

# Quantitative vertebral morphometry using neighbor-conditional shape models

Marleen de Bruijne <sup>a,b,\*</sup>, Michael T. Lund <sup>a</sup>, László B. Tankó <sup>b,c</sup>,  
Paola C. Pettersen <sup>c</sup>, Mads Nielsen <sup>a,b</sup>

<sup>a</sup> Department of Computer Science, University of Copenhagen, Denmark

<sup>b</sup> Nordic Bioscience A/S, Herlev, Denmark

<sup>c</sup> Center for Clinical and Basic Research, Ballerup, Denmark

Received 16 April 2007; received in revised form 16 July 2007; accepted 17 July 2007

Available online 26 July 2007

## 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 the shapes of all other vertebrae in the image. The difference 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 develops a patient-specific reference by combining population-based information on biological variation in vertebral shape and vertebra interrelations, and it provides a continuous measure of deformity.

The method is demonstrated on 282 lateral spine radiographs with in total 93 fractures. Vertebral fracture detection is shown to be in good agreement with semi-quantitative scoring by experienced radiologists and is superior to the performance of shape models alone. © 2007 Elsevier B.V. All rights reserved.

**Keywords:** Shape analysis; Conditional shape model; Shape regression; Vertebral fracture; Osteoporosis

## 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 two 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 (Sambrook and Cooper, 2006). 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 acute and chronic back pain, back deformity, and increased mortality (Nevitt et al., 1998; Center et al., 1999). 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 (Melton et al., 1999; Ismail et al., 1999). This makes the presence of vertebral fractures an important factor in clinical decision making and the primary endpoint in many clinical trials to assess osteoporosis incidence and monitor its progression.

Vertebral fractures are conventionally detected and graded on lateral X-rays by experienced radiologists using a semi-quantitative grading scheme proposed by Genant et al. (1993). 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

\* Corresponding author. Address: Department of Computer Science, University of Copenhagen, Denmark. Tel.: +45 3532 1445; fax: +45 3532 1401.

E-mail address: [marleen@diku.dk](mailto:marleen@diku.dk) (M. de Bruijne).

URL: <http://image.diku.dk/marleen> (M. de Bruijne).

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. This method was shown to have a good intra-observer agreement but may be difficult to apply and to standardize between different centers. As an alternative, 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 (see for instance Eastell et al., 1991; Melton et al., 1993; McCloskey et al., 1993; Grados et al., 2001; Lunt et al., 2002; Ferrar et al., 2005). These methods do not require a trained radiologist and may be easier to standardize between centers but are unable to capture subtle shape differences and suffer from variability in point placement. Studies have shown that a large number of fractures goes undiagnosed with current semi-quantitative and quantitative methods (Delmas et al., 2005; Schwartz and Steinberg, 2005). More precise and objective measures of vertebral deformity are therefore needed.

One shortcoming of the conventional methods to assess vertebral fractures is that the sparse representation of six points is unable to capture subtle shape changes. Smyth et al. (1999) used point distribution models (Cootes et al., 1995) to represent the full vertebral contour and observed a slight, but significant improvement in fracture detection with respect to conventional height measurements.

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 further distinguish normal biological shape variation from osteoporosis-related deformation. We use point distribution models to model variations in shape of individual vertebrae and conditional point distribution models to reconstruct the most likely shape of a vertebra given the known shape of another vertebra in the image. If both these models are built from a training set of normal, healthy spines this provides an estimate of what the vertebra shape would have been if it were normal. Pairwise predictions of normal shape are performed for all possible pairs of vertebra shapes in the image and the reliability of each of the shape reconstructions is estimated. Subsequently, all individual estimates are combined into one single, optimal estimate for each vertebra through a weighted summation. The difference between the final reconstruction and the true, segmented shape provides a measure of abnormality and thus of fracture severity. The proposed procedure of fracture quantification is illustrated in Fig. 1.

This paper is an extension of work previously published in de Bruijne et al. (2006) and Lund et al. (2005).

## 2. Shape reconstruction

This section describes the methods used for reconstructing the healthy vertebra shapes. The model of individual vertebrae, a standard PDM, is briefly described in Section 2.1. Section 2.2 presents the conditional shape model used for modeling interrelations between vertebrae in the image. Finally, Section 2.3 explains how the reliability of each shape can be determined and how this can be used to combine the individual reconstructions into one single estimate per vertebra.

### 2.1. Point distribution models

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 et al. (1995). PDMs model the shape probability distribution as a multivariate Gaussian in a subspace of reduced dimensionality. Shapes are defined by vectors containing the coordinates of a set of landmark points that correspond between different shape instances and that are typically located on the boundaries of the objects to model. A collection of training shapes is aligned using for instance Procrustes analysis (Goodall, 1991) to remove position, orientation, and possibly size variations, 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  $\Sigma$ , and the eigensystem of  $\Sigma$  are computed. The eigenvectors  $\phi_i$  of  $\Sigma$  provide the so-called *modes of shape variation* that describe a joint displacement of all landmarks. The eigenvectors corresponding to the largest eigenvalues  $\lambda_i$  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  $\Phi_t$  consists of the eigenvectors  $\phi$  corresponding to the  $t$  largest eigenvalues,  $\Phi_t = (\phi_1 | \phi_2 | \dots | \phi_t)$ ,  $\mathbf{b}$  is a vector of model parameters that specify the contribution of each of the modes, and  $\mathbf{r}$  is a vector of residual shape variation outside of the model subspace.

An example of the modes of variation of a vertebral shape model is given in Fig. 2.

### 2.2. 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

$$\mu = \mu_1 + \Sigma_{12} \Sigma_{22}^{-1} (S_2 - \mu_2) \quad (2)$$

$$K = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \quad (3)$$

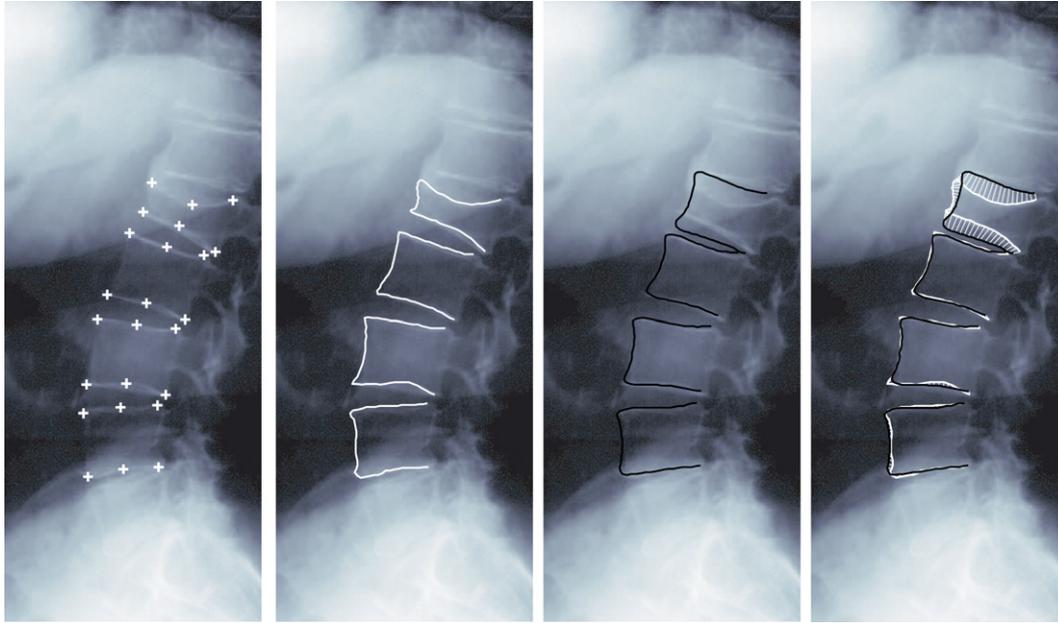


Fig. 1. Fracture quantification. From left to right: Original lateral X-ray image of a lumbar spine with one severe fracture. The pluses indicate the points to measure the anterior, middle, and posterior heights as annotated by a radiologist. Manual contour annotation. Reconstructed unfractured shapes for each of the vertebrae. The difference between the two shapes is a measure of vertebra abnormality.

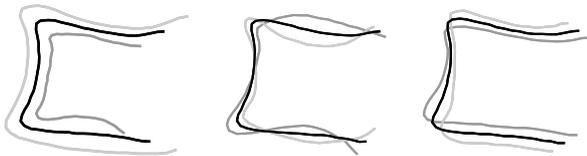


Fig. 2. The first three modes of variation of the shape model of individual shape. The black line is the mean shape, the gray lines are the mean shape plus and minus four standard deviations of the respective modes.

where  $\mu_1$  and  $\mu_2$  are the mean shapes of the training sets of  $S_1$  and  $S_2$ , and covariances  $\Sigma_{ij}$  are obtained from the combined covariance matrix

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \quad (4)$$

as

$$\Sigma_{ij} = \frac{1}{n-1} \sum_{k=1}^n (S_{ik} - \mu_i)(S_{jk} - \mu_j)^T,$$

with  $n$  the number of shapes in the training set. In this expression,  $\mu$  is the maximum likelihood estimate of  $S_1$  given  $S_2$ , and  $K$  is the variance in the estimate.

$\Sigma_{12}\Sigma_{22}^{-1}$  in Eq. (2) is the matrix of regression coefficients of  $(S_1 - \mu_1)$  on  $(S_2 - \mu_2)$ . Usually,  $\Sigma_{22}$  is not invertible owing to multi-collinearity in the landmark positions and unreliable because of chance covariance in a training set of limited size. Some regularization is therefore required. One option is to replace  $\Sigma_{22}$  by  $\Sigma_{22} + \gamma I$ , where  $\gamma$  is a positive and typically small constant. This approach is known as ridge regression (Hoerl and Kennard, 1970). As  $\gamma$  tends to infinity, the influence of the shape  $S_2$  decreases in

Eqs. (2) and (3), and the remaining model is the original model for  $S_1$  describing the shape variation independent of  $S_2$ . A suitable value for  $\gamma$  can be selected for instance using leave-one-out validation on a training set or by generalized cross validation (Golub et al., 1979).

One can choose to model both pose and shape of the unknown object with respect to the given object, or alternatively model the shape variation alone and leave out any possible correlations between shapes and relative position, scale, and rotation. In the first case the training shape pairs should be aligned together on the basis of the transformations that optimally align all predictor shapes in the training set. In the second case, the predictor and predicted shapes should be aligned independently. This is the approach we take in this paper. An example of the resulting conditional model is given in Fig. 3.

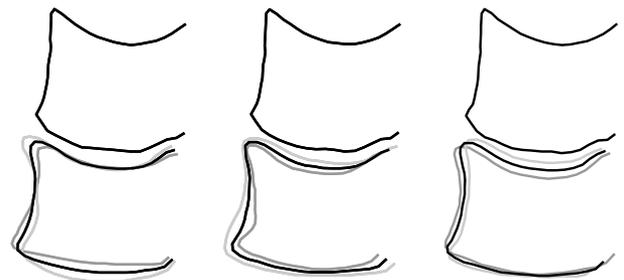


Fig. 3. The first three modes of variation of the conditional shape model of the lower vertebra given the known shape of the upper neighboring vertebra. The black line is the mean shape, the gray lines are the mean shape plus and minus four standard deviations of the respective modes.

### 2.3. Combining shape estimates

The pairwise shape reconstruction results in several shape estimates for each vertebra. Not all of these estimates will be equally accurate. For instance the vertebral shapes of two direct neighbors are expected to be more strongly correlated than two vertebrae that are further apart. In a fractured spine, the fractured vertebrae will likely produce inaccurate estimates of normal vertebra shape, even for their direct neighbors. We define the final shape estimate as a weighted combination of the individual estimates, where the weights express the degree of belief in each of the individual estimates.

We assume that the shape of normal vertebrae as they are observed in the image are produced by the underlying shape model of normal shapes, resulting in a multivariate Gaussian with variances  $\lambda_i$  in  $t$  directions, plus additional uncorrelated Gaussian noise with a variance  $\sigma_r^2$  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:

$$p(S|\theta) = c_s c_r \exp \left[ -\frac{1}{2} (M_s + M_r) \right] \quad (5)$$

$$c_s = \frac{1}{\sqrt{(2\pi)^t \prod_{i=1}^t \lambda_i}}, \quad M_s = \sum_{i=1}^t \frac{b_i^2}{\lambda_i} \quad (6)$$

$$c_r = \frac{1}{\sqrt{(2\pi)^n \sigma_r^{2n}}}, \quad M_r = \frac{|\mathbf{r}|^2}{\sigma_r^2} \quad (7)$$

where  $b_i$  are the model parameters from the PDM,  $\mathbf{r}$  is a vector of residuals, and  $M_s$  and  $M_r$  are the squared Mahalanobis distances of  $S$  to the mean of the shape model and the residual model respectively.

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 given shape  $S_2$  is a valid normal (unfractured) shape, and the variance in  $P(S_1|S_2)$  expresses the uncertainty in the estimate of  $S_1$  from the model conditional on  $S_2$ . The weight for the  $i$ th estimate of  $S_1$  is then given by

$$w_i = \frac{P(S_1|S_i)P(S_i)}{\sum_i P(S_1|S_i)P(S_i)}, \quad (8)$$

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 Eq. (5) and the mean and covariance of the training set for the predictor shapes for the model parameters  $\theta$ , whereas for  $P(S_1|S_i)$  the estimated shape for  $S_1$  and the mean and covariance of the conditional model must be substituted. Note, that the estimated shape for  $S_1$  is exactly the mean of the conditional model, and thus in this case the residuals  $r_i$  and shape parameters  $b_i$  are negligible and the estimate  $P(S_1|S_i)$

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 estimated shape and expresses the amount of correlation between the two shape models. In the pairwise vertebra reconstructions, the fact that direct neighbors contain the most useful information for reconstructing a shape is encoded in this term.

### 3. Vertebral fracture quantification

Using the methods of the previous section, we propose the following procedure for fracture quantification from segmented spine images:

- (1) Construct models of individual vertebral shape variation according to Section 2.1.
- (2) Construct covariance matrices of paired vertebral shape variation for all pairs of vertebrae in an image, according to Eq. (4).
- (3) For each pair of vertebrae in a new image:
  - align the predictor shape with the model;
  - perform shape regression using Eq. (1);
  - determine reliability weights according to Eq. (8);
  - swap roles of predictor/predicted shape.
- (4) Combine all estimates for each vertebra as a weighted sum.
- (5) Align reconstruction to true shape (if pose not included in the model).
- (6) Measure shape difference.

Various measures can be used to express the difference between the true shape and the reconstructed shape as a measure of vertebral deformity. In the experiments in this paper we investigate the following four measures:

- (1) The average landmark-to-landmark distance, measured between corresponding landmarks on the two shapes.
- (2) The average landmark-to-contour difference, measured by the differences from each landmark on the true shape to its closest point on the reconstruction.
- (3) The latter measured only for the landmarks of the true shape that lie inside the reconstruction.
- (4) The approximation error when the vector of differences between the two shapes is approximated by its projection on a PCA subspace derived from the differences between observed shapes and their reconstructions in a training set of normal spines. Since the PCA model is constructed from normal (unfractured) shapes, the deviation from the measured shape difference to the nearest plausible difference according to the model can be seen as a measure of abnormality. To ensure that the model describes only plausible, normal shape differences, projected differences are constrained to lie within

a Mahalanobis distance of  $k$  from the mean, as is normally done in for instance Active Shape Models (Cootes et al., 1995):

$$\mathbf{b}' = \begin{cases} \mathbf{b}, & M \leq k \\ \mathbf{b} \frac{k}{M}, & M > k \end{cases} \quad (9)$$

with  $\mathbf{b}$  the PC coefficients and  $M$  the Mahalanobis distance to the mean difference vector,  $M = \sqrt{\sum_i \frac{b_i^2}{\lambda_i}}$ . The approximation error is then measured as the sum of squares of residuals between the measured difference vector and its approximation.

## 4. Experiments

### 4.1. Data

Our database consists of 282 lateral, lumbar spine radiographs of 205 post-menopausal women selected from a cohort of Danish women that was followed for assessment of osteoporosis and atherosclerosis in the Prospective Epidemiological Risk Factors (PERF) study (Bagger et al., 2006). The images had been selected from the larger cohort as a subset having relatively many vertebral fractures and covering the complete spectrum of degree of calcification of the abdominal aorta. All X-rays were taken between 1990 and 2005 in several different centers in Denmark. To standardize the images to some extent, the focus-film distance was kept at approximately 1.2 m for all examinations and the central beam was directed to L2 when examining the lumbar spine. The patients were asked to hold their breath for the time of acquiring the radiograph. Only X-ray studies that covered the entire lumbar spine in one image were used. The resulting 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 one of three experienced radiologists. The fracture score assigned by the radiologist is used as the ground truth in this study.

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., 1993) method of semi-quantitative visual assessment in severity mild, moderate, or severe. 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 93 fractures (23 mild; 52 moderate; 18 severe) was identified in 80 spines; 202 of the spines were unfractured.

A set of leave-one-out experiments is performed in which both the individual shape models and the pairwise conditional models are trained on all available unfractured vertebra shapes except those of the patient under study.

### 4.2. 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. Separate models are constructed for L1–L4. Shapes are aligned individually so that only interrelations of shape and not pose are included in the model. Since vertebra size is expected to correlate with important shape information on fracture type and severity, Procrustes alignment is done using translation and rotation without scaling (Dryden and Mardia, 1998).

The number of modes  $t$  (Eq. (6)) is selected so that 95% of the variance is kept in the model. For each pairwise regression model the best regularization parameter  $\gamma$  is chosen out of 30 candidate values regularly sampled from 0.5 to 10, using leave-one-subject-out validation on the training set of unfractured vertebrae. The residual variance  $\sigma_r$  (Eq. (7)) is estimated from the reconstruction results on the training set using these values of  $\gamma$ .

### 4.3. Results

Some examples of reconstructions obtained are given in Fig. 4. Most of the normal shapes are reconstructed accurately. The second row in Fig. 4 contains two fractures, which results in two of the three estimates 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.

The average shapes and reconstructions of all of the unfractured vertebrae and all fractured vertebrae are shown in Fig. 5. The normal shapes are reconstructed very well, whereas the fractures on average reveal a large difference with the reconstructed normal shape. The average distances between the true shape and reconstructed healthy shape is 1.2 mm (average landmark-to-landmark distance  $1.2 \pm 0.5$ ; average landmark-to-contour distance  $0.8 \pm 0.4$ , square root of mean squared (RMS) landmark-to-landmark distance  $1.4 \pm 0.6$ ) for unfractured vertebrae and 3.5 mm for fractures (average landmark-to-landmark distance  $3.5 \pm 1.4$ ; average landmark-to-contour distance  $2.8 \pm 1.3$ , RMS landmark-to-landmark distance  $3.8 \pm 1.5$ ). The distances are slightly larger for L1 and L4 that have only one direct neighbor: L1  $1.21 \pm 0.60$ ; L2  $1.13 \pm 0.47$ ; L3  $1.18 \pm 0.49$ ; L4  $1.43 \pm 0.56$  (average landmark-to-landmark distances). The reconstruction accuracy of healthy L1 vertebra is not significantly lower than that of L2 ( $p = 0.08$  in a  $t$ -test) and L3 ( $p = 0.45$ ), but the difference with L4 is significant ( $p < 0.001$ ).

To further analyze the differences between the reconstructed shapes and the true shapes, a PCA was performed on the difference vectors. The resulting modes of variation, for normal and fractured vertebra, are shown

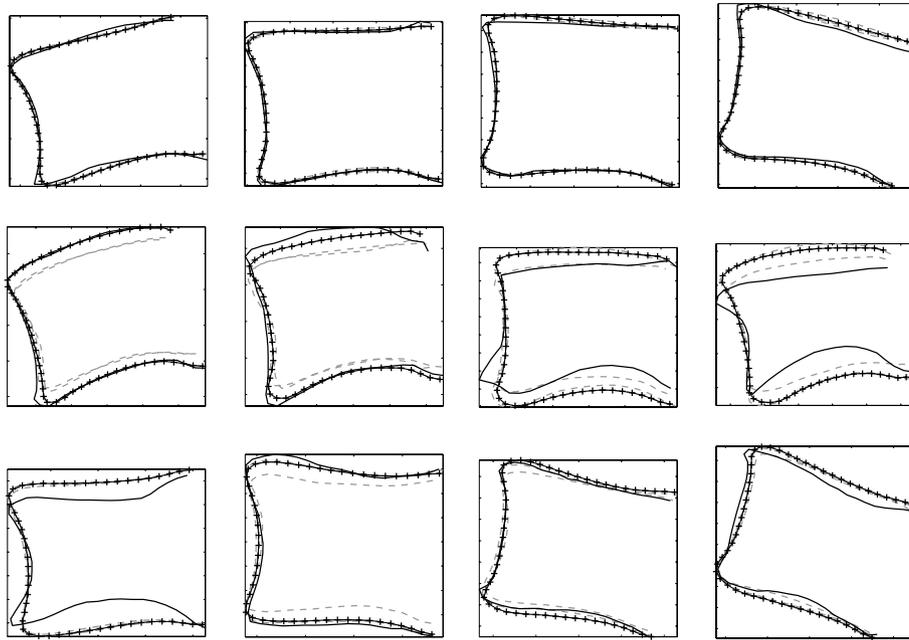


Fig. 4. Each row shows the vertebrae L1–L4 of the same image; the true shape (black line), three pairwise reconstructions (gray dashed lines), and the combined reconstruction (black line with pluses). The top row depicts a normal spine; row 2 contains 2 fractures (vertebra 3 and 4); row 3 was graded as all normals but vertebra 1 had a large shape difference in our method.

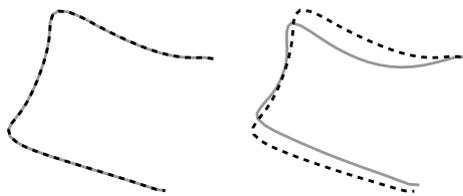


Fig. 5. The average shapes (gray lines) and their average reconstructions (black, dashed lines) for all unfractured shapes (left) and all fractures (right).

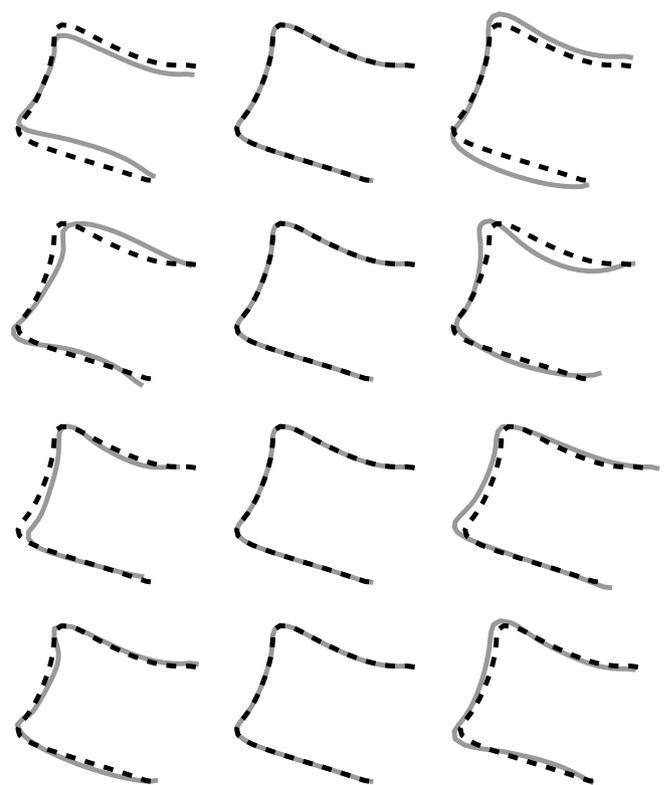


Fig. 6. Modes of variation of the differences between reconstructed normal shapes (black, dashed lines) and the true shapes (solid, gray lines) of all unfractured vertebrae. The four rows depict the first four modes of variation, ordered by decreasing variance. Left column: Mean difference minus four standard deviations of the mode of variation; Middle column: Mean difference (mean reconstructed shape and mean true shape); Right column: Mean difference plus four standard deviations of the mode of variation.

in Figs. 6 and 7. It can be seen that the variation in differences between the reconstruction and the true shape is much smaller for the unfractured cases. The PCA model of fractures shows clear examples of typical osteoporotic wedge, concave, and crush deformities as well as severe osteophytes. Although much less severe than in the fractures, the first mode of variation for the unfractured shapes does seem to indicate a compression deformation with respect to the reference shape similar to that of the first mode in the fractured shapes. The other modes of variation of normal vertebra shape with respect to the reconstructed reference seem to be related to tilting of the vertebra (mainly mode 2) and the formation of osteophytes (modes 2–4).

To investigate the added value of patient-specific shape modeling compared to modeling shape variation of the complete population, the performance in fracture recognition of the proposed method is compared with that of a model of vertebral shape alone, without the model of shape interrelations. Note, that the optimal reconstruction of unfractured shape according to the single shape model

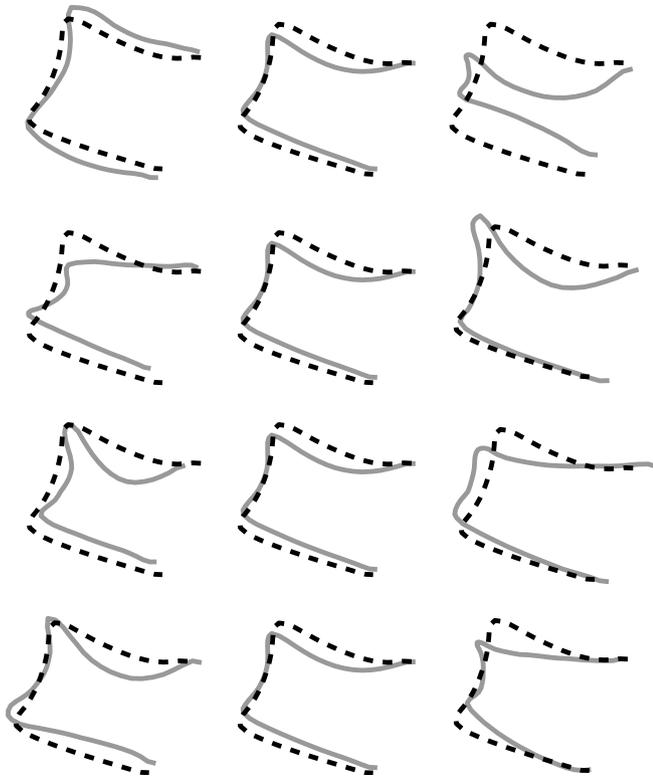


Fig. 7. Modes of variation of the differences between reconstructed normal shapes (black, dashed lines) and the true shapes (solid, gray lines) of all fractured vertebrae. The four rows depict the first four modes of variation, ordered by decreasing variance. Left column: Mean difference minus four standard deviations of the mode of variation; Middle column: Mean difference (mean reconstructed shape and mean true shape); Right column: Mean difference plus four standard deviations of the mode of variation.

would simply be the mean vertebral shape over the entire population. To incorporate shape variations as derived from the training set in the analysis a PCA is performed on the original, aligned shape vectors as well as on the difference vectors of aligned shape and reconstruction pairs. Both models are constructed from normal (unfractured) vertebrae and deviation from normality can therefore be measured as the residual after projection on the model subspace and constraining of the PC coefficients according to Eq. (9). The area under the ROC curves for fracture detection using these residuals with  $k \in \{1, 2, 3\}$  are given in Fig. 8. For a large range of model dimensionalities and cut-off criteria for the Mahalanobis distance, the model of differences between the reconstruction and true shape performs well and it consistently outperforms the model of shape alone. The classification accuracy and  $\kappa$ -statistic for agreement with the radiologists' scores (Landis and Koch, 1977) for the results with the largest area under the ROC curve, obtained by the shape and the difference models, are listed in Table 1.

The measured deviation from expected normal shape allows for detection of any type of deformity. Although this model already performs well in recognition of fractures, some of the cases with a large shape difference repre-

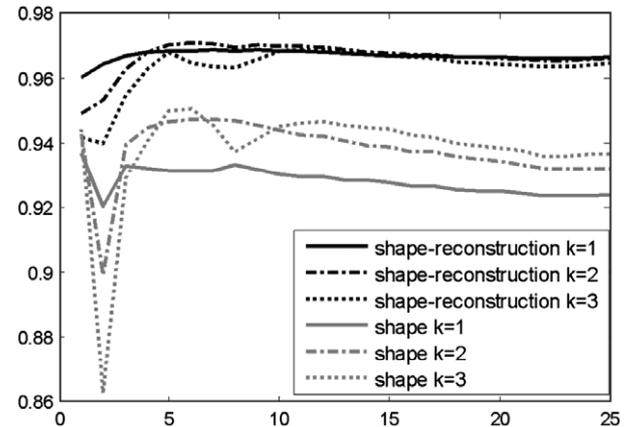


Fig. 8. Comparison of fracture recognition for the proposed model of differences between reconstruction and original shape and the individual shape model alone. Shown is the area under the ROC curve for fracture detection as a function of the number of modes in the model. The fracture measure used is the sum of squared residuals after projection on the PCA subspace with maximum Mahalanobis distance of 1 (solid lines), 2 (dash-dotted lines), and 3 (dotted lines). The black lines denote the model of differences between the true shapes and the reconstructed healthy shapes, the gray lines denote the model of true shapes alone.

Table 1

Fracture recognition performance: area under the ROC curve, accuracy (percentage of correctly classified shapes), agreement  $\kappa$  and its 95% confidence interval for the best performing shape and shape-reconstruction measures of Fig. 8 and for the direct distance measures between true shape and reconstruction

	ROC area	Accuracy	$\kappa$	95 % CI of $\kappa$
Shape	0.951	0.948	0.60	[0.50 0.70]
Shape-reconstruction	0.971	0.958	0.71	[0.63 0.79]
RMS lm-to-lm	0.966	0.957	0.68	[0.60 0.77]
RMS lm-to-contour	0.972	0.965	0.75	[0.67 0.83]
RMS lm-to-contour outside	0.979	0.967	0.76	[0.69 0.84]

sent not vertebral fractures but other deformations such as osteophytes (bony outgrowths). To make the model more specific to recognizing vertebral fracture and reduce the effect of osteophytes, in the following we ignore the points where the true shape is outside the reconstruction. We then use the RMS distance from all points where the reconstruction is outside the true shape to their closest points on the true shape. Fig. 9 summarizes the results of this fracture grading in all images. The fracture recognition results are listed in Table 1. The fracture measure was  $0.9 \pm 0.5$  mm on average for unfractured vertebrae and  $3.6 \pm 1.5$  mm for fractures. There is a fair correlation (0.81) between the shape distance and the fracture severity as indicated by the radiologists. The area under the ROC curve is 0.98; at a sensitivity of 98% the specificity is 92%. The areas under the ROC curve for mild, moderate, and severe fractures separately are 0.94, 0.99, and 1.00 respectively.

All results in Table 1 show 'moderate' to 'substantial' agreement with the expert score (Landis and Koch, 1977). The measures of landmark-to-contour distance

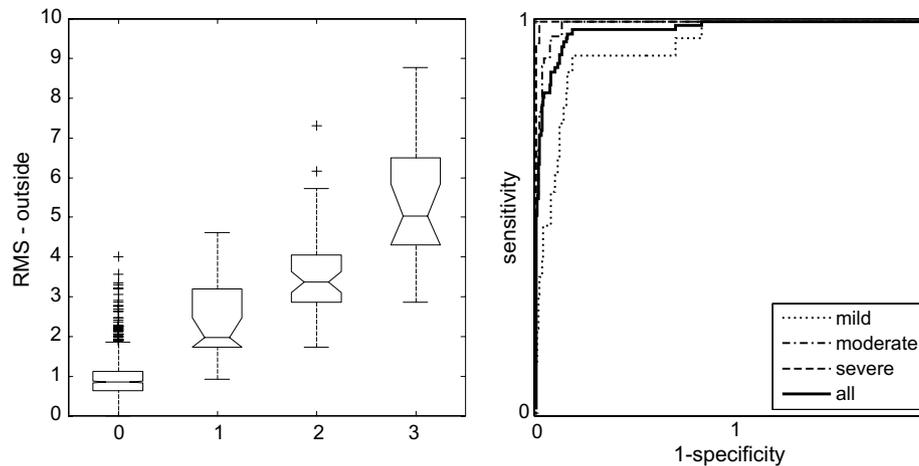


Fig. 9. Left: Box and whisker plot of the average distances between reconstruction and true shapes separated into fracture grade, ranging from normal (0) to severely fractured (3). The box has lines at the lower quartile, median, and upper quartile values. The whiskers extending from each end of the 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. Right: ROC curves of vertebral fracture detection.

between the observed and reconstructed shapes perform best and both are significantly better than the best measure of PCA residuals of individual vertebral shape ( $p = 0.03$  and  $p = 0.01$ ).

## 5. Discussion

We have proposed a method for reconstructing an expected shape and its allowed variation on the basis of a combination of pairwise predictions from the known shapes of neighboring objects. We have shown only its application in quantifying vertebral shape deformity, but the basic methods are more widely applicable and could be useful for instance in multi-object segmentation.

Overall, the proposed method is able to distinguish fractured vertebrae from normals and there is a fair correlation between the shape distance and the fracture severity as indicated by the radiologists. However, Fig. 9 reveals that the class of normal vertebrae contains a relatively large number of outliers that have a larger difference between the reconstruction and the true shape. One of these outliers is shown in the third row of Fig. 4. 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 reconstruction, more sensitive to subtle changes, can already identify. Visual inspection of the data set revealed that many of the unfractured vertebrae that had a large shape distance exhibit this type of deformation.

To maximize the amount of training data for each model, the models used in this study were trained on all vertebrae available except those that were identified as fractures, where semi-quantitative fracture grading by one radiologist was used as the gold standard. It is well known that the other vertebrae in spines with at least one fracture are more likely to be deformed as well and that radiologists often disagree whether a vertebra is fractured or not, which means that the shapes used for training will inevitably

include a small number of ambiguous cases and borderline fractures. This may also explain why the first mode of variation in Fig. 6 shows a variation (although very small) that is similar to the first mode of variation of the fractures in Fig. 7. This incorporation of mild disease in the model will have an effect mainly on the deformity measures that are derived from the residuals after projection on PCA models of normal vertebrae, for both the models of individual vertebral shapes and the models of differences between observed and reconstructed shapes. The direct distance between observed shape and maximum likelihood reconstruction, without taking the range of observed normal variations into account, is less likely to be affected as long as at least one of the vertebrae in the image is not deformed. However, it is still desirable to ensure that the models describe only normal, healthy shape variations and the models should preferably be trained on a separate set of healthy spines of which it is known that no fractures develop in the next years. Borderline cases should be left out, and a consensus fracture score of several radiologists is preferred. As an alternative, robust regression methods that reduce the effect of outlier shapes in the model could be useful (Rousseeuw and Leroy, 1987).

Apart from robust regression techniques, a large variety of multivariate regression methods is available and could be used instead of the ridge regression employed in this work. An obvious alternative to our approach of combining independent pairwise estimates would be to estimate the normal vertebra shape based on all other shapes in the image simultaneously, which would have the advantage that any correlations between the predictor shapes can be exploited. We have chosen for the pairwise approach as it allows for explicit weighting by reliability of the estimator shapes—which is especially useful in fractured spines—and natural optimization of regularization parameters per vertebra. In addition, the pairwise estimation suffers less from problems of multi-collinearity and the limited size of the

training set because a lower-dimensional model is estimated, and can without adaptation still be performed in low quality images in which annotations for one or several of the vertebrae may be missing.

In this work, the focus has been on producing accurate, patient specific models of normal shape variation, and shape differences between reconstructions and true shapes are summarized into one average distance measuring the deviation from normality. This enables detection of deformities and specification of the degree of abnormality, but not the type. In addition, any variation that may be normal but is not present or very rare in the training set will be indicated as abnormal. In our experiments this was the case with some vertebrae with severe osteophytes. We have shown that, to some extent, false positive fracture detections owing to osteophyte deformity can be countered by ignoring the points where the true shape protrudes outside its expected normal shape. Another option could be to increase the number of osteophytes in the training set of normal shapes, or weigh them stronger, to allow the model to better adhere to these special cases. A more elegant alternative would be to produce sufficiently large training sets of all deformities of interest, for instance osteophytes and wedge, biconcave, and crush compression fractures. The shape difference vectors—or a low-dimensional representation of them—can then be used to train a supervised classification scheme to discriminate between normal shapes and all types of deformities. Such explicit models of deformity 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. Preliminary results on supervised vertebral fracture classification are presented in [de Bruijne et al. \(2007\)](#).

The proposed method relies on full vertebra outlines, for which we in this paper used manual annotations by radiologists. This manual annotation procedure is of course time consuming and might hamper large-scale use of these methods. Several 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 ([Smyth et al., 1999](#); [Zamora et al., 2003](#); [de Bruijne and Nielsen, 2004](#); [Roberts et al., 2005](#); [Iglesias and de Bruijne, 2007](#)). 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.

In the presented setup, the expected normal shape of vertebrae is derived from neighboring vertebrae. Several studies indicate that the relative positioning of vertebral bodies plays an important role in fracture risk ([Sornay-Rendu et al., 2006](#); [Pettersen et al., 2007](#)). It may therefore be of interest to estimate the deviations from expected position and orientation as well. This can easily be achieved by adding the pose parameters to the shape vectors or by performing the Procrustes alignment that is done prior to

model construction on the pair of predictor and predicted shapes rather than on both shapes individually, as explained in [Section 2.2](#).

We have so far applied the methods to the lumbar part of the spine only, whereas many fractures occur in the lower or mid-thoracic region as well. We expect to obtain similar results if the thoracic and also the cervical part of the spine are included in the analysis—or perhaps even better, since there would be more shapes to base each reconstruction on. In our experiments, reconstruction accuracy for the vertebrae that had only one direct neighbor annotated, L1 and L4, was slightly lower. It is therefore desirable to include at least the two direct neighbors of a vertebra to assess in the analysis.

Finally, one obvious limitation of this work is that only information from one projection is used to assess the shape of a three-dimensional object. Volumetric imaging such as MRI may be more sensitive in detecting mild fractures ([Tomomitsu et al., 2005](#)). Variations in patient positioning and X-ray geometry may cause differences in the observed vertebral shape in the X-ray image. Differences in apparent size on the X-rays owing to a change in spine-to-film distance will be accounted for by the patient specific conditional models, but differences in projected shape can occur as a result of rotation of the vertebral body with respect to the image plane or as a result of the divergent X-ray beam. Some of the shape variations shown in [Fig. 6](#) may be caused by a tilting of the vertebra with respect to the other shapes in the image rather than a true variation in three-dimensional shape. Although these variations are present in the normal training set and are therefore recognized as normal variations and not deformity, it may be desirable to model such variations in the projected shape explicitly. This would be especially relevant in patients with scoliosis (lateral curving of the spine). In such cases, an additional anterior–posterior radiograph could be helpful in assessing the amount of scoliosis and thus the expected tilt. However, although not perfect, single projection lateral X-rays are still the method of choice for assessment of vertebral fractures and are widely used for this purpose.

## 6. Conclusions

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 estimated from its neighbors along with a model of normal variation. The patient-specific model of differences between the reconstruction and true shape obtains excellent fracture recognition rates (accuracy 96.7%) and it consistently outperforms a model of shape alone.

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 in individual patients and reduce the number of participants and follow-up time required in clinical trials assessing the efficacy of drug candidates.

## References

- Bagger, Y., Tankó, L., Alexandersen, P., Hansen, H., Qin, G., Christiansen, C., 2006. The long-term predictive value of bone mineral density measurements for fracture risk is independent of the site of measurement and the age at diagnosis: results from the prospective epidemiological risk factors study. *Osteoporos. Int.* 17 (3), 471–477.
- Center, J., Nguyen, T., Schneider, D., Sambrook, P., Eisman, J., 1999. Mortality after all major types of osteoporotic fracture in men and women: an observational study. *Lancet* 353 (9156), 878–882.
- Cootes, T., Taylor, C., Cooper, D., Graham, J., 1995. Active shape models – their training and application. *Comput. Vis. Image Underst.* 61 (1), 38–59.
- de Bruijne, M., Lund, M., Tankó, L., Pettersen, P., Nielsen, M., 2006. Quantitative vertebral morphometry using neighbor-conditional shape models. In: Larsen, R., Nielsen, M., Sporring, J. (Eds.), *MICCAI*, vol. 4190 of LNCS. Springer, pp. 1–8.
- de Bruijne, M., Nielsen, M., 2004. Image segmentation by shape particle filtering. In: Kittler, J., Petrou, M., Nixon, M. (Eds.), *International Conference on Pattern Recognition*. IEEE Computer Society Press, pp. 722–725.
- de Bruijne, M., Pettersen, P., Tankó, L., Nielsen, M., 2007. Vertebral fracture classification. In: Pluim, J., Reinhardt, J. (Eds.), *Med Imaging: Image Process*, . In: *Proceedings of SPIE*, vol. 6512. SPIE Press.
- Delmas, P., van de Langerijt, L., Watts, N., Eastell, R., Genant, H., Grauer, A., Cahall, D., IMPACTS Group, 2005. Underdiagnosis of vertebral fractures is a worldwide problem: the IMPACT study. *J. Bone Miner. Res.* 20 (4), 557–563.
- Dryden, I., Mardia, K., 1998. *Statistical Shape Analysis*. Wiley Series in Probability and Statistics.
- Eastell, R., Cedel, S., Wahner, H., Riggs, B., Melton, L.J., 1991. Classification of vertebral fractures. *J. Bone Miner. Res.* 6 (3), 207–215.
- Ferrar, L., Jiang, G., Adams, J., Eastell, R., 2005. Identification of vertebral fractures: an update. *Osteoporos. Int.* 16 (7), 717–728.
- Genant, H., Wu, C., van Kuijk, C., Nevitt, M., 1993. Vertebral fracture assessment using a semiquantitative technique. *J. Bone Miner. Res.* 8 (9), 1137–1148.
- Golub, G., Heath, M., Wahba, G., 1979. Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics* 21 (2), 215–223.
- Goodall, C., 1991. Procrustes methods in the statistical analysis of shape. *J. Roy. Stat. Soc. B* 53 (2), 285–339.
- Grados, F., Roux, C., de Vernejoul, M., Utard, G., Sebert, J., Fardellone, P., 2001. Comparison of four morphometric definitions and a semiquantitative consensus reading for assessing prevalent vertebral fractures. *Osteoporos. Int.* 12 (9), 716–722.
- Hoerl, A., Kennard, R., 1970. Ridge regression: biased estimation for nonorthogonal problems. *Technometrics* 12 (1), 55–67.
- Iglesias, J., de Bruijne, M., 2007. Semi-automatic segmentation of vertebrae in lateral X-rays using a conditional shape model. *Acad. Radiol.*, in press. doi:10.1016/j.acra.2007.06.003.
- Ismail, A., Cooper, C., Felsenberg, D., Varlow, J., Kanis, J., Silman, A., O'Neill, T., 1999. Number and type of vertebral deformities: epidemiological characteristics and relation to back pain and height loss. european vertebral osteoporosis study group. *Osteoporos. Int.* 9 (3), 206–213.
- Landis, J., Koch, G., 1977. The measurement of observer agreement for categorical data. *Biometrics* 33 (1), 159–174.
- Lund, M., de Bruijne, M., Tankó, L., Nielsen, M., 2005. Shape regression for vertebra fracture quantification. In: Fitzpatrick, M., Reinhardt, J. (Eds.), *Med Imaging: Image Process*, . In: *Proceedings of SPIE*, vol. 5747. SPIE Press, pp. 723–731.
- Lunt, M., Ismail, A., Felsenberg, D., Cooper, C., Kanis, J., Reeve, J., Silman, A., O'Neill, T., Group, E.P.O.S., 2002. Defining incident vertebral deformities in population studies: a comparison of morphometric criteria. *Osteoporos. Int.* 13 (10), 809–815.
- McCloskey, E., Spector, T., Eyres, K., Fern, E., O'Rourke, N., Vasikaran, S., Kanis, J., 1993. The assessment of vertebral deformity: a method for use in population studies and clinical trials. *Osteoporos. Int.* 3 (3), 138–147.
- Melton, L., Lane, A., Cooper, C., Eastell, R., O'Fallon, W., Riggs, B., 1993. Prevalence and incidence of vertebral deformities. *Osteoporos. Int.* 3 (3), 113–119.
- Melton, L., Atkinson, E., Cooper, C., O'Fallon, W., Riggs, B., 1999. Vertebral fractures predict subsequent fractures. *Osteoporos. Int.* 10 (3), 214–221.
- Nevitt, M., Ettlinger, B., Black, D., Stone, K., Jamal, S.A., Ensrud, K., Segal, M., Genant, H., Cummings, S., 1998. The association of radiographically detected vertebral fractures with back pain and function: a prospective study. *Ann. Intern. Med.* 128 (10), 793–800.
- Pettersen, P., de Bruijne, M., Chen, J., He, Q., Christiansen, C., Tankó, L., 2007. A computer-based measure of irregularity in vertebral alignment is a BMD-independent predictor of fracture risk in postmenopausal women. *Osteoporos. Int.*, in press. doi:10.1007/s00198-007-0388-y.
- Roberts, M., Cootes, T., Adams, J., 2005. Vertebral shape: automatic measurement with dynamically sequenced active appearance models. In: Duncan, J., Gerig, G. (Eds.), *MICCAI*, vol. 3750 of LNCS. Springer, pp. 733–740.
- Rousseeuw, P., Leroy, A., 1987. *Robust Regression and Outlier Detection*. John Wiley, New York.
- Sambrook, P., Cooper, C., 2006. Osteoporosis. *Lancet* 367 (9527), 2010–2018.
- Schwartz, E., Steinberg, D., 2005. Detection of vertebral fractures. *Curr. Osteoporos. Rep.* 3 (4), 126–135.
- Smyth, P., Taylor, C., Adams, J., 1999. Vertebral shape: automatic measurement with active shape models. *Radiology* 211 (2), 571–578.
- Sornay-Rendu, E., Allard, C., Munoz, F., Duboeuf, F., Delmas, P., 2006. Disc space narrowing as a new risk factor for vertebral fracture: the OFELY study. *Arthritis Rheum.* 54 (4), 1262–1269.
- Tomomitsu, T., Murase, K., Sone, T., Fukunaga, M., 2005. Comparison of vertebral morphometry in the lumbar vertebrae by T1-weighted sagittal MRI and radiograph. *Eur. J. Radiol.* 56, 102–106.
- Zamora, G., Sari-Sarraf, H., Long, R., 2003. Hierarchical segmentation of vertebrae from X-ray images. In: Sonka, M., Fitzpatrick, M. (Eds.), *Med Imaging: Image Process*, . In: *Proceedings of SPIE*, vol. 5032. SPIE Press, pp. 631–642.