Chapter 7

Registration and Averaging

With the methods described so far it is possible to create anatomical models of individuals. This is how conventional anatomical atlases, which are based on photographies or hand-drawings, were built. And already this can be very helpful. With computer based three dimensional neuro-anatomy, however, we can do much more. The atlas can be statistically secured, i.e. it can be assured that with respect to certain criteria the atlas represents an average. Additionally it is possible to import data from other individuals into the atlas, in order to interpret the data or to enrich the atlas. We will give examples in the applications Section 8.

7.1 Registration

Registration or alignment is a prerequisite whenever data sets from two different individuals are to be processed together. Alignment is the process of finding a transformation for one of the data set such that points of that data set are transformed to semantically corresponding points of the other data set. The question how to define such a correspondence in case the data sets come from different individuals and whether such a correspondence exists at all, can be very difficult however.

Registration algorithms can be classified by two essential properties: The set of allowed transformations and a criterion by which a particular transformation is chosen. The criterion can be an interactive manual specification by a user, or a computational quality measure that is to be maximized. An example would be an algorithm that chooses from all possible rigid transformations (rotation+translation) the one that maximizes an intensity correlation. In such cases it is appropriate to break down the second step mentioned above into a quality function and an optimization strategy. A survey of various techniques is given in [14].

7.1.1 Intensity Correlation

Intensity based measures are a natural quality criterion for an image registration method. One can expect that the same structure will have similar intensity values across different experiments or different individuals, if recorded under the same conditions. Therefore, a suitable quality measure would be:

$$Q(T) = - \int \left[A(\vec{x}) - B\left(T(\vec{x})\right) \right]^2 d\vec{x} \qquad (7.1)$$

where A and B are the image data sets and T is a transformation that is supposed to transform points in object A to corresponding points in object B.

In general for digital images the integral is discretized by taking the sum over all pixels. If sub-voxel accuracy is desired at least a suitable interpolation scheme has to used to evaluate B at locations which are not in its voxel centers.

$$Q_{\text{ssd}}(T) = -\sum_i \left[A(\vec{x}_i) - B\left(T(\vec{x}_i)\right) \right]^2 \tag{7.2}$$

If the intensity values for corresponding points in the two images are not the same, e.g. due to different recording conditions, but if still an (unknown) linear relation can be assumed, then often normalized cross correlation or the cross correlation coefficient is used:

$$Q_{\text{ncc}}(T) = \frac{\sum_i A(\vec{x}_i) B\left(T(\vec{x}_i)\right)}{\left[\sum_i B^2\left(T(\vec{x}_i)\right)\right]^{\frac{1}{2}}} \tag{7.3}$$

$$Q_{\text{ccoeff}}(T) = \frac{\sum_i \left[A(\vec{x}_i) - \mu_A \right]\left[B\left(T(\vec{x}_i)\right) - \mu_B \right]}{\left[\sum_i \left(A(\vec{x}_i) - \mu_A \right)^2 \left(B(T(\vec{x}_i)) - \mu_B \right)^2 \right]^{\frac{1}{2}}} \tag{7.4}$$

Finally, there are cases where the assumption of linear intensity correlation does not hold. This is typically the case when image data sets from different recording modalities are registered, like an MRI and a CT data set. In this case, structures being bright in one image could be dark in the other and vice versa. Still, however, it can be assumed that at least some of the structures are depicted in both images, and that there is a stochastical correlation between intensity values. A measure can be derived based on the joined entropy of the two images, which is assumed to be minimized by the optimal transformation. This leads to a measure called the *mutual information* (7.5) [82, 133, 34] or the *normalized mutual information* (7.6) [123].

$$Q_{\text{mi}}(T) = H(A) + H(B) - H(A, T(B)) \tag{7.5}$$

$$Q_{\text{nmi}}(T) = \frac{H(A) + H(B)}{H(A, T(B))} \tag{7.6}$$

$$H(A) = \sum_i p_A(g_i)\log(p_A(g_i))$$

$$H(A, T(B)) = \sum_{i,j} p_{AT(B)}(g_i, gj)\log(p_{AT(B)}(g_i, g_j)),$$

where $p_A(g_i)$ denotes the probability of intensity g_i to occur at a given point in A and $p_{AT(B)}(g_i, gj)$ denotes the probability of coincidence of intensity g_i in image A and g_j in the transformed image $T(B)$; i, j loop over all possible gray values.

Intensity correlations are often used in conjunction with global rigid transformations. An optimization procedure like a gradient-descent, conjugate gradient based method, or a simulated annealing method searches in the six-dimensional search space for the best transformation. To increase performance and avoid sensibility to local optima and noise, often a multi-scale method is chosen, that starts on a low-pass filtered version of the data set (compare Section 5.2.1). Most of the optimization methods used in practice do not guarantee that the global optimum is detected, therefore

often a suitable initial position has to be chosen manually. The extension to further degrees of freedom like isotropic scaling, non-isotropic scaling, or shearing is straight forward.

If local deformations are needed to eliminate differences, non-affine transformations must be chosen. Beside landmark based methods that will be discussed in the next section, a commonly used method is based on interpolation of translational transformations on a uniform grid, often applied hierarchically and with a precombined global rigid or affine transformation . A typical algorithm could read like this:

1. Compute global affine transformation $G : A \rightarrow B$

2. Subdivide the (cubical) domain of A into $2 \times 2 \times 2$ cells by defining a rectangular $3 \times 3 \times 3$ grid with one point in the center.

3. Define a translation vector $\vec{t}_j$ on the 27 grid points and interpolate trilinear within the 8 grid cells, resulting in a continuous translation field $\vec{t}(\vec{x})$. Find the vectors $\vec{t}_j$ that optimize the transformation $T(\vec{x}) = G(\vec{x}) + \vec{t}(\vec{x})$ with respect to a quality measure.

4. Continue to subdivide the grid cells and move the points until a certain quality or a predefined cell size is reached.

7.1.2 Landmark Based Methods

Other alignment methods are based on the definition of landmarks. These are discrete locations that can be identified in both data sets, either automatically or manually. Given a set of N landmarks $\vec{a}_i$ in data set A and the set of corresponding landmarks $\vec{b}_i$ in data set B, a natural way to define a rigid transformation is to use that rigid transformation that minimizes

$$Q(T) = \sum_i \left| T(\vec{a}_i) - \vec{b}_i \right|^2.$$

The translational part can be computed by shifting the barycenters of the two sets $\{\vec{a}_i\}$ and $\{\vec{b}_i\}$ onto each other. The rotational part of the optimal transformation can be found by solving a 3×3 eigenproblem [51, 52]. We define

$$M = \left(\sum_{i=1}^{N} \vec{a}_i \otimes \vec{b}_i \right).$$

Here $(\vec{a} \otimes \vec{b})_{\kappa\lambda} \equiv a_\kappa b_\lambda$, with $\kappa, \lambda = 1 \ldots 3$. Let $\mu_1 \geq \mu_2 \geq \mu_3$ be the sorted eigenvalues of the symmetric, positive definite 3×3 matrix $M^T M$. Let $\vec{p}_1, \vec{p}_2$ be associated normalized eigenvectors and

$$\vec{q}_j = M_i \vec{p}_j \quad (i \in 1, 2).$$

Then, with $\vec{p}_3 \equiv \vec{p}_1 \times \vec{p}_2$ and $\vec{q}_3 \equiv \vec{q}_1 \times \vec{q}_2$, we can compute the optimal rotation matrix

$$R = \sum_{l=1}^{3} \vec{q}_l \otimes \vec{p}_l.$$

The rigid transformation will in general not transform each point $\vec{a}_i$ exactly onto $\vec{b}_i$. If this is desired, i.e. if

$$T(\vec{a}_i) \stackrel{!}{=} \vec{b}_i \qquad (7.7)$$

is required, then the problem can be formulated as a scattered data interpolation problem. Different methods exist. For example a Delaunay triangulation can be computed (tetrahedrization in three dimensions) and a linear interpolation of the translation vector can be applied within each simplex. Higher order methods like simplex splines could be used as well. The quality of such a method depends to a great degree on the distribution of the points. In practice inhomogeneous and coarse distribution of points often lead to poor grid quality, i.e. many simplices with bad aspect ratio exists. This leads to unnatural transformation fields.

Another method is to define the interpolant as the linear combination of radial basis function centered at the landmarks. A set of basis functions which is very popular in medicine and biology are the so called thin-plate spline interpolation, sometimes also called Bookstein splines after [10].

Let $S : \Re \to \Re$ be a continuous, non-constant function. We then define a transformation field by

$$T(\vec{x}) = \sum_i \vec{\lambda}_i S(\left| \vec{x} - \vec{b}_i \right|) \qquad (7.8)$$

By inserting Equation (7.7) into Equation (7.8), we will get a linear system of equations. By solving it we can determine the coefficients $\vec{\lambda}_i$.

Bookstein [10] has used a basis function $S(x) = x^2 \log(x)$. He has motivated this with a physical analogy. If a two dimensional thin metal plate was forced to elevate to given z values at a discrete set of points, the elevation of the plate at any other point could be determined by interpolating the z values using the described method with these functions. The interpolant minimizes the overall bending energy. For the 3-dimensional case the function x^3 should be used [127]. Of course virtually any "reasonable" function would fulfill the interpolation property, see for example [60] and [17] for alternatives.

The transformations constructed this way are continuous, smooth, and they yield the desired transformation at the landmarks. In general, however, it is not guaranteed that they are invertible. It is well possible that two different points in data set A are mapped to the same point in B. At the same time potentially there are points in B that have no correspondence in A. In particular when exchanging the two data sets in Equations (7.7) and (7.8) one would expect to obtain the inverse transformation, but this is not the fact. Therefore, from a biological point of view the interpretation of the transformation as a point-wise correspondence has to be done with great care.

7.1.3 Labels

The data sets we deal with in this work often have no well identified point-like landmarks. Instead, corresponding voluminous structures can be identified. As described earlier these often will be segmented for quantification, or geometry reconstruction. Therefore, it is a natural idea, to use this information for the registration as well. As we will detail in the *Applications* chapter, we have used labels for registration in different ways.

In the first case, a transformation field was to be computed for the antennal lobe of the honey bee. The antennal lobe consist of roughly 180 so called glomeruli, which have almost spherical shape.

Some of them where segmented and labeled. From these the centers of gravity were calculated and used as landmarks. Then landmark based rigid and Bookstein warping was applied. We will show results in Figure 8.2 in Section 8.

The second case is a little more difficult. In that case image data sets from Drosophila flies were to be registered for comparison and averaging. Several non-spherical shapes were labeled. Reducing these to their centers of gravity would have been a tremendous loss of information. Therefore, we have developed a method which merely considers the complete structures as landmarks. It works as follows.

First, a global rigid transformation G was computed by using a quality function similar to Equation (7.1). However as explained earlier it makes no sense to interpret the difference of labels numerically, even if these are represented as numbers internally. Only a binary decision *equal* or *not equal* makes sense. If L_A and L_B are the labelings for data set A and B respectively, Q computes as

$$Q(T) = \int \Xi(\vec{x}) d\vec{x}, \tag{7.9}$$

where

$$\Xi(\vec{x}) = \begin{cases} 1 & ; & A(\vec{x}) = B(T(\vec{x})) \\ 0 & ; & A(\vec{x}) \neq B(T(\vec{x})). \end{cases}$$

Then another rigid registration T_i was computed for each structure individually. For any point belonging to a labeled structure we can compute the difference vector between the two transformations:

$$\vec{d}(\vec{x}) = T_i(\vec{x}) - G(\vec{x})$$

assuming that $\vec{x}$ belongs to structure i. The goal is now to find an interpolant for the correction vector field $\vec{d}$ in regions which are not labeled. Note that this approach is feasible since the differences of the individual transformations are relatively small in our application. If the rotational component deviated more then a few degrees only, interpolation based on translation vectors does not yield good results and quaternion based interpolation or similar methods have to be used, compare [117, 118, 4]. Our approach can be generalized to use these interpolation spaces as well.

We perform the interpolation of the translation vectors component-wise by applying the static heat transfer equation:

$$\Delta T = 0. \tag{7.10}$$

This equation has to be solved in the domain of unlabeled voxels U. The temperature T corresponds to the component to be interpolated. Since the translation is given for the labeled pixels, T is given at the boundary of our domain. Where U is bounded by the image domain, the global transformation is used to define proper boundary conditions. The partial differential equation (7.10) has been solved using an iterative solver. The resulting continuous translation field has been used to generate the average intensity map for non-affine registration in Figure 8.2.

7.1.4 Results and Discussion

The field of registration and warping is very broad and we have described only the aspects directly relevant for this work.

We have used the described methods for different applications. For rigid and affine registration of confocal data sets from different individuals, we have found that automatic intensity correlation based algorithms using the Equations (7.2) and (7.9) are robust and efficient. If a labeling is given computation costs can be saved by deriving the scaling factor from the volume ratios of corresponding structures.

The computation of a rigid transformation from corresponding sets of landmarks via Equation (7.1.2) is a very fast tool, if suitable landmarks are given. In our applications we have used it for the drosophila brains and the bee's antennal lobes, to compute a quick previews, or a starting positions for intensity correlation based optimization procedures. Here we have used the centers of gravity of labeled structures as landmarks.

To generate average intensity maps with a resolution on sub-structure scale, we have used non rigid transformations, based on the new method that we have proposed in Section 7.1.3. Result will be shown in Chapter 8.

Non-affine registration transformations based on thin-plate splines have been used to register antennal lobes with each other Equation (7.8). Again, centers of gravity of identifiable glomeruli where used as landmarks. Between these, the image data was deformed as described above. The goal is, that by registration with a fully labeled template, a full and reliable identification of all glomeruli is possible, potentially automatically. Final results are not yet available, but the first experiments look promising.

We want to emphasize a few more general observations and properties related to registration:

A registration implicitly or explicitly assumes that the two objects registered correspond. Whether this is correct or not is not a mathematical but a biological question. It also depends on the scale at which we are looking.

A registration is a prerequisite for comparison. On the other hand it eliminates differences. It is obvious that a rigid registration has to be performed, if the orientations of the specimen under the microscope have been different. If two preparations underwent a different histological procedure resulting in different amounts of shrinking, one may allow an additional uniform scaling. Then however, one must be aware of the fact that any biological caused size difference will also be eliminated. In case of non-linear transformations that are for example used to eliminate mechanical deformations that occurred during preparation the situation gets even more complicated. The two conclusions from this one has to keep in mind are: The more degrees of freedom a transformation has, the more complicated it can be to interpret the results with regards to comparison. And a registration transformation can in general not distinguish between differences resulting from preparational or other artifacts and those stemming from anatomical variability. It will in general eliminate both to the same degree.

Finally, it is worth noticing that if the problem of finding a perfect registration was solved, the problem of automatic segmentation and labeling would have been solved also. One could register any data set with a template data set, that was manually labeled once. This way one would obtain the correct labeling information for the individual data set.