Vertebral fracture classification

Marleen de Bruijne[a,b], Paola C. Pettersen[c], László B. Tankó[b], and Mads Nielsen[a,b]

[a]Department of Computer Science, University of Copenhagen, Copenhagen, Denmark
[b]Nordic Bioscience, Herlev, Denmark
[c]Center for Clinical and Basic Research, Ballerup, Denmark

ABSTRACT

A novel method for classification and quantification of vertebral fractures from X-ray images is presented. Using pairwise conditional shape models trained on a set of healthy spines, the most likely unfractured shape is estimated for each of the vertebrae in the image. The difference between the true shape and the reconstructed normal shape is an indicator for the shape abnormality. A statistical classification scheme with the two shapes as features is applied to detect, classify, and grade various types of deformities.

In contrast with the current (semi-)quantitative grading strategies this method takes the full shape into account, it uses a patient-specific reference by combining population-based information on biological variation in vertebra shape and vertebra interrelations, and it provides a continuous measure of deformity.

Good agreement with manual classification and grading is demonstrated on 204 lateral spine radiographs with in total 89 fractures.

Keywords: shape analysis, conditional shape model, classification, vertebral fracture, osteoporosis

1. INTRODUCTION

Among osteoporotic fractures, vertebral fractures are the most common and they occur in relatively young patients. The presence of vertebral fractures is known to be a strong indicator for the risk of future spine and hip fractures. This makes assessment of vertebral fracture important in clinical decision making and in clinical trials assessing the efficacy of drug candidates.

Conventionally, fractures are classified from lateral X-ray images by experienced radiologists using a semi-quantitative grading scheme.[1] Six points are placed on the corners and in the middle of the vertebra endplates, defining the anterior, middle and posterior heights. The fracture grade is then derived from these three height measures or from the ratios between the heights, in connection with subjective judgement of the radiological evidence by experienced radiologists. Alternatively, fully quantitative methods have been used that rely completely on the three height measures, possibly in comparison with population based measurements and/or normalized for inter-patient variability by comparison with measurements taken from a neighboring or reference vertebra.[1–5]

These coarse schemes, representing vertebral shape by three heights, are unable to capture subtle shape changes, and suffer from variability in point placement. Studies have shown that a large number of fractures goes undiagnosed with current methods.[6,7] More precise and objective measures of vertebral deformity are therefore needed.

We propose a method to detect, classify, and grade the severity of vertebral fractures from lateral X-ray images. Our method incorporates the following key features:

- Use of full vertebral contour
- Statistical model of shape variation of individual vertebrae over a (healthy) population
- Statistical model of interrelations between shapes of vertebrae within a (healthy) subject
- Supervised classification into different types of deformity

E-mail: marleen@diku.dk

Medical Imaging 2007: Image Processing, edited by Josien P. W. Pluim, Joseph M. Reinhardt,
Proc. of SPIE Vol. 6512, 651219, (2007) · 1605-7422/07/$18 · doi: 10.1117/12.706268

Using the complete contour instead of the conventional three height measures provides a rich morphological description and enables capturing of subtle shape differences. The statistical models that describe the normal, biological variation in vertebra shape identify abnormal shapes, and the models of interrelations between shapes of vertebrae within a patient allow adjustment of the models to individual patients. In the supervised classification, models of typical osteoporotic deformity are incorporated so as to further distinguish normal shape variability from osteoporosis-related deformation.

Variations in shape of individual vertebrae are modeled using a point distribution model.[8] For the vertebra interrelations within a patient we use conditional shape models to reconstruct the most likely shape of a vertebra given the known shape of a neighboring vertebra. If both these models are constructed from a training set of normal, healthy spines this provides an estimate of what the vertebra shape would have been if it were normal. In previous papers we have shown that vertebrae in healthy spines can be accurately reconstructed from their neighbors[9] and that individual pairwise reconstructions can be reliably combined into one reconstruction that is accurate also in the presence of deformities.[10] In the current paper we explore in more detail how the differences between the reconstruction and the true shape as observed in the image can be used to assess severity and subtype of fracture. Based on a training set of shapes with known fracture types, the presence, type, and severity of deformity is determined using discriminant analysis.

2. METHODS

This section describes the models used. The procedure to reconstruct unfractured vertebra shapes from neighboring vertebrae is described in more detail in an earlier paper.[10]

The variations of vertebra shape over a training set of examples of unfractured spines are modeled using the linear point distribution models (PDM) as proposed by Cootes and Taylor.[8] PDMs model the shape probability distribution as a multivariate Gaussian in a subspace of reduced dimensionality. Shapes are defined by the coordinates of a set of landmark points which correspond between different shape instances. A collection of training shapes are aligned using for instance Procrustes analysis[11] and a principal component analysis (PCA) is applied to the aligned shape vectors. To this end, the mean shape $\overline{\mathbf{x}}$, the covariance matrix Σ, and the eigensystem of Σ are computed. The eigenvectors ϕ_i of Σ provide the so-called *modes of shape variation* which describe a joint displacement of all landmarks. The eigenvectors corresponding to the largest eigenvalues λ_i account for the largest variation; a small number of modes usually captures most of the variation. Each shape $\mathbf{x}$ in the set can then be approximated by a linear combination of the mean shape and these modes of variation:

$$\mathbf{x} = \overline{\mathbf{x}} + \mathbf{\Phi_t}\mathbf{b} + \mathbf{r}$$

where $\mathbf{\Phi_t}$ consists of the eigenvectors corresponding to the t largest eigenvalues, $\mathbf{\Phi_t} = (\phi_1|\phi_2|\ldots|\phi_t)$, $\mathbf{b}$ is a vector of model parameters that weigh the contribution of each of the modes, and $\mathbf{r}$ is a vector of residual shape variation outside of the model subspace.

2.1. Modeling relations between vertebrae

We model the relations between two vertebrae in the same image as two conditional models, describing the expected shape and shape variation of the shape of the one vertebra if the shape of the other is known. The distribution $P(S_1|S_2)$, the probability distribution of a shape S_1 given a known other shape S_2, can be modeled as the Gaussian conditional density

$$P(S_1|S_2) = \mathcal{N}(\mu, K) \tag{1}$$

with

$$\mu = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(S_2 - \mu_2)$$
$$K = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

where μ_1 and μ_2 are the mean shapes of the training sets for S_1 and S_2, and covariances Σ_{ij} are obtained from the combined covariance matrix

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

as

$$\Sigma_{ij} = \frac{1}{n-1}\sum_n (S_{in} - \mu_i)(S_{jn} - \mu_j)^T.$$

Applied to pairwise vertebra shape prediction from neighboring vertebrae, S_2 is the predictor vertebra shape, S_1 is the shape to predict, μ is the maximum likelihood estimate of S_1 given S_2, and K is the variance in the estimate.

2.2. Combining shape estimates

The pairwise shape prediction of the previous subsection results in several shape estimates for each vertebra. Not all of these estimates will be equally accurate. For instance, one would expect that the vertebral shape correlation between two direct neighbors is stronger than between two vertebrae that are further apart. In a fractured spine, the fractured vertebra(e) will likely produce inaccurate estimates of normal vertebra shape, even for the direct neighbors. We will therefore define the final shape estimate as a weighted combination of the individual predictions, where the weights express the degree of belief in each estimate.

We assume that the observed vertebral shapes are produced by the underlying shape model of normal shapes, resulting in a multi-variate Gaussian with variances λ_i in t directions, plus additional uncorrelated Gaussian noise with a variance σ_r^2 in all directions which accounts for any residual shape differences. The probability density for a shape S is then given by the product of the Gaussian densities of the shape model and the residual model:

$$p(S|\theta) = c_s c_r \exp[-\frac{1}{2}(\mathrm{M}_s + \mathrm{M}_r)] \tag{2}$$

$$c_s = \frac{1}{\sqrt{(2\pi)^t \prod_{i=1}^t \lambda_i}} \quad , \quad \mathrm{M}_s = \sum_{i=1}^t \frac{b_i^2}{\lambda_i}$$

$$c_r = \frac{1}{\sqrt{(2\pi)^n \sigma_r^{2n}}} \quad , \quad \mathrm{M}_r = \frac{|\mathbf{r}|^2}{\sigma_r^2}$$

where b_i are the model parameters from the PDM and $\mathbf{r}$ is a vector of residuals.

The probability density for each conditional shape estimate can be expressed as $P(S_1|S_2)P(S_2)$, where $P(S_2)$ is the probability that the predictor shape S_2 is a valid normal (unfractured) shape, and the variance in $P(S_1|S_2)$ expresses the uncertainty in the prediction of S_1 from the model conditional on S_2. The weight for the ith prediction of S_1 is then given by

$$w_i = \frac{P(S_1|S_i)P(S_i)}{\sum_i P(S_1|S_i)P(S_i)}, \tag{3}$$

and the individual estimates are combined as a weighted sum.

$P(S_i)$ can be determined by substituting the predictor shape S_i for S in Equation 2 and the mean and covariance of the training set for the predictor shapes for θ, whereas for $P(S_1|S_i)$ the predicted shape for S_1 and the mean and covariance of the conditional model must be substituted.

2.3. Fracture Classification

The training set is separated in classes of normal vertebrae and vertebrae with different types of deformity, as indicated by an expert radiologists, and a classifier is subsequently trained to distinguish between the different classes.

A classifier can be written in terms of discriminant functions d_i.[12] An object with feature vector x is then assigned the class that corresponds to the largest $d_i(x)$. We search for the classifier that minimizes the 'risk' associated to misclassifications. If we assume a so-called 0-1 loss function — false positives and false negative detections are equally bad — the optimal discriminant function that minimizes the risk is

$$d_i(x) = p(x|w_k)P(w_k).$$

That is, each object should be assigned the class that maximizes the posterior probability. For a multivariate Gaussian distribution $p(x|w_k)$ this becomes

$$d_i(x) = lnP(w_i) - \frac{1}{2}ln|\Sigma_i| - \frac{1}{2}[(x-\mu_i)^T\Sigma_i^{-1}(x-\mu_i)],$$

which is known as the quadratic discriminant classifier as the boundaries between classes are quadratic.

The quadratic discriminant classifier is optimal if the classes are Gaussian distributed with arbitrary, unequal covariances. However, if the number of samples is small compared to the dimensionality of the feature space, the estimate of the individual class covariance matrices Σ_i from the data may be unreliable. It may therefore be beneficial to simplify the assumed distribution, thus reducing the number of parameters to be estimated. One assumption that is often used is that the classes are Gaussian distributed with arbitrary, but equal, covariances. The resulting optimal classifier is the linear discriminant classifier:

$$d_i(x) = lnP(w_i) - \frac{1}{2}[(x-\mu_i)^T\Sigma^{-1}(x-\mu_i)]$$

where Σ is the pooled covariance matrix, i.e. the average covariance matrix weighted by the class prior probabilities.

One can simplify the model even more and assume that the features are uncorrelated, which leads to a diagonal covariance matrix. If one would assume that the class priors are equal and, in addition, the covariance matrix is the identity (all features have equal variance), the optimal discriminant classifier turns into the nearest mean classifier.

Several hybrid approaches are possible in which the estimates of the covariances are regularized by adding for instance a small portion of the pooled covariance and/or the identity matrix.[13] Another possibility is to apply spatial smoothness constraints on the coefficients as in penalized discriminant analysis.[14] This may improve accuracy especially in situations in which there are many highly correlated features, such as in landmark based shape representations. In all cases, the optimal regularization parameter(s) can be selected using cross-fold validation on the training set.

In the experiments in this paper we have used a linear discriminant classifier that is further regularized by replacing

$$\Sigma' = (1-\alpha)\Sigma + \alpha\sigma^2 I \tag{4}$$

where I is the identity matrix, σ^2 is the mean variance, and α a regularization parameter with $0 <= \alpha <= 1$. Varying α makes the classifier vary between the non-regularized linear discriminant classifier and the nearest mean classifier weighted by the class prior probabilities.

3. EXPERIMENTS AND RESULTS

Experiments were performed on 204 spine radiographs from 204 women enrolled in a post-menopausal osteoporosis trial. The dataset is diverse, ranging from normal spines to spines with several severe fractures. The original radiographs have been scanned and the lumbar vertebrae L1 — L4 were annotated and graded by experienced radiologists. Fractures were identified and graded according to the Genant at al. method of semi-quantitative visual assessment[1] in severity mild, moderate, or severe and type wedge, biconcave, or crush fracture. According to the guidelines fractures are indicated as mild if one of the three heights is between 20% and 25% smaller than the maximum of the heights, moderate if the difference is between 25%, and 40% and severe if the difference is larger than 40%. A total of 89 fractures was found (22 mild, 49 moderate, 18 severe; 60 wedge, 23 biconcave, 6 crush). The average shapes and predictions of the unfractured vertebrae and each of the three fracture types is shown in Figure 1. The normal shapes are predicted very well, whereas all types of fractures reveal a large difference with the predicted normal shape.

A set of leave-one-out experiments was performed in which the prediction models and classifiers for each image were trained on the remaining 203 images. Classifications are performed using the Matlab pattern recognition toolbox *prtools*.[15] Pilot experiments showed that a (regularized) quadratic discriminant classifier had large

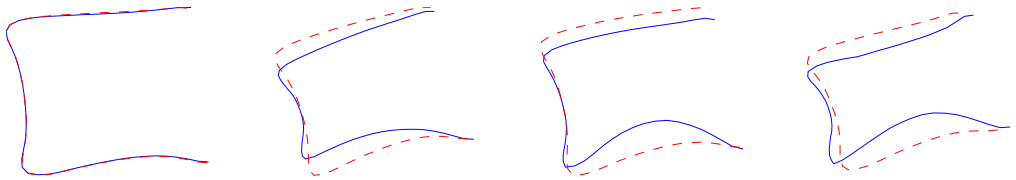

Figure 1. The average shapes (blue lines) and their average predictions (red, dashed lines) for different deformities. From left to right: Normals, wedge fracture, biconcave fracture, crush fracture.

error rates, likely owing to the relatively low number of fractures in the training set which makes it difficult to estimate individual class covariance matrices correctly. In the remainder of the experiments, the regularized linear discriminant classifier of Equation 4 is used with α varying between zero and one.

We compared performance of the following features based on the full contour or on the six point representation:

shape & pred all landmark coordinates of the aligned vertebral contours and their predictions of unfractured shape (the proposed approach)

shape all landmark coordinates of the aligned full vertebral contours

RMS the root mean squared distance from all points where the prediction is outside the true shape to their closest points on the true shape, as was used in[10]

six points all landmark coordinates of the six point shape representation

heights the anterior, middle, and posterior heights

heightratio the relative difference between the minimum and the maximum of the three heights, expressed as a percentage of the maximum height. This is the quantitative criterion used in semi-quantitative grading.

A/P & M/P ratios the ratios between anterior and posterior height and between middle and posterior height

The full contour features are tested both with 25 landmarks and with 50 landmarks in the contour. To reduce the influence of deformities such as osteophytes on the alignment, shapes are aligned using Procrustes alignment based on the four corner points only.

Figure 2 shows the classification error as a function of α for the different feature vectors, where classification error is defined as the percentage of misclassifications with respect to the radiologists' scoring. In nearly all cases, regularization helps improve the classification performance, and as expected, the classification into different fracture types requires more regularization than the two-class classification, as do the higher dimensional feature spaces (50 landmarks compared to 25 landmarks, shape and prediction compared to shape alone, shape measures compared to six point representations). The minimum classification error for each of the features is given in Table 1. The classification using the full contour features (shape and shape & pred) seems to perform slightly better in all cases than using the features derived from six points. In a McNemar's test[16,17] the difference is significant only between the shape based features and A/P & M/P ratios and RMS for the fracture vs normal classification, and between the shape based features and minimum/maximum height ratio and RMS for the classification in fracture types ($p < 0.05$).

In the remainder of the experiments, we use the contour features based on 25 landmarks only. Figure 3 gives the ROC curves for fracture detection on the basis of shape and prediction features. Overall, we are able to distinguish fractures from normals; classification is near-perfect for moderate and severe fractures with an area under the ROC (AUROC) of 1.00 and very good for mild fractures (AUROC 0.95).

Figure 4 shows two examples where our approach, using the reconstructed unfractured shape estimate, was able to detect a mild fracture that was classified as normal using shape alone.

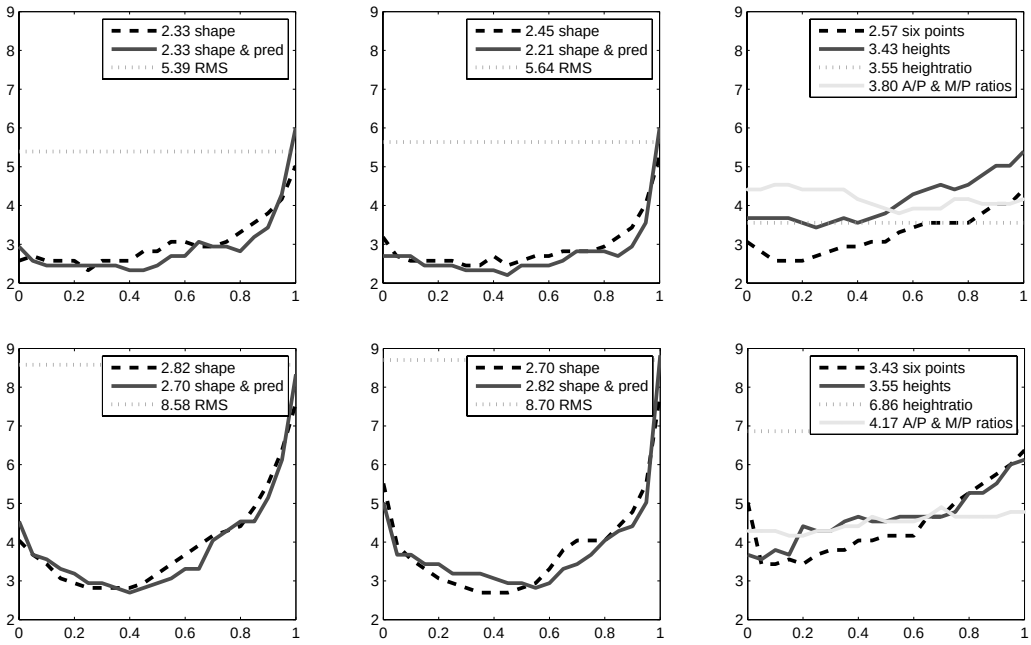

Figure 2. Classification error (percentage) as a function of regularization parameter α. The figure legend shows the minimum error achieved for each of the features. The first row provides the results for classification in fractured and unfractured vertebrae whereas the second row shows the results for classification in different types of deformity (no fracture, wedge, biconcave, or crush fracture). The figures on the left and in the middle depict the errors for the shape model based features: the full shapes ('shape', the full shape and predictions ('shape & pred') and the summarized distance between shape and prediction as in[10] ('RMS'). Left: models with 25 landmarks; Middle: models with 50 landmarks; Right: measures derived from the six point annotation.

	Fractured vs unfractured	Fracture type
shape & pred	2.33	2.70
shape	2.33	2.82
six points	2.57	3.42
heights	3.43	3.55
heightratio	3.55	**6.86**
A/P + M/P ratio	**3.80**	**4.17**

Table 1. Classification errors for the best performing regularization parameter α for each of the feature spaces. The results in bold are significantly different ($p < 0.05$) from the results of classification based on the shape and reconstructions ('shape & pred').

The posterior probability $1 - p(x|w_{\text{unfractured}})$ that a vertebra with features x is fractured can be used as a measure of fracture severity. We found a good correlation (0.91) between this measure and the fracture severity as indicated by the radiologists.

Finally, the confusion matrix for fracture type characterization on the basis of shape and prediction features is shown in Figure 5. The most frequent error is that wedge fractures are classified as normals in 10% of the cases.

4. DISCUSSION

The proposed method relies on full vertebra outlines, for which we used manual annotations by radiologists. This manual annotation procedure is of course time consuming and hampers large-scale use of these methods. Several

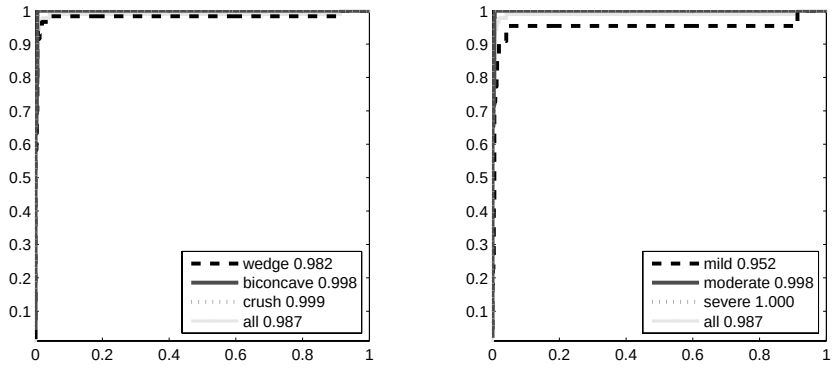

Figure 3. ROC curves for fracture detection, for different fracture type and severity. The numbers in the legend denote the areas under the ROC curve.

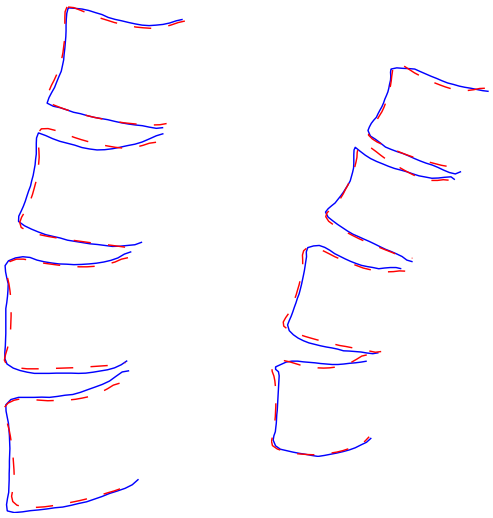

Figure 4. Examples where the patient specific model is able to detect a mild fracture that the radiologist had indicated, whereas the shape model of individual vertebrae alone cannot make the distinction. In both cases, the fracture is in L2 (the second vertebra from the top). The solid blue line denotes the true shape and the dashed, red line the reconstruction.

	N	W	B	C
normal (N)	723	1	3	0
wedge (W)	6	50	2	2
biconcave (B)	0	2	19	2
crush (C)	1	1	2	2

Figure 5. Confusion matrix for fracture type classification with shape and reconstruction. The rows give the true class labels and the columns the determined labels.

authors have previously proposed methods for automatic and semi-automatic spine segmentation from X-ray or dual X-ray absorptiometry (DXA) images with the aim of automating vertebral morphometry.[18–21] Although all these approaches work less good on fractured than on normal vertebrae, results are promising and it seems that at least semi-automatic segmentation would be feasible.

Previously, Smyth et al. used a Mahalanobis distance classifier — a quadratic discriminant classifier with assumed equal class priors — on a vertebra shape representation derived from DXA images to discriminate between fractured and unfractured vertebrae.[18] In a dataset with 128 normal images and 128 images with a fracture, they found a slight, but significant (p=0.03), improvement in fracture detection between the three height representation and the full contour representation. It is difficult to compare the results between these two studies, as the images, modality, and number of fractures are different. Overall, our results are comparable to or slightly better than those obtained by Smyth et al. with areas under the ROC curve of 0.987 vs. 0.957. In contrast to Smyth et al. we distinguish not only unfractured and fractured vertebrae but also different fracture types. This may be useful in the study of the relation between fracture type and symptoms or disease progression, although the need for this has not been clearly established.[22–24] In our study, separating into different fracture types seemed to also slightly improve the recognition of unfractured shapes.

In our experiments we found little difference between a full contour shape representation and the more complex shape and prediction representation. On of the reasons for this could be that the training set is relatively small, especially in the number of fractures, which makes it difficult to estimate large covariance matrices correctly. Another reason could be that our data is relatively homogeneous – although taken in different centers in Denmark, all images were from Danish women in the same age range, which makes the use of patient specific models less important. Previous work by other authors, comparing absolute height measures to their normalized counterparts, seemed to indicate that comparison with a neighboring or reference vertebra does improve results.

Interestingly, the percentage difference between minimum and maximum height, which is the quantitative measure that is used as the basis of the semi-quantitative grading, seems to perform worse in fracture detection (error rate 3.55%) than the shape based representations (2.33%) and even than the six point (2.57%) and three height representations (3.43%), which may indicate that the more complex representations do indeed capture some of the more subtle shape variation that the trained experts include in the subjective judgement.

The most frequent classification error is that wedge fractures are sometimes classified as normals. It must be noted here, that the models are trained on and compared to a surrogate gold standard: the radiologists' scoring performed by different radiologists, with only one scoring per case. In cases of mild shape abnormality, such as age-related bone remodeling that results in a slight 'wedging' of the vertebra, radiologists often don't agree whether it is a mild fracture or not. In clinical trials, it may therefore be of great value to report, instead of the categorical variable, a continuous measure of deformity such as the one proposed in this paper.

5. CONCLUSIONS

We present a shapemodel-based method for classifying and grading vertebral deformities. Statistical models are used of shape variations between and within subjects, for normal as well as fractured vertebrae. This helps discriminating between normal, biological variation and osteoporosis related deformity. The method shows good agreement with manual classification on a large set of images.

ACKNOWLEDGMENTS

We would like to thank our colleagues L.A. Conrad-Hansen (Nordic Bioscience A/S), and M.T. Lund, and R.L. Granlund (University of Copenhagen, Denmark), for providing the software to perform the manual segmentations.

REFERENCES

1. H. K. Genant, C. Y. Wu, C. van Kuijk, and M. C. Nevitt, "Vertebral fracture assessment using a semiquantitative technique.," *Journal of Bone and Mineral Research* **8**, pp. 1137–1148, Sep 1993.

2. H. K. Genant, M. Jergas, L. Palermo, M. Nevitt, R. S. Valentin, D. Black, and S. R. Cummings, "Comparison of semiquantitative visual and quantitative morphometric assessment of prevalent and incident vertebral fractures in osteoporosis The Study of Osteoporotic Fractures Research Group.," *Journal of Bone and Mineral Research* **11**, pp. 984–996, Jul 1996.
3. R. Eastell, S. L. Cedel, H. W. Wahner, B. L. Riggs, and L. J. Melton, "Classification of vertebral fractures.," *Journal of Bone and Mineral Research* **6**, pp. 207–215, Mar 1991.
4. E. V. McCloskey, T. D. Spector, K. S. Eyres, E. D. Fern, N. O'Rourke, S. Vasikaran, and J. A. Kanis, "The assessment of vertebral deformity: a method for use in population studies and clinical trials.," *Osteoporosis International* **3**, pp. 138–147, May 1993.
5. L. Ferrar, G. Jiang, J. Adams, and R. Eastell, "Identification of vertebral fractures: an update.," *Osteoporosis International* **16**, pp. 717–728, Jul 2005.
6. P. D. Delmas, L. van de Langerijt, N. B. Watts, R. Eastell, H. Genant, A. Grauer, D. L. Cahall, and I. M. P. A. C. T. S. Group, "Underdiagnosis of vertebral fractures is a worldwide problem: the IMPACT study," *Journal of Bone and Mineral Research* **20**, pp. 557–563, Apr 2005.
7. E. N. Schwartz and D. Steinberg, "Detection of vertebral fractures.," *Curr Osteoporos Rep* **3**, pp. 126–135, Dec 2005.
8. T. Cootes, C. Taylor, D. Cooper, and J. Graham, "Active shape models – their training and application," *Computer Vision and Image Understanding* **61**(1), pp. 38–59, 1995.
9. M. Lund, M. de Bruijne, L. Tankó, and M. Nielsen, "Shape regression for vertebra fracture quantification," in *Medical Imaging: Image Processing*, M. Fitzpatrick and J. Reinhardt, eds., *Proceedings of SPIE* **5747**, pp. 723–731, SPIE Press, 2005.
10. M. de Bruijne, M. Lund, L. Tankó, P. Pettersen, and M. Nielsen, "Quantitative vertebral morphometry using neighbor-conditional shape models," in *Medical Image Computing & Computer-Assisted Intervention*, R. Larsen, M. Nielsen, and J. Sporring, eds., *Lecture Notes in Computer Science* **4190**, pp. 1–8, Springer, 2006.
11. C. Goodall, "Procrustes methods in the statistical analysis of shape," *Journal of the Royal Statistical Society B* **53**(2), pp. 285–339, 1991.
12. R. Duda, P. Hart, and D. Stork, *Pattern Classification*, John Wiley & Sons, second ed., 2001.
13. J. Friedman, "Regularized discriminant analysis," *Journal of the American Statistical Association* **84**(405), pp. 165–175, 1989.
14. T. Hastie, A. Buja, and R. J. Tibshirani, "Penalized discriminant analysis," *The Annals of Statistics* **23**(1), pp. 73–102, 1995.
15. F. van der Heijden, R. Duin, D. de Ridder, and D. Tax, *Classification, Parameter Estimation and State Estimation: An Engineering Approach Using MATLAB*, wiley, 2004.
16. T. G. Dietterich, "Approximate statistical test for comparing supervised classification learning algorithms," *Neural Computation* **10**(7), pp. 1895–1923, 1998.
17. D. Altman, *Practical Statistics for Medical Research*, Chapman & Hall, 1991.
18. P. Smyth, C. Taylor, and J. Adams, "Vertebral shape: Automatic measurement with active shape models," *Radiology* **211**(2), pp. 571–578, 1999.
19. G. Zamora, H. Sari-Sarrafa, and R. Long, "Hierarchical segmentation of vertebrae from X-ray images," in *Medical Imaging: Image Processing*, M. Sonka and M. Fitzpatrick, eds., *Proceedings of SPIE* **5032**, pp. 631–642, SPIE Press, 2003.
20. M. de Bruijne and M. Nielsen, "Image segmentation by shape particle filtering," in *International Conference on Pattern Recognition*, J. Kittler, M. Petrou, and M. Nixon, eds., pp. III:722–725, IEEE Computer Society Press, 2004.
21. M. G. Roberts, T. F. Cootes, and J. E. Adams, "Vertebral shape: Automatic measurement with dynamically sequenced active appearance models.," in *Medical Image Computing & Computer-Assisted Intervention*, J. Duncan and G. Gerig, eds., *Lecture Notes in Computer Science* **3750**, pp. 733–740, Springer, 2005.
22. G. P. Lyritis, B. Mayasis, N. Tsakalakos, A. Lambropoulos, S. Gazi, T. Karachalios, M. Tsekoura, and A. Yiatzides, "The natural history of the osteoporotic vertebral fracture.," *Clin Rheumatol* **8 Suppl 2**, pp. 66–69, Jun 1989.

23. B. Ettinger, D. M. Black, M. C. Nevitt, A. C. Rundle, J. A. Cauley, S. R. Cummings, and H. K. Genant, "Contribution of vertebral deformities to chronic back pain and disability. the study of osteoporotic fractures research group.," *Journal of Bone and Mineral Research* **7**, pp. 449–456, Apr 1992.
24. A. A. Ismail, C. Cooper, D. Felsenberg, J. Varlow, J. A. Kanis, A. J. Silman, and T. W. O'Neill, "Number and type of vertebral deformities: epidemiological characteristics and relation to back pain and height loss. european vertebral osteoporosis study group.," *Osteoporosis International* **9**(3), pp. 206–213, 1999.