

CRITICAL REVIEW

G. Chintalapani, L. M. Ellingsen, O. Sadowsky, J. L. Prince, R. H. Taylor. **Statistical Atlases of Bone Anatomy: Construction, Iterative Improvement and Validation.** Medical Image Computing and Computer Assisted Intervention (MICCAI), 2007; 10 (Pt 1): 499-506.

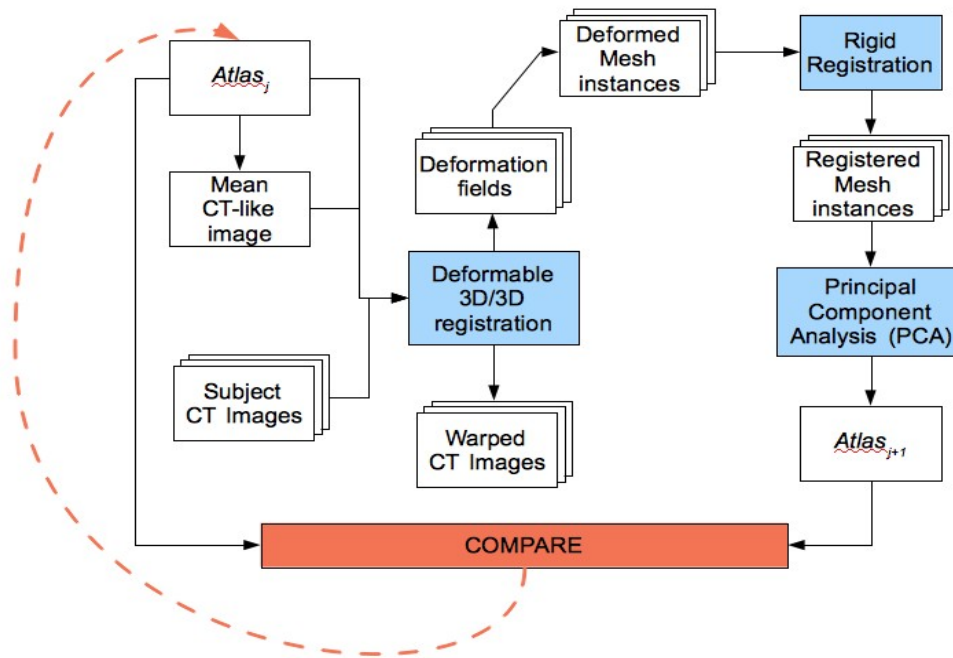
Statistical modeling and analysis of anatomical structures is an active topic of interest in medical imaging research. Using a large enough training sample, it is possible to construct statistical atlases that incorporate mean shape information as well as variational modes that describe the most common variations in a population. In this paper, Chintalapani et al. present a modular pipeline for building statistical atlases. The pipeline presented incorporates an iterative bootstrapping framework to fine-tune the atlas and to reduce its dependency on the choice of the initial template. They present their validation results on 110 male pelvis CT images.

By making use of an iterative approach, the pipeline presented in this paper eliminates the dependency on the choice of the initial template. This is a significant improvement since it obviates the need to manually analyze a large data set to pick a template that is an average image. A method for comparing statistical atlases and for quantifying their accuracies is presented, which allows for optimizing the parameters of the pipeline.

Their atlas building pipeline includes several algorithmic steps that can be summarized as follows:

1. Select a template CT image and manually segment it to create a tetrahedral mesh.
2. Deformably register the remaining CT images (subjects) onto the template image and find the deformation fields.

3. Interpolate the deformation field at each vertex of the template mesh to create a mesh instance for each subject.
4. Rigidly register all the meshes.
5. Perform PCA on the registered mesh instances to calculate the mean shape mesh \bar{M} and the variational modes D_k . This results in the first atlas, $Atlas_0$.
6. Create a CT-like image from the mean shape mesh \bar{M} and the mean density polynomial.



7. The same procedure outlined above is now performed using the obtained atlas and the CT-like image produced in the above step as follows:
8. Repeat this bootstrapping measure until the result of the comparison between the two atlases is below a certain threshold. Two metrics are presented in the paper for comparing shapes:
 - a) Vertex-vertex correspondence error: the average Euclidean distance between two corresponding vertices. This method assumes that the two meshes are similar enough that there is a one-to-one correspondence between the vertices in the two meshes.
 - b) Surface-surface distance: the average Euclidean distance from the vertices of the model

surface to the closest points on the subject surface.

After the statistical atlas is obtained, any shape instance S can be expressed as a linear combination of the anatomical modes of variation using the mode weights λ_k as

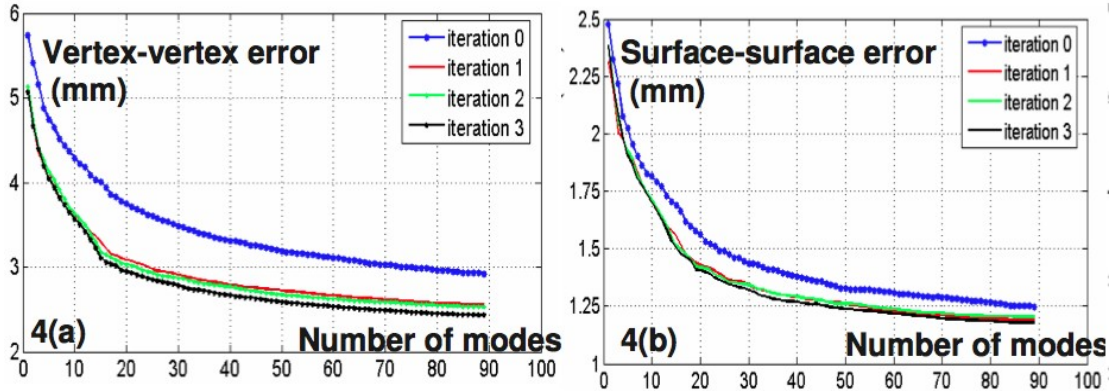
$$S = \bar{M} + \sum_k \lambda_k D_k$$

Rewriting this equation as $S = \bar{M} + D\lambda$, where $D = [D_1 \ D_2 \ \dots \ D_N]$ and $\lambda \in \mathbb{R}^N$, it is possible to solve for the mode weights for a given shape instance S^{true} by

$\lambda^{est} = D^{-1}(S^{true} - \bar{M})$. The first n dominant eigenvalues can then be used to estimate this shape by

$$S^{est} = \bar{M} + \sum_{k=1}^n \lambda_k^{est} D_k$$

In order to assess the effect of the number of modes used and the number of bootstrapping iterations carried out, the authors used 90 CT images in the data set for training and estimated the remaining 20 images using the equations described above. They used both of the metrics described to compare S^{true} and S^{est} . Their results are presented below.



Figures taken from G. Chintalapani, L. M. Ellingsen, O. Sadowsky, J. L. Prince, R. H. Taylor. Statistical Atlases of Bone Anatomy: Construction, Iterative Improvement and Validation. MICCAI, 2007.

Based on these results, the authors have concluded that the process seems to have converged after the second iteration. Even though including a large number of modes of variation in the statistical atlas results in more accurate shape estimates, it is computationally very expensive to do so. The

authors picked the surface to surface distance threshold to be approximately 1.5 mm and concluded that keeping only the first 15 modes of variation results in accurate enough atlases.

In order to analyze the effect of the size of the training population on the atlas accuracy, the authors randomly selected n meshes, where $n = 20, 30, \dots, 90$, and created statistical atlases using these meshes, averaged across 20 trials for each n . They then estimated the shapes of the 20 images that were left out in the previous experiment using the obtained atlases and calculated the vertex to vertex correspondence error between the estimated shape and the true shape.

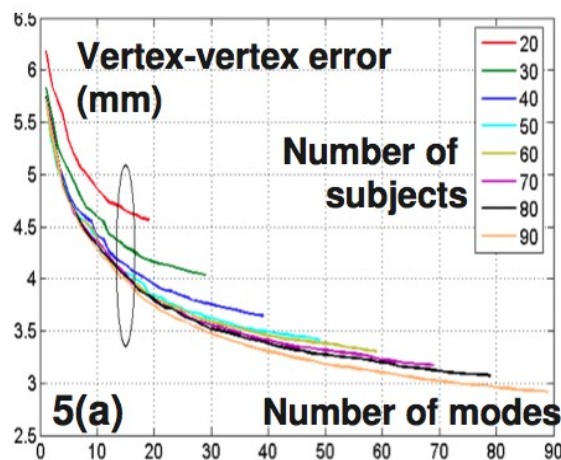


Figure taken from G. Chintalapani, L. M. Ellingsen, O. Sadowsky, J. L. Prince, R. H. Taylor. Statistical Atlases of Bone Anatomy: Construction, Iterative Improvement and Validation. MICCAI, 2007.

These results show that beyond 50 images, the accuracy of the atlas increases only minimally.

The authors conclude that 40 to 50 images are sufficient to create an accurate atlas.

There is a very short discussion regarding whether the atlas obtained using this iterative framework is a global or local optimum. The authors do not analyze the theory behind the atlas construction pipeline to address this issue, but rather mention several experiments they have conducted where they compared the mean meshes they obtained using the pipeline to “ground truth” meshes that were obtained as a result of manual segmentation. One limitation of this verification procedure that is pointed out in the paper is the lack of a firm consensus manual segmentation. Since manual segmentation is subject to random errors, it is necessary to evaluate the algorithm against a consensus

from multiple segmentations.

Even though the authors mention that the bootstrapping process eliminates the dependency on the initial template, there is no discussion of how the parameters of the algorithm (i.e., training population size, number of iterations and variational modes required for an accurate atlas) are affected by the choice of the template. If an outlier patient image is used as the template, it is likely that the convergence process will require more iterations, and the other parameters of the algorithm might need to be adjusted as well. Whether the optimal parameters presented (initial population consisting of 50 images, 15 variational modes and 2 iterative steps) would always result in atlases with similar accuracy is unclear from the paper.

The authors describe two metrics for comparing two atlases, vertex to vertex correspondence error and surface to surface distance. However, they do not mention the situations where one of these metrics would be more suitable to use than the other. Their reason for using the surface-surface distance metric and thresholding this measure at 1.5 mm to find the optimal number of variational modes to use is not explained.

Additional possible future directions building on this work include experimenting with different registration methods to assess the influence of the registration method used on the convergence of the iterative framework, and to use this pipeline to construct statistical atlases of other anatomical structures.

Since our project involves using the pipeline explained in this paper, reading this article has allowed us to gain further insight into the specifics of the atlas building procedure. The two atlas comparison metrics presented will be very useful for us to assess the accuracy of our results. Additionally, using the figures in this paper, we can estimate the number of images we need to use, the number of iterations we need to carry out, and the number of variational modes we need to use in our project.