the transform parameters. The image structure tensor is used in the definition of these probabilities to account for matching ambiguities.

Table 1. Summary of the proposed method

1: Define the anchor locations $\boldsymbol{a}_k$ in I.
2: **for** $r = 0$ **to** $R - 1$
3:     Compute the structure tensor $\mathcal{S}$ of image I at scale r.
4:     Compute the LEPT weights $\boldsymbol{w}(\boldsymbol{x})$ at scale level r.
5:     Select the K best matches $C = \{(\boldsymbol{m}_i, \boldsymbol{s}_{i,j}, \pi_{ij})\}$ between I and $T^{(r-1,N_{it})} \circ J$.
6:     **for** $i = 1$ **to** N_{it} **do**
7:         *E-Step* Update the probabilities: $C^{(r,i)} \leftarrow \text{Update}(C^{(r,i-1)}; T^{(r,i-1)}, \mathcal{S})$.
8:         *M-Step* Optimize the affine corrections $\delta L^{(r,i)} \leftarrow \text{Optimize}(T^{(r,i)}; C^{(r,i)}, \mathcal{S})$.
9:         Composition of the affine components: $\exp(L_k^{(r,i)}) \leftarrow \exp(L_k^{(r,i-1)}) \exp(\delta L_k^{(r,i)})$.
10:     Interpolation of the transform: $T^{(r,i)} \leftarrow \text{Interpolate}(L^{(r,i)}, \boldsymbol{w}(\boldsymbol{x}))$.
11:     **end for**
12: **end for**

2.1 Log-Euclidean Polyaffine Transforms

Log-euclidean polyaffine transforms (LEPT) are defined as a weighted composition of N affine transforms in the logarithm domain [2]. In other words, under the action of an LEPT, the point $\boldsymbol{x}$ is transformed to $\boldsymbol{x} + T(\boldsymbol{x})$ with

$$T(\boldsymbol{x}) = \exp\left(\sum_{k=1}^{K} w_k(\boldsymbol{x}) L_k \tilde{\boldsymbol{x}}\right), \tag{1}$$

where $\exp(.)$ is the exponential-map, $L_k \in \mathbb{R}^{3 \times 4}$ is the principal logarithm of an affine matrix in homogenous coordinates (with the last line ignored), $w_k(\boldsymbol{x})$ are the weights which depend on the anchor location $\boldsymbol{a}_k$, and $\tilde{\boldsymbol{x}} = [\boldsymbol{x}, 1]^T$. LEPT have remarkable properties. They are invertible and their inverse is a LEPT [2].

In this paper, the location of the anchors are determined a priori based on the local intensity heterogeneity of the image. More precisely, a measure of heterogeneity $h_I(\boldsymbol{x})$ is computed at every $\boldsymbol{x}$ (see Sec. 2.3). The anchors are defined as the K-means centroids of the point cloud spanned by all $\boldsymbol{x}$ s.t. $h_I(\boldsymbol{x}) > \bar{h}_I(\boldsymbol{x})$, where $\bar{h}_I(\boldsymbol{x})$ is a sliding mean of heterogeneity. Hence, the anchors will tend to concentrate more in areas with high local contrast. Given K anchors $\boldsymbol{a}_k$, we want to optimize the $12K$ parameters corresponding to the L_k's.

2.2 Block Matching

Block matching is used to establish a dense correspondence between points in image I and J. As a similarity measure, the correlation coefficient (CC) is used because of its invariance under any linear mapping of intensities. Let B_m (resp. B_s) be the block centered at $\boldsymbol{m}_i$ (resp. $\boldsymbol{s}_{i,j}$) in image I (resp. J), the CC is:

$$\rho(\boldsymbol{m}_i, \boldsymbol{s}_{i,j}) = \frac{\mathrm{Cov}(B_m, B_s)}{\sqrt{\mathrm{Var}(B_m)\mathrm{Var}(B_s)}}.$$

For each block in I, the CC of the best match $\rho_{\max}(\boldsymbol{m}_i)$ is compared to the mean of all CC for this block $\bar{\rho}(\boldsymbol{m}_i)$. The N points with the highest ratio $\rho_{\max}/\bar{\rho}$ are selected. The prior matching probability is modeled as a normal distribution of the root mean squared error ($\mathrm{MSE}_n = 2 - 2\rho$) between normalized blocks (zero mean, unit variance):

$$\pi_{ij} = \frac{1}{\sqrt{2\pi\sigma_e^2}} \exp\left(-\frac{2 - 2\rho(\boldsymbol{m}_i, \boldsymbol{s}_{i,j})}{2\sigma_e^2}\right).$$

Block matching is not robust to noise and suffers from the aperture problem. To deal with this issue, M matches in J are recorded for each of the N selected points in I, and we will choose $N \gg K$. The remaining of this paper explains how the LEPT can be robustly estimated based on this set C of matching pairs.

2.3 Image Structure Tensor

The location of the matches can be ambiguous due to the regularity of the intensity profile around the point. This problem can be addressed by anisotropically weighting the error with the structure tensor [5].

The structure tensor at $\boldsymbol{x}_k$, $S(\boldsymbol{x}_k) \in \mathbb{R}^3$ is defined as the autocorrelation of the intensity gradient $\boldsymbol{\nabla}I(\boldsymbol{x})$ in a neighborhood Ω_k:

$$S(\boldsymbol{x}_k) = E_{\Omega_k}\left\{\boldsymbol{\nabla}I(\boldsymbol{x})\boldsymbol{\nabla}I(\boldsymbol{x})^T\right\} = \frac{1}{|\Omega_k|} \sum_{\boldsymbol{x} \in \Omega_k} \boldsymbol{\nabla}I(\boldsymbol{x})\boldsymbol{\nabla}I(\boldsymbol{x})^T.$$

The normalized structure tensor, $\mathcal{S}(\boldsymbol{x}) = S(\boldsymbol{x})/||S(\boldsymbol{x})||_2$, is positive semidefinite with a maximum eigenvalue of 1 and an eigenvector aligned with $\nabla I(\boldsymbol{x})$ in Ω_k.

Consequently, the norm of $\boldsymbol{e}' = \mathcal{S}\boldsymbol{e}$ is more affected by the component of $\boldsymbol{e}$ parallel to $\nabla I(\boldsymbol{x})$. In other words, the overall error is less affected by matching

ambiguities due to regular structures than by errors made in the direction of $\nabla I(\boldsymbol{x})$.

Besides weighting the errors, the structure tensor is also used to define anchor locations. Indeed, its highest eigenvalue λ_3 is significantly higher than zero only for heterogenous areas, making it natural to define $h_I(\boldsymbol{x}) = \lambda_3(\boldsymbol{x})$.

2.4 Transformation Estimation

Given the set C of matching pairs with prior probabilities π_{ij}, we propose to estimate the transform T by maximizing the joint log-likelihood of C and T [7,1]:

$$T^* = \arg\max_T E\left\{\log P\left(C, T\right)\right\} = \arg\max_T E\left\{\log P\left(C|T\right)\right\} + E\left\{\log P\left(T\right)\right\}. \quad (2)$$

The first term of (2) tends to honor the detected correspondences. The second term is a prior on T. This term favors some transforms over others, based on intrinsic properties of the transforms only. This is a statistical interpretation of the regularization energy used in [4,6,9]. We will use the following prior:

$$\log P(T) = -\lambda^2 \sum_{k,l} a_{k,l} ||L_k - L_l||^2 + \text{cst},$$

where $||.||^2$ is the Frobenius norm and $a_{k,l} = \sum_{\boldsymbol{x}} w_k(\boldsymbol{x}) w_l(\boldsymbol{x}) \left(\frac{1}{\sum_{\boldsymbol{x}} w_k(\boldsymbol{x})} + \frac{1}{\sum_{\boldsymbol{x}} w_l(\boldsymbol{x})}\right)$ is an overlapping coefficient. This expression states that transforms are more likely if nearby affine components (components that share a common affecting area) are close to each other. Interestingly, this term is equivalent to the regularization energy defined in [6].

An efficient method to optimize (2) is EM-ICP [7]. This algorithm consists in alternatively optimizing the criterion for C considering T fixed (*E-step*) and for T considering C fixed (*M-step*). If enough matching pairs are selected, EM-ICP is very robust to noise. It is thus well suited to cope with the block matching issues. The E-step simply results in computing the matching probabilities:

$$\overline{(C)_{ij}} = \frac{\pi_{ij} p(\boldsymbol{s}_{i,j}|\boldsymbol{m}_i, T)}{\sum_k \pi_{ik} p(\boldsymbol{s}_{i,j}|\boldsymbol{m}_k, T)}, \quad (3)$$

where the expression of $p(\boldsymbol{s}_{i,j}|\boldsymbol{m}_i, T)$ is accounts for the matching ambiguities:

$$p(\boldsymbol{s}_{i,j}|\boldsymbol{m}_i, T) = \exp\left(-\frac{||\mathcal{S}(\boldsymbol{m}_i)(\boldsymbol{m}_i - T * \boldsymbol{s}_{i,j})||^2}{2\sigma_n^2}\right).$$

The M-step then consists in optimizing the parameters of T:

$$T^* = \arg\max_T \sum_{i=1}^{N} \sum_{j=1}^{M} \overline{(C)_{ij}} \log p(\boldsymbol{s}_{i,j}|\boldsymbol{m}_i, T) - \lambda^2 \sum_{k,l=1}^{K} a_{k,l} ||L_k - L_l||^2. \quad (4)$$

Unlike affine transforms, this equation does not have an obvious solution for the parameters L_k, due to the exponential map of (1). Therefore, we propose a first order approximation in which $T(\boldsymbol{x})$ is approximated by:

$$T(\boldsymbol{x}) \approx \mathbb{I} + \sum_{k=1}^{N} w_k(\boldsymbol{x}) L_k \tilde{\boldsymbol{x}}.$$

In that case, optimizing (4) amounts to the quadratic programming (QP):

$$l^* = \arg\min_{l} (Hl - D)^T S (Hl - D) + \lambda^2 l^T A l, \tag{5}$$

where:

$l \in \mathbb{R}^{12K}$ is the vector of the elements of L_k taken row-wise and concatenated,

$H \in \mathbb{R}^{3N \times 12K}$ is the interpolation matrix. Each 3×12 block $[H]_{ij}$ corresponds to point $\boldsymbol{m}_i = (x_i, y_i, z_i)$ and the jth anchor $\boldsymbol{a}_j$ and is equal to $I_3 \otimes (w_j(\boldsymbol{m}_i) \tilde{\boldsymbol{m}}_i)^T$, where I_3 is the 3×3 identity matrix and $\otimes$ stands for the Kronecker product,

$D \in \mathbb{R}^{3N}$ is the vector obtained by taking, for each $\boldsymbol{m}_i$, the barycenter of its matches $\overline{\boldsymbol{s}}_i$ weighted by $\overline{(C)}_{ij}$ (the equivalence between criterion (4) and the use of barycenters is justified in [7].),

$S \in \mathbb{R}^{3N \times 3N}$ is the block diagonal matrix of the structure tensor. Each 3×3 block on the diagonal corresponds to the structure tensor at point $\boldsymbol{m}_i$,

$A \in \mathbb{R}^{12K \times 12K}$ is the prior matrix with $(A)_{k \neq l} = -a_{k,l}$ and $(A)_{kk} = \sum_l a_{k,l}$.

In practice, the weights w_i and overlapping coefficients $a_{k,l}$ are thresholded out, so that H, S and A are sparse. Taking the derivative of (5) w.r.t. l and setting it to 0 yields the linear system:

$$(H^T S H + \lambda^2 A) l = H^T S D. \tag{6}$$

Interestingly, the regularization term is optimized simultaneously with the similarity term, unlike [6] where an ad-hoc one step gradient descent is performed on the regularization energy. Alternating the estimation of matching probabilities (3) and the estimation of parameters (6) until convergence results in a globally optimal LEPT that best fits the observations of the block matching.

2.5 Weights: the Kriging Estimator

In [6,2], the weights $w_k(\boldsymbol{x})$ are simply a normalized Gaussian function of the distance between the point and the anchor. Here we use the Kriging estimator (KE) to define these weights, as in [8]. The KE has the advantage of adapting the weights to the spatial distribution of anchors in a statistically sound way.

Let us interpret LEPT as random fields of matrix logarithms $L(\boldsymbol{x})$ for which L_k are observations at locations $\boldsymbol{a}_k$. In this interpretation, the weights are the coefficients of a linear estimator of L at $\boldsymbol{x}$. KE is a best linear unbiased estimator
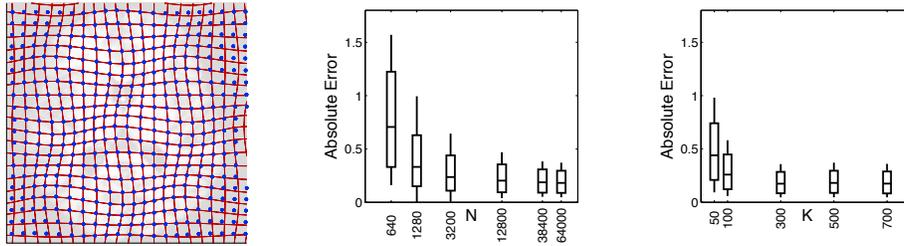
Fig. 1. (Left) LEPT (displayed as checkpoints) are able to accurately recover the synthetic field (deformed grid) within the brain volume (Middle) Influence of N on the recovering accuracy (the $x-$axis is logarithmic) (Right) Influence of K on the recovering accuracy.

for the field $L(\boldsymbol{x})$, given a (presumably valid) model of its spatial correlation encoded in a variogram: $\gamma(\boldsymbol{x}, \boldsymbol{y}) = E\left\{|L(\boldsymbol{x}) - L(\boldsymbol{y})|^2/2\right\}$. Given $(\boldsymbol{\gamma}_a(\boldsymbol{x}))_i = \gamma(\boldsymbol{a}_i, \boldsymbol{x})$, the weights $\boldsymbol{w}(\boldsymbol{x}) = (w_1(\boldsymbol{x})...w_K(\boldsymbol{x}))^T$ are obtained by solving the linear system:

$$\begin{bmatrix} \boldsymbol{w}(\boldsymbol{x}) \\ \mu \end{bmatrix} = \Gamma(\boldsymbol{x})^{-1}\boldsymbol{\gamma}_a(\boldsymbol{x}), \text{ with } \Gamma(\boldsymbol{x}) = \begin{bmatrix} (\gamma(\boldsymbol{a}_i, \boldsymbol{a}_j)) & \mathbf{1} \\ \mathbf{1}^T & 0 \end{bmatrix} \in \mathbb{R}^{(K+1)\times(K+1)}$$

where μ is the Lagrange multiplier ensuring unbiasedness of the estimate. In this paper an exponential isotropic variogram is used: $\gamma(\boldsymbol{x}, \boldsymbol{y}) = 1 - e^{-||\boldsymbol{x}-\boldsymbol{y}||/t}$.

3 Experiments and Results

The method was tested on a dataset of ten brain T1-MRI (resolution: $256\times256 \times176$). A synthetic experiment was first carried out. Inter-subject registration was then performed. Unless otherwise mentioned, parameter values are: $K=500$, $N=(20+5r)^3$ at scale r, $R=5$, $M=20$, $N_{it}=5$, $t=40$, $\lambda^2=0.3$, $\sigma_e=\sigma_n=5$. For block matching, blocks of 5^3 voxels and searching region of 9^3 voxels are used.

3.1 Synthetic Experiments

The knowledge of a ground true deformation helps understanding how the algorithm behaves. A synthetic field $T(\boldsymbol{x}) = \sin(\frac{\pi x}{50})\cos(\frac{\pi y}{50})(3,3,3)^T$ was applied to the image. Both adaptive LEPT and diffeomorphic demons accurately recovered T within the brain volume, with mean absolute error of 0.21 and 0.12 respectively (Fig. 1) and were invertible (min. jacobian of 0.67 and 0.65 respectively).

The registration was then performed for different number N of sparse matches and K of anchors. The evolution of the accuracy with N (Fig. 1) suggests that the method is not sensitive to the number of selected matches as long as this number is high compared to K. For lower N, the affine transforms can no more be robustly estimated and the performances collapse. The evolution of the accuracy with K (Fig. 1) tells us that no loss of accuracy is observed between
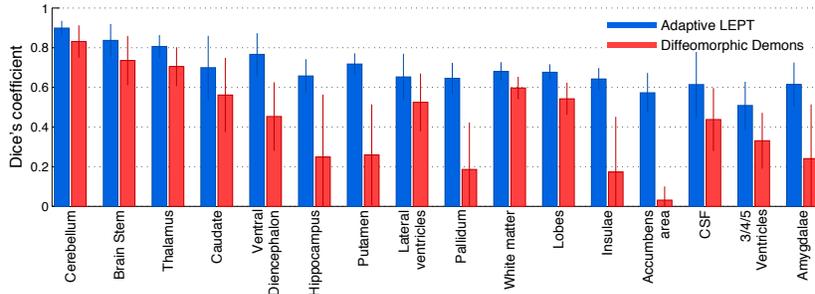
Fig. 2. Dice's coefficient with a 95% CI. Adaptive LEPT aligns structures better than diffeomorphic demons. The difference is especially significant for low contrast structures such as the putamen and the insluae.

$K = 700$ and $K = 300$, suggesting that the number of selected matches is the bottleneck here. For lower K, the performance decreases, but the method does not diverge, since all affine components can still be robustly estimated.

3.2 Inter-Subject Registration

The method was then applied to register ten multiple-sclerosis (MS) patients to one healthy subject. Each brain was skull stripped and affinely registered to the subject. Images were manually segmented by an expert in a validated protocol, providing an external validation criterion for the registration. More precisely, the Dice's coefficient between the subject label image and the label image of the patients after alignment were computed for each tissue (Fig. 2).

As a whole, adaptive LEPT aligns tissues better than diffeomorphic demons. For high contrast regions (*e.g.* lateral ventricles), both methods achieve comparable results. However, diffeomorphic demons tend to excessively favor these regions at the cost of a very poor alignment of low contrast structures (*e.g.* putamen). For these structures, the difference of performance between the two algorithms is strongly significant.

Fig. 3 depicts the aligned contours for one slice, along with the deformed grid. Again, the Jacobian of the field never fell under 0 (min. jacobian of 0.12). However, a strong pinching effect appeared in the lobe regions. These regions are typically subject specific and, while good pairs of local matches can be detected, they may be misleading in the transform estimation. Finally, note that the non-rigid deformation of the structures (as seen *e.g.* by the bending of the lateral ventricles) would not be captured if a single affine region was defined for them as in the previous LEPT registration framework.

4 Conclusion and Future Work

This paper introduced a registration algorithm that adaptively fits a LEPT based on a set of sparse matches. Results on inter-subject registration show that LEPT
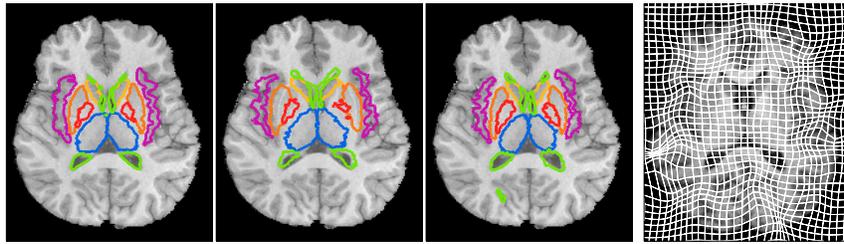
Fig. 3. (Left to right) Subject with labeled tissues, projected labels of the patient's tissues after diffeomorphic demons registration and after adaptive LEPT registration, patient image with the deformed grid overlaid.

are able to capture the local affine deformations occurring at small scales. In a future work, we want to investigate an adaptive way of incrementally defining the anchor locations such as in [3]. The choice of location could include a cost related to the confidence of the estimation in order to avoid the pinching effect observed in Fig. 3.

Acknowledgements

References

1. Akselrod-Ballin, A., Bock, D., Reid, R., Warfield, S.: Improved registration for large electron microscopy images. In: IEEE ISBI. pp. 434–437. IEEE (2009)
2. Arsigny, V., et al.: A Fast and Log-Euclidean Polyaffine Framework for Locally Linear Registration. JMIV 33(2), 222–238 (2009)
3. Buerger, C., Schaeffter, T., King, A.: Hierarchical adaptive local affine registration for fast and robust respiratory motion estimation. MedIA (2011)
4. Cachier, P., Bardinet, E., Dormont, D., Pennec, X., Ayache, N.: Iconic feature based nonrigid registration: the pasha algorithm. CVIU 89(2-3), 272–298 (2003)
5. Clatz, O., Delingette, H., Talos, I., Golby, A., Kikinis, R., Jolesz, F., Ayache, N., Warfield, S.: Robust nonrigid registration to capture brain shift from intraoperative mri. IEEE TMI 24(11), 1417–1427 (2005)
6. Commowick, O., Arsigny, V., Isambert, A., Costa, J., Dhermain, F., Bidault, F., Bondiau, P., Ayache, N., Malandain, G.: An efficient locally affine framework for the smooth registration of anatomical structures. MedIA 12(4), 427–441 (2008)
7. Granger, S., Pennec, X.: Multi-scale EM-ICP: A fast and robust approach for surface registration. ECCV pp. 69–73 (2002)
8. Ruiz-Alzola, J., Westin, C., Warfield, S., Alberola, C., Maier, S., Kikinis, R.: Non-rigid registration of 3d tensor medical data. MedIA 6(2), 143–161 (2002)
9. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Diffeomorphic demons: Efficient non-parametric image registration. NeuroImage 45(1), S61–S72 (2009)