Quo Vadis, Atlas-Based Segmentation?

Torsten Rohlfing[1], Robert Brandt[2], Randolf Menzel[3],
Daniel B. Russakoff[4], and Calvin R. Maurer, Jr.[5]

[1]Neuroscience Program, SRI International, Menlo Park, CA, USA

[2]Mercury Computer Systems GmbH, Berlin, Germany

[3]Institut für Neurobiologie, Freie Universität Berlin, Berlin, Germany

[4]Department of Neurosurgery and Computer Science Department, Stanford University, Stanford, CA, USA

[5]Department of Neurosurgery, Stanford University, Stanford, CA, USA

Send correspondence to: Dr. Torsten Rohlfing, SRI International, Neuroscience Program, 333 Ravenswood Ave, Menlo Park, CA 94025-3493, USA, phone: (+1) 650 859 3379, fax: (+1) 650 859 2743, eMail: torsten@synapse.sri.com

Contents

Quo Vadis, Atlas-Based Segmentation?

Torsten Rohlfing [1]*, Robert Brandt* [2]*, Randolf Menzel* [3]*,*
Daniel B. Russakoff [4]*, Calvin R. Maurer, Jr.* [5]

11.1 Segmentation Concepts

There are many ways to segment an image, that is, to assign a semantic label to each of its pixels or voxels. Different segmentation techniques use different types of image information, prior knowledge about the problem at hand, and internal constraints of the segmented geometry. The most suitable method for a given case depends on the image data, the objects imaged, and the type of desired output information.

Purely intensity-based classification methods [31, 82, 87] work locally, typically one voxel at a time, by clustering the space of voxel values (i.e., image intensities). The clusters are often determined by an unsupervised learning method, for example k-means clustering, or derived from example segmentations [47]. Each cluster is identified with a label, and each voxel is assigned the label of the cluster corresponding to its value. This assignment is independent of the voxel's spatial location. Clustering methods obviously require that the label for each voxel is determined by its value. Extensions of clustering methods that avoid overlapping clusters work on vector-valued data, where each voxel carries a vector of intensity values. Such data is routinely generated by multi-spectral magnetic resonance (MR) imaging, or by combination of images of the same object acquired from different imaging modalities in general.

There are, however, many applications where there is no well-defined relationship between a voxel's value(s) and the label that should be assigned to it. This observation is fairly obvious when we are seeking to label *anatomical structures* rather than *tissue*

[1]Neuroscience Program, SRI International, Menlo Park, CA, USA

[2]Mercury Computer Systems GmbH, Berlin, Germany

[3]Institut für Neurobiologie, Freie Universität Berlin, Berlin, Germany

[4]Department of Neurosurgery and Computer Science Department, Stanford University, Stanford, CA, USA

[5]Department of Neurosurgery, Stanford University, Stanford, CA, USA

types. It is clear, for example, that different structures composed of the same tissue (e.g., different bones) cannot be distinguished from one another by looking at their intensity values in an image. What distinguishes these structures instead is their location and their spatial relationship to other structures. In such cases, spatial information (e.g., neighborhood relationships) needs to be taken into consideration and included in the segmentation process.

Level set methods [40, 71, 81, 92] simultaneously segment all voxels that belong to a given anatomical structure. Starting from a seed location, a discrete set of labeled voxels is evolved according to image information (e.g., image gradient) and internal constraints (e.g., smoothness of the resulting segmented surface). Snakes or active contours [91] use an analytical description of the segmented geometry rather than a discrete set of voxels. Again, the geometry evolves according to the image information and inherent constraints.

In addition to geometrical constraints, one can take into account neighborhood relationships between several different structures [80, 90]. A complete description of such relationships is an *atlas*. In general, an atlas incorporates the locations and shapes of anatomical structures, and the spatial relationships between them. An atlas can, for example, be generated by manually segmenting a selected image. It can also be obtained by integrating information from multiple segmented images, for example from different individuals. We shall discuss this situation in more detail in Section 11.4.3.

Given an atlas, an image can be segmented by mapping its coordinate space to that of the atlas in an anatomically correct way, a process commonly referred to as registration. Labeling an image by mapping it to an atlas is consequently known as atlas-based segmentation, or registration-based segmentation. The idea is that, given an accurate coordinate mapping from the image to the atlas, the label for each image voxel can be determined by looking up the structure at the corresponding location in the atlas under that mapping. Obviously, computing the coordinate mapping between the image and atlas is the critical step in any such method. This step will be discussed in some detail in Section 11.3.

A variety of atlas-based segmentation methods have been described in the literature [3, 11, 12, 15, 16, 22, 24, 26, 42, 45]. The characterizing difference between most of these methods is the registration algorithm that is used to map the image coordinates onto those of the atlas. One important property, however, is shared among all registration methods applied for segmentation: as there are typically substantial shape differences between different individuals, and therefore between an individual and

an atlas, the registration must yield a non-rigid transformation capable of describing those inter-subject deformations.

In this chapter, we in particular take a closer look at an often neglected aspect of atlas-based segmentation, the selection of the atlas. We give an overview of different strategies for atlas selection, and demonstrate the influence of the selection method on the accuracy of the final segmentation.

11.2   From Subject to Atlas: Image Acquisition and Processing

We illustrate the methods and principles discussed in this chapter by segmenting confocal microscopy images from 20 brains of adult, honeybee workers. The methods and principles apply equally well to commonly acquired clinical images (e.g., MR images of the human brain). Confocal laser scanning microscopy is a type of fluorescence microscopy, where a focused laser beam deflected by a set of xy-scanning mirrors excites the fluorescently stained specimen (i.e., the dissected brain). The emitted fluorescence is then recorded by inserting a so-called "confocal pinhole" into the microscope's optical path. This pinhole ensures that only light from the focal plane reaches the detector, thus enabling the formation of an image that can be considered an optical section through the specimen. By moving the position of the specimen along the optical axis of the microscope a three-dimensional (3-D) image is generated [8, 74, 94]

The staining of the bee brains depicted in this chapter followed an adapted immunohistological protocol. Dissected and fixated brains were incubated with two primary antibodies (nc46, SYNORF1) that detect synapse proteins [30, 50]. Because cell bodies in insects reside separately from fibers and tracts, this staining ensures response from those regions in the tissue that exhibit high synaptic densities, i.e., neuropil, while somata regions remain mostly unstained. A Cy3-conjugated secondary antibody sensitive to the constant part of the primary antibody was subsequently used to render labeled regions fluorescent. After dehydration and clearing, the specimen were mounted in double-sided custom slides.

The brains were imaged with a confocal laser scanning microscope (Leica TCS 4D). The chromophor was excited with an ArKr laser, and the fluorescence was detected using a longpass filter. The intensity of the fluorescence was quantized with a resolution of 8 bits. Due to the size of the dissected and embedded brain (about

Table 11.1: Anatomical structures of the bee brain with abbreviations used in this chapter.

Abbreviation	Structure	Abbreviation	Structure
PL-SOG	protocerebral lobes	r-medBR	right medial basal ring
CB	central body	l-latBR	left lateral basal ring
l-Med	left medulla	r-latBR	right lateral basal ring
r-Med	right medulla	l-medColl	left medial collar
l-Lob	left lobula	r-medColl	right medial collar
r-Lob	right lobula	l-latColl	left lateral collar
l-AL	left antennal lobe	r-latColl	right lateral collar
r-AL	right antennal lobe	l-medLip	left medial lip
l-vMB	left mushroom body	r-medLip	right medial lip
r-vMB	right mushroom body	l-latLip	left lateral lip
l-medBR	left medial basal ring	r-latLip	right lateral lip

$2.5 \times 1.6\,\mathrm{mm}$ laterally and about $0.8\,\mathrm{mm}$ axially), it cannot be imaged in a single scan. Therefore we used multiple image-stack acquisition (3D-MISA) [94]. The entire brain was scanned in 2×3 partially overlapping single scans, each using 512×512 pixels laterally and between 80 and 120 sections axially. The stacks were combined into a single 3-D image using custom software or a script running in Amira (see next paragraph). Because of the refractive index mismatch between the media in the optical path, images exhibit a shortening of distances in axial direction that was accounted for by a linear scaling factor of 1.6 [7].

Post-acquisition image processing was done with the Amira 3-D scientific visualization and data analysis package (ZIB, Berlin, Germany; Indeed – Visual Concepts GmbH, Berlin, Germany; TGS Inc., San Diego, CA). Image stacks were resampled laterally to half of the original dimensions in order to increase display speeds and allow interactive handling of the data. The final image volume contained 84–114 slices (sections) with thickness $8\,\mu\mathrm{m}$. Each slice had 610–749 pixels in x direction and 379–496 pixels in y direction with pixel size $3.8\,\mu\mathrm{m}$. In most cases no further image processing was necessary. In a few cases unsharp masking filters were applied in order to enhance contours.

Subsequently, for each brain an atlas of the neuropil areas of interest was generated by tracing them manually on each slice. We distinguished 22 major compartments, 20 of which are bilaterally symmetric on either brain hemisphere [43]. The paired structures we labeled were medulla; lobula; antennal lobe; ventral mushroom body consisting of peduncle, α- and β-lobe; and medial and lateral lip, collar and basal ring neuropil. The unpaired structures we identified were the central body with its upper
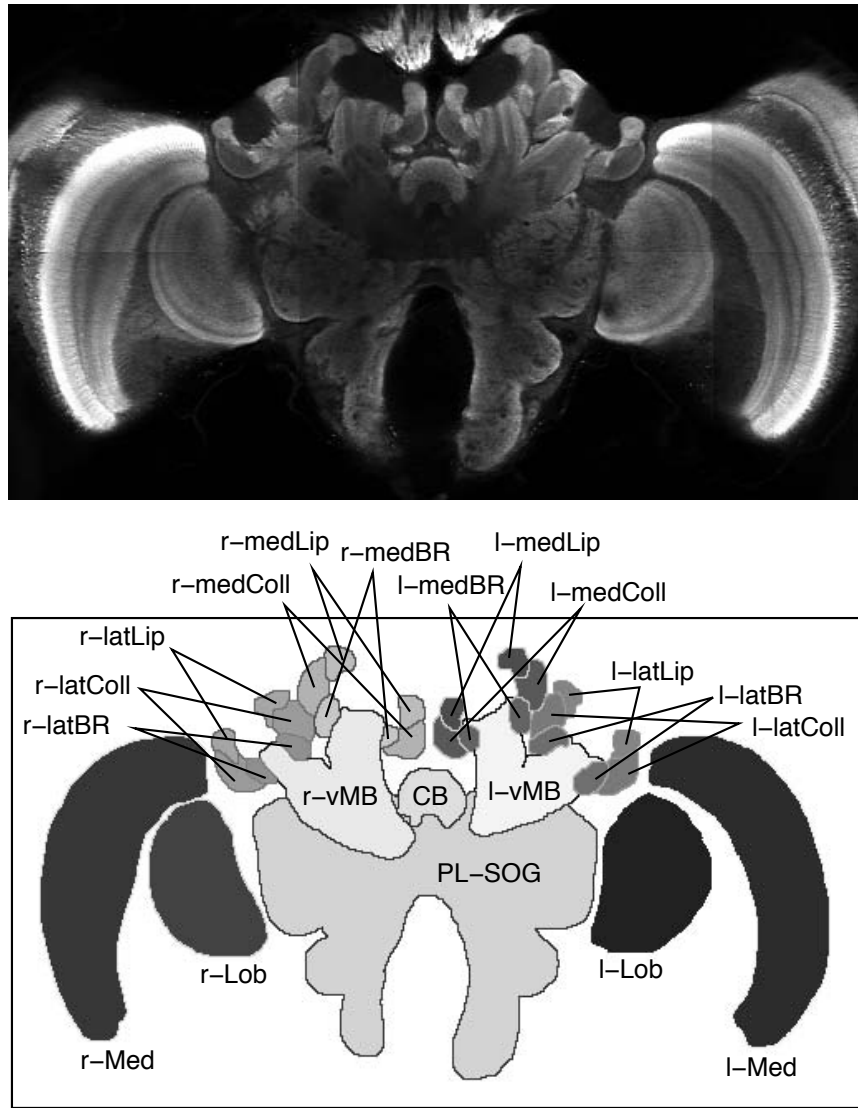
Figure 11.1: Example of bee brain confocal microscopy (*top*) and corresponding label image as defined by manual segmentation (*bottom*). Following radiological convention for axial slices, the image is seen from the cranial direction. Every gray level in the label image represents a different anatomical structure. Due to limitations of reproduction different gray levels may look alike. The correspondence between anatomical structures and abbreviations is listed in Table 11.1. Note that two structures, the left and right antennal lobes (l-AL and r-AL), are not visible in this slice, but can be seen in Fig. 11.2.

Figure 11.2: Three-dimensional rendering of a segmented bee brain. *From top to bottom*: View from frontal, top, and back, respectively. Note that the two symmetrical blue structures in the lower part of the brain in the frontal view (*top*) are the left and right antennal lobes (l-AL and r-AL) that were not depicted in Fig. 11.1.
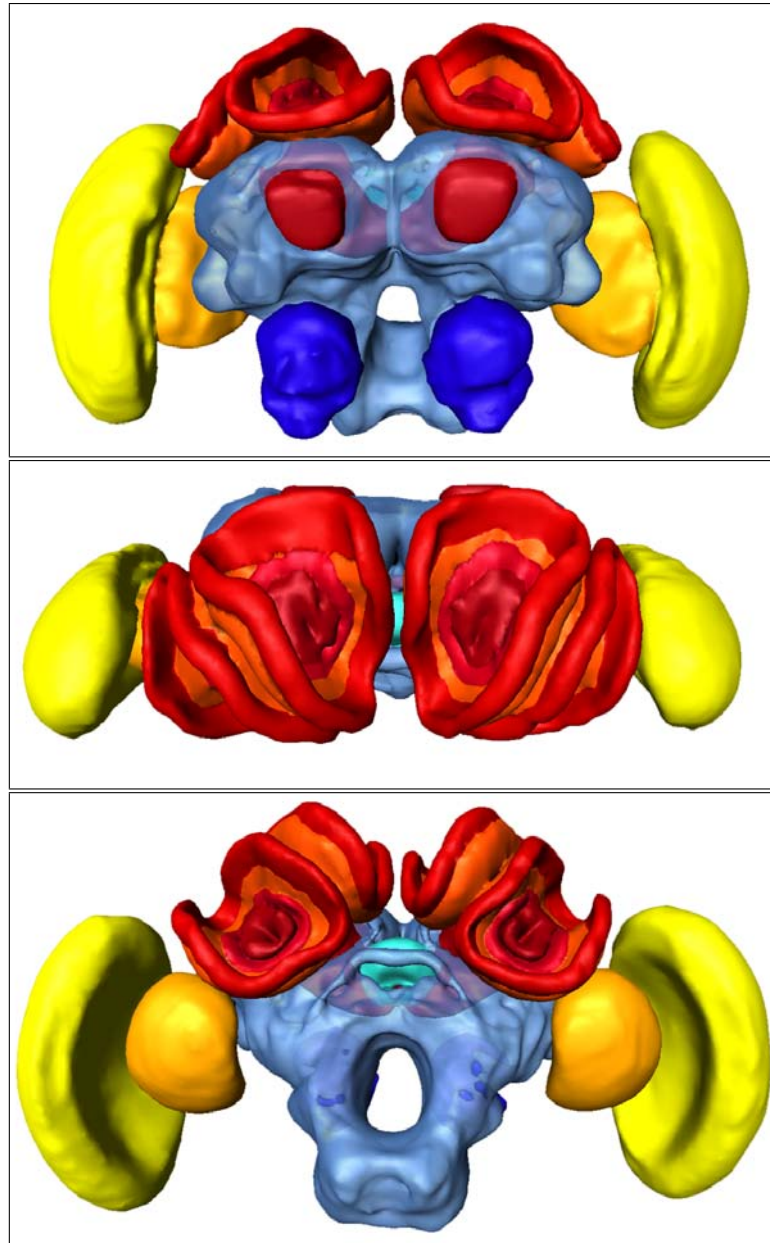
and lower division and the protocerebral lobes including the subesophageal ganglion. Examples of confocal microscopy and label images are shown in Fig. 11.1. Three-dimensional surface renderings of the segmented bee brain are shown in Fig. 11.2. The labeled structures and the abbreviations used for them in this chapter are listed in Table 11.1.

11.3   Fundamentals of Atlas-Based Segmentation

Mathematically speaking, an atlas A is a mapping $A : \mathbb{R}^n \to \Lambda$ from n-dimensional spatial coordinates to labels from a set of classes Λ. It is conceptually very similar to an image in the same coordinate space, which is a mapping from $\mathbb{R}^n$ to the space of gray values, a subset of $\mathbb{R}$. An atlas can therefore itself be considered as a special type of image, that is, a label image. In order to segment a new image R using an atlas A, we need to compute a coordinate mapping between them, that is, we need to register one image to the other. The coordinate mapping must be anatomically correct for the segmentation to be accurate.

An atlas is often generated by (manually) segmenting an actual image, say F. Therefore, we typically have access not only to a spatial map of labels, the actual atlas, but also to a corresponding realization using at least one particular imaging modality. In case multiple co-registered images from different modalities form the basis of the atlas, there may even be multiple instances of actual images. An example of an atlas and a corresponding microscopy image is shown in Fig. 11.1. This dual character is relevant insofar as, while fundamentally possible, registration of an image to the label representation of an atlas is a much harder problem than registration to the corresponding original image.

Let us consider two 3-D scalar images, $R : \mathbb{R}^3 \mapsto \mathbb{R}$ and $F : \mathbb{R}^3 \mapsto \mathbb{R}$. We assume that each point in one image has a corresponding equivalent in the other. For any two images, this correspondence is mathematically represented as a coordinate transformation $\mathbf{T}$ that maps the image coordinates of R onto those of F. For a given location $\mathbf{x}$ in the domain of R, we find the corresponding location in the domain of F as $\mathbf{T}(\mathbf{x})$. If F is associated with an atlas A, then we can find the correct label for any location $\mathbf{x}$ in R through the mapping

$$\mathbf{x} \mapsto A(\mathbf{T}(\mathbf{x})). \tag{11.1}$$

The transformation $\mathbf{T}$ is parameterized by a p-dimensional parameter vector $\mathbf{p} \in \mathbb{R}^p$. The process of finding the vector $\mathbf{p}$ that describes the "correct" transformation is

known as image registration. One of the images, R, remains fixed during registration, while the other, F, is transformed in space. The fixed image R is commonly referred to as the "reference image", the transformed image F is called the "floating image".

The terminology used in the remainder of this chapter is as follows. We refer to the already segmented image as the *atlas image* and the image to be segmented as the *raw image*. The coordinates of the raw image are mapped by registration onto those of the atlas image and thereby provide a segmentation of the former. In the context of non-rigid registration, the atlas image is to be deformed while the raw image remains fixed. The correspondence between the common terms for both images in image registration and in the present context is such that the atlas image acts as the *floating image* during registration while the raw image acts as the *reference* (or *target*) *image*.

11.3.1 Entropy-based Image Similarity

It is not usually known *a priori*, what the correct mapping between the two images R and F is. Instead, the correctness of any given transformation is usually quantified by a so-called similarity measure. This measure is a scalar function $S : \mathbb{R}^p \mapsto \mathbb{R}$ designed so that higher values of S correspond to better matches. That is, if for two parameter vectors, $\mathbf{p}_1$ and $\mathbf{p}_2$, we have $S(\mathbf{p}_1) > S(\mathbf{p}_2)$, then the mapping $\mathbf{T}_1$ parameterized by $\mathbf{p}_1$ is assumed to be "more correct" than the mapping $\mathbf{T}_2$ described by $\mathbf{p}_2$. Again, since the correct mapping is not known, S can only be a more or less suitable approximation to the true correctness. The registration is performed by finding the parameter vector $\mathbf{p}$ that maximizes S.

A similarity measure that has been empirically found to be particularly well-suited for many registration applications is mutual information (MI) [39, 83, 86]. It is based on the information-theoretic entropy concept and is defined as

$$S_{\mathrm{MI}} = H_R + H_F - H_{RF},\tag{11.2}$$

where H_R is the entropy of image R, H_F is the entropy of image F, and H_{RF} is the joint entropy of corresponding voxel pairs between the two images. A modification proposed by Studholme *et al.* [77], normalized mutual information (NMI), has been found to be slightly more robust

$$S_{\mathrm{NMI}} = \frac{H_R + H_F}{H_{RF}}.\tag{11.3}$$

There are many different implementations of both MI and NMI, using different numerical methods to estimate the image entropies. While some use continuous methods

such as Parzen windowing [83, 86], others estimate the entropies from discrete two-dimensional histograms [39, 76]. The latter techniques are more easily implemented and more common.

11.3.2 Rigid Registration

The first iteration of inter-subject registration, such as registration of an image to an atlas, usually aims at correcting for gross differences in position, orientation, and size between the individual images. Consequently, we initially apply a 9 degree-of-freedom (DOF) affine registration algorithm that performs appropriate translation, rotation, and scaling. A technique described by Studholme *et al.* [76] has been found to produce highly accurate (rigid) transformation in an independent, blinded evaluation study [88]. The algorithm optimizes the NMI similarity measure described above using a simple but robust multiresolution optimization strategy with hierarchically re-sampled images. For details about our particular implementation of this algorithm, the interested reader is referred to Refs. [52, 53, 54, 66].

11.3.3 Non-Rigid Registration

There is typically considerable inter-individual variability in the shapes of anatomical structures in the brains of humans and animals. Figure 11.3 illustrates this for the microscopy images of bee brains that we are using to demonstrate the methods in this chapter. For MR images of human brains, Fig. 11.4 provides an analogous illustration. Therefore, in order to be effective, any registration-based segmentation method requires a registration algorithm that can compensate not only for different pose and size, but also for inter-individual shape differences between raw image and atlas (i.e., a non-rigid registration algorithm).

Many different non-rigid registration methods have been published [25, 34, 67]. Some of these, such as methods based on optical flow [78] and most methods using elastic [42, 9] or fluid models [10, 35], typically require both images to be from the same imaging modality to be able to identify corresponding features. Note that the motion model, i.e., fluid or elastic, does not require single modality images. However, most algorithms based in which fluid or elastic differential equations govern the transformation combine these with image similarity terms that are equivalent to the mean squared difference of image intensities.

Unfortunately, the particular nature of the microscopy images in our example ap-

Reference Brain Image

Floating Image #1	Floating Image #2	Floating Image #3

After Affine Registration

After Non-rigid Registration

Deformed Coordinate Grid

Deformation Vector Field
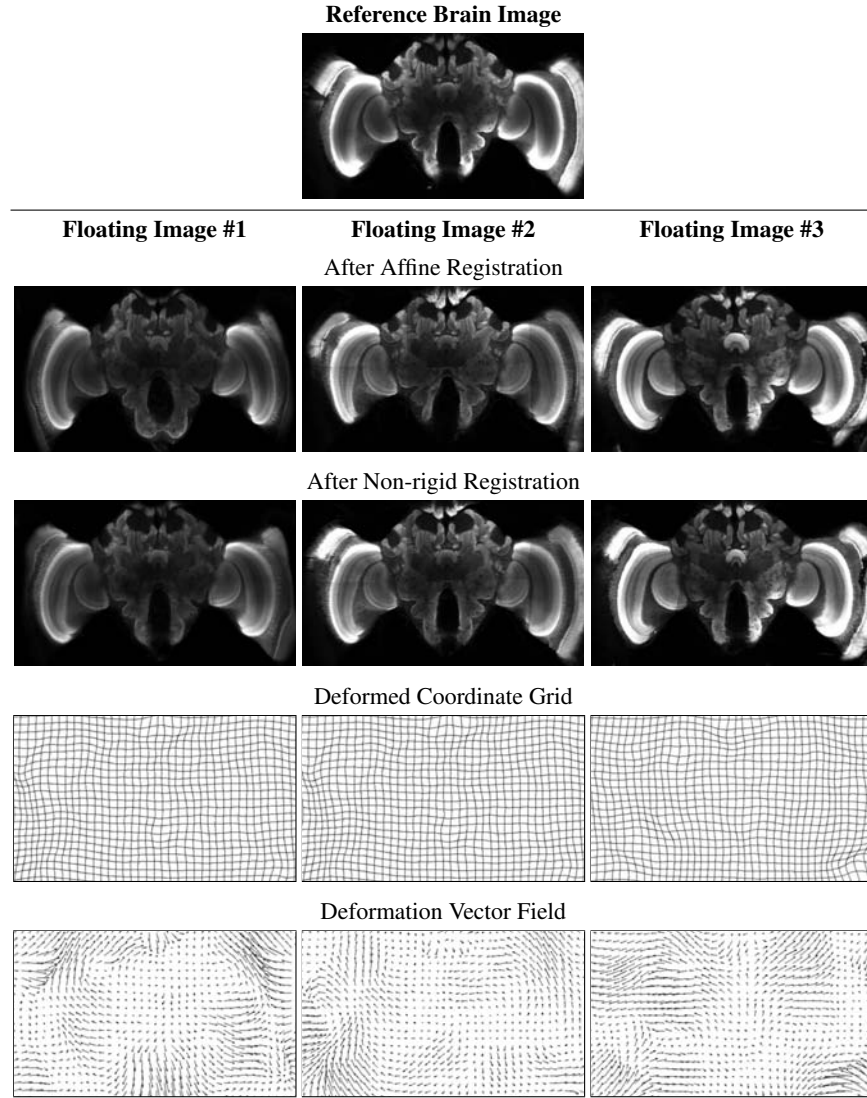
Figure 11.3: Illustration of inter-subject differences between several individual bee brains. *Top:* Central axial slice from a 3-D microscopy image used as the reference image for this example. *Second row:* Corresponding slice from three other bee brains after affine registration. *Third row:* Corresponding slices after non-rigid registration. *Fourth row:* Deformed coordinate grids. *Fifth row:* Deformation vector fields. Note that only the 2-D projection of the 3-D deformed coordinate grid and vector field are shown.

Reference Brain Image

| Image #1 | Image #2 | Image #3 | Image #4 | Image #5 |

After Affine Registration

After Non-rigid Registration

Deformed Coordinate Grid

Deformation Vector Field
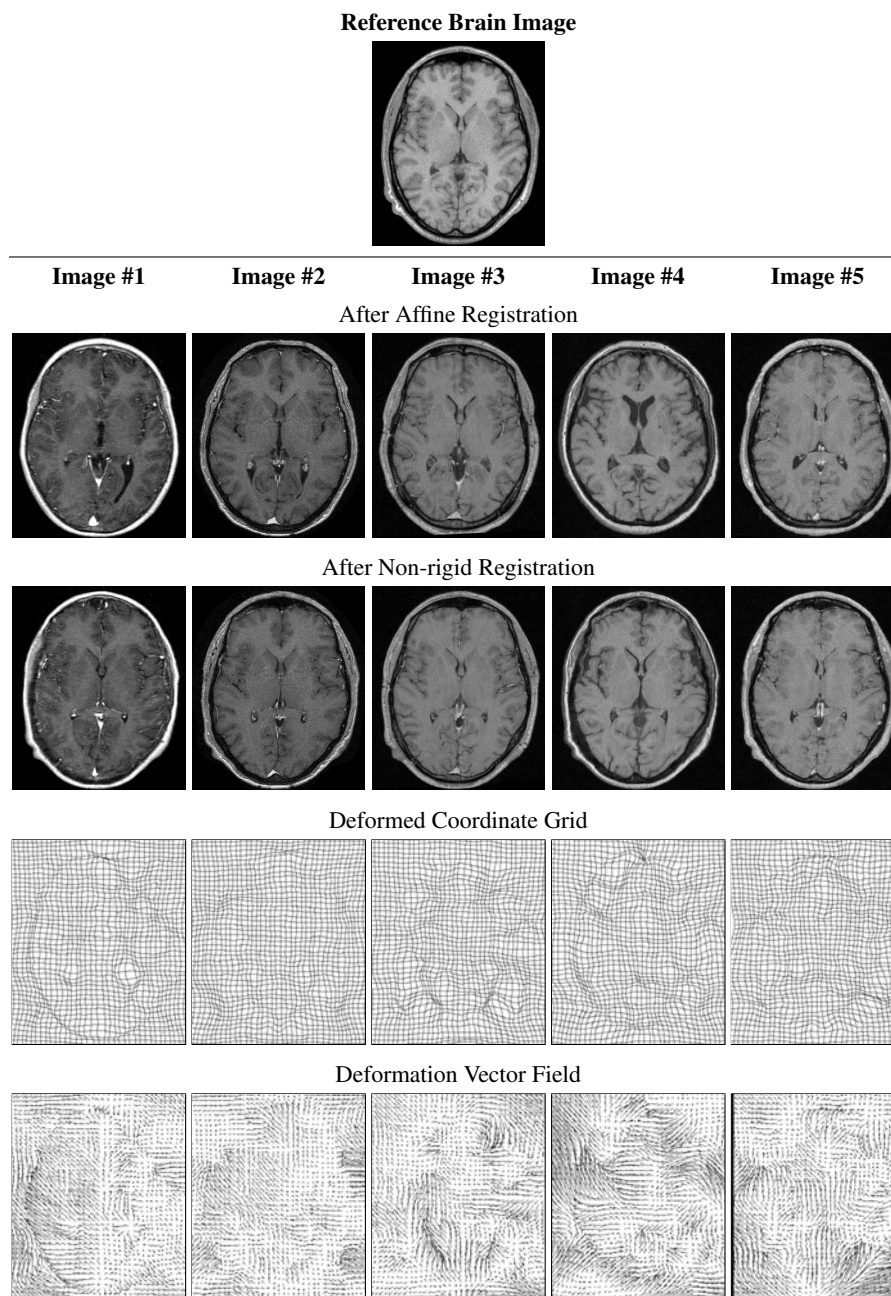

Figure 11.4: Illustration of inter-subject differences between several individual human brains. The organization of this figure is identical to Fig .11.3.

plication prohibits the use of any such method. While strictly these images are all generated by the same imaging process, they are subject to imaging artifacts that vary from acquisition to acquisition. Sources of these artifacts include individual concentration differences of the chromophor, fluctuation of laser intensity, and increasing location-dependent absorption with increasing tissue depth. We review below a registration algorithm that has been successfully applied to microscopy images of bee brains, as well as many other applications [17, 41, 55, 57, 56, 59, 62, 69, 72].

A non-rigid registration algorithm that inherently supports images originating from multiple imaging modalities was described by Rueckert *et al.* [69]. It uses the same NMI similarity measure as the affine algorithm mentioned above. The transformation model is a free-form deformation [73] that is defined on a data-independent, uniformly spaced control point grid (CPG) Φ covering the reference image. The CPG consists of discrete control points $\phi_{i,j,k}$, where $-1 \leq i < n_x - 1$, $-1 \leq j < n_y - 1$, and $-1 \leq k < n_z - 1$. Points with i, j, or k equal to either 0 or $n_x - 3$ ($n_y - 3$ and $n_z - 3$ for j and k) are located on the edge of the image data. The spacings between the control points in x, y, and z are denoted by δ_x, δ_y, and δ_z, respectively. For any location (x, y, z) in the domain of Φ, the transformation $\mathbf{T}$ is computed from the positions of the surrounding $4 \times 4 \times 4$ control points:

$$\mathbf{T}(x, y, z) = \sum_{l=0}^{3} \sum_{m=0}^{3} \sum_{n=0}^{3} B_l(u) B_m(v) B_n(w) \phi_{i+l, j+m, k+n}.$$

Here, i, j, and k denote the index of the control point cell containing (x, y, z), and u, v, and w are the relative positions of (x, y, z) inside that cell in the three spatial dimensions:

$$i = \left\lfloor \frac{x}{\delta_x} \right\rfloor - 1, \; j = \left\lfloor \frac{y}{\delta_y} \right\rfloor - 1, \; k = \left\lfloor \frac{z}{\delta_z} \right\rfloor - 1,$$

and

$$u = \frac{x}{\delta_x} - \left\lfloor \frac{x}{\delta_x} \right\rfloor, v = \frac{y}{\delta_y} - \left\lfloor \frac{y}{\delta_y} \right\rfloor, w = \frac{z}{\delta_z} - \left\lfloor \frac{z}{\delta_z} \right\rfloor.$$

The functions B_0 through B_3 are the approximating third-order spline polynomials [33]:

$$
\begin{aligned}
B_0(t) &= \left(-t^3 + 3t^2 - 3t + 1 \right)/6, \\
B_1(t) &= \left(3t^3 - 6t^2 + 4 \right)/6, \\
B_2(t) &= \left(-3t^3 + 3t^2 + 3t + 1 \right)/6, \\
B_3(t) &= t^3/6.
\end{aligned}
$$

The degrees of freedom of a B-spline based transformation $\mathbf{T}$, and thus the elements of the parameter vector $\mathbf{p}$, are the coordinates of the control points $\phi_{i,j,k}$.

The optimum parameters of the non-rigid registration transformation $\mathbf{T}$ are determined by a line search algorithm similar to the steepest descent method [48]. The target function of the optimization is the NMI similarity of the reference and the transformed floating image. We start by computing a discrete approximation of the gradient of the target function with respect to the parameters of the transformation $\mathbf{T}$. This is achieved by a simple finite difference scheme. Despite the high-dimensional parameter space, gradient approximation can be performed very efficiently; due to the compact support of the B-spline functions, each parameter of $\mathbf{T}$ influences only a small volume in image space (i.e., the local $4 \times 4 \times 4$ control point neighborhood). When moving any single control point, all voxels of the floating image outside this area remain in the same location. Their contribution to the similarity measure can therefore be precomputed and reused [75].

In order to capture large deformations as well as small ones, the algorithm incorporates a multiresolution deformation strategy based on multilevel B-splines [33]. After finishing optimization at one control point resolution, the spacing between the control points is reduced by a factor of 2 before registration continues. The positions of the control points in the refined grid are determined in a way that exactly preserves the current deformation [19, 62]

Using adaptive grid refinement [59, 72] and a parallel multiprocessor implementation [60], we are able to keep computation times within reasonable bounds. For example, we can complete a non-rigid registration of an image to an atlas, each about the size as described earlier in this chapter, within about ten minutes on a modern PC (Intel Pentium 4, 3.0 GHz, hyperthreading enabled).

11.3.4   Regularization of the Non-Rigid Transformation

Confocal microscopy imaging is a substantially less controlled image formation process than typical medical imaging modalities. Varying concentrations of the chromophor within one structure, laser power fluctuations, tiling artifacts, and absorption of emitted light from deep structures lead to substantial imaging artifacts. As illustrated in Fig. 11.5, these artifacts can cause severe problems for the non-rigid registration, leading to grossly incorrect coordinate transformations. These can to some extent be prevented by regularizing the image similarity cost function with an additional constraint term that controls the geometric properties of the coordinate mapping. The

(a) Reference image   (b) Floating image after rigid registration   (c) Floating image after unconstrained non-rigid registration   (d) Floating image after constrained non-rigid registration
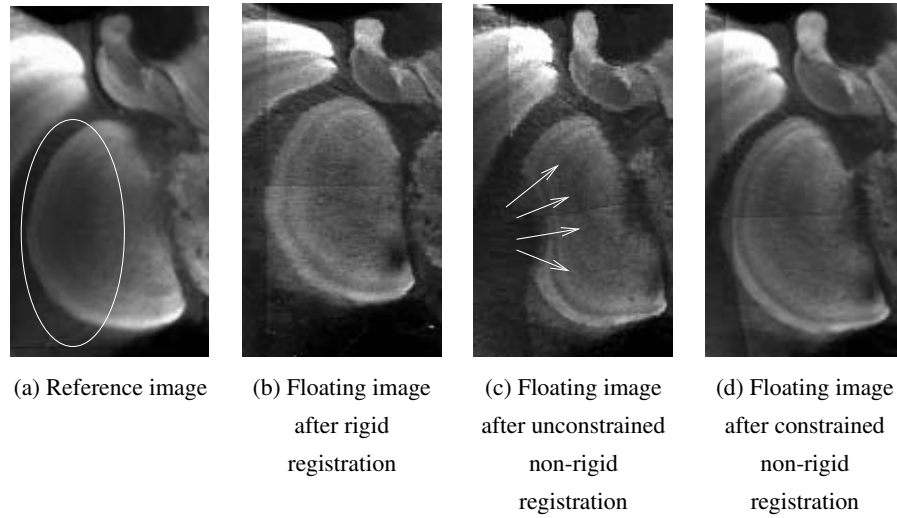
Figure 11.5: Illustration of the importance of constraining non-rigid registration. These microscopy images are magnified to focus on the area of the right lobula (compare Fig. 11.1 for an anatomical overview). In the reference image (*a*), the lobula appears substantially darker on the lateral side (*ellipse*). In the rigidly registered floating image (*b*) from another individual the lobula has a more homogeneous intensity. Without smoothness constraint, intensity-based non-rigid registration (*c*) computes a grossly incorrect deformation (*arrows*). A constrained non-rigid registration (*d*) does not have this problem.

total optimization function thus becomes a weighted sum of the data-dependent image similarity and the regularization constraint term:

$$E_{\text{total}} = (1 - w)E_{\text{NMI}} + wE_{\text{constraint}}. \tag{11.4}$$

In detail, we constrain the deformation to be smooth by adding a biharmonic penalty term, which is based on the energy of a thin plate of metal that is subjected to bending deformations [4, 84]. The penalty term is composed of second-order derivatives of the deformation, integrated over the domain D of the transformation $\mathbf{T}$ as follows:

$$E_{\text{constraint}} = \int_D \left(\frac{\partial^2 \mathbf{T}}{\partial x^2}\right)^2 + \left(\frac{\partial^2 \mathbf{T}}{\partial y^2}\right)^2 + \left(\frac{\partial^2 \mathbf{T}}{\partial z^2}\right)^2 +$$
$$2\left[\left(\frac{\partial^2 \mathbf{T}}{\partial x \partial y}\right)^2 + \left(\frac{\partial^2 \mathbf{T}}{\partial y \partial z}\right)^2 + \left(\frac{\partial^2 \mathbf{T}}{\partial z \partial x}\right)^2\right] \, d\mathbf{x}. \tag{11.5}$$

Since the 3-D spline is the tensor product of independent 1-D polynomial functions, its second-order derivative with respect to one variable, x, is easily computed as follows:

$$\frac{\partial^2}{\partial x^2}\mathbf{T}(x, y, z) = \frac{1}{\delta_x^2}\sum_{l=0}^{3}\sum_{m=0}^{3}\sum_{n=0}^{3}\left(\frac{d^2}{du^2}B_l(u)\right)B_m(v)B_n(w)\,\phi_{i+l,j+m,k+n}. \tag{11.6}$$

Computation of the derivatives of $\mathbf{T}$ is in fact very similar to computing $\mathbf{T}$ itself. Depending on the derivation variable, the spline polynomials B_0 through B_3 in the respective dimension are simply replaced by their respective derivatives. These derivatives are easily computed analytically. Mixed second-order derivatives with respect to two different variables are computed by substituting two spline polynomials with their respective first-order derivatives, e.g.,

$$\frac{\partial^2}{\partial x \partial y}\mathbf{T}(x, y, z) =$$
$$\frac{1}{\delta_x \delta_y}\sum_{l=0}^{3}\sum_{m=0}^{3}\sum_{n=0}^{3}\left(\frac{d}{du}B_l(u)\right)\left(\frac{d}{dv}B_m(v)\right)B_n(w)\,\phi_{i+l,j+m,k+n}. \tag{11.7}$$

Using the above derivative terms, the continuous integral in Eq. (11.5) is approximated as a discretely sampled sum over a set of points, for example the $N_D = n_x \times n_y \times n_z$ voxels in the reference image.

11.4 Atlas Selection Strategies

This section will take a closer look at possible choices for atlases in atlas-based segmentation. Usually, this aspect of atlas-based segmentation receives little attention.

Selection Strategy	No. of Atlases per Raw Image	Type of Atlas	Assignment of Atlas to Raw Image
IND	single	individual	fixed
SIM	single	individual	variable
AVG	single	average	fixed
MUL	multiple	individual	fixed

Table 11.2: Categorization of atlas selection strategies by number, type, and assignment of atlases. See Sections 11.4.1 through 11.4.4 for details and Fig. 11.6 for a schematic overview of the different methods.

Yet, the decision what atlas to use has a substantial impact on the segmentation accuracy, and simple methods are not always the best as we will see below.

We describe and compare here four different atlas-based segmentation strategies with different atlas selections: segmentation with one single individual atlas, segmentation with varying single individual atlases, segmentation with an average shape atlas, and simultaneous segmentation with multiple atlases. These four strategies can be categorized according to the number of atlases used per raw image (one or multiple), the type of atlas used (individual or average), and the assignment of atlases to raw images (fixed, i.e., same atlas(es) for all raw images, or variable, i.e., different atlas image selected for each raw image). A schematic graphical comparison of the four methods is given in Fig. 11.6, and a textual summary of the categorization can be found in Table 11.2.

For each strategy, the resulting segmentation accuracy was evaluated. Automatic segmentations were compared to a manual gold standard segmentation. A detailed description of the methods used for validation and accuracy evaluation, together with the results for the four atlas selection strategies, are presented in Section 11.5.

11.4.1   Segmentation with a Fixed, Single Individual Atlas

The most straight forward strategy for selection of an atlas is to use one individual segmented image. The selection can be random, or based on heuristic criteria such as image quality, lack of artifacts, or normality of the imaged subject. This strategy is by far the most commonly used method for creating and using an atlas [27]. It requires only one atlas, which greatly reduces the preparation effort as compared to the more complex methods described below.

IND: Segmentation using a single individual atlas.

SIM: Segmentation using the "most similar" individual atlas.

AVG: Segmentation using an average shape atlas.

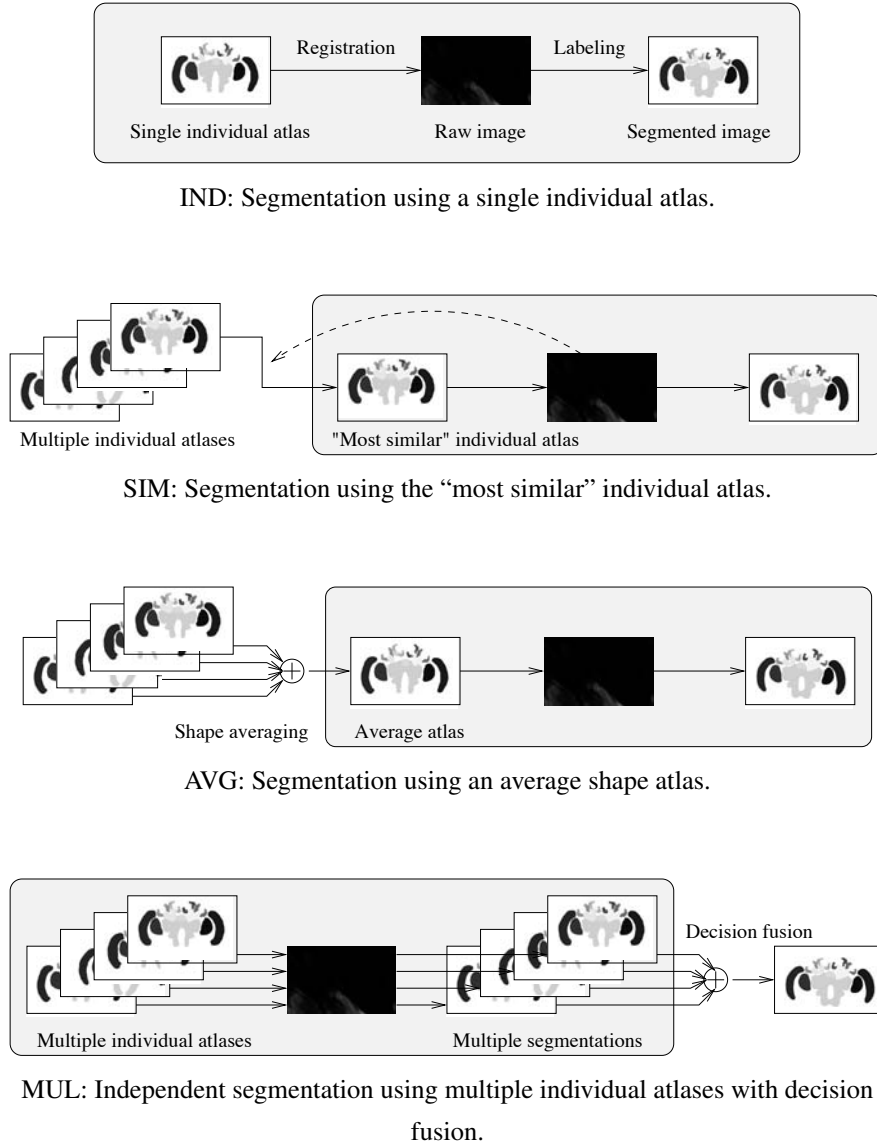MUL: Independent segmentation using multiple individual atlases with decision fusion.

Figure 11.6: Depiction of the atlas selection strategies discussed in this chapter. The strategies are also categorized in Table 11.2. Note that the basic atlas-based segmentation with a single atlas (IND, *gray box*) occurs in different stages in the other three strategies, in the MUL case replicated for each atlas.

Out of the 20 bee brains in our population, we picked the one that was judged to have the best image quality and least artifacts. We then used this atlas brain to segment the remaining 19 brain images. Each of the 19 raw images was registered non-rigidly to the microscopy image of the atlas, and labeled using the accordingly transformed atlas label image.

11.4.2   Segmentation with the Best Atlas for an Image

Suppose that instead of a single atlas, we have several atlases that originate from several different subjects. For each image that we are segmenting, there is one atlas that will produce the best segmentation accuracy among all available atlases. It is obviously desirable to use this optimum atlas, which is most likely a different atlas for each image.

The problem is that we do not know what the correct segmentation for an unsegmented image is. Therefore, we can only hope to find a more or less successful heuristic for selecting the best atlas for a given image. There are at least two easily accessible characteristic numbers that describe the similarity between an image and an atlas. One is the final value of the registration criterion, or image similarity measure, after either affine or non-rigid registration. The other is the magnitude of the deformation (i.e., non-rigid transformation) that is required to map the coordinates of the image onto that of the atlas.

Based on these two concepts, we have compared four different criteria for selecting the single atlas that is most likely to produce the best segmentation of a given raw image. These criteria are:

- **NMI affine: Image similarity after affine registration.**   The atlas image with the highest NMI similarity to the raw image after *affine* registration is selected and used for its segmentation. This criterion requires only an affine registration to be computed between the raw image and each of the atlases. It is therefore considerably less computationally expensive than the remaining three criteria described below.

- **NMI non-rigid: Image similarity after non-rigid registration.**   The atlas with the highest NMI value after *non-rigid* registration is selected and used for segmentation.

- **DEF avg: Average deformation of the atlas over all voxels.**   After non-rigid registration, the magnitude of the deformation between the raw image and each

individual atlas is computed and averaged over all voxels. The atlas with the smallest average deformation is selected and used for segmentation. Whereas the above criteria are based on intensity similarity, this criterion is based on geometric (i.e., shape) similarity.

- **DEF max: Maximum deformation of the atlas over all voxels.** This criterion is identical to the previous one, except that it uses the maximum deformation over all voxels rather than the average. This criterion pays more attention to outliers. The idea is that atlases that match well overall may be substantially inaccurate in some regions.

Segmentations were generated for each of the 20 bee brains, with the remaining 19 brains as possible atlas candidates in each case. For each raw image, one of the atlas candidates was chosen using each of the criteria above. The accuracy of a segmentation was computed as the SI between the segmentation and the manual gold standard.

Figure 11.7 shows a graph of the percentages of structures segmented with varying levels of accuracy. For comparison, this graph includes results achieved when using the best atlas according to the *a posteriori* SI values for each raw image (leftmost column, *Best SI*). In other words, this column shows the best possible result that can be achieved using only a single individual atlas, where the selection of this atlas is governed by the knowledge of the resulting segmentation accuracy (SI value). Obviously, this is not a strategy that is available in practice. However, it provides the upper bound for the segmentation accuracy that can be achieved using any criterion for selection of the best atlas for a given raw image.

Among the four criteria that do not depend on the *a posteriori* accuracy evaluation and thus a gold standard, the NMI image similarity after non-rigid registration performed slightly better than the other three. It was therefore selected as the criterion used for the SIM atlas selection strategy in the comparison to the other three strategies later in this chapter (Section 11.5.4).

11.4.3   Segmentation with an Average Shape Atlas

As we stated in the previous chapter, atlas-based segmentation is an easier task if the atlas is similar to the image that is to be segmented. Smaller magnitudes of the deformation between image and atlas that the non-rigid registration algorithm has to determine typically result in a higher accuracy of the matching. If the atlas is itself derived from an individual subject, then the risk is high that this individual is an outlier
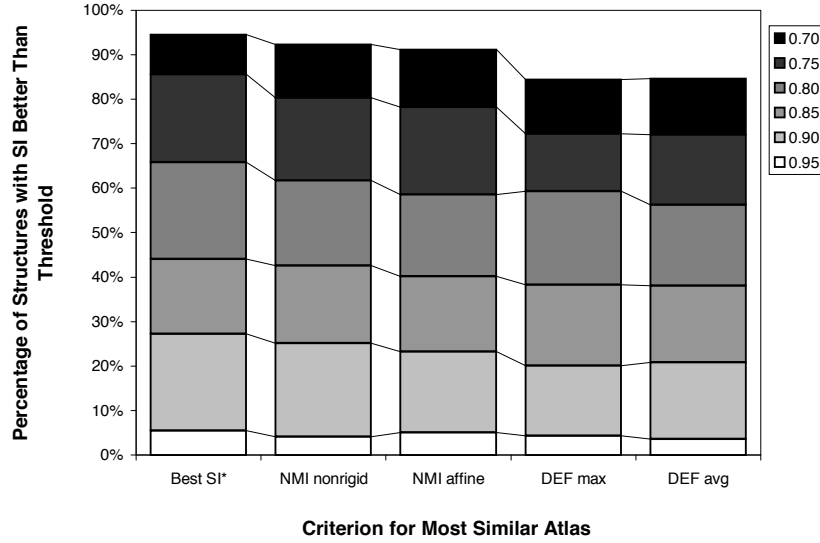
Figure 11.7: Percentages of structures segmented with accuracy better than given SI thresholds using different "most similar" atlas selection criteria. Each column represents one criterion (see text for details). The stacked bars from bottom to top show the percentages of structures that were segmented with SI better than 0.95 through 0.70. For comparison, the leftmost column shows the results when the atlas with the best *a posteriori* SI segmentation result is used for each raw image. This is the upper bound for the accuracy achievable with any criterion for selection of the best single individual atlas.

in the population. In such a case segmenting other subjects using the atlas becomes a more difficult problem. A better atlas would be one that is as similar to as many individuals as possible. Such an atlas can be generated by creating an average over many individuals.

For the human brain, such an average atlas is available from the Montreal Neurological Institute (MNI) as the BrainWeb phantom [5, 13]. Note, however, that the BrainWeb phantom is an atlas of brain *tissue types*, so as we discussed in Section 11.1, it is not as useful for atlas-based segmentation as an atlas of brain structures. For the human heart, an average atlas derived from cardiac MR images [49] has been used for atlas-based segmentation [38]. Similarly, an average atlas of the lung has been derived from CT images [36].

For demonstration in this chapter, we have therefore generated an average shape atlas of the structures of the bee brain using a technique outlined below [55].

11.4.3.1   Iterative Shape Averaging

One way of obtaining an average shape atlas from a population of subjects is to gener-
ate an active shape model (ASM). In short, an ASM provides a statistical description
of a population of subjects by means of an average shape and the principal modes
of variation [14, 37]. Generating an ASM typically requires the identification of cor-
responding landmarks on all individuals, a tedious and error-prone process despite
recent success in automating this step using non-rigid registration [21]. Other meth-
ods are based entirely on non-rigid registration, such as active deformation models
(ADM) [20, 68] and a method described by Guimond *et al.* [23]. Most of these, how-
ever, require not only non-rigidly registering several individuals, but also inverting the
transformations between them.

In general, non-rigid transformations are not easily inverted. Even for bijective
mappings there is typically no closed-form inverse. An iterative method for generating
average shape images that does not require inverse computations was first suggested
by Ashburner [2]. It also does not require the explicit computation of an average
transformation, the definition of which is not trivial for non-linear mappings. Instead,
the average deformation is generated by the iterative process itself. This technique
was later extended to segmented atlas images and applied to generate an average shape
atlas of the bee brain [55].

The central idea is to first map all original individual images onto to a common
reference and then generate an average image. After that, the original images are
mapped onto the average, and a new average image is generated. This process pro-
duces a sequence of average images that converges to an average shape image. Note
that convergence and average shape in this context are not defined in a strict math-
ematical sense. However, the result of this iteration is sufficient for the purpose of
obtaining an average atlas for atlas-based segmentation, as we will demonstrate be-
low.

In the first step of the iteration, there is not yet an average image to register the
original images to. One of the latter is therefore chosen as the reference for the initial
registration. In order to avoid bias of the averaging iteration by the choice of the
initial reference, the first registration is affine only, thus correcting for pose and size
differences but leaving object shape unchanged. For the subsequent steps, the average
image resulting from the previous registration step is chosen as the reference image of
non-rigid registrations, while the individual images are used as floating images, one
at a time. As a result, all floating images are independently mapped into the same
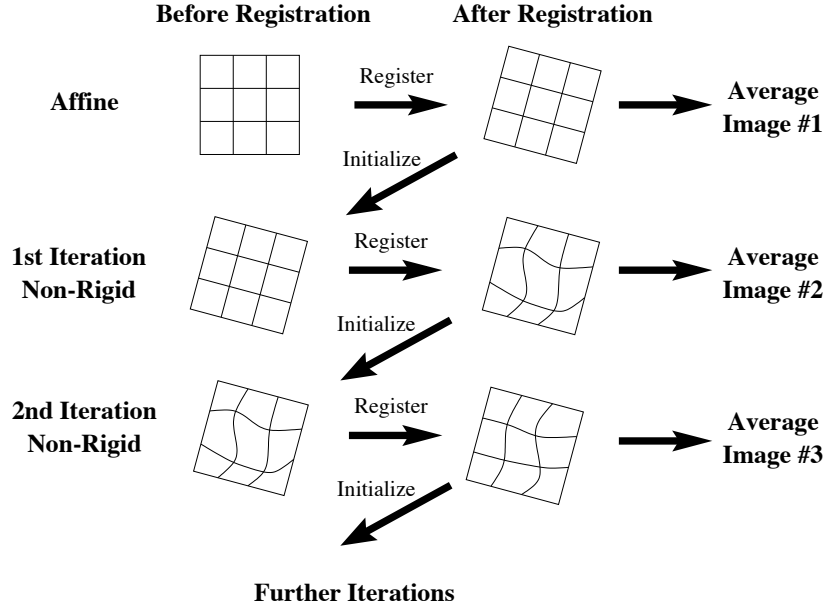
Figure 11.8: Propagation of transformations through the iterative shape averaging algorithm. For each individual image, the transformation (affine or non-rigid) used to generate the current average image is propagated to the next iteration as the initial estimate for the next transformation. [reproduced from [55]]

reference space, thus enabling the generation of the next average image.

11.4.3.2 Propagation of Transformations

It is well known that intensity-based image registration algorithms, both rigid and non-rigid, fail when started with an initial transformation estimate outside the "capture range" of the desired local optimum of the similarity measure. A wise choice of the initial transformation is therefore beneficial for the robustness of the registration method.

During the iterative averaging process, there are only minor changes in the overall shape of the average brain from one iteration to the next. Consequently, for all images n and all iterations i, the transformation $\mathbf{T}_n^{(i+1)}$ differs from the preceding $\mathbf{T}_n^{(i)}$ only by a small additional deformation. A similar situation, although for different reasons, is encountered when registering images from a time series to a common reference; temporally consecutive images typically differ from each other by a smaller amount than they differ from the common reference. In the context of temporal image
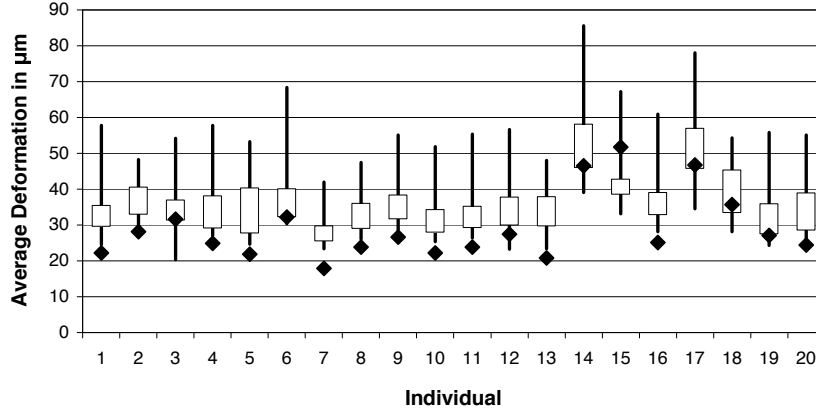
Figure 11.9: Comparison of deformation magnitudes between subjects vs. between a subject and the average shape atlas. The diamonds show the average deformation (over all foreground voxels) in μm when registering the respective raw image to the average shape atlas. The vertical lines show the range of average deformations when registering the respective raw image to the remaining 19 individual atlas images. The boxes show the 25th and 75th percentiles of the respective distributions.

sequence registration, a framework to incorporate deformations from previous steps into the current registration was recently proposed [62, 63].

For the iterative average image generation described here, we follow a similar approach. Our registration algorithm at each iteration takes as the initial transformation estimate the mapping found during the previous iteration (Fig. 11.8). This is the mapping used to generate the current average image. For the transition from affine to non-rigid registration, incorporation of the previous transformation is achieved by initializing the control point grid with control point positions transformed according to the individual affine transformations. For the transition from one non-rigid iteration to the next, the deformation is taken as is and used as the starting point for the optimization.

11.4.3.3   Distance to the Average Shape

Let us recall the rationale behind the creation of our average shape atlas: by minimizing the deformation required to map the atlas onto a given individual, the segmentation accuracy would be improved. So does the atlas produced by the method outlined above in fact minimize this deformation? Indeed, Fig. 11.9 illustrates that the differences be-

tween a raw image and an individual atlas are on average substantially larger than the differences between a raw image and the average atlas. Most raw images are more similar in shape to the average shape atlas than to any (or at least the majority) of the remaining 19 individual atlas images. Since the individuals registered to the average shape atlas were the same that built the atlas in the first place, this finding is not too surprising. However, it was important to show that at least for the "training set", our shape averaging does in fact produce a reasonable approximation to the population average shape.

11.4.3.4 Noise and Artifacts in the Average Atlas

In Fig. 11.10, we compare some individual images that were used to build the average shape atlas with the average shape atlas itself. It is easy to see that, in addition to representing an average shape, the average atlas also comes with an average microscopy image. The latter is easily generated by averaging the gray values of the appropriately deformed original microscopy images. The average image shows substantially reduced noise and imaging artifacts. It also shows a more homogeneous distribution of the chromophor compared to the individual brains. All these properties can potentially make registration of the atlas to a given raw image easier, and thus may aid in further improving segmentation accuracy.

11.4.4   Multi-Atlas Segmentation: A Classifier Approach

We can look at an atlas combined with a coordinate mapping from a raw image as a special type of *classifier*. The input of the classifier is a coordinate within the domain of the raw image. The classifier output, determined internally by transforming that coordinate and looking up the label in the atlas at the transformed location, is the label that the classifier assigns to the given raw image coordinate.

As we have briefly mentioned before, using a different atlas leads to a different segmentation of a given raw image. From a classifier perspective, we can therefore say that different atlases generate different classifiers for the same raw image. In the pattern recognition community, it has been well-known for some time that multiple independent classifiers can be combined, and together consistently achieve classification accuracies, which are superior to that of any of the original classifiers [29].

Successful applications of multiple classifier systems have been reported in recognizing handwritten numerals [32, 89] and in speech recognition [1, 70]. In the medical

Average Shape Atlas

Average Microscopy Image   Corresponding Label Field

Individual Microscopy Images, Mapped to Average Shape

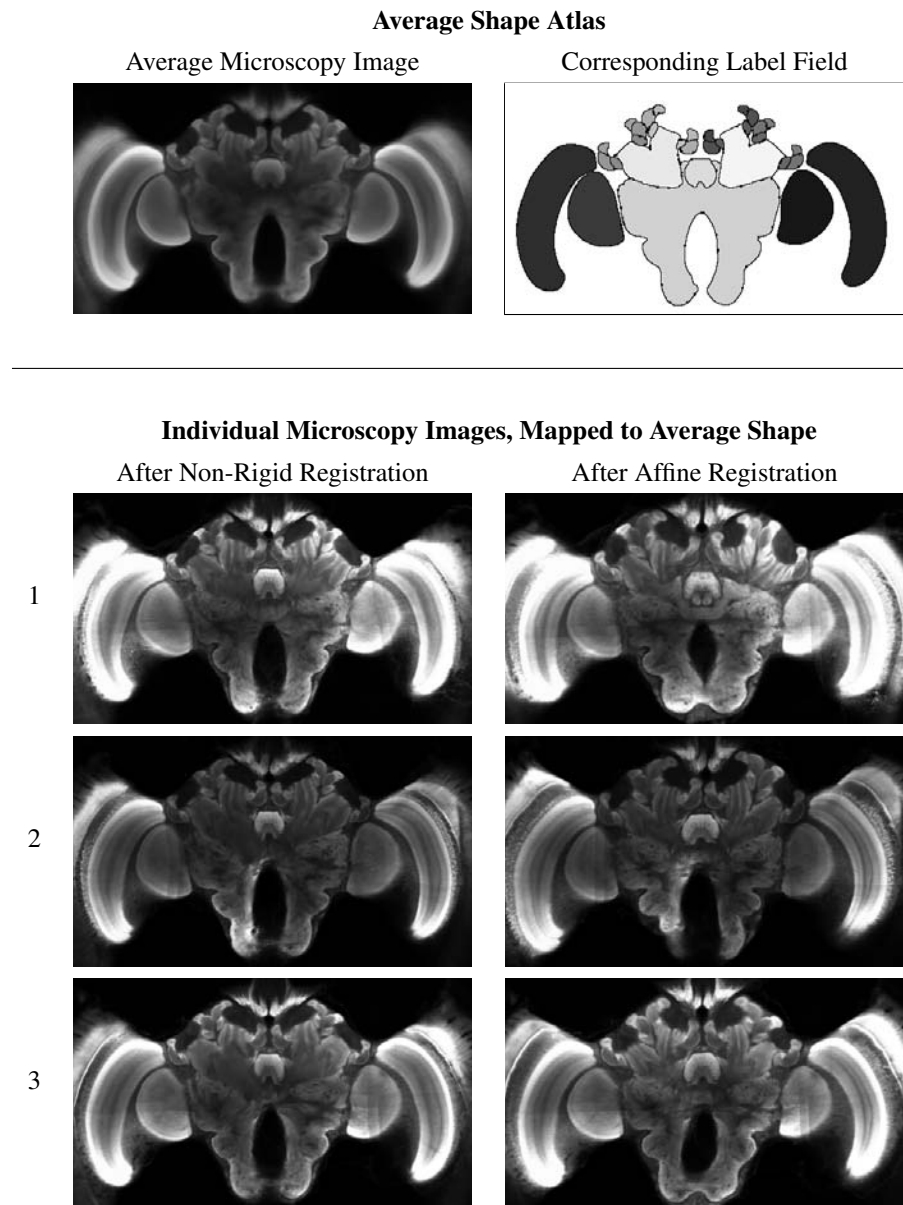After Non-Rigid Registration   After Affine Registration

Figure 11.10: Comparison of individual microscopy images and average shape atlas. Note that the average microscopy image of the average shape atlas also has a better signal-to-noise ratio and generally fewer artifacts than the original individual microscopy images (three randomly selected examples shown as rows 1 through 3 in this figure).

image analysis field, this principle has been applied for example to multi-spectral segmentation [46] and to computer-aided diagnosis of breast lesions [18, 51]. In related work, Thompson and Toga [79] used a database of atlases to generate a probabilistic segmentation of a subject.

The particular beauty of applying a multi-classifier framework to atlas-based segmentation is that multiple independent classifiers arise naturally from the use of multiple atlases. In fact, multiple classifiers also arise from using the same atlas with a different non-rigid registration method. However, adding an additional atlas is typically easier to do than designing an additional image registration algorithm. One could, however, also apply the same basic registration algorithm with a different regularization constraint weight (see Section 11.3.4), which would also lead to slightly different segmentations.

For the demonstration in this chapter, we performed a leave-one-out study with only one registration method, but a population of independent atlases. Each of the 20 bee brains was taken as the raw image and automatically segmented using every one of the remaining 19 brains as an atlas. This resulted in 19 segmentations per brain. These 19 segmentations were then combined into a final segmentation.

The most straight forward method for combining multiple classifications into one is the so-called "Vote Rule" decision fusion [28]. For each voxel in the raw image, the outputs of the individual atlas-based classifiers are determined. Their "votes" are then counted, and the label that received that highest number of votes is assigned to the voxel. It is worth, however, to take a closer look at the way an atlas-based classifier works: by looking up a label according to a transformed image coordinate. The label map is discrete, arranged on a 3-D grid of labeled voxels. Yet the coordinates of the raw image voxels that we are trying to label hardly ever directly fall on grid points in the atlas. Therefore, looking up the correct label requires some sort of interpolation. The simplest label interpolation method is nearest neighbor (NN) interpolation, resulting in a single unique label per atlas-based classifier. These can easily be combined using vote fusion as described above.

A slightly more complex interpolation technique that can be applied to labels is partial volume (PV) interpolation as introduced by Maes *et al.* [39]. Here, the labels of all eight neighbors of the interpolated coordinate are determined and weighted with the trilinear interpolation coefficients of their respective grid nodes. Therefore, the output of an atlas-based classifier using PV interpolation is a vector of weights between zero and one, which are assigned to each of the possible labels. One can interpret the

weights as the confidence of the classifier in the respective label being the correct answer. These weighted decisions from all classifiers can be combined by so-called "Sum Rule" fusion [28]. The weights for each label are added over all classifiers, and the label with the highest sum is taken as the combined decision.

11.5   Quantifying Segmentation Accuracy

In addition to presenting selected algorithms for atlas-based segmentation, this chapter provides a quantitative comparison among different methods. For each segmentation that we perform, its accuracy is computed. The accuracies achieved for each image are then compared among different methods in order to illustrate quality differences and to identify superior algorithms.

Computing the accuracy of a segmentation requires a gold standard, or ground truth. That is, the correct segmentation needs to be known for an image in order to be able to compute the accuracy of an automatically generated segmentation of that image. While not at all guaranteed to be correct, it is commonly accepted today to use a manual segmentation by a human expert, supported by advanced semi-automatic labeling techniques such as intelligent scissors [44], as the gold standard that automatic segmentation methods are measured against.

11.5.1   Similarity Index

Figure 11.11 provides a visual impression of the segmentation result for two representative slices from one segmented bee brain image. However, in order to effectively compare different segmentation methods, we need to quantify the segmentation accuracy. One possible measure of segmentation quality is the *similarity index* (SI) [93]. For a structure s, the SI is computed from the set $V_{\mathrm{auto}}^{(s)}$ of voxels in s according to the automatic segmentation and the set $V_{\mathrm{manual}}^{(s)}$ of voxels in s according to the (gold standard) manual segmentation:

$$\mathrm{SI}(s) = \frac{2\left|V_{\mathrm{manual}}^{(s)} \cap V_{\mathrm{auto}}^{(s)}\right|}{\left|V_{\mathrm{manual}}^{(s)}\right| + \left|V_{\mathrm{auto}}^{(s)}\right|}. \tag{11.8}$$

For perfect mutual overlap of both segmentations, manual and automatic, the SI has a value of 1. Lesser overlap results in smaller values of SI. No overlap between the segmentations results in an SI value of 0. A major advantage of the SI measure is that
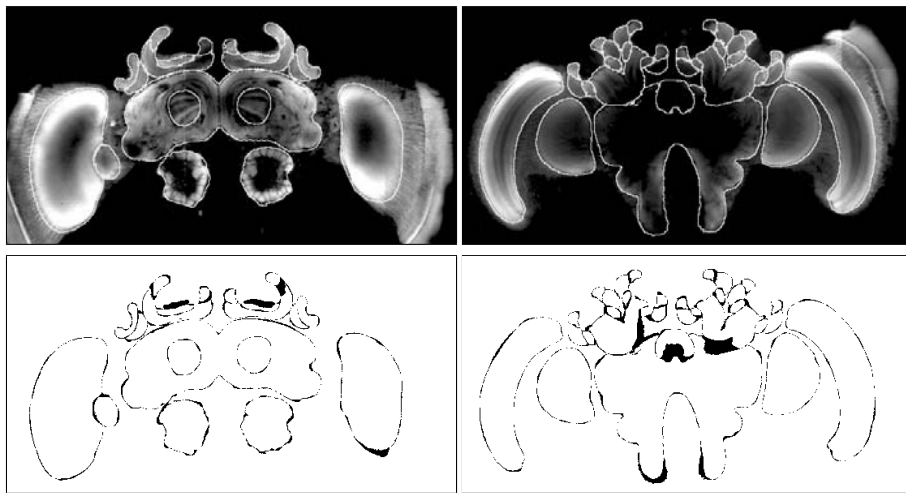
Figure 11.11: Example of segmentation using non-rigid image registration (MUL atlas selection paradigm). The two columns show axial images at two different slice locations. *Top row:* Overlays of segmentation contours (shown in white) after nonrigid image registration. *Bottom row:* Difference images between manual and automatic segmentation. Voxels with different labels assigned by manual and automatic segmentation are shown in black.
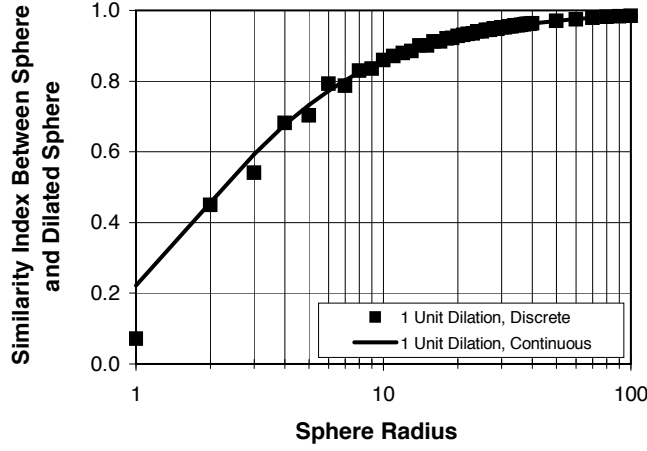
Figure 11.12: Dependence of SI values on size for spherical objects. The squares show SI values computed from discrete numerical simulation of dilation by one voxel. The solid line shows SI values for the continuous case (Eq. (11.9)). Note that while the units on the horizontal axis are voxels for the discrete case, they are arbitrary units for the continuous case.

it is sensitive to both over-segmentation and under-segmentation, that is, it recognizes both false positives and false negatives among the voxels of a given structure.

11.5.2   Bias from Structure Volume

In order to understand the SI values computed later in this chapter and to compare them with other published values, we investigated the dependence of SI values on object size. We performed a numerical simulation in which discretely sampled spheres of various radii were dilated by one or two voxels and the SI values between the original and dilated spheres were computed. The resulting SI values are plotted versus object radius in Fig. 11.12. It is also easy to derive a closed-form expression for the continuous case. The SI between two concentric spheres, one with radius R and the other dilated by d, i.e., with a radius of $R + d$, is

$$\mathrm{SI} = \frac{2(R/d)^3}{2(R/d)^3 + 3(R/d)^2 + 3(R/d) + 1} \ . \tag{11.9}$$

The SI values for the discrete and continuous cases are almost identical (Fig. 11.12). The SI value between a sphere and a concentric dilated sphere approximates the SI value for a segmentation error consisting of a uniform thickness misclassification on
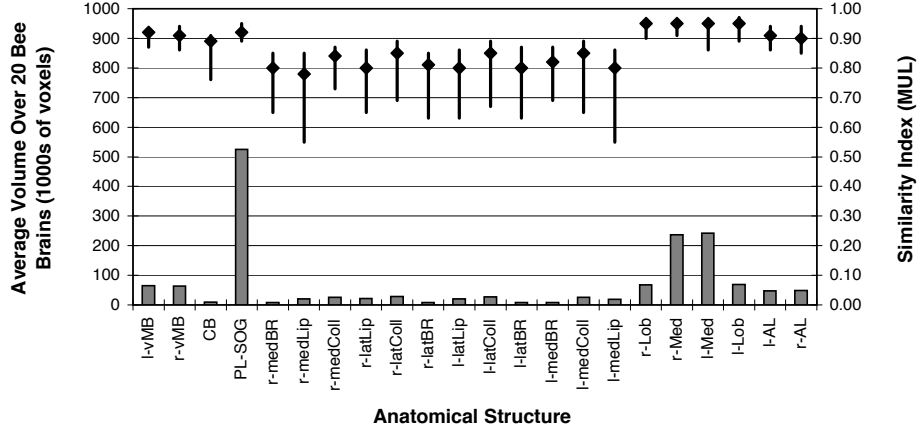
Figure 11.13: Volumes of anatomical structures and corresponding segmentation accuracies. The gray bars show the volumes (in numbers of voxels) of the 22 anatomical structures, averaged over the 20 bee brains. The black vertical lines show the range of SI values achieved by the automatic segmentation (MUL paradigm) over all segmented raw images. The diamond shows the median over all segmented raw images.

the perimeter of a spherical object. Inspection of Fig. 11.12 and Eq. (11.9) shows that SI depends strongly on object size and is smaller for smaller objects. A one voxel thick misclassification on the perimeter of a spherical object with a radius of 50 voxels has an SI value of 0.97, but for a radius of 10 voxels the SI value is only 0.86. Thus it is not surprising that Dawant *et al.* [16] reported mean SI values of 0.96 for segmentation of the human brain from MR images and mean SI values of only 0.85 for segmentation of smaller brain structures such as the caudate.

In Fig. 11.13, the average volumes of the anatomical structures in the bee brain images under consideration are shown with the actual segmentation accuracies achieved for them using one of the segmentation methods discussed later (MUL). It is easy to see that the larger a structure, the more accurately it was typically segmented by the atlas-based segmentation. This confirms the theoretical treatment above and illustrates the varying bias of the SI metric when segmenting structures of different sizes.

11.5.3 Bias from Structure Shape

A simple numerical measure that characterizes the shape of a geometrical object is its surface-to-volume ratio (SVR). For a discrete set of labeled voxels in a segmented structure, we can approximate the SVR ρ as the ratio of the number of surface voxels
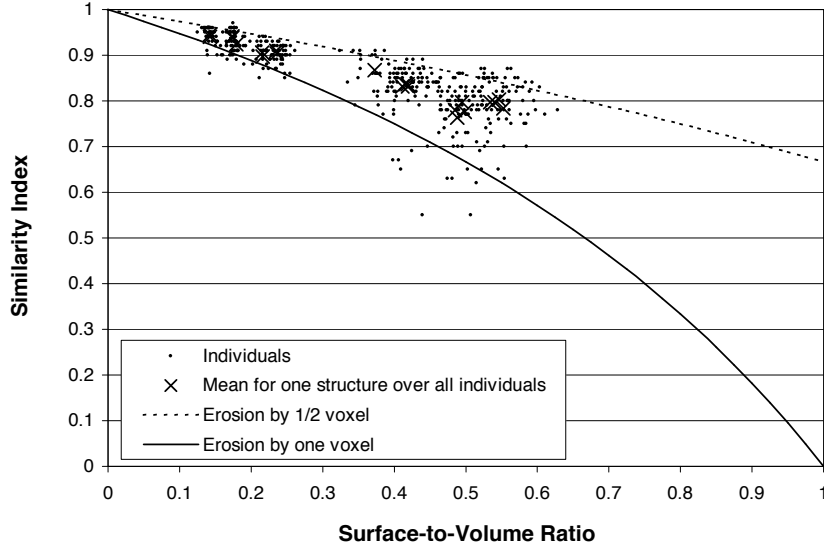
Figure 11.14: Similarity index vs. surface-to-volume ratio. Each dot represents one structure in one brain (418 structures in total). The average over all individuals for one structure is marked by a $\times$. The solid and dashed lines show the numerical relationship between SVR and SI for erosion by one and 1/2 voxel, respectively.

N_s to the total number of voxels N_t, that is,

$$\rho \approx \frac{N_s}{N_t} \tag{11.10}$$

A surface voxel is easily defined as one that has a neighbor with a label different from its own. When the entire surface of a structure is misclassified, this can be seen as an erosion of the structure by one voxel. The SI value computed between the original structure and the eroded structure represents the SI resulting from a segmentation that misclassifies exactly all surface voxels. From the structure's SVR ρ and its total volume V, this SI can be computed as

$$SI = \frac{2V(1-\rho)}{V + (1-\rho)V} = \frac{1-\rho}{1-\rho/2}. \tag{11.11}$$

Similarly, we can estimate the result of an erosion by two voxels, corresponding to a segmentation error of two voxels. Figure 11.14 shows the SVR values computed for all structures in all brains in our 20 bee brains, plotted versus the SI values of the automatic segmentations. The figure also shows two curves that represent the numerical simulation of misclassification by one and 1/2 voxels, respectively, based on Eq. (11.11).
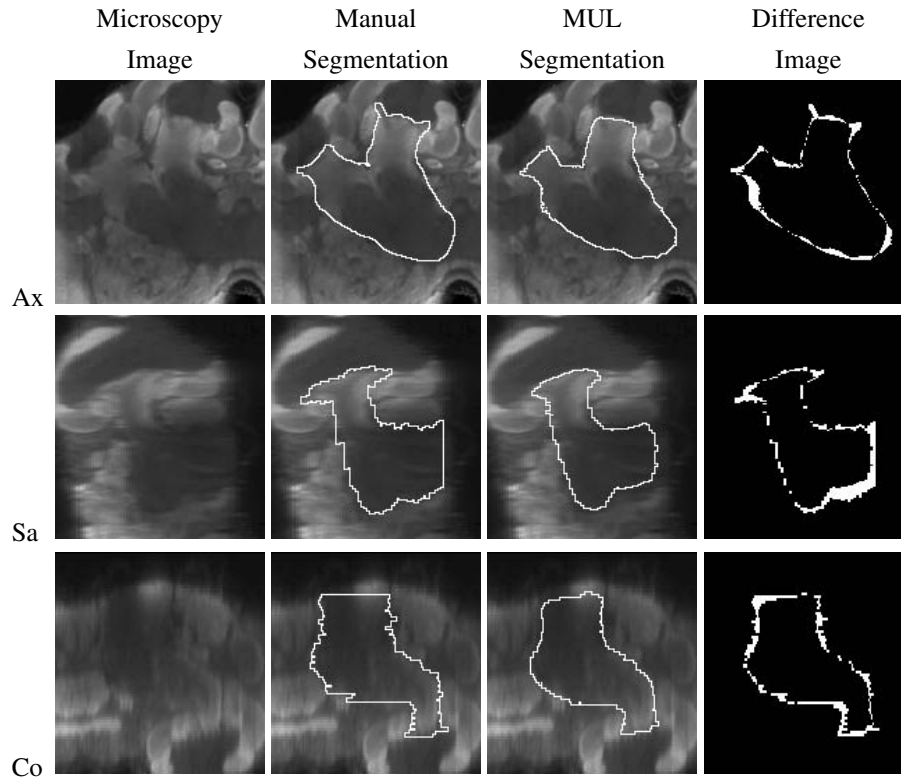
Figure 11.15: A typical segmented structure: right ventral mushroom body (SI = 0.86). *Columns from left to right*: microscopy image, contour from manual segmentation, contour from automatic segmentation (MUL paradigm), and difference image between manual and automatic segmentation. The white pixels in the difference image show where manual and automatic segmentation disagree. *Rows from top to bottom*: axial, sagittal, and coronal slices through the right ventral mushroom body.
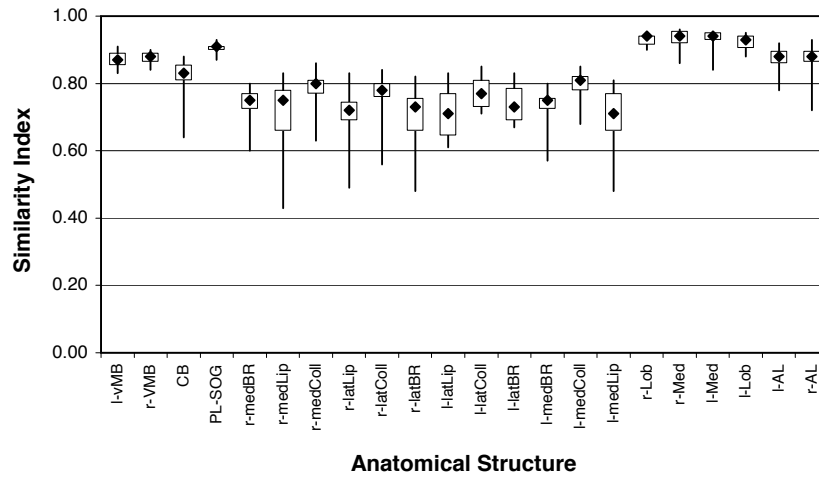
Figure 11.16: SI by label for segmentation using a single individual atlas (IND atlas selection strategy). The diamond shows the median value over 19 segmented images. The vertical lines show the range of values, while the boxes show the 25th and 75th percentiles of the respective distributions. [modified from [56]]

For a typical segmentation result of a single structure, a detailed comparison of manual and automatic segmentation is shown in Fig. 11.15. The structure shown here, a right ventral mushroom body, is typical in that its volume and its surface-to-volume ratio are close to the respective means over all structures (volume 141,000 pixels vs. mean 142,000 pixels; SVR 0.24 vs. mean 0.36). The segmentation accuracy for the segmentation shown was SI = 0.86, which is the median SI value over all structures and all brains.

11.5.4   Comparison of Atlas Selection Strategies

The results achieved using the different atlas selection strategies outlined above are visualized in Figs. 11.16 through 11.19. Each graph shows a plot of the distribution of the SI segmentation accuracies over 19 segmentations, separated by anatomical structure. There were 19 segmentations per strategy as one out of the 20 available bee brain images served as the fixed individual atlas for the IND strategy. Therefore, this brain was not available as the raw image for the remaining strategies, in order to avoid bias of the evaluation.

A comparison of all four strategies is shown in Fig. 11.20. It is easy to see from the latter figure that the widely used IND strategy produced the least accurate results
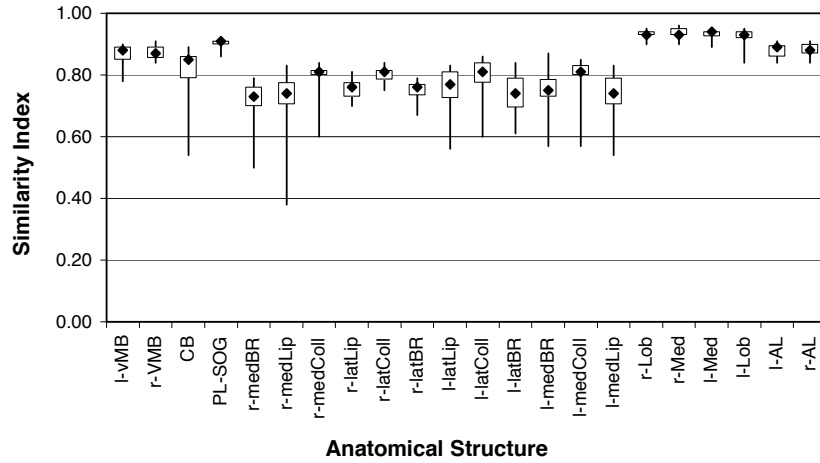
Figure 11.17: SI by label for segmentation using the most similar single individual atlas (SIM atlas selection strategy). For a description of the graphical representation see Fig. 11.16. [modified from [56]]
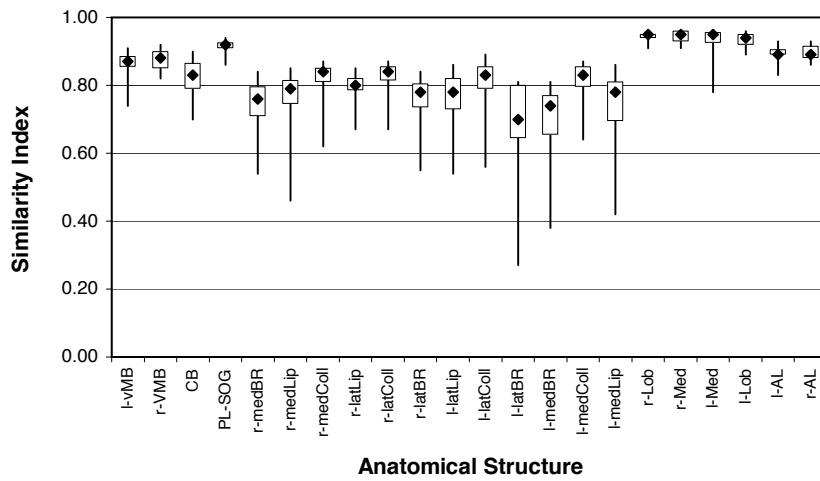

Figure 11.18: SI by label for segmentation using a single average shape atlas (AVG atlas selection strategy). For a description of the graphical representation see Fig. 11.16. [modified from [56]]
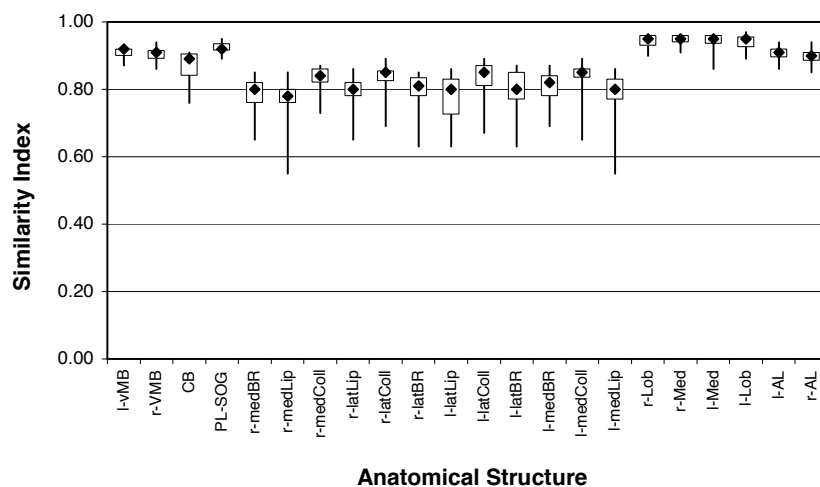
Figure 11.19: SI by label for segmentation by combining multiple independent atlas-based segmentations (MUL atlas selection strategy). For a description of the graphical representation see Fig. 11.16. [modified from [56]]
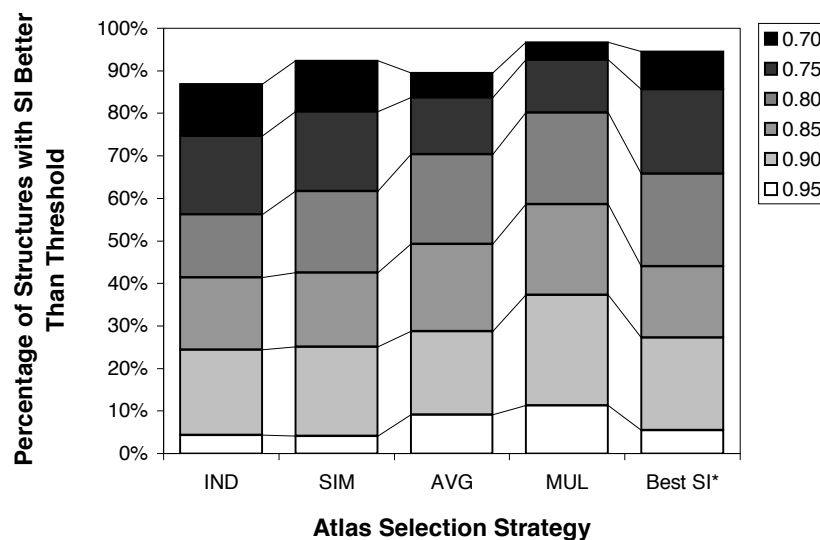

Figure 11.20: Percentage of registration-based segmentations with similarity index SI better than given threshold plotted by atlas selection strategy. The series labeled "Best SI" is the upper bound of all strategies working with a single individual atlas (see text for details).

of all strategies. Only slightly better results were achieved by selecting a different individual atlas for each raw image, based on the NMI after non-rigid registration criterion discussed in Section 11.4.2. The AVG strategy, segmentation using an average shape atlas, outperformed both the IND and SIM strategies, but was itself clearly outperformed by the MUL strategy. Our results therefore show that the multi-classifier approach to atlas-based segmentation produced substantially more accurate segmentations than the other three strategies. This finding is in fact statistically significant when performing a t-test on the SI values for all structures over all segmentations, which confirms the experience of the pattern recognition community that multiple classifier systems are generally superior to single classifiers [29, 89].

Another interesting finding is that both the AVG and the MUL strategies performed better than the theoretical upper bound of any strategy working with only a single individual atlas (series labeled "Best SI" in Fig. 11.20). We note that "Best SI" is the upper bound not only for any method with the best atlas for each raw image, but also for any possible selection of one atlas for all raw images. Therefore, it is also the upper bound of the IND and SIM strategies, which in our study can consequently never outperform the AVG or MUL strategies.

11.6   More on Segmentation with Multiple Atlases

We saw in the previous section that a multi-classifier approach to atlas-based segmentation outperforms atlas-based segmentation with a single atlas, be it an individual atlas, an average atlas, or even the best out of a database of atlases. Compared to that, the insight underlying the SIM ("most similar") atlas selection strategy was that different atlases lead to segmentations of different accuracies. Combined, both observations lead to an even more interesting concept: combination of multiple atlas-based segmentations, weighted by estimates of their individual segmentation accuracy.

In other words, if we had estimates of how well each atlas-based classifier is performing, then we could be more confident in decisions of those classifiers that perform well, compared to the decisions of those that do not. One would hope that by concentrating on more accurate classifiers in the ensemble, the classification accuracy would be further improved.

The performance of each atlas-based classifier is obviously not known in general, due to the lack of a ground truth. However, several methods have been proposed that can estimate the performance parameters, for example using expectation maxi-

mization (EM) methods. Two of these are outlined below, one based on a per-label binary performance model [85], and another based on a simultaneous multi-label performance model [64, 65].

For the description of both methods, we assume that an image with N voxels is segmented by K different (atlas-based) classifiers. For each voxel x, we denote with $e_k(x)$ the decision by classifier k, which is one of the labels assigned in the segmentation. For the sake of simplicity of the presentation, we assume classifiers that internally use NN interpolation for atlas lookup and therefore only produce one unique label as their output. If the (unknown) ground truth for voxel x is i, we say that x is in class i and write this as $x \in C_i$.

11.6.1   A Binary Classifier Performance Model

An EM algorithm described by Warfield *et al.* [85] estimates the classifier performance for each label separately. The method is based on the common performance parameters p (sensitivity) and q (specificity), i.e., the fractions of true positives and true negatives among the classified voxels. The parameters p and q are modeled independently for each classifier k and each class C_i (label in the segmentation) as the following conditional probabilities:

$$p_i^{(k)} = P(e_k(x) = i | x \in C_i) \text{ and } q_i^{(k)} = P(e_k(x) \neq i | x \notin C_i). \qquad (11.12)$$

From these definitions, an EM algorithm that estimates p and q from the classifier decisions can be derived as described by Warfield *et al.* [85]. From the computed classifier performance parameters for each label, a contradiction-free final segmentation E at voxel x can be computed as

$$E(x) = \arg \max_i P(x \in C_i | e_1(x), \dots, e_K(x)). \qquad (11.13)$$

Here, the probability $P(x \in C_i | \mathbf{e})$ follows from the classifiers' decisions and their performance parameters using Bayes' rule. For details on the application of this algorithm to classifier fusion, see Ref. [64].

11.6.2   A Multi-Label Classifier Performance Model

In a generalization of the Warfield algorithm to multi-label segmentations [64], the classifier parameters p and q are replaced by a matrix of label cross-segmentation coefficients $\lambda_{i,j}^{(k)}$. These describe the conditional probabilities that for a voxel x in

class C_i the classifier k assigns label $j = e_k(x)$, that is

$$\lambda_{i,j}^{(k)} = P(e_k(x) = j | x \in C_i). \tag{11.14}$$

This formulation includes the case that $i = j$, i.e., the classifier decision for that voxel was correct. Consequently, $\lambda_{i,i}^{(k)}$ is the usual sensitivity of classifier k for label i. We also note that for each classifier k the matrix $(\lambda_{i,j}^{(k)})_{i,j}$ is a row-normalized version of the "confusion matrix" [89] in Bayesian multi-classifier algorithms. This matrix, when filled with proper coefficients, expresses prior knowledge about the decisions of each classifier. Again, the coefficients can be estimated iteratively from the classifier decisions by an EM algorithm.

In the "E" step of the EM algorithm, the unknown ground truth segmentation is estimated. Given the current estimate for the classifier parameters (λ) and the classifier decisions $e_k(x)$, the likelihood of voxel x being in class C_i is

$$W(x \in C_i) = \frac{P(x \in C_i) \prod_k \lambda_{i,e_k(x)}^{(k)}}{\sum_{i'} \left[P(x \in C_{i'}) \prod_k \lambda_{i',e_k(x)}^{(k)} \right]}. \tag{11.15}$$

Note that W is a function of two parameters, x and i. The "M" step of the algorithm estimates the classifier parameters (λ) that maximize the likelihood of the current ground truth estimate determined in the preceding "E" step. Given the previous estimates W of the class probabilities, the new estimates for the classifier parameters are computed as follows:

$$\hat{\lambda}_{i,j}^{(k)} = \frac{\sum_{x:e_k(x)=j} W(x \in C_i)}{\sum_x W(x \in C_i)}. \tag{11.16}$$

11.6.3   Results of Performance-Based Multi-Atlas Segmentation

The accuracy of the performance parameter estimation using both EM algorithms is shown in Fig. 11.21. We computed the actual performance parameters for each atlas-based classifier by comparing its output with the manual segmentation. These *a posteriori* performances (conceptually equivalent to the *recognition rate* of the classifier) were then plotted versus the estimates computed by either of the EM methods. It is easy to see that there is a very good agreement between the actual and the the estimated parameters, with a slightly higher predictive accuracy following the binary performance model. The Pearson correlation coefficient between true and estimated performances was 0.94 for the binary expert model, and 0.87 for the multi-label expert model. The increased quality of the parameter estimation using the binary per-

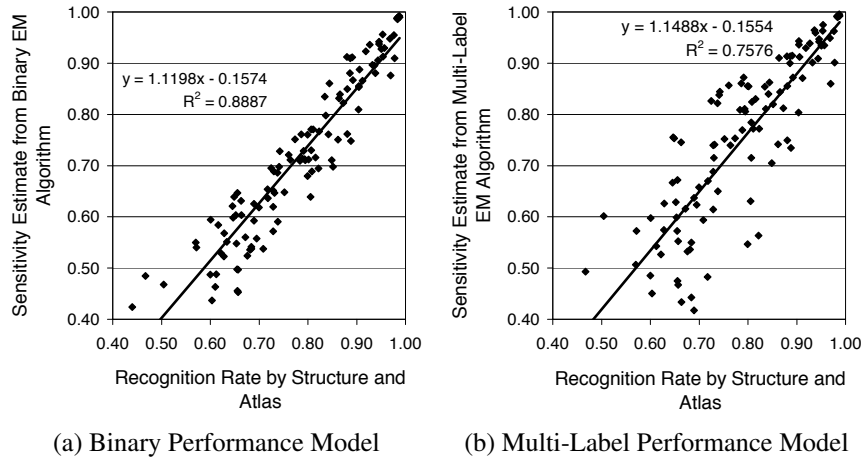(a) Binary Performance Model   (b) Multi-Label Performance Model

Figure 11.21: Accuracy of classifier performance parameter estimation using EM algorithms.

formance model can be explained by the substantially larger number of degrees of freedom in the multi-label model, due to the inter-label crosstalk coefficients.

As Fig. 11.22 illustrates, the accuracy of a multi-classifier segmentation can be improved considerably when the performance parameters of the individual classifiers are estimated and taken into account. Overall, the estimation method using a multi-label performance parameter model was slightly less accurate in estimating the actual parameters, but produced a slightly better segmentation accuracy than the method based on a binary performance model.

11.7   Conclusion

This chapter has shed light on some often overlooked aspects of atlas-based segmentation methods. We have compared four different strategies for atlas selection and demonstrated that the accuracy of atlas-based segmentation can be improved substantially by moving beyond the use of a single, individual atlas.

Recently published works on atlas creation and atlas-based segmentation make increasing use of standard atlases that incorporate properties of a population of subjects [36, 38, 49]. Our results confirm that this is likely beneficial for improved segmentation accuracy and robustness. However, our results also suggest that the benefits of applying a multi-classifier strategy are well worth the extra effort.
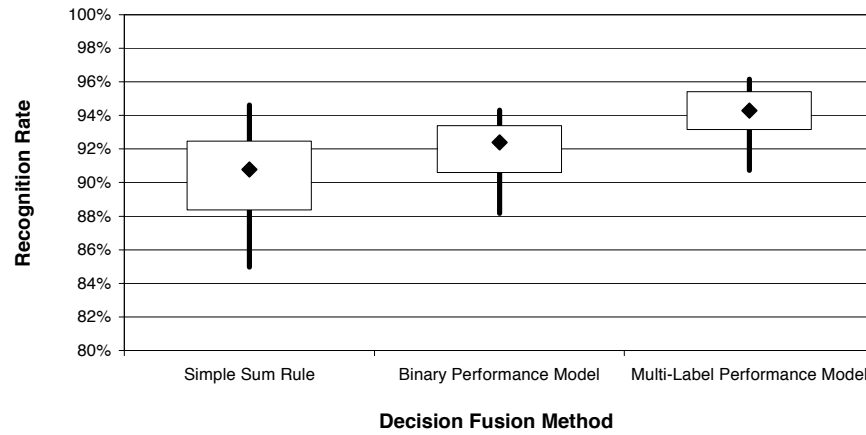
Figure 11.22: Recognition rates (foreground only) of multiple classifier systems based on simple sum fusion, the binary performance model, and the multi-label performance model.

On a more application-specific note regarding the accuracy of atlas-based segmentation of bee brains, we observe that the mean SI value of segmentations produced using the MUL method in this chapter is 0.86, which, given the small size of most of the structures in the bee brains considered, is comparable to the values reported by Dawant *et al.* and supports the visual assessment observation (Fig. 11.11) that the automatic segmentations described here differ from manual segmentations on average by approximately one voxel. In fact, Zijdenbos *et al.* [93] state that "SI > 0.7 indicates excellent agreement" between two segmentations. This criterion (SI > 0.7) is satisfied by virtually all (97 percent) contours generated by our segmentations using the MUL method (Fig. 11.19). Furthermore, since the image quality of confocal microscopy images is inferior to clinical MR and CT images in many ways, we believe that our registration-based segmentation method represents a satisfactory intermediate solution to a segmentation problem that is appreciably harder than that of segmenting commonly used images of the human brain.

Acknowledgments

11.8   Brainstorming Questions

(Q1) What are the advantages of atlas-based segmentation over other segmentation techniques?

(Q2) Why is the described non-rigid registration method superior to other techniques?

(Q3) What is the best value for the smoothness constraint weight of the non-rigid registration (Section 11.3.4)?

(Q4) What if in Section 11.4.2 the atlas most similar to the raw image were selected using the following criterion I invented: . . . ?

(Q5) When combining multiple atlas-based segmentations, what is the practical difference between NN interpolation with vote fusion and PV interpolation with sum fusion?

(Q6) If the MUL atlas selection strategy is so much better than the others, then why is it not always used?

(Q7) How does atlas-based segmentation compare to manual segmentation?

(Q8) Are there parallels to the multi-atlas segmentation method in pattern recognition?

(Q9) Could an Active Shape Model be used as an atlas for segmentation?

(Q10) Why does the binary classifier performance model predict actual performance more accurately, yet the multi-label performance model gives better combined classification results?

(A1) Which segmentation technique is the best strongly depends on the application. Atlas-based segmentation is most useful when the spatial location of a voxel, together with its gray value, determines which class it belongs to. Conversely,

atlas-based segmentation should not be applied for problems where the classification depends purely on a voxel's gray value, for example when classifying different tissue types.

(A2) It is not in general superior. Again, the most suitable technique depends on the application. The reason for using this particular method is that it does not require single-modality image data. While strictly all images in the application presented here are confocal microscopy images, the imaging process causes very different gray value distributions for different individuals. The NMI image similarity measure is well-suited for this kind of situation, while other measures (Mean Squared Difference, Cross Correlation) are not.

(A3) Unfortunately, there is no *a priori* best value for this weight. There is always a trade-off between the smoothness and regularity of the transformation (higher weights), and good fitting of the data (lower weights). Choosing the best constraint weight is subject to active research, and is by no means a solved problem. For the bee brain microscopy images, we have found that the best results were achieved with a value of 0.1, but in other applications with different data and different objectives [61], the "best" values were very different. It is also possible to vary the constraint weight during the optimization to achieve better results [58].

(A4) Sorry, but – no! No criterion for the selection of the most similar atlas can produce segmentations of higher accuracy than a criterion that has knowledge of the true segmentation. This is precisely what the "Best SI" criterion is all about. So in our study at least, whatever the criterion, the SI values would never be better than using the criterion that selects based on the SI values themselves (which, let us mention it again, are not available in a real-world application anyway). Therefore, there is no criterion that would not still be outperformed by the MUL and even the AVG segmentation strategy.

(A5) By integrating the trilinear interpolation coefficients, PV interpolation leads to smoother transitions from one label to another and avoids jagged edges when interpolating oblique slices from an atlas. In general, we found NN and vote fusion to perform better for large numbers of segmentations (more than 30), since a distribution of samples from a large number of atlases approximates a continuous spatial distribution. PV with sum fusion seemed to perform slightly better for smaller numbers of segmentations. On a related note, due to the real-valued

weights, it is much less likely to see a voxel tied between two or more labels using PV than using NN, so PV can help to reduce "rejected" classifications.

(A6) The MUL strategy has two drawbacks: it requires multiple atlases – a problem that it shares with the SIM and the AVG strategies. Furthermore, the raw image needs to be non-rigidly registered separately to each atlas, which again is also true for some criteria that could be used in a SIM strategy. Non-rigid registration is computationally expensive, so the bottom line is: for a MUL segmentation strategy, you need more atlases and more CPU time.

(A7) In short: we do not know. One important question is, how different is an automatic segmentation from a manual one, compared to the typical difference between two (or more) manual segmentations. The latter, the inter-observer difference, can only be determined by repeated manual segmentation of the same image. This is costly and time consuming and was therefore not done for the data we have used in this chapter.

(A8) A closely related technique in pattern recognition is the concept of "bagging" [6] (short for bootstrap aggregation), a method for generating multiple, somewhat independent instantiations of a classifier by using different training sets. Multi-atlas segmentation is similar to this method in that it also generates multiple atlas-based classifiers using multiple training sets. In our case, the different training sets are the atlases, and the resulting classifiers are described by the coordinate transformations after registration of each atlas to the raw image.

(A9) Yes. In fact, it would be interesting to use a dynamic atlas during the registration stage rather than a static one. The dynamic atlas could be the average shape, with the most significant modes of variations added to it. The weights of all modes of variation could be incorporated as additional degrees of freedom into the registration process. So in some sense, this procedure would be similar to using an average shape atlas by itself, but with the additional benefit that the atlas could adapt to the individual that is being segmented. Another way to look at it is as a registration from two sides, where one image is deformed using a free-form deformation, whereas the deformation of the other, the atlas, is governed by the more specific ASM.

(A10) The multi-label performance model has substantially more free parameters than the binary model, and we only looked at one of them per segmented structure

(i.e., sensitivity). While the binary model can only predict the probability of an error, the multi-label model actually makes a prediction of what the error will be. That is, the multi-label model allows us to model situations where, for example, misclassification of a voxel of class A into class B is much more likely than into class C. The binary model would only predict the likelihood of the misclassification, but would not provide any information as to the nature of the error.

Bibliography

[1] Altincay, H. and Demirekler, M., An information theoretic framework for weight estimation in the combination of probabilistic classifiers for speaker identification, Speech Communication, Vol. 30, No. 4, pp. 255–272, 2000.

[2] Ashburner, J., *Computational Neuroanatomy*, Ph.D. thesis, University College London, 2000.

[3] Baillard, C., Hellier, P. and Barillot, C., Segmentation of brain 3D MR images using level sets and dense registration, Medical Image Analysis, Vol. 5, No. 3, pp. 185–194, 2001.

[4] Bookstein, F. L., 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.

[5] Online: `http://www.bic.mni.mcgill.ca/brainweb/`.

[6] Breiman, L., Bagging predictors, Machine Learning, Vol. 24, No. 2, pp. 123–140, 1996.

[7] Bucher, D., Scholz, M., Stetter, M., Obermayer, K. and Pflüger, H.-J., Correction methods for three-dimensional reconstructions from confocal images: I. tissue shrinking and axial scaling, Journal of Neuroscience Methods, Vol. 100, pp. 135–143, 2000.

[8] Cheng, P. C., Lin, T. H., Wu, W. L., and Wu, J. L., eds., *Multidimensional Microscopy*, Springer-Verlag, New York, 1994.

[9] Christensen, G. E. and Johnson, H. J., Consistent image registration, IEEE Transactions on Medical Imaging, Vol. 20, No. 7, pp. 568–582, 2001.

[10] Christensen, G. E., Rabbitt, R. D. and Miller, M. I., Deformable templates using large deformation kinematics, IEEE Transactions on Image Processing, Vol. 5, No. 10, pp. 1435–1447, 1996.

[11] Collins, D. L. and Evans, A. C., Animal: Validation and applications of nonlinear registration-based segmentation, International Journal of Pattern Recognition and Artificial Intelligence, Vol. 11, No. 8, pp. 1271–1294, 1997.

[12] Collins, D. L., Holmes, C. J., Peters, T. M. and Evans, A. C., Automatic 3-D model-based neuroanatomical segmentation, Human Brain Mapping, Vol. 3, No. 3, pp. 190–208, 1995.

[13] Collins, D. L., Zijdenbos, A. P., Kollokian, V., Sled, J. G., Kabani, N. J., Holmes, C. J. and Evans, A. C., Design and construction of a realistic digital brain phantom, IEEE Transactions on Medical Imaging, Vol. 17, No. 3, pp. 463–468, 1998.

[14] Cootes, T. F., Taylor, C. J., Cooper, D. H. and Graham, J., Active shape models – Their training and application, Computer Vision and Image Understanding, Vol. 61, No. 1, pp. 38–59, 1995.

[15] Crum, W. R., Scahill, R. I. and Fox, N. C., Automated hippocampal segmentation by regional fluid registration of serial MRI: Validation and application in Alzheimer's Disease, NeuroImage, Vol. 13, No. 5, pp. 847–855, 2001.

[16] Dawant, B. M., Hartmann, S. L., Thirion, J. P., Maes, F., Vandermeulen, D. and Demaerel, P., Automatic 3-D segmentation of internal structures of the head in MR images using a combination of similarity and free-form transformations: Part I, methodology and validation on normal subjects, IEEE Transactions on Medical Imaging, Vol. 18, No. 10, pp. 909–916, 1999.

[17] Denton, E. R. E., Sonoda, L. I., Rueckert, D., Hill, D. L. G., Leach, M. O. and Hawkes, D. J., Comparison and evaluation of rigid, affine, and nonrigid registration of breast MR images, Journal of Computer Assisted Tomography, Vol. 23, No. 5, pp. 800–805, 1999.

[18] De Santo, M., Molinara, M., Tortorella, F. and Vento, M., Automatic classification of clustered microcalcifications by a multiple expert system, Pattern Recognition, Vol. 36, No. 7, pp. 1467–1477, 2003.

[19] Forsey, D. R. and Bartels, R. H., Hierarchical B-spline refinement, ACM SIGGRAPH Computer Graphics, Vol. 22, No. 4, pp. 205–212, 1988.

[20] Frangi, A. F., Rueckert, D., Schnabel, J. A. and Niessen, W. J., Automatic 3D ASM construction via atlas-based landmarking and volumetric elastic registration, in Insana, M. F. and Leahy, R. M., eds., *Information Processing in Medical*

Imaging: 17th International Conference, IPMI 2001, Davis, CA, USA, June 18-22, 2001, Proceedings, Vol. 2082 of *Lecture Notes in Computer Science*, pp. 78–91, Springer-Verlag, Berlin Heidelberg, 2001.

[21] Frangi, A. F., Rueckert, D., Schnabel, J. A. and Niessen, W. J., Automatic construction of multiple-object three-dimensional statistical shape models: application to cardiac modeling, IEEE Transactions on Medical Imaging, Vol. 21, No. 9, pp. 1151–1166, 2002.

[22] Gee, J. C., Reivich, M. and Bajcsy, R., Elastically deforming a three-dimensional atlas to match anatomical brain images, Journal of Computer Assisted Tomography, Vol. 17, No. 2, pp. 225–236, 1993.

[23] Guimond, A., Meunier, J. and Thirion, J.-P., Average brain models: A convergence study, Computer Vision and Image Understanding, Vol. 77, No. 2, pp. 192–210, 2000.

[24] Hartmann, S. L., Parks, M. H., Martin, P. R. and Dawant, B. M., Automatic 3-D segmentation of internal structures of the head in MR images using a combination of similarity and free-form transformations: Part II, validation on severely atrophied brains, IEEE Transactions on Medical Imaging, Vol. 18, No. 10, pp. 917–926, 1999.

[25] Hill, D. L. G., Batchelor, P. G., Holden, M. and Hawkes, D. J., Medical image registration, Physics in Medicine and Biology, Vol. 46, pp. R1–R45, 2001.

[26] Iosifescu, D. V., Shenton, M. E., Warfield, S. K., Kikinis, R., Dengler, J., Jolesz, F. A. and McCarley, R. W., An automated registration algorithm for measuring MRI subcortical brain structures, NeuroImage, Vol. 6, No. 1, pp. 13–25, 1997.

[27] Kikinis, R., Shenton, M. E., Iosifescu, D. V., McCarley, R. W., Saiviroonporn, P., Hokama, H. H., Robatino, A., Metcalf, D., Wible, C. G., Portas, C. M., Donnino, R. M. and Jolesz, F. A., A digital brain atlas for surgical planning, model-driven segmentation, and teaching, IEEE Transactions on Visualization and Computer Graphics, Vol. 2, No. 3, pp. 232–241, 1996.

[28] Kittler, J. and Alkoot, F. M., Sum versus vote fusion in multiple classifier systems, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 25, No. 1, pp. 110–115, 2003.

[29] Kittler, J., Hatef, M., Duin, R. P. W. and Matas, J., On combining classifiers, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 20, No. 3, pp. 226–239, 1998.

[30] Klagges, B. R. E., Heimbeck, G., Godenschwege, T. A., Hofbauer, A., Pflugfelder, G. O., Reifegerste, R., Reisch, D., Schaupp, M. and Buchner, E., Invertebrate synapsins: a single gene codes for several isoforms in Drosophila, The Journal of Neuroscience, Vol. 16, pp. 3154–3165, 1996.

[31] Kovacevic, N., Lobaugh, N. J., Bronskill, M. J., B., L., Feinstein, A. and Black, S. E., A robust method for extraction and automatic segmentation of brain images, NeuroImage, Vol. 17, No. 3, pp. 1087–1100, 2002.

[32] Lam, L. and Suen, C. Y., Optimal combinations of pattern classifiers, Pattern Recognition Letters, Vol. 16, No. 9, pp. 945–954, 1995.

[33] Lee, S., Wolberg, G. and Shin, S. Y., Scattered data interpolation with multilevel B-splines, IEEE Transactions on Visualization and Computer Graphics, Vol. 3, No. 3, pp. 228–244, 1997.

[34] Lester, H. and Arridge, S. R., A survey of non-linear medical image registration, Pattern Recognition, Vol. 32, No. 1, pp. 129–149, 1999.

[35] Lester, H., Arridge, S. R., Jansons, K. M., Lemieux, L., Hajnal, J. V. and Oatridge, A., Non-linear registration with the variable viscosity fluid algorithm, in Kuba, A., Samal, M. and Todd-Pokvopek, A., eds., *Information Processing in Medical Imaging: 16th International Conference, IPMI'99, Visegrád, Hungary, June/July 1999. Proceedings*, Vol. 1613, pp. 238–251, Springer-Verlag, Heidelberg, 1999.

[36] Li, B., Christensen, G. E., Hoffman, E. A., McLennan, G. and Reinhardt, J. M., Establishing a normative atlas of the human lung: Intersubject warping and registration of volumetric CT images, Academic Radiology, Vol. 10, No. 3, pp. 255–265, 2003.

[37] Lorenz, C. and Krahnstöver, N., Generation of point-based 3D statistical shape models for anatomical objects, Computer Vision and Image Understanding, Vol. 77, pp. 175–191, 2000.

[38] Lorenzo-Valdés, M., Sanchez-Ortiz, G. I., Mohiaddin, R. and Rueckert, D., Atlas-based segmentation and tracking of 3D cardiac MR images using non-rigid registration, in Dohi, T. and Kikinis, R., eds., *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2002: 5th International Conference, Tokyo, Japan, September 25-28, 2002, Proceedings, Part I*, Vol. 2488 of *Lecture Notes in Computer Science*, pp. 642–650, Springer-Verlag, Heidelberg, 2002.

[39] Maes, F., Collignon, A., Vandermeulen, D., Marchal, G. and Suetens, P., Multimodality image registration by maximisation of mutual information, IEEE Transactions on Medical Imaging, Vol. 16, No. 2, pp. 187–198, 1997.

[40] Malladi, R., Sethian, J. A. and Vemuri, B. C., Shape modelling with front propagation: A level set approach, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 17, No. 2, pp. 158–175, 1995.

[41] Maurer, Jr., C. R., Hill, D. L. G., Martin, A. J., Liu, H., McCue, M., Rueckert, D., Lloret, D., Hall, W. A., Maxwell, R. E., Hawkes, D. J. and Truwit, C. L., Investigation of intraoperative brain deformation using a 1.5 Tesla interventional MR system: Preliminary results, IEEE Transactions on Medical Imaging, Vol. 17, No. 5, pp. 817–825, 1998.

[42] Miller, M. I., Christensen, G. E., Amit, Y. and Grenander, U., Mathematical textbook of deformable neuroanatomies, Proceedings of the National Academy of Sciences of the U.S.A., Vol. 90, No. 24, pp. 11944–11948, 1993.

[43] Mobbs, P. G., Brain structure, in Kerkut, G. A. and Gilbert, L. I., eds., *Comprehensive insect physiology biochemistry and pharmacology, Vol. 5: Nervous system: structure and motor function*, pp. 299–370, Pergamon Press, Oxford, New York, Toronto, Sydney, Paris, Frankfurt, 1985.

[44] Mortensen, E. N. and Barrett, W. A., Interactive segmentation with intelligent scissors, Graphical Models and Image Processing, Vol. 60, No. 5, pp. 349–384, 1998.

[45] Musse, O., Heitz, F. and Armspach, J.-P., Fast deformable matching of 3D images over multiscale nested subspaces. application to atlas-based MRI segmentation, Pattern Recognition, Vol. 36, No. 8, pp. 1881–1899, 2003.

[46] Paclik, P., Duin, R. P. W., van Kempen, G. M. P. and Kohlus, R., Segmentation of multi-spectral images using the combined classifier approach, Image and Vision Computing, Vol. 21, No. 6, pp. 473–482, 2003.

[47] Park, H., Bland, P. H. and Meyer, C. R., Construction of an abdominal probabilistic atlas and its application in segmentation, IEEE Transactions on Medical Imaging, Vol. 22, No. 4, pp. 483–492, 2003.

[48] Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P., *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge University Press, Cambridge, UK, 2 edition, 1992.

[49] Rao, A., Sanchez-Ortiz, G. I., Chandrashekara, R., Lorenzo-Valdés, M., Mohiaddin, R. and Rueckert, D., Construction of a cardiac motion atlas from MR using non-rigid registration, in Magnin, I. E., Montagnat, J., Clarysse, P., Nenonen, J. and Katila, T., eds., *Functional Imaging and Modeling of the Heart – Second International Workshop, FIMH 2003, Lyon, France, June 5-6, 2003, Proceedings*, Vol. 2674 of *Lecture Notes in Computer Science*, pp. 141–150, Springer-Verlag, Heidelberg, 2003.

[50] Reichmuth, C., Becker, S., Benz, M., Reisch, D., Heimbeck, G., Hofbauer, A., Klagges, B. R. E., Pflugfelder, G. O. and Buchner, E., The sap47 gene of Drosophila melanogaster codes for a novel conserved neuronal protein associated with synaptic terminals, Molecular Brain Research, Vol. 32, pp. 45–54, 1995.

[51] Rogova, G. L. and Stomper, P. C., Information fusion approach to microcalcification characterization, Information Fusion, Vol. 3, No. 2, pp. 91–102, 2002.

[52] Rohlfing, T., *Multimodale Datenfusion für die bildgesteuerte Neurochirurgie und Strahlentherapie*, Ph.D. thesis, Technische Universität Berlin, 2000.

[53] Rohlfing, T., Efficient voxel lookup in non-uniformly spaced images using virtual uniform axes, in Sonka, M. and Hanson, K. M., eds., *Medical Imaging: Image Processing*, Vol. 4322 of *Proceedings of the SPIE*, pp. 986–994, 2001.

[54] Rohlfing, T., Incremental method for computing the intersection of discretely sampled m-dimensional images with n-dimensional boundaries, in Sonka, M. and Fitzpatrick, J. M., eds., *Medical Imaging: Image Processing*, Vol. 5032 of *Proceedings of the SPIE*, pp. 1346–1354, 2003.

[55] Rohlfing, T., Brandt, R., Maurer, Jr., C. R. and Menzel, R., Bee brains, B-splines and computational democracy: Generating an average shape atlas, in Staib, L., ed., *IEEE Workshop on Mathematical Methods in Biomedical Image Analysis*, pp. 187–194, IEEE Computer Society, Los Alamitos, CA, Kauai, HI, 2001.

[56] Rohlfing, T., Brandt, R., Menzel, R. and Maurer, Jr., C. R., Segmentation of three-dimensional images using non-rigid registration: Methods and validation with application to confocal microscopy images of bee brains, in Sonka, M. and Fitzpatrick, J. M., eds., *Medical Imaging: Image Processing*, Vol. 5032 of *Proceedings of the SPIE*, pp. 363–374, 2003.

[57] Rohlfing, T., Brandt, R., Menzel, R. and Maurer, Jr., C. R., Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains, NeuroImage, Vol. 21, No. 4, pp. 1428–1442, 2004.

[58] Rohlfing, T., Maurer, C. R., Bluemke, D. A. and Jacobs, M. A., An alternating-constraints algorithm for volume-preserving non-rigid registration of contrast-enhanced MR breast images, in Gee, J. C., Maintz, J. B. A. and Vannier, M. W., eds., *Biomedical Image Registration – Second International Workshop, WBIR 2003, Philadelphia, PA, USA, June 23-24, 2003*, Vol. 2717 of *Lecture Notes in Computer Science*, pp. 291–300 , Springer-Verlag, Berlin Heidelberg, 2003.

[59] Rohlfing, T. and Maurer, Jr., C. R., Intensity-based non-rigid registration using adaptive multilevel free-form deformation with an incompressibility constraint, in Niessen, W. and Viergever, M. A., eds., *Proceedings of Fourth International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2001)*, Vol. 2208 of *Lecture Notes in Computer Science*, pp. 111–119, Springer-Verlag, Berlin, 2001.

[60] Rohlfing, T. and Maurer, Jr., C. R., Non-rigid image registration in shared-memory multiprocessor environments with application to brains, breasts, and bees, IEEE Transactions on Information Technology in Biomedicine, Vol. 7, No. 1, pp. 16–25, 2003.

[61] Rohlfing, T., Maurer, Jr., C. R., Bluemke, D. A. and Jacobs, M. A., Volume-preserving nonrigid registration of MR breast images using free-form deformation with an incompressibility constraint, IEEE Transactions on Medical Imaging, Vol. 22, No. 6, pp. 730–741, 2003.

[62] Rohlfing, T., Maurer, Jr., C. R., O'Dell, W. G. and Zhong, J., Modeling liver motion and deformation during the respiratory cycle using intensity-based free-form registration of gated MR images, in Mun, S. K., ed., *Medical Imaging: Visualization, Display, and Image-Guided Procedures*, Vol. 4319 of *Proceedings of the SPIE*, pp. 337–348, 2001.

[63] Rohlfing, T., Maurer, Jr., C. R., O'Dell, W. G. and Zhong, J., Modeling liver motion and deformation during the respiratory cycle using intensity-based free-form registration of gated MR images, Medical Physics, in print.

[64] Rohlfing, T., Russakoff, D. B. and Maurer, Jr., C. R., Performance-based classifier combination in atlas-based image segmentation using expectation-maximization parameter estimation, IEEE Transactions on Medical Imaging, Vol. 23, No. 8, pp. 983–994, 2004.

[65] Rohlfing, T., Russakoff, D. B. and Maurer, Jr., C. R., Extraction and application of expert priors to combine multiple segmentations of human brain tissue, in Ellis, R. E. and Peters, T. M., eds., *Proceedings of Sixth International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Lecture Notes in Computer Science, pp. 587–585, Springer-Verlag, Berlin Heidelberg, 2003.

[66] Rohlfing, T., West, J. B., Beier, J., Liebig, T., Taschner, C. A. and Thomale, U.-W., Registration of functional and anatomical MRI: Accuracy assessment and application in navigated neurosurgery, Computer Aided Surgery, Vol. 5, No. 6, pp. 414–425, 2000.

[67] Rueckert, D., Nonrigid registration: Concepts, algorithms, and applications, in Hajnal, J. V., Hill, D. L. G. and Hawkes, D. J., eds., *Medical Image Registration*, pp. 281–301, CRC Press, Boca Raton, FL, 2001.

[68] Rueckert, D., Frangi, A. F. and Schnabel, J. A., Automatic construction of 3-D statistical deformation models of the brain using nonrigid registration, IEEE Transactions on Medical Imaging, Vol. 22, No. 8, pp. 1014–1025, 2003.

[69] Rueckert, D., Sonoda, L. I., Hayes, C., Hill, D. L. G., Leach, M. O. and Hawkes, D. J., Nonrigid registration using free-form deformations: Application to breast MR images, IEEE Transactions on Medical Imaging, Vol. 18, No. 8, pp. 712–721, 1999.

[70] Saranli, A. and Demirekler, M., A statistical unified framework for rank-based multiple classifier decision combination, Pattern Recognition, Vol. 34, No. 4, pp. 865–884, 2001.

[71] Sarti, A., de Solórzano, C. O., Locket, S. and Malladi, R., A geometric model for 3-D confocal image analysis, IEEE Transactions on Biomedical Engineering, Vol. 47, No. 12, pp. 1600–1609, 2000.

[72] Schnabel, J. A., Rueckert, D., Quist, M., Blackall, J. M., Castellano-Smith, A. D., Hartkens, T., Penney, G. P., Hall, W. A., Liu, H., Truwit, C. L., Gerritsen, F. A., Hill, D. L. G. and Hawkes, D. J., A generic framework for non-rigid registration based on non-uniform multi-level free-form deformations, in Niessen, W. and Viergever, M. A., eds., *Proceedings of Fourth International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2001)*, Vol. 2208 of *Lecture Notes in Computer Science*, pp. 573–581, Springer-Verlag, Berlin, 2001.

[73] Sederberg, T. W. and Parry, S. R., Free-form deformation and solid geometric models, Computer Graphics, Vol. 20, No. 4, pp. 151–160, 1986.

[74] Stevens, J. K., Mills, L. R. and Trogadis, J. E., eds., *Three-Dimensional Confocal Microscopy: Volume Investigation of Biological Specimens*, Academic Press, London, 1994.

[75] Studholme, C., Constable, R. T. and Duncan, J. S., Accurate alignment of functional EPI data to anatomical MRI using a physics-based distortion model, IEEE Transactions on Medical Imaging, Vol. 19, No. 11, pp. 1115–1127, 2000.

[76] Studholme, C., Hill, D. L. G. and Hawkes, D. J., Automated three-dimensional registration of magnetic resonance and positron emission tomography brain images by multiresolution optimization of voxel similarity measures, Medical Physics, Vol. 24, No. 1, pp. 25–35, 1997.

[77] Studholme, C., Hill, D. L. G. and Hawkes, D. J., An overlap invariant entropy measure of 3D medical image alignment, Pattern Recognition, Vol. 32, No. 1, pp. 71–86, 1999.

[78] Thirion, J.-P., Image matching as a diffusion process: An analogy with Maxwell's demons, Medical Image Analysis, Vol. 2, No. 3, pp. 243–260, 1998.

[79] Thompson, P. M. and Toga, A. W., Detection, visualization and animation of abnormal anatomic structure with a deformable probabilistic brain atlas based on random vector field transformations, Medical Image Analysis, Vol. 1, No. 4, pp. 271–294, 1997.

[80] Tsai, A., Wells, W., Tempany, C., Grimson, E. and Willsky, A., Coupled multi-shape model and mutual information for medical image segmentation, in Taylor, C. and Noble, J. A., eds., *Information Processing in Medical Imaging*, Vol. 2732 of *Lecture Notes in Computer Science*, pp. 185–197, Springer-Verlag, Berlin Heidelberg, 2003, 18th International Conference, IPMI 2003, Ambleside, UK, July 2003.

[81] Tsai, A., Yezzi, Jr., A., Wells, W., Tempany, C., Tucker, D., Fan, A., Grimson, W. E. and Willsky, A., A shape-based approach to the segmentation of medical imagery using level sets, IEEE Transactions on Medical Imaging, Vol. 22, No. 2, pp. 137–154, 2003.

[82] Vannier, M. W., Pilgram, T. K., Speidel, C. M., Neumann, L. R., Rickman, D. L. and Schertz, L. D., Validation of magnetic resonance imaging (MRI) multispec-tral tissue classification, Computerized Medical Imaging and Graphics, Vol. 15, No. 4, pp. 217–223, 1991.

[83] Viola, P. A., Alignment by maximization of mutual information, International Journal of Computer Vision, Vol. 24, No. 2, pp. 137–154, 1997.

[84] Wahba, G., *Spline Models for Observational Data*, Vol. 59 of *CBMS-NSF Regional Conference Series*, SIAM, 1990.

[85] Warfield, S. K., Zou, K. H. and Wells, W. M., Validation of image segmenta-tion and expert quality with an expectation-maximization algorithm, in Dohi, T. and Kikinis, R., eds., *Proceedings of Fifth International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), Part I*, Vol. 2488 of *Lecture Notes in Computer Science*, pp. 298–306, Springer-Verlag, Berlin Heidelberg, 2002.

[86] Wells, W. M., Viola, P. A., Atsumi, H., Nakajima, S. and Kikinis, R., Multi-modal volume registration by maximization of mutual information, Medical Image Analysis, Vol. 1, No. 1, pp. 35–51, 1996.

[87] Wells, III., W. M., Grimson, W. E. L., Kikinis, R. and Jolesz, F. A., Adaptive segmentation of MRI data, IEEE Transactions on Medical Imaging, Vol. 15, No. 4, pp. 429–442, 1996.

[88] West, J. B., Fitzpatrick, J. M., Wang, M. Y., Dawant, B. M., Maurer, Jr., C. R., Kessler, R. M., Maciunas, R. J., Barillot, C., Lemoine, D., Collignon, A., Maes, F., Suetens, P., Vandermeulen, D., van den Elsen, P. A., Napel, S., Sumanaweera, T. S., Harkness, B., Hemler, P. F., Hill, D. L. G., Hawkes, D. J., Studholme, C., Maintz, J. B. A., Viergever, M. A., Malandain, G., Pennec, X., Noz, M. E., Maguire, Jr., G. Q., Pollack, M., Pelizzari, C. A., Robb, R. A., Hanson, D. and Woods, R. P., Comparison and evaluation of retrospective intermodality brain image registration techniques, Journal of Computer Assisted Tomography, Vol. 21, No. 4, pp. 554–566, 1997.

[89] Xu, L., Krzyzak, A. and Suen, C. Y., Methods of combining multiple classifiers and their applications to handwriting recognition, IEEE Transactions on Systems, Man and Cybernetics, Vol. 22, No. 3, pp. 418–435, 1992.

[90] Yang, J., Staib, L. H. and Duncan, J. S., Neighbor-constrained segmentation with a 3D deformable model, in Taylor, C. and Noble, J. A., eds., *Information Processing in Medical Imaging*, Vol. 2732 of *Lecture Notes in Computer Science*, pp. 198–209, Springer-Verlag, Berlin Heidelberg, 2003, 18th International Conference, IPMI 2003, Ambleside, UK, July 2003.

[91] Yezzi, Jr., A., Kichenassamy, S., Kumar, A., Olver, P. and Tannenbaum, A., A geometric snake model for segmentation of medical imagery, IEEE Transactions on Medical Imaging, Vol. 16, No. 2, pp. 199–209, 1997.

[92] Zeng, X., Staib, L. H., Schultz, R. T. and Duncan, J. S., Segmentation and measurement of the cortex from 3-D MR images using coupled-surfaces propagation, IEEE Transactions on Medical Imaging, Vol. 18, No. 10, pp. 927–927, 1999.

[93] Zijdenbos, A. P., Dawant, B. M., Margolin, R. A. and Palmer, A. C., Morphometric analysis of white matter lesions in MR images: Method and validation, IEEE Transactions on Medical Imaging, Vol. 13, No. 4, pp. 716–724, 1994.

[94] Zuschratter, W., Steffen, T., Braun, K., Herzog, A., Michaelis, B. and Scheich, H., Acquisition of multiple image stacks with a confocal laser scanning microscope, in *Proceedings of Threedimensional and Multidimensional Image Acquisition and Processing V*, Vol. 3261, pp. 177–186, Proceedings of SPIE, 1998.