Piecewise Affine Registration of Biological Images

Alain Pitiot[1,2], Grégoire Malandain[1], Eric Bardinet[3,1], and Paul M Thompson[2]

[1] Epidaure, INRIA, 2004 route des lucioles BP 93, 06 902 Sophia-Antipolis, France
[2] LONI, UCLA School of Medicine, Los Angeles, CA 90095, USA
[3] CNRS UPR640-LENA, 47 blvd. de l'hôpital, 75651 Paris, France

Abstract. This manuscript tackles the registration of 2D biological images (histological sections or autoradiographs) to 2D images from the same or different modalities (*e.g.*, histology or MRI). The process of acquiring these images typically induces composite transformations that can be modeled as a number of rigid or affine local transformations embedded in an elastic one. We propose a registration approach closely derived from this model. Given a pair of input images, we first compute a dense similarity field between them with a block matching algorithm. A hierarchical clustering algorithm then automatically partitions this field into a number of classes from which we extract independent pairs of sub-images. Finally, the pairs of sub-images are, independently, affinely registered and a hybrid affine/non-linear interpolation scheme is used to compose the output registered image. We investigate the behavior of our approach under a variety of conditions, and discuss examples using real biomedical images, including MRI, histology and cryosection data.

1 Introduction

A key component of medical image analysis, image registration essentially consists of bringing two images, acquired from the same or different modalities, into spatial alignment. For instance, monomodal registration of a population's MRIs can be used to build anatomical atlases [1,2], while mono- or multi-modal registration of the same patient's data can help determine the nature of an anomaly or monitor the evolution of a tumor or other disease process.

In particular, pair-by-pair registration of a series of 2-D biological images (histological sections or autoradiographs) enables the reconstruction of a 3D biological image. Subsequent fusion with 3D data acquired from tomographic imaging modalities (*e.g.* MRI) then allows the tissue properties to be studied in an adequate anatomic framework, using *in vivo* reference data.

More formally, given two input images, registering the floating (*i.e.*, moveable) image to the reference (*i.e.*, fixed) one entails finding the transformation that minimizes the dissimilarity between the transformed floating image and the reference. As such, it can be decomposed into 3 elements: a transformation space, a similarity metric and an optimization algorithm (see [3] for an extensive review of medical image registration). For iconic (*i.e.*, intensity-based) methods, optimal similarity measures can be derived from a careful analysis of the expected relationships between the input images [4], with different hypotheses leading to different measures. Similarly, *a priori* knowledge about the acquisition process for biological images may allow the transformation space to be modeled more accurately.

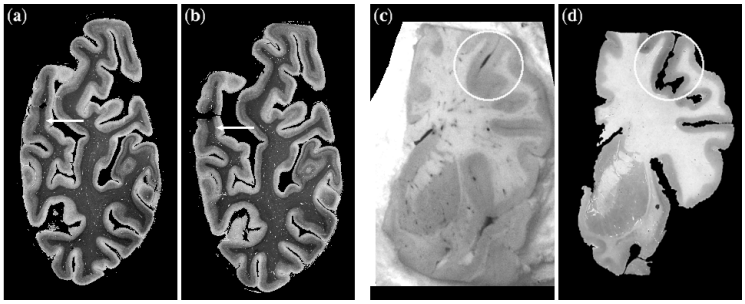

Fig. 1. Two consecutive myelin-stained histological sections of the human brain (a & b); Human brain cryosection (c) and its associated Nissl-stained section (d).

In our case, the cutting process, successive chemical treatments, and the glass mounting step that a slab of tissue undergoes during a histological preparation yield a fairly flexible global transformation that is however locally affine for some identifiable components of the section. In brain sections for instance, each gyrus (compare white arrows in Figure 1.a & b, and white circles in c & d) undergoes an affine transformation (due to successive manipulations) relatively independent from those of other gyri. Consequently, even though a large variety of transformation spaces have been discussed in the literature (among others, one finds linear (rigid, affine) and non-linear transformations (polynomial [5], elastic [6,7] or fluid [8]), their functional form may not reflect our specific needs.

Note that the utility of the above-stated transformation model extends to medical as well as biological images (our primary motivation here). For instance, abdominal or torso MRIs often include rigid structures such as bones (ribs, vertebrae, etc.), deformable organs (liver, heart, etc.), and elastic tissues. Two abdominal MRIs of the same patient are then linked by a complex transformation which can be rigid in some regions (for bones) but potentially exhibits large local dilations (in deformable organs). Global rigid or affine transformations obviously cannot handle such a case adequately. Also, a single rigid transformation would not correctly register all the vertebrae along the spinal column simultaneously. Furthermore, high degree of freedom (*e.g.*, fluid) transformations could correctly map one image onto the other, but they may not ensure that specific components (*e.g.*, bones) will be only rigidly transformed.

To alleviate these issues, a few authors have developed local registration techniques, where the input images are divided into a number of smaller sub-images, and a transformation is associated with each. An automatic hierarchical elastic image registration technique is presented in [9]. The initial 2-D images are partitioned into quadtree structures. At each level of the quadtree, the floating sub-images are independently registered to their counterparts in the reference image, before being merged via thin-plate spline interpolation. However, this technique cannot apply a transform selectively to a specific region whose boundary does not coincide with the quadtree grid. In [10], Little *et al.* describe an approach where a user selects a number of pairs of corresponding rigid structures in the input images along with associated linear transformations (also given by the user).

A number of pairs of landmarks further constrain a hybrid affine/non-linear interpolation scheme that acts as a local registration algorithm. This method, even though it fits our needs, still relies on interactive specification of the components to be rigidly matched.

This paper addresses the problem of *automatically* registering two images, when the images consist of a number of independent components, subject to linear transformations. Figure 2 illustrates our approach. First a correspondence field (section 2.1) is computed between the two images. From a hierarchical clustering of this field (section 2.2), we automatically extract pairs of sub-images (with an assumed linear transformation relation between the pairs). These pairs are then, separately, rigidly or affinely registered. Finally, the hybrid affine/non-linear interpolation scheme described in [10] is used to compose the registered floating image (section 2.3). Results on phantom and real data are presented in section 3.

2 Method

The first step of our approach consists of *automatically* partitioning the input floating and reference images, I_F and I_R, into a number of pairs of corresponding sub-images, where each sub-image is associated with an independent image component (in terms of transformation). We approach this segmentation issue as a process of partitioning a correspondence field computed from I_F to I_R. Our method is motivated by the following observation. When both images are composed of pairs of independent components, where each component is subject to some linear transformation, the associated correspondence field should exhibit rather homogeneous characteristics within each component, and heterogeneous ones across them.

2.1 Computing the correspondence field

We use a block-matching algorithm [11] to compute the correspondence field. We associate with I_F and I_R two sparse regular rectangular lattices $L_F = [1 \dots w_F] \times [1 \dots h_F]$ and $L_R = [1 \dots w_R] \times [1 \dots h_R]$. From these we discard, for histological sections, sites that lie on the background. Note that we may choose to associate a site to each pixel of the input images, in which case w_F, h_F and w_R, h_R would be the width and height of I_F and I_R.

The correspondence field then associates to each site (i, j) in L_F a 2-D spatial similarity distribution and an "optimal" displacement, computed as follows. For each site (i, j) in L_F, we consider in I_F a neighborhood $B^{i,j}_{I_F}$ of the pixels associated with (i, j) (usually a square neighborhood of constant size called a "block"), whose centroid is denoted by $p^{i,j}_F$. We then compute the similarity measures (given a similarity metric sim) between block $B^{i,j}_F$ and every block $B^{k,l}_R$ in I_R associated to sites (k, l) in the corresponding neighborhood $N_{i,j}$ of (i, j) in L_R (the "exploration neighborhood"). For every site (i, j) in L_F, we then get a 2-D spatial similarity distribution (the values $sim(B^{i,j}_F, B^{k,l}_R)$ defined in the neighborhood $N_{i,j}$ of (i, j) in L_R), and an "optimal" displacement $d^{i,j}$ defined by $d^{i,j} = p^{(k,l)_{max}}_R - p^{i,j}_F$ where $(k, l)_{max}$ is the site of L_R that

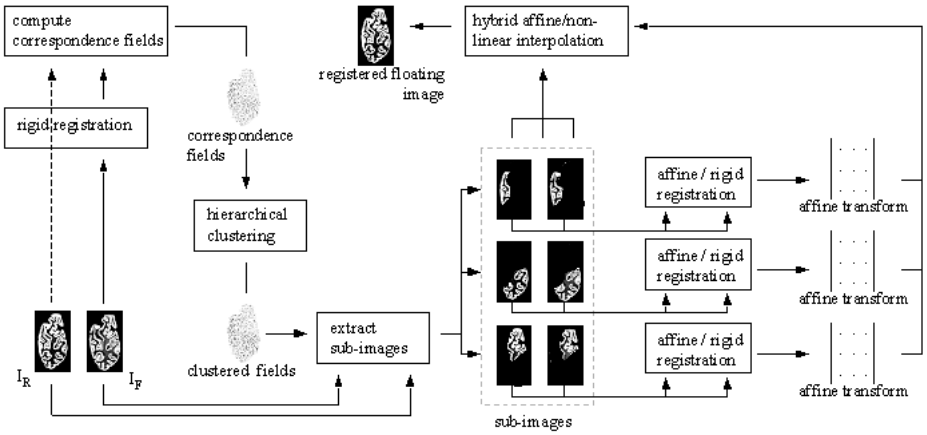

Fig. 2. Overview of our local registration approach

is associated to the block $B_R^{k,l}$ in $N_{i,j}$ which is the most similar to $B_F^{i,j}$, *i.e.* $(k,l)_{max} = \arg\max_{k,l} sim(B_F^{i,j}, B_R^{k,l})$.

The quality of both the similarity distribution and the displacement field is essentially determined by three parameters: the block size, the similarity metric and the size of the exploration neighborhood in L_R. The similarity metric and block size must reflect the expected relationship between the intensity distributions of blocks in the floating and reference images, and the scale of the features of interest within those blocks [12,1]. For most of our experiments, we used the correlation coefficient [13] (which assumes an affine relation between the intensity distributions of the blocks). The size of the exploration neighborhood is linked to the expected magnitude of the residual displacements after global alignment. It conditions the extent to which our registration algorithm can recover large deformations. In our case, it is input by the user (typically 15 pixels here).

As a pre-processing step, we first rigidly register I_F to I_R to remove the global rigid transform (that uniformly affects all components) from the subsequently computed correspondence fields. We use the fully automated intensity-based registration algorithm presented in [11], where a *robust* multi-scale block-matching strategy was introduced.

2.2 Extracting the image components

Clustering the correspondence field We are looking for a *hierarchical clustering* of L_F, that is, a sequence of partitions in which each partition is nested into the next partition in the sequence [14]. Cluster analysis (unsupervised learning) essentially consists of sorting a series of multi-dimensional points into a number of groups (clusters) to maximize the intra-cluster degree of association and minimize the inter-cluster one. It is particularly well-suited here as it behaves adequately even when very little is known about the category structure of the input set of points.

Our clustering method is adapted from the standard agglomerative hierarchical clustering algorithm [15]:

step 1: initialize a cluster list by placing each site of L_F in an individual cluster, and let the distance between any two of these clusters be the distance between the sites they contain.

step 2: find the closest pair of clusters, remove them from the cluster list, merge them into a new single cluster and add the new cluster to the cluster list.

step 3: compute the distances between the newly formed cluster and the other ones in the cluster list.

step 4: repeat steps 2 and 3 until the desired number of clusters is reached.

The number of clusters can either be specified by the user (our case here), or by using pre-indicators like the Davies-Bouldin index [16] or the cophenetic correlation coefficient [14] to assist this choice.

Computation of these cluster distances is the pivotal element of the clustering algorithm. The distance should be consistent with the model we choose for the input images and the relationships we expect between them.

To define a distance on clusters, we first need a distance on sites. Given two sites, t and u, their distance is defined as a linear combination of three distances:

$$D_{site}(t,u) = \alpha\ D_p(p_F^t, p_F^u) + \beta\ D_d(d^t, d^u) + \gamma\ D_\rho(\rho^t, \rho^u) \tag{1}$$

$D_p(p_F^t, p_F^u)$ is the geodesic distance between the block centroids. It ensures that close blocks are more likely to be clustered than blocks that are far apart (model constraint), while adequately representing proximity from an anatomical point of view when the input images contain several pieces of tissue.

$D_d(d^t, d^u)$ is the distance between the block displacement vectors, we choose the modulus of their difference, *i.e.* $\|d^t - d^u\|$.

$D_\rho(\rho^t, \rho^u)$ is the distance between the block similarity distributions. Clearly, the "optimal" displacement may sometimes disagree with the actual movement (because of noise, decoys, etc.). The similarity distribution is then better-suited to capture fine structural elements and discriminate between blocks. We use a normalized version ρ of these distributions to ensure that they all have the same overall unit mass. For each site t in L_F and each site u in its exploration neighborhood, ρ is given by: $\rho^t(p_R^u - p_F^t) = sim(B_F^t, B_R^u) / \sum_v sim(B_F^t, B_R^v)$. As a distance between distributions, we chose the Earth mover's distance [17], a discrete solution to the discrete Monge-Kantorovich mass-transfer problem [18]. Given the "ground distance" (the distance between elements of the distribution, the Euclidean distance in our case), the Earth mover's distance (EMD) between two distributions becomes the minimal total amount of work (= mass × distance) it takes to transform one distribution into the other. As argued by Rubner *et al.* [17], this boils down to a bipartite network flow problem, which can be modeled with linear programming and solved by a simplex algorithm.

α, β and γ are real-valued positive weights. We choose for β a value substantially smaller than that of α and γ (typically, $\beta = \frac{\alpha}{50}$).

Once we have a distance between blocks, a cluster distance can be defined. We adapted the standard complete link distance [14] to additionally take into account the transformations that can be estimated on the clusters already formed. Namely, when the size of a cluster reaches a given threshold (we usually take 20,

even though experiments showed that the value of that threshold does not really impact the quality of the clustering), a rigid or affine transformation can be estimated, in a robust fashion, from the associated set of "optimal" displacement vectors (*e.g.*, the quaternion-based approach presented in [19]). The decision to merge two clusters can then be biased by the agreements between the directions of the "optimal" displacement vectors of one cluster with the estimated transformation of the other as this might indicate that they belong to the same component.

Given a site t in L_F, its distance to a transformation T is defined by the coherence between T and the "optimal" displacement vector associated with t. Consequently, the transformation-based distance between a site t and a cluster C is defined by the distance between t and T^C, the transformation estimated over C: $D_{tran}(t, C) = \|T^C(p_F^t) - (p_F^t + d_F^t)\|$ if T^C is defined, and 0 otherwise.

Given two clusters of sites $C^a = \{a_1, \ldots, a_{n_a}\}$ and $C^b = \{b_1, \ldots, b_{n_b}\}$, with associated estimated transformations T^a and T^b respectively, the cluster distance between them is defined as the longest distance from any block of C^a to any block of C^b (complete-link) *plus* the sum of the "transformation distances" wherever they can be computed:

$$D_{cluster}(C^a, C^b) = \max_{i,j} D_{site}(a_i, b_j) + \sum_j D_{tran}(b_j, C^a) + \sum_i D_{tran}(a_i, C^b) \quad (2)$$

Extracting the sub-images Let N_C be the final number of clusters, $C = \left\{C^1, \ldots, C^{N_C}\right\}$ the cluster partition of L_F, and $\left\{c_1^i, \ldots, c_{n_i}^i\right\}$ the n_i sites of the i^{th} cluster C^i. We want to build a set of N_C sub-images $\left\{I_F^i\right\}_{i=1}^{N_C}$, each of them associated with a single cluster. Our clustering method does not ensure that the borders between clusters are sufficiently precise to adequately represent the sub-images' borders. Moreover, as we are going to use these sub-images to find local transformations, it is often better to choose larger supports to avoid boundary effects.

Consequently, rather than build a partition of I_F from the partition of L_F, we build a covering of I_F, *i.e.*, a set of sub-images that could overlap. To do so, we aggregate in I_F^i the pixels of I_F in the vicinity of the sites of the cluster C^i. We get: $I_F^i = \{(x, y) \in I_F \text{ such that } D((x, y), p_F^{c_j^i}) \leq radius \text{ for some } c_j^i \in C^i\}$ In practice we use the L_∞ distance. Then, if the blocks associated to the sites are of size $n \times n$, taking $radius = n/2$ we get $I_F^i = \bigcup_j B_{I_F}^{s_j^i}$. In our experiments, to ensure a large support for the sub-images, we chose $radius = 3n/4$.

The corresponding reference sub-images I_R^i are built identically, but with the centroids $p_R^{(k,l)_{max}}$ of the most similar blocks (see Section 2.1): $I_R^i = \{(x, y) \in I_R \text{ such that } D((x, y), p_F^{c_j^i} + d^{c_j^i}) \leq radius, \text{ for some } c_j^i \in C^i\}$. Again, we use the L_∞ distance here, with $radius = n$ (a larger extent than that of the floating sub-image) to ensure that I_F^i can be effectively registered against I_R^i.

2.3 Composing the registered floating image

Once we have extracted the reference and floating sub-images, we use the robust block-matching algorithm described in [11] to register them, independently, pair

by pair. For each pair $\left\{I_R^l, I_F^l\right\}$, $l \in 1 \ldots N_C$, we obtain an affine transform T^l. Note that as these registrations are robust, the sub-images do not need to perfectly correspond to the anatomically separate components.

We then compose the final registered floating image using the Little *et al.* method (see [10] for details). Their approach applies user-provided affine transforms to user-defined structures and ensures a smooth interpolation in between them. In our application, the set of floating sub-images forms a covering of the input floating image, so we have to erode the sub-images to leave space for interpolation. Furthermore, the floating sub-images must be cut to ensure that they do not overlap, once transformed, as this may also impair the interpolation scheme.

Note that the entire registration process could easily be included within an iterative multi-scale framework to achieve a better trade-off between accuracy and complexity. Such a framework could also handle both large-scale and small-scale components.

3 Results

3.1 Synthetic experiments

Synthetic fields enable us to evaluate the quality of the clustering algorithm independently from textural issues and aside of considerations regarding the similarity distribution. That is, in simulated data, we can assume that the "optimal" displacement field is the correct one. Due to a lack of space, we outline the quantitative results only briefly here. Please refer to `http://www-sop.inria.fr/epidaure/personnel/apitiot/WBIR/index.html` for the corresponding figures and tables.

The first experiment was designed to illustrate how our approach behaves on images consisting of several *separate* independent components. We considered a spine-like structure with 4 components (4 squares) to which 4 different affine transformations were applied. We computed for each component an exact field and then corrupted it with a uniform noise: for each component, $\alpha\%$ of the displacement vectors were selected for perturbation. Our algorithm managed to perfectly cluster the components even for very high values of noise ($\alpha = 75\%$), mostly due to the use of the geodesic distance. Additionally, the estimated transformations were within 5% of the desired ones up until $\alpha = 75\%$. A second experiment demonstrated the behavior of our approach for a *connected* component. We used similar settings for 3 more structures with 2 components each (a small square inside a larger one, and two vertical halves of a large square). Since the geodesic distance did not help here, we obtained inferior performances (20% more wrongly clustered vectors at $\alpha = 50\%$). This may also be because the displacement vectors were quite similar on both sides of the border between the components. Throughout, clustering results with the transformation-based distance were consistently better than without, and even more so when the noise level increased (40% more wrongly clustered vectors at $\alpha = 25\%$, 50% at $\alpha = 75\%$). Incidentally, a perfect clustering is not necessary since the subsequently extracted

sub-images are "larger" than the clusters and their independent registrations are robust.

We used phantom images and the same experimental settings to evaluate how our clustering algorithm performed under controlled textural conditions. Namely, pairs of synthetic images were created with artifical textures (grayscale cloud patterns) and the same transformations applied as above. We observed better performances when the EMD distribution distance was used than when it was not (15% better on average). The similarity distribution distance helps the clustering algorithm to form, at early stages, sensible clusters. These are then agglomerated with the aid of the robust estimation of the associated transformations. Again, results were better with the cluster transformations estimation (9% better). However the difference in performance was less obvious than before: as the optimal displacement field is very noisy, the transform distance only marginally helps the clustering process. A more thorough validation using complete series of histological sections is in progress.

3.2 Biomedical Images

Figure 3 displays the results of our local registration for a pair of myelin-stained histological sections (myelin-stained coronal section through the occipital cortex, first row), and for a Nissl stained human brain section and its associated cryosection image (second row). It demonstrates the behavior of the piecewise approach (1) when reconstructing a histological block from a series of stained sections, and (2) in a multi-modal registration problem. These examples present two classic registration difficulties: in the first case, a gyrus (top left corner) was detached during the histological preparation and manually realigned in an unsatisfactory fashion; in the second case, many gyri were separated during histological treatment.

The reference images are those of Figure 1. For each pair of images, we show in Figure 3 the reference image with the superimposed edges of the globally affinely registered floating image (a), the floating image with clustered optimal displacement field (b), the image of a regular grid convected by the associated hybrid affine/non-linear transformation (c), the locally registered floating image (d), and the reference image with the superimposed edges of the locally registered floating image (e).

To better demonstrate how our approach can register multiple component images, we applied only rigid transforms to the sub-images of the myelin-stained sections. The clustering algorithm adequately isolated in a separate sub-image the floating gyrus which was subsequently correctly registered to its counterpart in the reference image. Our technique also successfully compensated for smaller rotations applied by the operator in the glass mounting step. An affine transform would of course further decrease the discrepancy between the pairs of sub-images. However, in the general case, when one suspects only a rigid transformation between sub-images, opting for an affine registration would only introduce unnecessary over-parameterization. Among other disadvantages this could substantially alter textures.

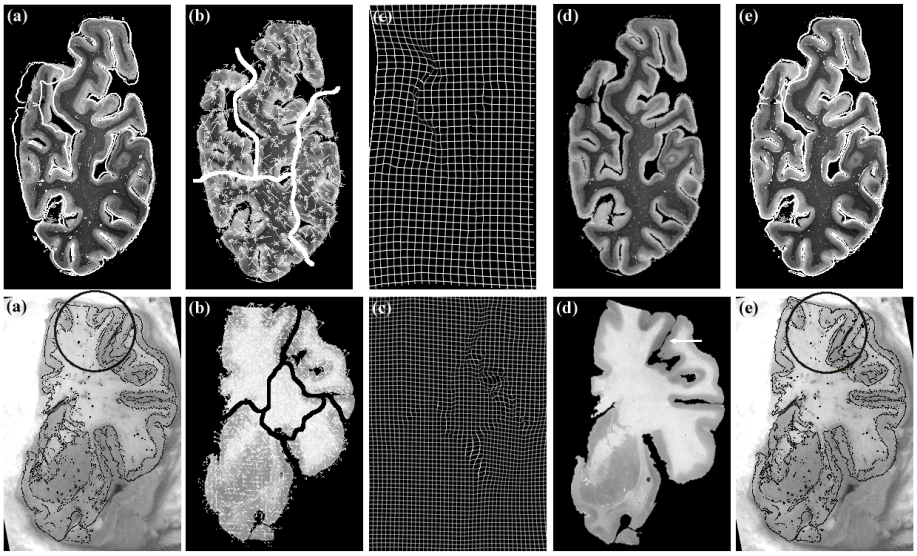

Fig. 3. Registration of two consecutive myelin-stained histological sections of the human brain (first row) and of a Nissl-stained human brain section and its associated cryosection (second row).

For the multi-modal example, we used affine registration and specified a larger number of clusters (6) to take into account the many inadequately manually realigned components. Our algorithm successfully closed most of the gyri (see white arrow in Figure 3.d and compare circled area in 3.a and 3.e) while minimizing the overall amount of deformation.

4 Conclusion and Perspectives

We have presented a fully automated local registration method, capable of dealing with a variety of 2-D images. It builds complex spatial transformations by elastically interpolating between rigid or affine transforms that are locally defined on pairs of sub-images. Consequently, we minimize the overall number of degrees of freedom of the transformation, following the guidelines of the parsimony principle (see [20] for a discussion of some issues associated with high-dimensional transformations). These sub-images represent geometrically coherent components (in our biomedical applications, they are even anatomically coherent components). They are automatically extracted from an initial correspondence field computed between the images. All user interaction is avoided, by contrast with other approaches [10].

The use of a hierarchical clustering approach and a similarity distribution distance proved promising: while the distribution distance can effectively deal with noise and textural issues to discriminate between image blocks, our clustering algorithm manages to extract the expected sub-images.

Results on real data showed that the proposed method is adequate for several specific problems in biomedical imaging. Finally, even though the presented

method works in 2-D, it could readily be extended to 3-D (or n-D) with a close-to-linear increase in processing time.

References

1. Collins, D.L., Zijdenbos, A.P., Paus, T., Evans, A.C.: Use of registration for cohort studies. In Hajnal, J., Hawkes, D., Hill, D., eds.: Medical Image Registration. (2003)
2. Thompson, P.M., Woods, R.P., Mega, M.S., Toga, A.W.: Mathematical/computational challenges in creating deformable and probabilistic atlases of the human brain. Human Brain Mapping **9** (2000) 81–92
3. Maintz, J.B.A., Viergever, M.A.: A survey of medical image registration. Medical Image Analysis **2** (1998) 1–36
4. Roche, A., Malandain, G., Ayache, N.: Unifying Maximum Likelihood Approaches in Medical Image Registration. International Journal of Imaging Systems and Technology: Special Issue on 3D Imaging **11** (2000) 71–80
5. Woods, R.P., Grafton, S.T., Holmes, C.J., Cherry, S.R., Mazziotta, J.C.: Automated image registration: I. General methods and intrasubject, intramodality validation. J. Comput. Assist. Tomogr. **22** (1998) 141–154
6. Davatzikos, C.: Spatial transformation and registration of brain images using elastically deformable models. CVIU **66** (1997) 207–222
7. Gee, J.C., Reivich, M., Bajcsy, R.: Elastically Deforming 3D Atlas to Match Anatomical Brain Images. J. Comput. Assist. Tomogr. **17** (1993) 225–236
8. Christensen, G.E.: Consistent linear-elastic transformations for image matching. In: Proc. of IPMI. Volume 1613 of LNCS., Springer (1999) 224–237
9. Likar, B., Pernus, F.: Registration of Serial Transverse Sections of Muscular Fibers. Cytometry **37** (1999) 93–106
10. Little, J.A., Hill, D.L.G., Hawkes, D.J.: Deformations Incorporating Rigid Structures. Computer Vision and Image Understanding **66** (1997) 223–232
11. Ourselin, S., Roche, A., Subsol, G., Pennec, X., Ayache, N.: Reconstructing a 3D Structure from Serial Histological Sections. Image and Vision Computing **19** (2001) 25–31
12. Dengler, J.: Estimation of discontinuous displacement vector fields with the minimum description length criterion. In: Proc. of CVPR, IEEE (1991) 276–282
13. Penney, G.P., Weese, J.W., Little, J.A., Desmedt, P., Hill, D.L.G., Hawkes, D.J.: A comparison of similarity measures for use in 2D-3D medical image registration. In: Proc. of MICCAI. Volume 1496 of LNCS., Springer (1998) 1153–1161
14. Backer, E.: Computer-assisted reasoning in cluster analysis. Prentice Hall (1995)
15. Johnson, S.C.: Hierarchical Clustering Schemes. Psychometrika **2** (1967) 241–254
16. Davies, D.L., Bouldin, D.W.: A cluster separation measure. IEEE Trans. PAMI **1** (1979) 224–227
17. Rubner, Y., Tomasi, C., Guibas, L.: A metric for distributions with applications to image databases. In: Proc. of ICCV, Bombay, India, IEEE (1998)
18. Haker, S., Angenent, S., Tannenbaum, A.: Minimizing Flows for the Monge-Kantorovich Problem. SIAM Journal of Math Analysis (2003) [to appear].
19. Horn, B.K.P.: Closed-form solution of absolute orientation using unit quaternions. Journal of the Optical Society of America A **4** (1987) 629–642
20. Kjems, U., Hansen, L.K., Chen, C.T.: A non-linear 3d brain co-registration method. In Hansen, P.C., ed.: Proceedings of the Interdisciplinary Inversion Workshop 4, Lyngby, Denmark, IMM, Technical University of Denmark (1996)