

Quantitative Vertebral Morphometry Using Neighbor-Conditional Shape Models

Marleen de Bruijne¹, Michael T. Lund¹,
László B. Tankó², Paola C. Pettersen², and Mads Nielsen¹

¹ IT University of Copenhagen, Denmark

² Center for Clinical and Basic Research, Ballerup, Denmark
marleen@itu.dk

Abstract. A novel method for vertebral fracture quantification from X-ray images is presented. Using pairwise conditional shape models trained on a set of healthy spines, the most likely normal vertebra shapes are estimated conditional on all other vertebrae in the image. The differences between the true shape and the reconstructed normal shape is subsequently used as a measure of abnormality. 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.

The method is demonstrated on 212 lateral spine radiographs with in total 78 fractures. The distance between prediction and true shape is 1.0 mm for unfractured vertebrae and 3.7 mm for fractures, which makes it possible to diagnose and assess the severity of a fracture.

1 Introduction

Osteoporosis is a common skeletal disorder characterized by a decrease in bone mass, leading to bone fragility and an increased risk of fractures. It is a major public health problem; one out of every three women and one out of eight men over the age of 50 is expected to have an osteoporosis-related fracture in the remainder of their lives. Any bone can be affected but the fractures typically occur in the hip, spine, and wrist. Of these, hip fractures are the most serious in terms of morbidity and mortality. Vertebral fractures can be asymptomatic, but can also have serious consequences, including severe back pain and deformity. Furthermore, vertebral fractures are the most common osteoporotic fracture, they occur in younger patients, and their presence is known to be a good indicator for the risk of future spine and hip fractures. This makes vertebral fracture assessment suitable as an outcome variable in clinical trials to diagnose osteoporosis incidence and progression.

Vertebral fractures are conventionally detected and graded on lateral X-rays. Numerous methods have been proposed for this purpose, typically relying on a subjective judgement of vertebral appearance by an expert radiologist in connection with six-point morphometry. In the latter, 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 derived from these heights or from the ratios between these heights, 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 (see for instance [1, 2, 3, 4]).

Several authors have proposed methods for automatic spine segmentation from X-ray or dual X-ray absorptiometry (DXA) images with the aim of automating vertebral morphometry [5, 6, 7, 8]. In this work, we start from the segmented vertebrae and aim at improving the diagnosis of fractures with respect to the current height based measurements. Currently, the diagnosis of a vertebral fracture in clinical trials relies on a 20% decrease in body height of a vertebra. Due to the categorical nature of this study parameter, clinical trials continue to demand a large number of participants and long follow-up time.

One shortcoming of the conventional models is that the sparse representation of six points is unable to capture subtle shape changes. Smyth et al. [5] used point distribution models to represent the full contour of normal and fractured vertebrae and detected fractures using a Mahalanobis distance classifier on this representation. A slight, but significant improvement with respect to conventional height measurements was reported.

We propose to model not only the shape variation over a population for individual vertebrae, but to also model the interrelations between vertebrae in the same subject. This additional prior information allows adjustment of the models to individual patients so as to distinguish normal biological shape variation from osteoporosis-related deformation. We use conditional shape models to predict the most likely shape of a vertebra given the known shape of a neighboring vertebra. If the 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 a previous paper we showed that vertebrae in healthy spines can be accurately reconstructed from their neighbors[9].

In the current paper we show how the reliability of predicted shapes can be estimated and we use this to combine pairwise predictions of all vertebrae in the image into one single, optimal prediction for each vertebra. The difference between the prediction and the true, segmented shape provides a (continuous) measure of fracture severity.

2 Shape Estimation

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 [10]. 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 $\bar{\mathbf{x}}$, 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:

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

where Φ_t consists of the eigenvectors corresponding to the t largest eigenvalues, $\Phi_t = (\phi_1 | \phi_2 | \dots | \phi_t)$, \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 Shapes

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

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

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

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

$\Sigma_{12}\Sigma_{22}^{-1}$ is the matrix of regression coefficients of $S_1 - \mu_1$ on $S_2 - \mu_2$. Usually, Σ_{22} is not invertible owing to multi-collinearity in the landmark positions and unreliable due to chance covariance in a limited training set. Some regularization is therefore required. One option is to replace Σ_{22} by $\Sigma_{22} + \gamma I$, where γ is a positive and typically small constant. This approach is known as ridge regression [12]. As γ tends to infinity, the influence of the shape S_2 decreases, and the remaining model is the original model for S_1 , describing the shape variation independent of S_2 .

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.

One can choose to model both pose and shape with respect to the given vertebra, which means that the training shape pairs should be aligned together, on the basis of the transformations that optimally align the predictor vertebra in all training shapes. Alternatively, one could model the shape variation alone and leave out any possible correlations between shapes and relative position, scale, and rotation. In that case, the vertebrae should be aligned independently.

2.2 Estimator Uncertainty

The pairwise shape prediction 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 its 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:

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

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

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

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

2.3 Combining Shape Estimates

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 i th prediction of S_1 is then given by

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

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. Note, that the regression models will typically produce shapes that are inside the model subspace — apart from some minor differences owing to the cut off at t eigenmodes — and thus in this case the residuals r_i and shape parameters b_i are negligible and the estimate reduces to a constant that is proportional to the inverse of the total variance in the conditional model. This constant is independent of both the predictor and the predicted shape and expresses the amount of correlation between the two shape models. In the pairwise vertebra predictions, the fact that direct neighbors contain the most useful information for predicting a shape is encoded in this term.

3 Vertebral Fracture Quantification

To summarize, we propose the following procedure for fracture quantification from segmented spine images:

- construct models of vertebral shape variation for all pairs of vertebrae in an image, according to Section 2
- for each pair of vertebrae in a new image:
 - align the predictor shape with the model
 - perform shape regression using Equation 1
 - determine reliability weights according to Equation 3
- combine all predictions for each vertebra as a weighted sum
- if pose is not included in the model, align prediction to true shape
- measure fracture severity

Various measures can be used to express the difference between the true shape and the predicted shape as a measure of vertebral deformity. In this work we ignore the points where the true shape is outside the prediction, which is usually caused by osteophytes. We then use the root mean squared (RMS) distance from all points where the prediction is outside the true shape to their closest points on the true shape.

4 Experiments

Our database currently consists of 212 lateral, lumbar spine radiographs taken from different programs of osteoporosis and atherosclerosis screening in post-menopausal women. The dataset is diverse, ranging from normal spines to spines with several severe fractures. The original radiographs have been scanned at a resolution of either 300 or 570 dpi and the lumbar vertebrae L1 — L4 were annotated and graded by experienced radiologists.

The outlines of the vertebrae were drawn manually and the corner points and mid-points of the vertebral endplates were indicated. Fractures were identified and graded according to the Genant et al. method of semi-quantitative visual assessment [1] in severity mild, moderate, or severe and type wedge, biconcave, or crush fracture. A total of 78 fractures was identified in 64 spines; 148 of the spines were unfractured.

A set of leave-one-out experiments is performed in which the models are trained on all available healthy vertebra shapes except those of the patient under study. To avoid including mild fractures in the model, shapes of which the difference between the maximum and minimum heights exceeds 15% of the maximum height are excluded from training.

4.1 Parameter Settings

A total of 52 landmarks is placed along the upper, anterior, and lower boundary of each vertebra, interpolated equidistantly between the four vertebral corners. The corners are defined as the points on the contour closest to the corner points that were used for standard six-point morphometry. Shapes are aligned individually using translation and rotation without scaling, since vertebra size is expected to correlate with important

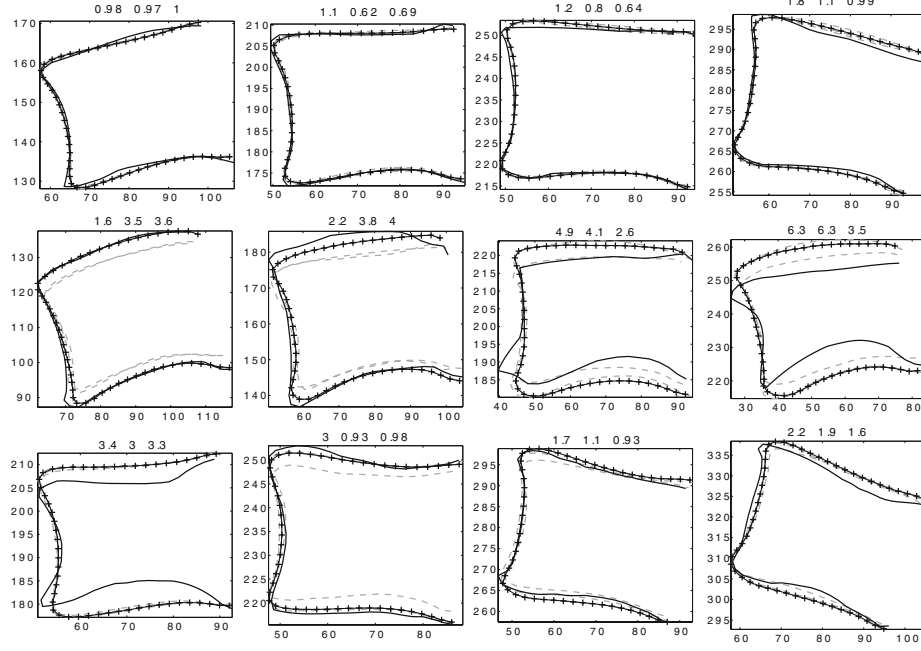


Fig. 1. Each row shows the vertebrae L1 — L4 of the same image; the true shape (black line), three pairwise predictions (gray dashed lines), and the combined prediction (black line with pluses). The numbers above the plots give the RMS distance to the true shape for each of the three individual predictions. The top row depicts a normal spine; row 2 contains 2 fractures (vertebra 3 and 4); row 3 was graded as all normals but vertebra 1 had a large shape difference in our method.

shape information on fracture type and severity. The number of modes is selected so that 95% of the variance is kept in the model. The residual variance and the regularization parameters γ were optimized separately for each pairwise regression model using leave-one-out validation on the training set of normal vertebrae.

4.2 Results

Some examples of predictions obtained are given in Figure 1. Most of the normal shapes are predicted accurately. The second row in Figure 1 contains two fractures, which results in two of the three predictions being consistently smaller than the true shape. However, the model is able to detect that those shapes are less likely and the weighted estimate for the normal vertebrae is still close to the correct shape.

Figure 2 summarizes the prediction results of all 212 images, separated into fracture type and grade. The RMS shape-to-contour distance between the predicted shape and the true contour was 1.0 mm on average for unfractured vertebrae and 3.67 mm for fractures. The area under the ROC curve is 0.93; at a sensitivity of 95% the specificity is 84%. The areas under the ROC curve for mild, moderate, and severe fractures separately are 0.92, 0.96, and 0.99 respectively.

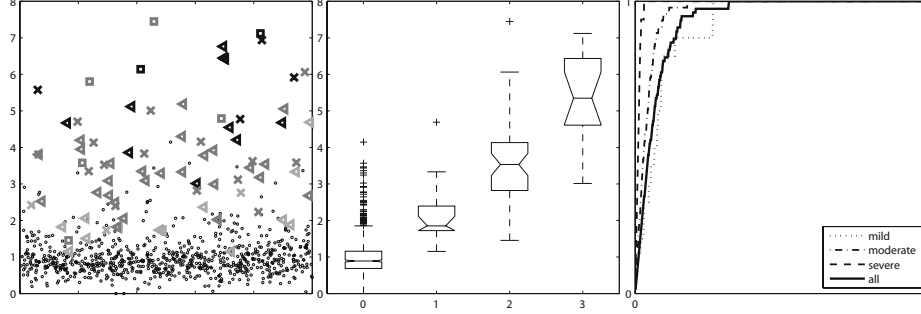


Fig. 2. (a) Measured RMS distances in mm between the predicted shapes and true shapes. To allow discrimination of similar measures the distances are set out against a random value on the horizontal axis. The symbols denote the type of vertebral deformity: Normal (circles); wedge fracture (triangles); biconcave fracture (crosses); and crush fracture (squares). For the fractures, the grey value denotes the fracture severity: mild (light grey); moderate (darker grey); and severe (black) **(b)** Matlab box and whisker plot of the average distances between predicted and true shapes separated into fracture grade, ranging from normal (0) to severely fractured (3). The box has lines at the lower quartile, median, and upper quartile values. The whiskers extending from each end of the box show the extent of the rest of the data. Data with values outside 1.5 times the interquartile range from the median are considered to be outliers and are denoted by pluses. **(c)** ROC curves of vertebral fracture detection.

Overall, we are able to distinguish fractures from normals and there is a fair correlation (0.81) between the shape distance and the fracture severity as indicated by the radiologists. However, Figure 1 reveals that the class of normal vertebrae contains a relatively large number of outliers that have a larger difference between the prediction and the true shape. One of these outliers is shown in the third row of Figure 1. The leftmost vertebra may be a very mild ‘fracture’ that falls outside the capture range of the standard semi-quantitative morphometry and that the shape prediction, more sensitive to subtle changes, can already identify. This should be further validated using for instance longitudinal studies. Visual inspection of the dataset revealed that many of the unfractured vertebrae that had a large shape distance exhibit this type of deformation.

5 Discussion and Conclusion

We propose a shape model based approach to vertebral fracture quantification in which an observed vertebral shape is compared to its reconstructed normal shape as can be predicted from its neighbors. Compared to the current standard of semi-quantitative morphometry which is based on three height measurements per vertebra, this method provides a richer description of deformation and may be able to detect more subtle shape changes while maintaining specificity. This could lead to earlier diagnosis and reduce the number of participants and/or the follow-up time required in clinical trials assessing the efficacy of drug candidates.

Currently, the shape differences are summarized into one average distance. This enables detection of deformities and specification of the degree of abnormality, but not

the type. If a sufficiently large training set of different types of deformity is available, it may be useful to take the shape difference vectors — or a low-dimensional representation of them — as input in a supervised classification scheme. Such explicit models of deformities could also be incorporated in the step of combining different shape estimates, such that the estimate of uncertainty is not based only on the distance to known normal shapes but also on the distance to known deformities.

In the current setup, vertebral shapes were reconstructed with an accuracy of on average 1.0 mm for normal vertebrae and 3.7 mm for fractures, leading to fracture detection with a sensitivity of 95% at a specificity of 84%.

References

1. Genant, H.K., Wu, C.Y., van Kuijk, C., Nevitt, M.C.: Vertebral fracture assessment using a semiquantitative technique. *J Bone Miner Res* **8**(9) (1993) 1137–1148
2. Genant, H.K., Jergas, M., Palermo, L., Nevitt, M., Valentin, R.S., Black, D., Cummings, S.R.: Comparison of semiquantitative visual and quantitative morphometric assessment of prevalent and incident vertebral fractures in osteoporosis The Study of Osteoporotic Fractures Research Group. *J Bone Miner Res* **11**(7) (1996) 984–996
3. McCloskey, E.V., Spector, T.D., Eyres, K.S., Fern, E.D., O'Rourke, N., Vasikaran, S., Kanis, J.A.: The assessment of vertebral deformity: a method for use in population studies and clinical trials. *Osteoporos Int* **3**(3) (1993) 138–147
4. Ferrar, L., Jiang, G., Adams, J., Eastell, R.: Identification of vertebral fractures: an update. *Osteoporos Int* **16**(7) (2005) 717–728
5. Smyth, P., Taylor, C., Adams, J.: Vertebral shape: Automatic measurement with active shape models. *Radiology* **211**(2) (1999) 571–578
6. Zamora, G., Sari-Sarrafa, H., Long, R.: Hierarchical segmentation of vertebrae from X-ray images. In Sonka, M., Fitzpatrick, M., eds.: *Med Imaging: Image Process.* Volume 5032 of *Proc of SPIE.*, SPIE Press (2003) 631–642
7. de Bruijne, M., Nielsen, M.: Image segmentation by shape particle filtering. In Kittler, J., Petrou, M., Nixon, M., eds.: *ICPR, IEEE Computer Society Press* (2004) III:722–725
8. Roberts, M.G., Cootes, T.F., Adams, J.E.: Vertebral shape: Automatic measurement with dynamically sequenced active appearance models. In Duncan, J., Gerig, G., eds.: *MICCAI.* Volume 3750 of *LNCS.*, Springer (2005) 733–740
9. Lund, M., de Bruijne, M., Tankó, L., Nielsen, M.: Shape regression for vertebra fracture quantification. In Fitzpatrick, M., Reinhardt, J., eds.: *Med Imaging: Image Process.* Volume 5747 of *Proc of SPIE.*, SPIE Press (2005) 723–731
10. Cootes, T., Taylor, C., Cooper, D., Graham, J.: Active shape models – their training and application. *Comput Vis Image Underst* **61**(1) (1995) 38–59
11. Goodall, C.: Procrustes methods in the statistical analysis of shape. *Journal of the Royal Statistical Society B* **53**(2) (1991) 285–339
12. Hoerl, A., Kennard, R.: Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* **12**(1) (1970) 55–67