Evaluation of 3D Correspondence Methods for Model Building

Martin A. Styner[1], Kumar T. Rajamani[1], Lutz-Peter Nolte[1], Gabriel Zsemlye[2] and, Gábor Székely[2], Chris J. Taylor[3], and Rhodri H. Davies[3]

[1] M.E. Müller Institute for Surgical Technology and Biomechanics, University of Bern, P.O.Box 8354, 3001 Bern, Switzerland Martin.Styner@MEMcenter.unibe.ch
[2] Computer Vision Lab, Gloriastrasse 35, ETH-Zentrum, 8092 Zürich, Switzerland
[3] Division of Imaging Science and Biomedical Engineering, Stopford Building, Oxford Road, University of Manchester, Manchester, M13 9PT, UK *

Abstract. The correspondence problem is of high relevance in the construction and use of statistical models. Statistical models are used for a variety of medical application, e.g. segmentation, registration and shape analysis. In this paper, we present comparative studies in three anatomical structures of four different correspondence establishing methods. The goal in all of the presented studies is a model-based application. We have analyzed both the direct correspondence via manually selected landmarks as well as the properties of the model implied by the correspondences, in regard to compactness, generalization and specificity. The studied methods include a manually initialized subdivision surface (MSS) method and three automatic methods that optimize the object parameterization: SPHARM, MDL and the covariance determinant (DetCov) method. In all studies, DetCov and MDL showed very similar results. The model properties of DetCov and MDL were better than SPHARM and MSS. The results suggest that for modeling purposes the best of the studied correspondence method are MDL and DetCov.

1 Introduction

Statistical models of shape show considerable promise as a basis for segmenting, analyzing and interpreting anatomical objects from medical datasets [5, 14]. The basic idea in model building is to establish, from a training set, the pattern of legal variation in the shapes and spatial relationships of structures for a given class of images. Statistical analysis is used to give a parameterization of this variability, providing an appropriate representation of shape and allowing shape

* We are thankful to C. Brechbühler for the SPHARM software and to G. Gerig for support and insightful discussions. D. Jones and D. Weinberger at NIMH (Bethesda, MD) provided the MRI ventricle data. J. Lieberman and the neuro-image analysis lab at UNC Chapel Hill provided the ventricle segmentations. This research was partially funded by the Swiss National Centers of Competence in Research CO-ME (Computer assisted and image guided medical interventions). The femoral head datasets were provided within CO-ME by F. Langlotz.

constraints to be applied. A key step in building a model involves establishing a *dense* correspondence between shape boundaries over a reasonably large set of training images. It is important to establish the *correct* correspondences, otherwise an inefficient parameterization of shape will be determined. The importance of the correct correspondence is even more evident in shape analysis, as new knowledge and understanding related to diseases and normal development is extracted based on the established correspondence [10, 21]. Unfortunately there is no generally accepted definition for anatomically meaningful correspondence. It is thus difficult to judge the correctness of an established correspondence.

In 2D, correspondence is often established using manually determined landmarks [1], but this is a time-consuming, error-prone and subjective process. In principle, the method extends to 3D, but in practice, due to very small sets of reliably identifiable landmarks, manual landmarking becomes impractical. Most automated approaches posed the correspondence problem as that of defining a parameterization for each of objects in the training set, assuming correspondence between equivalently parameterized points. In this paper we compare methods introduced by Brechbühler[2], Kotcheff[16] and - [8]. A fourth method is based on manually initialized subdivision surfaces similar to Wang[24]. These methods are presented in more detail in sections 2.1-2.4. Similar approaches have also been proposed e.g. Hill[11] and Meier [18]. Christensen[4], Szeliski[22] and Rueckert[20] describe conceptionally different methods for warping the space in which the shapes are embedded. Models can then be built from the resulting deformation field [13, 9, 20]. Brett[3], Rangarajan[19] and Tagare[23] proposed shape features (e.g. regions of high curvature) to establish point correspondences.

In the remainder of the paper, we first present the studied correspondence methods and the measures representing the goodness of correspondence in order to compare the methods. In the result section we provide the qualitative and quantitative results of the methods applied on three populations of anatomical objects (left femoral head, left lateral ventricle and right lateral ventricle).

2 Methods

Alignment - As a prerequisite for any shape modeling, objects have to be normalized with respect to a reference coordinate frame. A normalization is needed to eliminate differences across objects that are due to rotation and translation. This normalization is achieved in studies based on the SPHARM correspondence (section 2.2) using the Procrustes alignment method without scaling. In the study based on the MSS correspondence (section 2.1) the alignment was achieved using manually selected anatomical landmarks. MDL and DetCov can align the object via direct pose optimization, an option not used in this paper.

Principal Component Analysis (PCA) model computation - A training population of n objects described by individual vectors x_i can be modeled by a multivariate Gaussian distribution. Principal Component Analysis (PCA) is performed to define axes that are aligned with the principal directions. First the mean vector $\bar{x}$ and the covariance matrix D are computed from the set of object

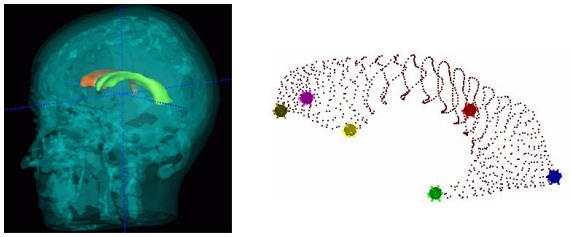


Fig. 1. Left: Visualization of the left and right lateral ventricle in a transparent human head. Right: The manually selected landmarks on the left ventricle template.

vectors(1). The sorted eigenvalues λ_i and eigenvectors p_i of the covariance matrix are the principal directions spanning a shape space with $\bar{x}$ at its origin. Objects x_j in the shape space are described as a linear combination of the eigenvectors on $\bar{x}$ (2). The shape space here is defined within $[-3 \cdot \sqrt{\lambda_i} \ldots 3 \cdot \sqrt{\lambda_i}]$.

$$\bar{x} = \frac{1}{n}\sum_{1}^{n} x_i;\ D = \frac{1}{n-1}\sum_{1}^{n}(x_i - \bar{x}) \cdot (x_i - \bar{x})^T \tag{1}$$

$$P = \{p_i\};\ D \cdot p_i = \lambda_i p_i;\ x_j = \bar{x} + P \cdot b \tag{2}$$

2.1 MSS: Manually initialized subdivision surfaces

This method is the only semi-automatic one, all others are fully automatic. The correspondence starts from a set of predefined anatomical landmarks and anatomically meaningful curves determined on the segmented objects using a interactive display (e.g. spline on the crista intertrochanterica). After a systematic discretization of the higher dimensional landmarks, a sparsely sampled point set results, which is triangulated in a standardized manner and further refined via subdivision surfaces. The correspondence on the 0^{th} level meshes is thus given by the manually placed control curves, on the subsequent levels by the subdivision rule: the triangles are split to four smaller ones, the new vertices are the midpoints of the pseudo-shortest path between the parent vertices. This path is the projection of the edges connecting in three-space the parent vertices to the original surface. The direction of the projection is determined by the normals of the neighboring triangles. This method was successfully applied on organs with a small numbers of anatomical point-landmarks.

2.2 SPHARM: Uniform area parameterization aligned to first order ellipsoid

The SPHARM description was introduced by Brechbühler[2] and is a parametric surface description that can only represent objects of spherical topology. The spherical parameterization is computed via optimizing an equal area mapping of

the 3D quadrilateral voxel mesh onto the sphere and minimizing angular distortions [2]. The basis functions of the parameterized surface are spherical harmonics. SPHARM can be used to express shape deformations [15], and is a smooth, fine-scale shape representation, given a sufficiently small approximation error. Based on an uniform icosahedron-subdivision of the spherical parameterization, we obtain a Point Distribution Model (PDM) directly from the coefficients via linear mapping [15]. The correspondence of SPHARM is determined by aligning the parameterization so that the ridges of the first order ellipsoid coincide. It is evident that the correspondence of objects with rotational symmetry in the first order ellipsoid is ambiguously defined.

2.3 DetCov: Determinant of the covariance matrix

Kotcheff et al [16] and later - [6] propose to use an optimization process that assigns the best correspondence across all objects of the training population, in contrast to MSS and SPHARM, which assign inherently a correspondence to each individual object. This view is based on the assumption that the correct correspondences are, by definition, those that build the optimal model given the training population. For that purpose they proposed to use the determinant of the covariance matrix as an objective function. The disadvantages of the original implementation was the computationally expensive genetic optimization algorithm, and the lack of a re-parameterization scheme. The implementation in this paper is different and is based on the optimization method by - [6], which efficiently optimizes the parameterization of the objects. This same optimization scheme was also used for the MDL criterion described in the next section. DetCov has the property to minimize the covariance matrix and so explicitly favors compact models.

2.4 MDL: Minimum Description Length

- [6, 8] built on the idea of the DetCov method, but proposed a different objective function for the optimization process using on the Minimum Description Length (MDL) principle. The DetCov criterion can be viewed as a simplification of the MDL criterion. The MDL principle is based on the idea of transmitting a dataset as an encoded message, where the code originates from some pre-arranged set of parametric statistical models. The full transmission then has to include not only the encoded data values, but also the coded model parameters. Thus MDL balances the model complexity, expressed in terms of the cost of transmitting the model parameters, against the quality of fit between the model and the data, expressed in terms of the coding length. The MDL objective function has similarities to the one used by DetCov [6]. The MDL computations for all our studies were initialized with the final position of the DetCov method.

2.5 Measures of Correspondence Quality

In this section we present the measures of the goodness of correspondence used in this paper. Such measures are quite difficult to define, since there is no gen-

eral agreement on a mathematical definition of correspondence. All methods in this paper produce correspondences that are fully continuous and have an inherent description of connectivity without any self-crossings. Measures of goodness evaluating a method's completeness and continuity (e.g. as suggested in Meier et al. [18]) are thus not applicable here.

We propose the use of four different measures, each biasing the analysis to its viewpoint on what constitutes correct correspondence. The first goodness measure is directly computed on the corresponding points as differences to manually selected anatomical landmarks. The three other ones are of indirect nature, since they are computed using the PCA model based on the correspondence. Further details not discussed in this paper about the following methods can be found in [7]. These three model based methods are in brief:

- Generalization: The ability to describe instances outside of the training set.
- Compactness: The ability to use a minimal set of parameters.
- Specificity: The ability to represent only valid instances of the object.

Distance to manual landmarks as gold standard - In medical imaging human expert knowledge is often used as a substitute for a gold standard, since ground truth is only known for synthetic and phantom data, but not for the actual images. In the evaluation of correspondence methods this becomes even more evident, because the goal is not clearly defined, in contrast to other tasks such as the segmentation of anatomical structures. We propose to use a small set of anatomical landmarks selected manually by a human expert on each object as a comparative evaluation basis. We computed the mean absolute distance (MAD) between the manual landmarks and each method's points corresponding to the same landmarks in a template structure. For comparison, we report the reproducibility error of the landmark selection.

Model compactness - A compact model is one that has as little variance as possible and requires as few parameters as possible to define an instance. This suggests that the the compactness ability can be determined as the cumulative variance $C(M) = \sum_{i=1}^{M} \lambda_i$, where λ_i is the i^{th} eigenvalue. $C(M)$ is measured as a function of the number of shape parameters M. The standard error of $C(M)$ is determined from training set size n_s: $\sigma_{C(M)} = \sum_{i=1}^{M} \sqrt{2/n_s}\lambda_i$

Model generalization - The generalization ability of a model measures its capability to represent unseen instances of the object class. This is a fundamental property as it allows a model to learn the characteristics of an object class from a limited training set. If a model is overfitted to the training set, it will be unable to generalize to unseen examples. The generalization ability of each model is measured using leave-one-out reconstruction. A model is built using all but one member of the training set and then fitted to the excluded example. The accuracy to which the model can describe the unseen example is measured. The generalization ability is then defined as the approximation error (MAD) averaged over the complete set of trials. It is measured as a function $G(M)$ of the number of shape parameters M used in the reconstruction. Its standard error

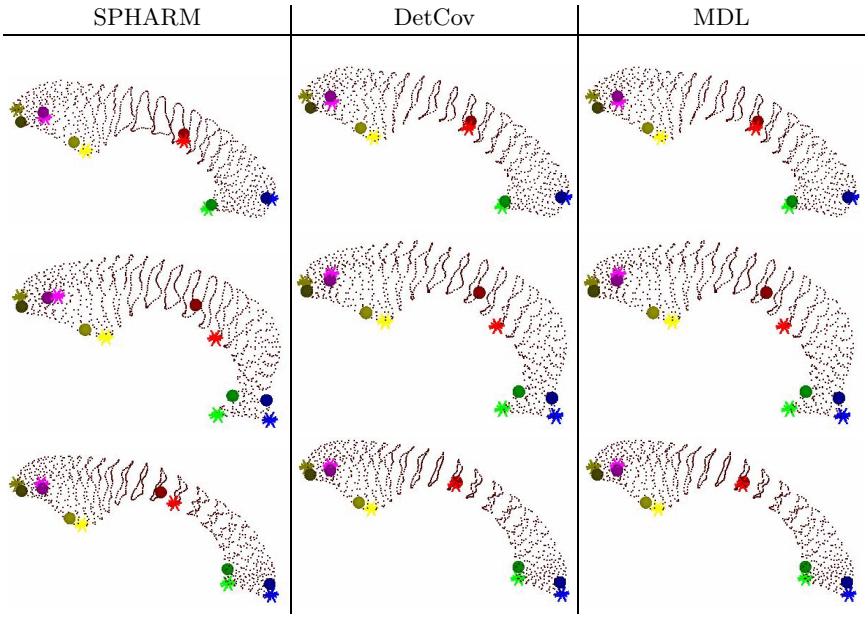


Fig. 2. Visualization of the correspondences of a set of landmarks from the template (see Figure 1) in three selected objects from the two ventricle populations using the different methods. The manually determined landmarks are shown as star-symbols and the SPHARM, DetCov and MDL corresponding locations are shown as spheres.

$\sigma_{G(M)}$ is derived from the sampling standard deviation σ and the training set size n_s as: $\sigma_{G(M)} = \sigma/\sqrt{n_s - 1}$

Model specificity - A specific model should only generate instances of the object class that are similar to those in the training set. It is useful to assess this qualitatively by generating a population of instances using the model and comparing them to the members of the training set. We define the quantitative measure of specificity $S(M)$ (again as a function of M) as the average distance of uniformly distributed, randomly generated objects in the model shape space to their nearest member in the training set. The standard error of $S(M)$ is given by: $\sigma_{S(M)} = \sigma/\sqrt{N}$, where σ is the sample standard deviation of $S(M)$ and N is the number of random samples (N was chosen $10'000$ in our experiments). The distance between two objects is computed using the MAD.

A minimal model specificity is important in cases when newly generated objects need to be correct, e.g. for model based deformation or shape prediction. Model specificity is of lesser importance in the case of shape analysis since no new objects are generated.

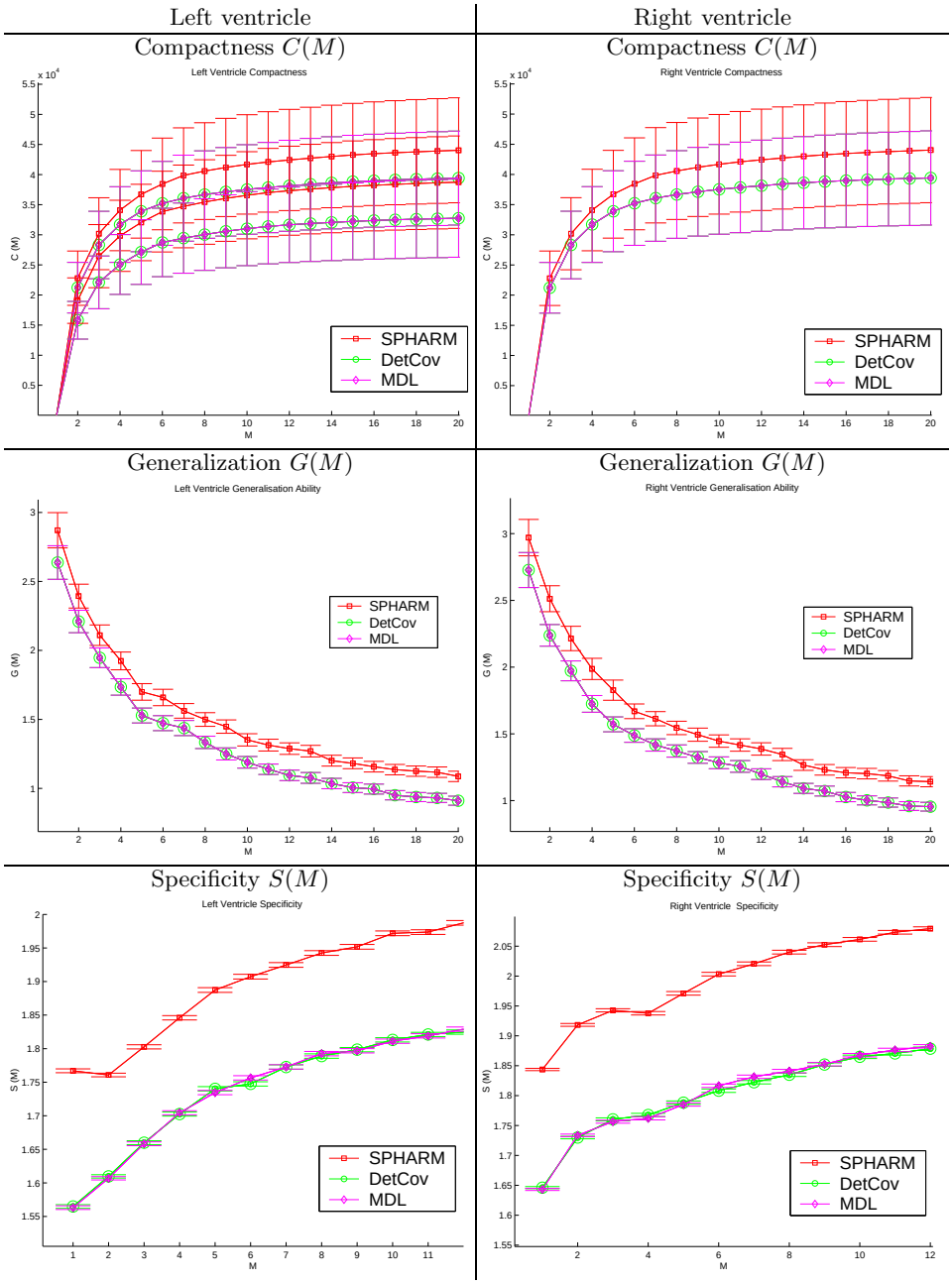


Fig. 3. Table with errors graphs of compactness ($C(M)$), generalization ($G(M)$) and specificity ($S(M)$) for the two ventricle studies (left column: left lateral ventricle, right column: right lateral ventricle). The plot view is zoomed to a M value below 30, since for higher M the plot values did not change.

Method	Left ventricle			Right ventricle		
	Mean	Max	Min	Mean	Max	Min
SPHARM	4.47 mm	6.57 mm	1.72 mm	4.32 mm	6.70 mm	1.11 mm
DetCov	4.00 mm	6.16 mm	1.50 mm	4.28 mm	6.69 mm	1.10 mm
MDL	4.00 mm	6.15 mm	1.48 mm	4.28 mm	6.68 mm	1.10 mm

Table 1. Table with mean, maximal and minimal MAD between the manual landmarks and the studied methods for the ventricle studies. It is clearly visible that there is little change between DetCov and MDL. On the left side, DetCov and MDL have better results than SPHARM. For comparison, the mean landmark selection error was 1.9mm.

3 Results on 3D anatomical structures

In the following sections we present the results of the application of the studied correspondence methods to 3 different population: a left femoral head population of 36 subjects, and a left and a right lateral ventricle population of each 58 subjects. The application of the model constructed from the femoral head populations is the femoral model-based segmentation from CT for patients undergoing total hip replacement. The application of the two ventricle populations is shape analysis for finding population differences in schizophrenia. In this document, we focus only on the correspondence issue. It is noteworthy that the studied populations are comprised not only of healthy subjects, but also of patients with pathologically shaped objects.

3.1 Lateral ventricles

This section describes the studies of the left and the right lateral ventricle structure (see Figure 1) in a population of 58 subjects. The segmentation was performed with a single gradient-echo MRI automatic brain tissue classification [17]. Postprocessing with 3D connectivity, morphological closing and minor manual editing provided simply connected 3D objects. The manual landmarks were selected by an expert with an average error of 1.9mm per landmark.

In Figure 2 the results of the correspondence methods in three exemplary cases are shown. The first row shows the good correspondence with the manual landmarks, as it is seen in the majority of the objects in this study. The second row shows the frequent case in the remaining objects, in which all three methods have a rather large difference to the manual landmarks. In most cases of disagreement with the manual landmarks, all methods produced similar results. The last row shows the rare case in which SPHARM is clearly further away from the landmarks than DetCov and MDL. The opposite case was not observed.

Table 1 displays the landmark errors and Figure 3 displays the error plots, which both suggest that DetCov and MDL produce very similar results. Both show smaller errors than SPHARM.

Fig. 4. Visualization of bad alignment in the first femoral head study on 3 example objects seen from the same viewpoint. A large rotational alignment error is clearly visible for a rotation around the long axis of the first order ellipsoid, which is close to the femoral neck-axis.

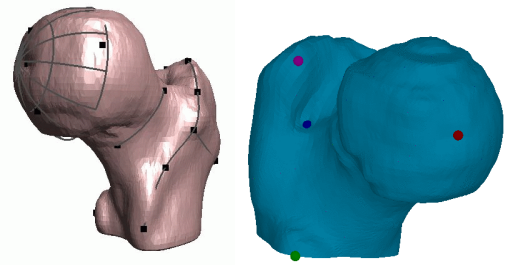
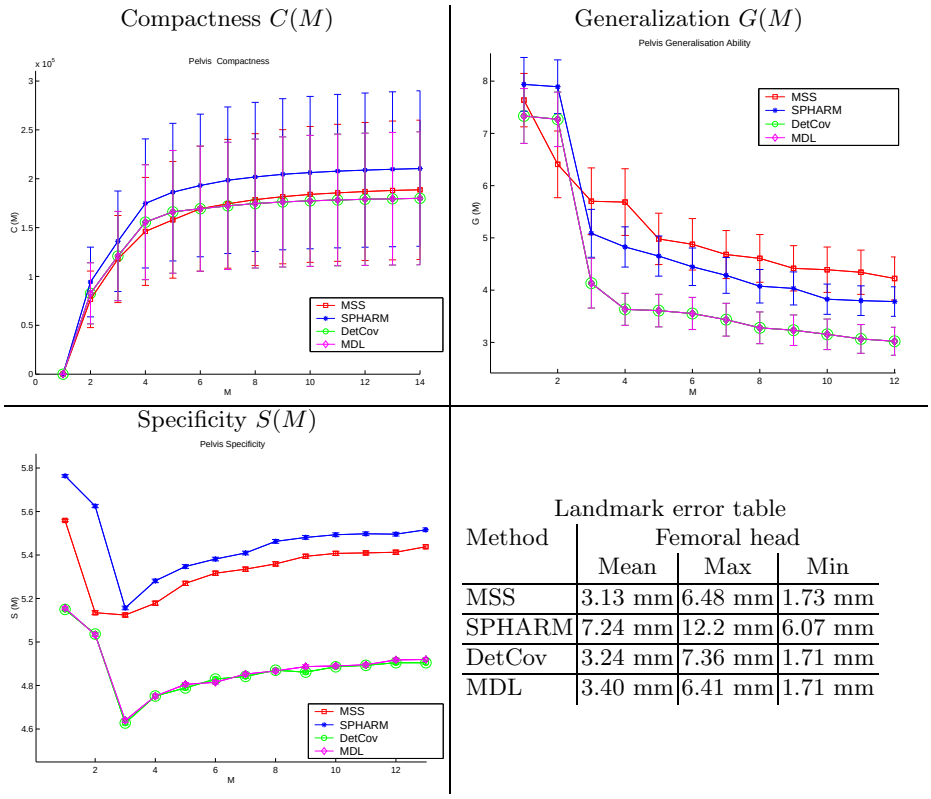


Fig. 5. Left: Display in MSS tool with a single femur head object and manually placed anatomical curves (anterior viewpoint). Right: Visualization of the femoral head template (posterior viewpoint) and four of its anatomical landmarks (Fovea, center lesser trochanter, tip greater trochanter).

3.2 Femoral head

This section describes the results on a population of objects from the head region of the femoral bone. The segmentations were performed from CT-images with a semi-automated slice-by-slice explicit snake algorithm [12]. The correspondence study was done in 2 steps. Initially we only computed SPHARM, DetCov and MDL on all available cases (30 total). We realized that the SPHARM correspondence was not appropriate, so the results of the following computation were meaningless (discussed further down). In a second step, we selected only those cases, which contained the lesser trochanter in the dataset. For these cases (16) we then computed MSS, SPHARM, DetCov and MDL.

The first study was based on the full 30 cases including 14 datasets with missing data below the calcar. The distal cut of the femoral bone was performed through the calcar perpendicular to the bone axis. The alignment was performed using the Procrustes alignment based on the SPHARM correspondence. MSS was not computed in this case. We observed a bad alignment due to the SPHARM correspondence. In Figure 4 we visualize this inappropriate alignment in three cases. As a consequence the DetCov and MDL results were inappropriate. The



Landmark error table

Method	Femoral head		
	Mean	Max	Min
MSS	3.13 mm	6.48 mm	1.73 mm
SPHARM	7.24 mm	12.2 mm	6.07 mm
DetCov	3.24 mm	7.36 mm	1.71 mm
MDL	3.40 mm	6.41 mm	1.71 mm

Fig. 6. Top row, Bottom row left: Table with Errors graphs of compactness ($C(M)$), generalization ($G(M)$) and specificity ($S(M)$) for the femoral head study. Bottom row right: Table with mean, maximal and minimal MAD between the manual landmarks and the studied methods for the femoral head study. There is little change between DetCov and MDL. SPHARM shows clearly the worst results of all studied methods. For comparison, the mean landmark selection error was 2.5mm.

bad SPHARM correspondence resulted from a rotational symmetry along the long first order ellipsoid axis, which is close to the neck-axis. Due of the bad correspondence, we do not present here the error analysis of these cases.

The second study was based on a subset of the original population comprising only of those datasets that include also the lesser trochanter. The distal cut of the femoral bone was performed by a plane defined using the lesser trochanter center, major trochanter center and the intertrochanteric crest. For the MSS method the anatomical landmarks for the subdivision surfaces were chosen as follows: the fovea, the half-sphere approximating the femoral head, the circle approximating the orthogonal cross-section of the femoral neck at its thinnest location, the intertrochanteric crest and the lower end of the major trochanter. The landmarks for the MSS alignment were the lesser trochanter, the femoral

head center, and the center of the circle approximating the neck at its smallest perimeter. Each landmark was selected on the respective 3D femur model either directly on the reconstructed bone surface or using 3D spherical primitives. The manual landmarks for the comparison were selected by a different expert with an average error of 2.5mm per landmark. The landmarks sets for the MSS and the comparison were not exclusive, due to the scarceness of good landmarks.

All correspondences were based on the MSS alignment. We observed that the SPHARM correspondence was visually better behaved in this study due to the inclusion of the lesser trochanter, which eliminated some problems with the rotational symmetry. However, figure 6 shows that the landmark errors for SPHARM alignment are clearly the worst of the studied methods. MSS shows the best average agreement with the manual landmarks, which is not surprising since the landmarks contained points used also to construct MSS. MDL was surprisingly better than both DetCov and MSS in regard to the minimal and maximal MAD, although the MAD differences are rather small. In figure 6 it is clearly visible that MDL and DetCov have similar and better modeling properties than SPHARM and MSS. Only for $G(M)$ MSS is better than SPHARM.

4 Conclusions

In this paper, we have presented a comparison of the SPHARM, DetCov, MDL and MSS correspondence methods in three populations of anatomical objects. The goal in all of the presented studies is a model-based application. We have analyzed both the direct correspondence via manually selected landmarks as well as the properties of the model implied by the correspondences, in regard to compactness, generalization and specificity.

The results for SPHARM of the first femoral head study revealed that in case of rotational symmetry in the first order ellipsoid, independent of the higher order terms, the correspondence is inappropriate. Since correspondence and alignment are dependent on each other, such a bad correspondence cannot be significantly improved using methods like MDL and DetCov. In all studies, DetCov and MDL showed very similar results. The model properties of DetCov and MDL were better than both SPHARM and MSS. The findings suggest that for modeling purposes the best of the studied correspondence method are MDL and DetCov.

The manual landmark errors are surprisingly large for all methods, even for the MSS method, which is based on landmarks. This finding is due to the high variability for the definition of anatomical landmarks definition by human experts, which is usually in the range of a few millimeters.

In the lateral ventricle studies we plan to do the following shape analysis with the model built on the MDL correspondence. Other current research in our labs suggest that the shape analysis could gain statistical significance by using MDL rather than SPHARM. In the femoral head study, we plan to use the MDL model for shape prediction in the shape space. The results of the specificity error is in this study very relevant, since it is desired to generate 'anatomically correct' objects from the shape space.

References

1. Bookstein, F.L.: Morphometric Tools for Landmark Data: Geometry and Biology, Cambridge University Press (1991)
2. Brechbühler, C., Gerig, G., Kübler, O.: Parameterization of Closed Surfaces for 3-D Shape Description. Comp. Vision and Image Under. **61** (1995) 154–170
3. Brett, A.D., Taylor, C.J.: Construction of 3D Shape Models of Femoral Articular Cartilage Using Harmonic Maps. MICCAI (2000) 1205–1214
4. Christensen, G., Joshi,S., Miller, M.: Volumetric Transformation of Brain Anatomy. IEEE Trans. Med. Imag., **16** 6 (1997) 864–877
5. Cootes, T, Hill, A., Taylor, C.J., Haslam, J.: The Use of Active Shape Models for Locating Structures in Medical Images. Img. Vis. Comp. **12** (1994) 355-366
6. Davies, Rh.H, Twining, C.J., Cootes, T.F., Waterton, J. C., Taylor, C.J.: 3D Statistical Shape Models Using Direct Optimization of Description Length. ECCV (2002) I.
7. Davies, Rh.H: Learning Shape: Optimal Models for Analysing Natural Variability. Dissertation University of Manchester, 2002.
8. Davies, Rh.H, Twining, C.J., Cootes, T.F., Waterton, J. C., Taylor, C.J.: A Minimum Description Length Approach to Statistical Shape Model. IEEE TMI **21** (2002)
9. Fleute, M., Lavallee, S.: Building a Complete Surface Model from Sparse Data Using Statistical Shape Models, MICCAI (1998) 879-887
10. Gerig, G., Styner, M.: Shape versus Size: Improved Understanding of the Morphology of Brain Structures, MICCAI (2001) 24-32
11. Hill, A., Thornham ,A., Taylor ,C.J.: Model-Based Interpretation of 3D Medical Images. Brit. Mach. Vision Conf. BMCV, (1993) 339-348
12. Hug, J., Brechbühler, C., Székely, G.: Tamed Snake: A Particle System for Robust Semi-automatic Segmentation. MICCAI (1999) 106–115
13. Joshi, Banerjee, Christensen, Csernansky, Haller, Miller, Wang:Gaussian Random Fields on Sub-Manifolds for Characterizing Brain Surfaces. IPMI (1997) 381–386
14. McInerney, T., Terzopoulos, D.: Deformable Models in Medical Image Analysis: A Survey. Med. Image Analysis **1** 2 (1996) 91–108
15. Kelemen, A., Székely, G., Gerig, G.: Elastic Model-Based Segmentation of 3D Neuroradiological Data Sets. IEEE Trans. Med. Imag. **18** (1999) 828–839
16. Kotcheff, A.C.W., Taylor, C.J.: Automatic Construction of Eigenshape Models by Direct Optimization. Med. Image Analysis **2** 4 (1998) 303–314
17. Van Leemput, K., Maes, F., Vandermeulen, D., Suetens, P.: Automated Model-based Tissue Classication of MR Images of the Brain, IEEE TMI **18** (1999) 897–908
18. Meier, D., Fisher, E.: Parameter Space Warping: Shape-Based Correspondence Between Morphologically Different Objects. Trans. Med. Imag. **12** (2002) 31–47
19. Rangarajan, A., Chui, H., Bookstein, F.L.: The Softassign Procrustes Matching Algorithm, IPMI (1997) 29–42
20. Rueckert, D., Frangi, A.F., Schnabel, J.A.: Automatic Construction of 3D Statistical Deformation Models Using Non-rigid Registration. MICCAI (2001) 77–84
21. Styner, M., Gerig, G., Lieberman, J., Jones, D., Weinberger, D.: Statistical Shape Analysis of Neuroanatomical Structures Based on Medial Models. Med. Image Anal.
22. Szeliski, R., Lavallee, S.: Matching 3-D anatomical surfaces with non rigid deformations using octree-splines, Int. J. Computer Vision, **18** 2 (1996) 290–200
23. Tagare, H.: Shape-Based Nonrigid Correspondence with Application to Heart Motion Analysis. IEEE Trans. Med. Imag. **18** 7 (1999) 570–580
24. Wang, Y., Peterson B. S., Staib, L. H.,: Shape-based 3D Surface Correspondence Using Geodesics and Local Geometry. CVPR **2** (2000) 644-651