

Piecewise affine initialized spline-based patient-specific registration of a high-resolution ear model for surgical guidance

Michel A. Audette¹, Rupert Brooks², Robert Funnell³,
Gero Strauss⁴, and Tal Arbel²

¹ ICCAS, University of Leipzig, Germany michel.audette@iccas.de

² Center for Intelligent Machines, McGill University, Montreal, Canada

³ AudiLab, McGill University, Montreal, Canada

⁴ University Hospital Leipzig, Germany

Abstract. Image guidance of ear surgery would enable an ENT surgeon to navigate about the components of the middle and inner ear, but the elaboration of anatomical models for this application is limited by the resolution of CT and its inability to distinguish among soft tissues. As a result, it is impossible to identify manually some tissues in clinical data, while visible tissues can only be identified with significant overhead.

We propose a method for producing patient-specific description of the middle and inner ear on the basis of the minimally supervised registration of a high resolution model elaborated from micro-MR to patient CT, where the transformation among the model and the patient data is determined in a component-wise coarse-to-fine strategy. The first two stages feature a rough alignment on the basis of a few homologous point pairs, followed by a refinement based on a global affine transformation determined by mutual information. The middle stage involves a piecewise affine registration where each local affine transformation is given the global transformation as a starting point and is determined by mutual information over an appropriate anatomical mask. The final registration of each component is produced by mutual information-based thin-plate splines, whose anchor points overlap the affine-transformed mask.

1 Introduction

Image guidance of ear surgery would enable an ENT surgeon to navigate about the components of the middle and inner ear, and in particular avoid critical tissues such as the facial nerve, but the elaboration of anatomical models for this application is limited by the resolution of CT and its inability to distinguish among soft tissues. As a result, the tissues that are visible can only be identified with significant overhead, and the descriptiveness of the resulting models is limited by the relatively coarse voxel sampling, even with state-of-the-art clinical CT, in relation to the scale of the components of the anatomy. Moreover, it is impossible to identify manually some of the relevant tissues in routine clinical data, such as the chorda tympani nerve.

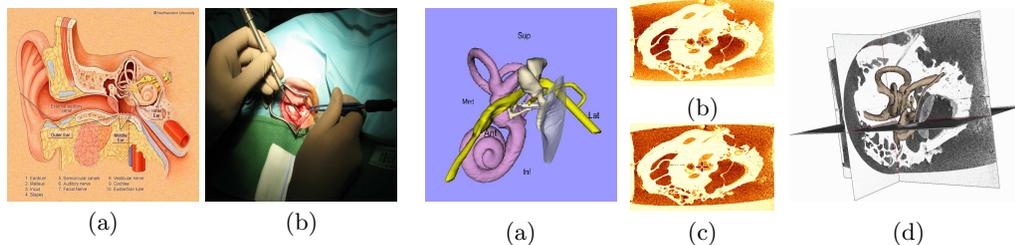


Fig. 1. Illustration of ear anatomy and surgery: (a) outer, middle and inner ear anatomy, reproduced with permission from T.C. Hain [1];(b) post-auricular incision.

Fig. 2. High resolution ear model derived from micro-MR data: (a) model as it appears in interactive web-site; micro-MR data: (b) raw data and (c) intensity non-uniformity corrected data; (d) orthogonal planes depiction of micro-MR data overlaid with ear model.

Ear surgery typically begins with a post-auricular incision, as shown in figure 1, which may lead to the repair of a tympanic membrane, the replacement of ossicular bones by a prosthesis, the resection of a choleostoma, or a combination of these interventions. A related intervention involves drilling the mastoid bone, behind the ear, and resecting a choleostoma present in it.

The presence of pathology complicates the application of statistical shape models [2] in patient-specific segmentation, as the notion of an average shape is compromised by the random nature of tumour. Furthermore, existing registration methods in use with anatomical models, typically featuring a mutual information similarity measure and global affine-initialized spline-based transformation, do not apply readily to an anatomy that features many components, ideally transforming independently from each other, in contrast with brain or breast registration. To further complicate matters, some components, such as ossicular bones, may in fact be missing, as a result of a previous operation.

In spite of the limited resolution, segmentations of clinical CT scans have been used to visualize and model the ear. For example, Seeman et al. [3] demonstrated manual segmentation of middle and inner ear structures from high resolution CT, and suggested a combination of surface rendering of soft tissues and volume rendering of bone. Most use of clinical scans for the ear has involved manual or simple threshold-based segmentation, although Xianfen [4] used a combination of manual and 3D level-set segmentation on CT data.

Non-clinical imaging modalities such as histology, micro-CT and high-field micro-MR imaging permit much better identification of small middle and inner ear structures and have been used by a number of groups for generic anatomical modeling. Folowosele et al. [5] demonstrated the descriptive quality of high-field magnetic resonance imaging in producing models of the middle and inner ear. One of us published results [6] of expert segmentation of high-field micro-MR data [7], and it is the model refined from the triangulated boundaries of this segmentation that is applied clinically in this paper.

We propose a method for producing a patient-specific description of the middle and inner ear on the basis of the minimally supervised registration of a high resolution model elaborated from micro-MR to patient CT, where the transformation between the model and the patient data is determined in a component-wise coarse-to-fine strategy. To our knowledge, this paper is the first to advocate this approach to addressing the clinical requirement of ENT surgeons for reliable patient-specific anatomical description of the ear for surgical planning. Moreover, the middle resolution of our coarse-to-fine strategy coincides with an anatomically motivated piecewise affine transformation, and ours is one of few methods that uses such a computation stage prior to a freeform transformation.

Our main assumption is that CT data of currently available image resolution ($.2mm \times .2mm \times .3mm$), whose manual segmentation comes at the cost of considerable overhead and is challenged by fine structures such as the chorda tympani, contain sufficient image information to determine a registration with a descriptive micron-scale model, especially if the transformation is modeled appropriately for the anatomy. Another important assumption is that this registration process can be initialized on the basis of minimal supervision, namely through a few homologous point pairs, while also taking into account user-provided anatomical information, such as the absence of a relevant anatomical component. Lastly, our perspective towards the usefulness of the piecewise affine transformation is inspired by Frangi [8], who modeled global motion of the left and right ventricles of the heart separately in the freeform estimation of cardiac motion, and by Pitiot [9], who viewed the deformation between cryosection slices as piecewise affine in the reconstruction of the brain.

This paper is divided into the following sections. In the Methods section, we first provide details about the high resolution model of the middle and inner ear, and the micro-MR data from which it was obtained. We describe the minimally supervised process of registering the micro-MR data to patient CT, and applying this transformation to the ear model. We validate the registration and illustrate a co-registered CT and model in the Results section, and finally, discuss the results and future directions in the Conclusions section.

2 Methods

2.1 Application of a micron-scale model of the middle and inner ear

The ear model that is used in our research features triangulated boundaries of the components of the left middle and inner ear of an individual, and has been manually segmented by ear researchers from micro-MR data of $78.125 \mu m$ isotropic voxel. The model can be seen in figure 2 (a), and the axial plane of the raw micro-MR data is seen in (b). As can be seen in the figure, these data exhibit considerable intensity inhomogeneity due to magnetic susceptibility, and this issue was addressed with automatic inhomogeneity correction software [10], to improve the correlation between voxel intensity and underlying tissue, and to make the image-based registration more reliable later on.

Our philosophy is to require only limited user supervision in initializing the search for the appropriate transformation. The objective of our method is to determine the transformation $\mathbf{T}(x, y, z)$ between the generic, micro-MR model data and the patient CT data:

$$\mathbf{x}_P = \mathbf{T}(\mathbf{x}_M) \mathbf{x}_M, \quad (1)$$

where \mathbf{x}_P and \mathbf{x}_M express patient and model coordinates respectively, and $\mathbf{T}(\mathbf{x}_M)$ describes a transformation that may vary according to model coordinates. This transformation can then be applied to the high quality middle and inner ear model derived from the micro-MR data.

Generally, registration can be viewed in terms of 3 stages: choice of transformation \mathbf{T} , choice of similarity criterion, and optimization over some parameter space for the best \mathbf{T} that fulfills some objective function based on similarity. The method of choice for multi-modality registration is mutual information maximization [11], as it is based on the most general assumption between the two modalities, i.e.: that modality A is predictive of modality B in the information-theoretic sense, without there necessarily existing a functional relationship between the two. Methods of this kind define the mutual information between two images A and B as:

$$\mathcal{I}(A, B) = \sum_a \sum_b p(a, b) \log \frac{p(a, b)}{p(a)p(b)}, \quad (2)$$

where $p(a, b)$ is the intensity joint probability distribution of the images and $p(a)$ and $p(b)$ are the corresponding marginal distributions. The transformation that is sought is the one that maximizes this measure of image similarity, over the space of admissible transformations.

The choice of transformation must reflect realistic assumptions about the nature of the spatial relationship between the two sets of data. Given that the anatomy that we are registering is composed of several parts that naturally transform component-wise across individuals, it makes sense to view the transformation as a generalization of a piecewise affine transform, i.e.: a piecewise affine registration followed by a spline-based freeform transformation. The main difficulty is representing the transformation in a manner that is *sufficiently descriptive to be applied across patients, while limiting the possibility of the optimization becoming mired in local optima associated with a free-form transformation.*

One must bear in mind that the number of degrees of freedom of a freeform transformation can easily dwarf that of a piecewise affine transformation required over a volume like this one, especially if one were to fit a regular grid over the volume spanned by the middle and inner ear with the density required of such a descriptive mapping. In our case, the piecewise affine mapping is determined by 7 parameters: rigid transformation plus one scale, also known as Procrustes transformation, per component, times the number of components. The Procrustes registration is chosen over other options, namely 9 or 12 parameters, because it produces numerically stable results [12], so as to provide a reliable initialization to the last stage. Moreover a *global affine transformation*

provides no initialization to finding spline coefficients in a manner that it does to a piecewise affine transformation, so that if a piecewise affine assumption is appropriate for the spatial correspondence, then this middle stage is more consistent with a coarse-to-fine approach to such an inverse problem, as used in stereo vision for example [13], than proceeding directly from a coarse level coinciding with the global affine to a spline-based fine level.

Our basic approach is then:

1. A global rigid registration based on 5 or more homologous pairs [12]:

$$\mathbf{T}_{GR,HP}(\mathbf{x}) \equiv \mathbf{R}_{HP}\mathbf{x} + \mathbf{t}_{HP}, \quad (3)$$

where \mathbf{R}_{HP} and \mathbf{t}_{HP} are a rotation matrix and a translation vector determined by homologous point pairs, and points are chosen for their identifiability and for spanning a sufficiently large volume that subsumes the anatomy.

2. A global affine registration based on mutual information maximization:

$$\mathbf{T}_{GA,MI}(\mathbf{x}) \equiv s_{MI}\mathbf{R}_{MI}\mathbf{x} + \mathbf{t}_{MI}, \quad (4)$$

whose starting point is the result of step 1, and where s is a scale factor.

3. A component-wise piecewise affine registration, also based on mutual information, each of whose component transformation is initialized by the result of step 2:

$$\mathbf{T}_{PA,MI}(\mathbf{x}) \equiv \{s_{MI,k}\mathbf{R}_{MI,k}\mathbf{x} + \mathbf{t}_{MI,k} \mid k=\{TM, FN, CT, VO, O\}\}, \quad (5)$$

where k designates an anatomical component: tympanic membrane (TM), facial nerve (FN), chorda tympani (CT), vestibular organ (VO), ossicles (O).

4. A thin-plate spline-based mutual information registration $\mathbf{T}_{TPS,MI}(\mathbf{x})$, which captures the remaining deformation between each affine-transformed component and the corresponding component in the patient data set:

$$\mathbf{T}_{final}(\mathbf{x}) \equiv \mathbf{T}_{PA,MI}(\mathbf{x}) + \mathbf{T}_{TPS,MI}(\mathbf{x}). \quad (6)$$

2.2 Registration: initialization by homologous pairs and mutual information-based global affine and piecewise affine stages

Some level of supervision is necessary to provide the registration with *knowledge of which side of the head is involved, left or right*, so as to first apply a reflection about the $y - z$ plane to our high resolution model of a left ear if confronted with a right ear procedure. Requiring the user to provide the system with a small number of homologous points, previously identified on the model, as seen in figure 3, is consistent with our definition of minimal user supervision.

The homologous points seen here coincide with the points of attachment of malleus to the tympanic membrane, of the incus to the malleus, and of the malleus to the stapes, as well as the centre of the cochlea, and the meeting point of the semi-circular canals on the vestibular organ. If ossicles are missing

due to a previous resection, landmark-based initialization is still possible, but would require other recognizable, geometrically well defined points in the middle ear. The homologous pairs determine a well-tested rigid registration based on singular value decomposition [12]. If there are doubts about the validity of any pair, the computation of the transformation by robust statistics is also feasible.

Furthermore, we are also planning to incorporate into our interface provision for additional user interaction, namely indicating the absence of relevant anatomy where previously resected and providing seed points for segmenting a choleostoma, which after all cannot be predicted by statistical model.

The global mutual information registration method is as proposed by Maes [14], and determines the optimal 7 parameters coinciding with rotation, translation and scale. The search for optimal \mathbf{R} , s , \mathbf{t} is given the starting point of the transformation $\mathbf{T}_{GR,HP}(\mathbf{x})$ in expression 3.

The piecewise affine method is given the starting point of $\mathbf{T}_{GA,MI}(\mathbf{x})$ in expression 4, and makes use of anatomical masks computed from the triangulated surfaces of each component, stored as Stereo Lithography (.stl) file, in particular the vertices and normal orientation of each face. We estimate the surface normals at the vertices by averaging the normals of the faces coincident with each vertex. We then bucket-sort the vertices, storing the position and normal of each.

Thereafter we label as *inside or outside the surface* every voxel tested in the model image, within the $x - y - z$ span of triangulated boundary of the anatomical component, by looking at the sign of the projection of the vector from each nearby boundary vertex $\mathbf{v}_{M,i}$ to the position (world coordinates) $\mathbf{x}_{M,j}$ of that voxel. For a vertex with an outward normal, the vector $\mathbf{x}_{M,j} - \mathbf{v}_{M,i}$ from the vertex to a contained point should have a projection on its outward normal that is negative, and we seek to enforce this negative projection as a consensus over a small neighbourhood of the vertices nearest to each tested point.

Moreover, for these masks to be useful in restricting the search for their respective optimal local transformation, they must be dilated from the set of voxels strictly contained by each anatomical boundary. The dilation must be sufficient

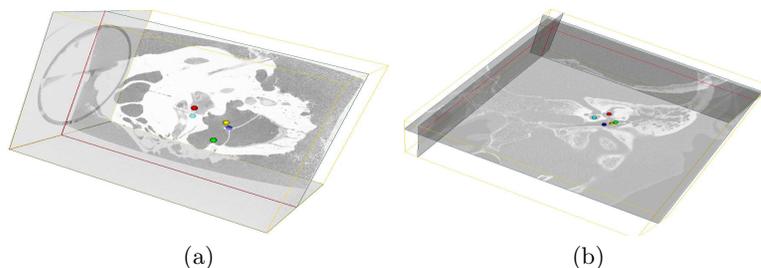


Fig. 3. Illustration of homologous points used to provide a rough initial rigid transformation: selected in (a) microCT data, and (b) in patient CT data, overlaid on semi-transparent triplanar views of the corresponding data. These points are chosen for their identifiability as landmarks.

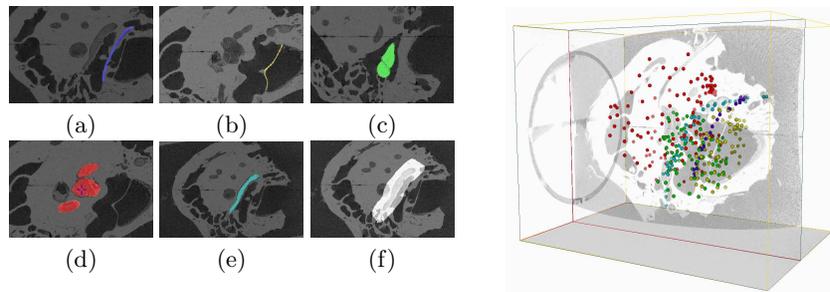


Fig. 4. Anatomical masks for piecewise affine registration, overlaid on micro-MR data (axial view): (a) chorda tympani, (b) tympanic membrane, (c) plate spline anchor points, in ossicles), (d) vestibular organ, and (e) facial nerve. (f) Dilated mask from facial nerve in (e). **Fig. 5.** Visualization of thin-plate spline model space, with colour code as in figure 4.

to contain the equivalent structure in the patient data putatively transformed by the global affine transformation. Figure 4 illustrates some masks coinciding with contained voxels and their dilated results. 24 iterations of a structuring element of 6 voxels is applied to the exact anatomical masks is used to produce the dilated masks.

2.3 Thin-plate spline registration based on mutual information

The last registration stage is a parametric approach, where the mutual information cost function is implemented using the efficient Parzen-windowed formulation of Thevenaz and Unser [15]. The deformation is modeled using a thin plate (TP) spline [16, 17] defined over a set of landmark points, as seen in figure 5. These points are chosen manually to span the structures of interest in the model well. This is done once on the model, and not on patient data: it does not factor into the supervision required of the user in the future.

Thin plate splines can suffer from the problem of creating non-diffeomorphic warps, that is folding or tearing of the space being deformed. Folds or rips occur when a part of the mesh of landmark points is "turned inside-out" by the transformation. This problem is avoided heuristically by visually selecting points so that they do not form any long, needle-like tetrahedra – as these are more likely to be flipped. The spline is parameterized by the positions of these landmark points in the moving image [18]. The mutual information is optimized over these parameters using a quasi-Newton optimization approach [19].

A similar multimodal registration approach was described in [20], but this approach used a B-spline deformation model rather than a thin plate spline model. The thin plate spline model is justified for this application for a number of reasons. The structures of interest are quite complicated in shape, and a detailed deformation field is necessary to accurately register them. Since the B-spline model is defined on a regularly spaced grid, in order to have fine detail in one area, it is necessary to have fine detail everywhere. This would greatly

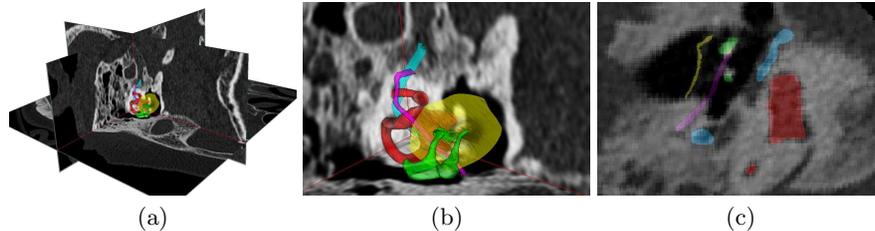


Fig. 6. Illustration of registered model overlaid on patient data set: (a) semi-transparent surface rendering of registered model over orthogonal plane of patient data; (b) zoom image of anatomy in (a); semi-transparent illustration of planar view, showing all 5 components on patient data.

increase the number of parameters, slowing the optimization. The additional control points would also make the B-spline more flexible, and additional regularization would be needed to keep it from generating spurious deformations.

3 Results

Figure 6 presents a visualization of preliminary results of the piecewise affine initialized TP spline registration of the model applied to a patient data set. In the 3D views, the semi-transparent surfaces can be seen overlapping locally dark or light well circumscribed areas of the CT data, especially in the zoomed image. For example, the vestibular surface, shown in red, can be seen overlapping well over dark voxel areas in the 3D-projected planar views of the patient CT. The ossicles' intersection with the orthogonal planes, in contrast, coincide with light areas. The 2D visualisation presents a similar picture, where in particular the dark areas under the vestibular organ and facial nerve are visible. However, the chorda tympani, in purple, is not actually visible in much of the CT data.

Figure 7 overlays registered results on manually labeled results, as visualized by semi-transparently rendered surfaces. The registered facial nerve is clearly longer than the manually labeled surface, due to the lack of image data to constrain the registration lengthwise, in contrast with other anatomical components that are well circumscribed everywhere. In part, this is due to lack of contrast with other soft tissues within the cranium. The chorda tympani was not visible enough in CT to enable a manual segmentation.

The sensitivity of the method is featured in table 1. It is difficult to define a fair measure of a true negative in this case, so we have not compiled specificity statistics. The registration method produces a somewhat lesser sensitivity than expected for the facial nerve and tympanic membrane, given qualitatively encouraging visual results. First, it remains to be seen how these errors compare to intra- and inter-rater variability, given the small scale of the structures in relation to the voxel size and sensitivity to choice of threshold for some tissues. Explanations for each error must be examined on a component-by-component basis. In the case of the facial nerve, errors tend to occur near the ends, where

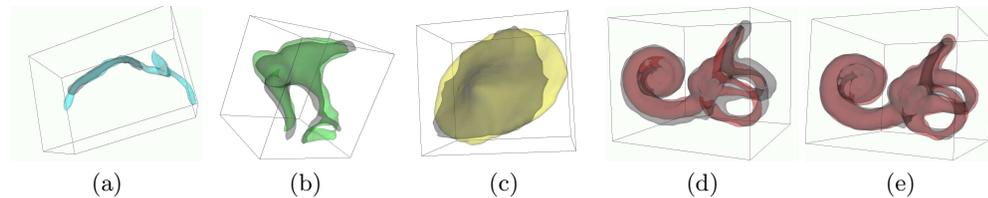


Fig. 7. Illustration of semi-transparent registered model surfaces, shown with surfaces of manually labeled tissues rendered in grey: (a) facial nerve; (b) ossicles; (c) tympanic membrane; (d)-(e) evolution of vestibular organ: (d) $\mathbf{T}_{GA,MI}$ and (e) \mathbf{T}_{final} .

component & stage	$\mathbf{T}_{GR,HP}$	$\mathbf{T}_{GA,MI}$	$\mathbf{T}_{PA,MI}$	\mathbf{T}_{final}
vestibular organ	54.6	54.9	59.8	86.7
ossicles	49.5	41.3	61.4	79.1
facial nerve	30.8	32.7	34.7	53.6
tympanic membrane	10.3	8.1	15.2	44.9

Table 1. Sensitivity values (%) for registration stages.

the local affine transformation is most challenged, for lack of image contrast in CT. In order to make the TP spline elastic enough to overcome this error, one would introduce spurious deformations in the intracranial area. Moreover, currently our method treats all ossicles as one component, but would benefit from registering the malleus, incus, and stapes individually. Last, the tympanic membrane features low intensity data. Here, the mutual information parameters can be better adapted for this small intensity range over all stages. In addition, a 9 or 12-parameter affine stage can be considered prior to the TP spline stage.

4 Discussion

This paper presented preliminary results of a method for applying a piecewise affine-initialized thin-plate spline transformation to a high-quality model of the middle and inner ear, towards the surgical guidance of ear surgery. We demonstrated a computationally stable coarse-to-fine approach where piecewise affine and local TP-spline stages coincided with significant improvements in the registration. The results indicate that with further refinements, minimally supervised model computation for the guidance of ear surgery is feasible.

Our method currently does not eliminate the possibility of overlap between registered components after the piecewise affine registration. In other words, this registration stage does not enforce diffeomorphism. It is however feasible to achieve such a mapping with a simple heuristic: the overlapping region between registered components can be ascertained, and the average of the competing affine-transformed positions could be assigned to each voxel in that area. Thereafter, each non-overlapping voxel would be given a new position that is a weighted sum of this average and its original transformation applied to its coordi-

nates, where the weight would vary with distance to the overlap. A best-fit affine transform could then be derived from these heuristically estimated positions .

References

1. Hain, T.: Information about dizziness, balance and hearing. <http://www.dizziness-and-balance.com/disorders/hearing/hearing.html>
2. Rueckert, D., et al.: Automatic construction of 3d statistical deformation models using non-rigid registration. In: *Med. Imag. Comp. & Comp. Ass. Interv.* (2001) 77–84
3. Seemann, M., et al.: Evaluation of the middle and inner ear structures: comparison of hybrid rendering, virtual endoscopy and axial 2d source images. *Eur. Radiol.* **9** (1999) 1851–1858
4. Xianfen, D., et al.: 3d semi-automatic segmentation of the cochlea and inner ear. In: *Conf. Proc. IEEE Eng. Med. Biol. Soc.* (2005) 6285–6288
5. Folowosele, F., et al.: 3d imaging and modeling of the middle and inner ear. In: *Proc. of SPIE, Medical Imaging: Visualization, Image-Guided Procedures and Display.* (2004) 508–515
6. Nicholson, D., et al.: A randomized controlled study of a computer-generated three-dimensional model for teaching ear anatomy. *Med. Educ.* **40** (2006) 1081–1087
7. UNC: Dept. cell & developmental biology: The vertebrate ear and temporal bone. http://cbaweb2.med.unc.edu/henson_mrm/
8. Frangi, A., et al.: Automatic construction of multiple-object three-dimensional statistical shape models: Application to cardiac modelling. *IEEE Trans. Med. Imaging* **21**(9) (2002) 1151–1166
9. Pitiot, A., et al.: Piecewise affine registration of biological images for volume reconstruction. *Medical Image Analysis* **10**(3) (June 2006) 465–483
10. Sled, J., et al.: A nonparametric method for automatic correction of intensity nonuniformity in mri data. *IEEE Trans. Med. Imaging* **17**(1) (1998) 87–97
11. Viola, P., Wells, W.: Alignment by maximization of mutual information. *Int. J. Computer Vision* **24**(2) (1997) 137–154
12. Sibson, R.: Studies in the robustness of multidimensional scaling: Procrustes statistics. *J. Roy. Statis. Soc. Ser. B* **40**(2) (1978) 234–238
13. Marr, D., Poggio, T.: A computational theory of human stereo vision. *Proc. Royal Society of London B* **204** (1979) 301–328
14. Maes, F., et al.: Multimodality image registration by maximization of mutual information. *IEEE Trans. Med. Imag.* **16**(2) (1997) 187–198
15. Thévenaz, P., Unser, M.: Optimization of mutual information for multiresolution image registration. *IEEE Trans. Image Proc.* **9**(12) (2000) 2083–2099
16. Bookstein, F.L.: Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Trans. Patt. Anal. & Mach. Intell.* **11**(6) (1989) 567–585
17. Davis, M.H., et al.: A physics-based coordinate transformation for 3-d image matching. *IEEE Trans. Med. Imag.* **16**(3) (1997) 317–328
18. Brooks, R., Arbel, T.: Improvements to the itk::KernelTransform and subclasses. *Insight Journal* (March 2007) DSpace handle <http://hdl.handle.net/1926/494>.
19. Zhu, C., et al.: L-BFGS-B: Algorithm 778: L-BFGS-B, FORTRAN routines for large scale bound constrained optimization. *ACM Transactions on Mathematical Software* **23**(4) (1997) 550–560
20. Mattes, D., et al.: Pet-ct image registration in the chest using free-form deformations. *IEEE Trans. Med. Imag.* **22**(1) (2003) 120–128

MICCAI Workshop on

Image Guidance and Computer Assistance for Soft-Tissue Interventions

New York, September 10, 2008

Edited by:
Philippe Cattin
Marco Feuerstein
Martin Groher
Theo van Walsum
Stefan Weber
Stephan Zidowitz

MICCAI Workshop on

Image Guidance and Computer Assistance for Soft-Tissue Interventions

New York, September 10, 2008

Edited by:

Philippe Cattin
Marco Feuerstein
Martin Groher
Theo van Walsum
Stefan Weber
Stephan Zidowitz