STATISTICAL DEFORMATION MODELS FOR HIP FRACTURE DISCRIMINATION: VALIDATION BY MODEL COMPARISON

Museyko O.[1], Bousson V.[2], Adams J.[3], Laredo J.-D.[2], Engelke K.[1]

[1]Institute of Medical Physics (IMP), University of Erlangen-Nuremberg, Germany; [2]Service de Radiologie OstéoArticulaire, Hôpital Lariboisière, Paris, France; [3]Clinical Radiology, The Royal Infirmary, University of Manchester, United Kingdom

1. ABSTRACT

Using statistical shape models to classify shapes where shape differences are not readily observed may be complicated, one example is classification of proximal femur shape with respect to the fracture risk. Apart from the issues with model quality, the classifications made may depend on the shape model due to the real difference between model populations. We propose few simple similarity measures on the space of models and use them to check if the model (dis)similarity can explain difference, resp. conformity, in hip fracture discrimination from these models.

2. MATERIALS AND METHODS

Study population: 46 women with fresh hip fractures (imaged at the non-fractured site) and 56 controls from two imaging centers (EFFECT study [1]).

QCT: 120 kV, 170 mAs, slice thickness: 1 or 1.25 mm, medium reconstruction kernel.

Image analysis: MIAF-Femur was used for 3D segmentation and analysis. Osteophytes were excluded in the segmentation step. In the fractured group, left femurs were mirrored to make pooled analysis of both sides possible. Image segmentation masks (Fig. 1) were used for the shape analysis. First, a rigid preregistration step using the binary masks was performed in order to eliminate differences in the position of the femurs. Then, all binary masks were non-rigidly registered to one arbritrarily selected reference mask to obtain the shape models.

Statistical deformation models: 20 datasets from each group were selected for the test sample set, the rest was used to form several random statistical deformation models [2]. Models included either unfractured subjects only (type 1), or a combination of control and fractured subjects (type 2). Each model consisted of a mean shape and principal components of the covariance matrix of shapes, i.e., sets of registration displacement vectors at each voxel relative to the reference dataset. Components with the low energy (i.e., small corresponding eigenvalues) were excluded.

2. METHODS (cont.)

Model similarity measures: We concentrate on the comparison of the principal components and assume the equality of the model mean shapes for the sake of simplicity. Principal components form an orthogonal basis of a subspace within a high-dimensional Euclidian space of all shapes ***S***. With each component we associate a principal vector of the length equal to the corresponding eigenvalue. Two different models $U=\{u_i\}$ and $V=\{v_i\}$ correspond to two subspaces of ***S*** which are not coinciding in general. It means that some variations caught in one model may not be explained by the second one. Let us define $D_V(u)$, a part of variation along the principal axis u in U which is not explainable by V:

$$D_V(u)=\frac{\left\|u-\sum_i \frac{u\cdot v_i}{\lambda_i^2}v_i\right\|}{\|u\|},$$

where $\lambda_i=\|v_i\|$ is the *i*-th eigenvector of V. In other words, we measure the length of the vector $u^\perp=u-\sum_i \frac{u\cdot v_i}{\lambda_i^2}v_i$ equal to the difference between u and its best approximation u^V within V, normalized by the magnitude of the original variation, $\|u\|$. The overall unexplainable variation of the second model is the sum of unexplainable variations along all components relative to the total model variation:

$$D(U,V)=\frac{\sum_i \|u_i^\perp\|}{\sum_i \mu_i},\quad \mu_i=\|u_i\|.$$

Even if the principal axes of two models coincide, variations for a common principal axis may be different. Let us measure the discrepance between the variation of u_i (its projected part u_i^V) and the variation in V in this direction by the scale factor β_i bringing u_i^V onto the ellipsoid of variations in V: $\beta_i^2\sum_j \frac{|u_i\cdot v_j|^2}{\lambda_j^4}=1$. The fracture of the sum of the variation discrepances along all axes within the total model variation is the corresponding measure for whole U:

$$M(U,V)=\frac{\sum_i |\beta_i-1|\left(\sum_j \frac{|u_i\cdot v_j|^2}{\lambda_j^2}\right)^{1/2}}{\sum \lambda_j}.$$

Model comparison: for each model, components significant with respect to the fracture discrimination on the test set were determined. Models which had very similar or, vice versa, most discrepant discrimination power were compared using the proposed measures. Finally, the model for the test set was built and checked against each of the initial models.

3. RESULTS

[Work in progress]

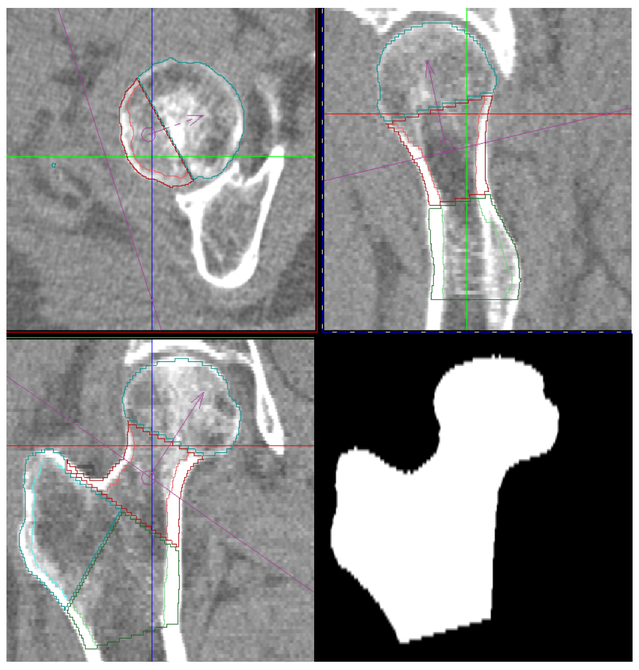

Fig. 1: Multiplanar reconstructions of 3D CT dataset and coronal reformation of segmentation mask (lower right)

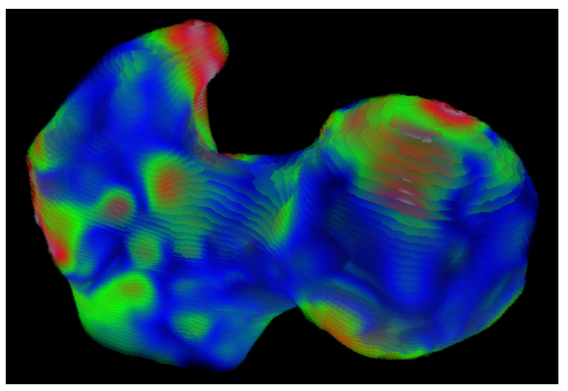

Fig. 2: Example of shape variation. The largest changes of the voxel position are in red, the smallest are in blue.

To visualize the changes in shape that correspond to a particular principal component, we used shape models in the generative mode. For this, a mean shape was first computed by setting weighting factors for all principal components to 0 (no variations). Then, setting the factor for the principal component of interest to non-zero while keeping others fixed to 0, we obtained a shape with changes along the corresponding principal component. The difference between these two shapes is the offset vector field associated with the particular component. Larger changes in shape correspond to longer vectors. Thus, the absolute values of vectors constitute a 'topographic' map of shape variations. One such map is visualized on Fig. 2.

4. CONCLUSIONS

Summarizing, we have proposed simple measures of similarity between statistical shape models and evaluated their performance by comparing the computed model difference with the correspondent observed difference in their discrimination power.

Further questions we would like to address next include: given several shape models and a target population, rather different from the model populations, is it possible to combine the models and estimate whether the result is more suitable/similar to the target population?

[1] V. D. Bousson et al., "In vivo discrimination of hip fracture with quantitative computed tomography: results from the prospective European Femur Fracture Study (EFFECT)", J Bone Miner Res, vol. 26, 881-93.

[2] D. Rueckert et al "Automatic construction of 3D statistical deformation models of the brain using nonrigid registration", IEEE TME, 8(22), pp. 1014-1025, 2003.