

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
27 November 2008 (27.11.2008)

PCT

(10) International Publication Number
WO 2008/141996 A2

(51) International Patent Classification:
G06K 9/64 (2006.01) **G06K 9/46** (2006.01)

(21) International Application Number:
PCT/EP2008/055956

(22) International Filing Date: 15 May 2008 (15.05.2008)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0709599.5 18 May 2007 (18.05.2007) GB
0710540.6 1 June 2007 (01.06.2007) GB
0718369.2 20 September 2007 (20.09.2007) GB

(71) Applicant (for all designated States except US): **NORDIC BIOSCIENCE A/S** [DK/DK]; Herlev Hovedgade 207, DK-2730 Herlev (DK).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **DE BRUIJNE, Marleen** [DK/DK]; Fiolstræde 25, 1., DK-1171 Copenhagen (DK). **IGLESIAS, Juan Eugenio** [ES/US]; 945 Weyburn Terrace 423, Los Angeles, California 90024 (US).

(74) Agent: **CHANDRANI, Vandita**; Beck Greener, Fulwood House, 12 Fulwood Place, London London WC1V 6HR (GB).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report

(54) Title: SEMI-AUTOMATIC CONTOUR DETECTION

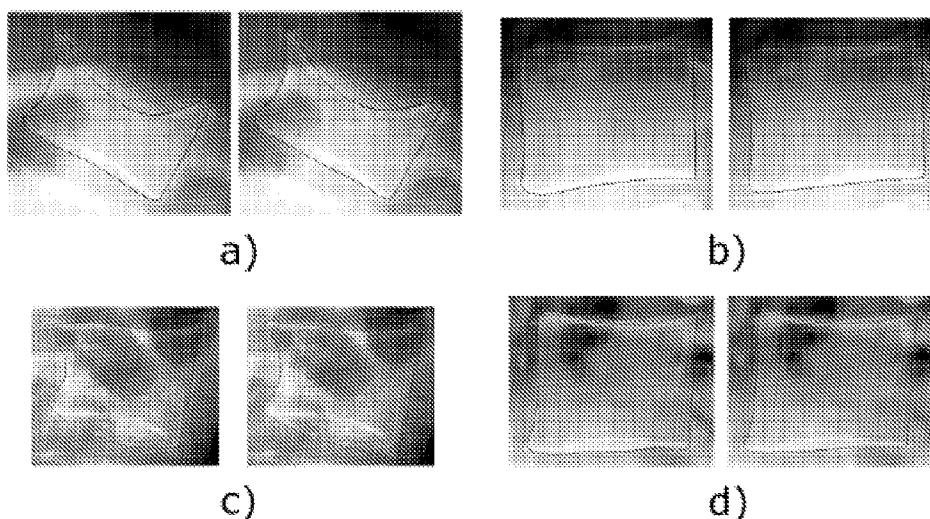


Figure 5

(57) Abstract: A method of locating a contour of a structure in an image by processing said image including the structure is provided. A starting set of digital data representative of the image including the structure is taken, the structure in said image having annotated on it from three to ten landmark positions. A statistical model of said structure to the landmark positions annotated on the image is fitted and an initial estimate of the contour of the structure made. Using grey level information derived from points adjacent the estimated contour the contour boundary is iteratively moved to produce a final estimate of the contour of the structure.

WO 2008/141996 A2

Semi-Automatic Contour Detection

The present invention relates to a method of locating
5 the contour of a structure in an image.

Osteoporosis is a bone disease in which the bone mineral density (BMD) is reduced, bone micro-architecture is disrupted, and the amount and variety of non-
10 collagenous proteins in bone is altered. Bones affected by the disease are more likely to fracture. Osteoporosis is defined by the World Health Organisation (WHO) as either a bone mineral density 2.5 standard deviations below peak bone mass (20-year-old sex-matched healthy
15 person average) as measured by dual X-ray absorptiometry (DXA), or any fragility fracture. Due to its hormonal component, more women, particularly after menopause, suffer from this disease than men.

20 Osteoporotic fractures are those that occur under slight amounts of stress that would not normally lead to fractures in non-osteoporotic people. Typical fractures occur in the vertebral column, hip and wrist. Vertebral fractures are the most common ones. Occurring in younger
25 patients, they are a good indicator of the risk of future spine and hip fractures. These two are the most serious cases, leading to limited mobility and possibly disability. Hip fractures, in particular, usually require prompt surgery, which has other important risks
30 associated, such as deep vein thrombosis and pulmonary embolism. Although osteoporosis patients have an increased mortality rate due to the complications of fractures, most patients die with the disease rather than of it.

Vertebral fractures are conventionally detected and graded on lateral X-rays. Apart from the subjective judgement of the image by a radiologist, a standard six-point morphometry is commonly used. In this technique, 5 six landmarks are placed on the corners and in the middle point of both vertebra endplates, defining the anterior, middle and posterior heights. These measurements can be used to calculate a fracture grade. In current clinical trials, a fractured vertebra is defined as the one for 10 which one of the three heights is at least 20% larger than any other.

The six-point representation captures most of the important information in the image, but it is unable to 15 capture certain structures (such as osteophytes) or subtle shape variations. Use of a full contour would overcome this problem. However, manually annotating the full contour of the vertebrae represents a huge amount of work. Accordingly, many methods have been proposed to segment 20 vertebrae automatically or semi-automatically.

Gardner et al (SPIE Vol. 2710, February 1996, pp 996 - 1008) discloses a semi-automated system, based on active contour (snake) modelling of vertebrae. Points on the 25 vertebral boundary are specified by a user on a digital image of a vertebra. The selected boundary points then act as physical constraints on an active contour that is automatically fitted to the vertebra boundary. Various parameters are monitored and adjusted to provide an 30 appropriate degree of freedom for the snake. Too much freedom can result in an improbable shape as the snake may be misled by an unclear edge of the vertebra and too little freedom can result in sharp turns in the contour, for example the corners of the vertebra or an osteophyte, 35 being missed.

Semi-automatic and automatic segmentation of the vertebrae in lateral x-ray images is a difficult task because of the nature of the images. Lateral x-ray images show the superimposition of many layers, making it difficult to distinguish the region of interest - a sagittal plane along the spine. This is the reason why many classical segmentation approaches, such as those based on region growing, fail.

Active shape models have the advantage that they make use of prior knowledge of vertebral shape and appearance, and therefore do not need to rely as much on the information in the images.

In Smyth et al (Radiology, May 1999, pp 571-578), a point distribution model (PDM) was used to represent the full contour of vertebrae, and a classifier to separate normal and fractured vertebrae. An improvement was observed in the performance of the full contour representation compared to that of the six landmarks. A similar conclusion was drawn in de Bruijne et al ("Vertebral Fracture Classification," in *Medical Imaging: Image Processing*, J.P. Pluim and J.M. Reinhardt, eds., *Proceedings of SPIE 6512*, SPIE Press, 2007).

Zamora et al (Medical Imaging: Image Processing, Vol 5032 of Proceedings of SPIE, SPIE Press, 2003, pp 631-642) used an active shape model (ASM) including grey level edge information, initialising it with a customised implementation of the generalised Hough transform. They applied the method on x-ray images and achieved errors lower than 6.4mm in 50% of the lumbar images. Smyth et al also used ASM methods to segment vertebrae in dual energy x-ray absorptiometry (DXA). They require the user to

annotate the mid-points of the bottom of L4, top of T12
and top of T7. They achieved a root mean square (RMS)
point-to-line error lower than 1.23mm in 95% of the cases
in healthy vertebrae, and lower than 2.24mm in 92% of the
5 fractures.

De Bruijne et al applied a fully automated method
based on shape particle filtering, lowering the average
point-to-line error to 1.mm. Roberts et al (Volume 2750
10 of LNCS, Springer, 2005, pp 733-840) incorporated an
active appearance model (AAM) and a dynamic ordering
algorithm to segment the vertebrae in DXA images.
Requiring the user to mark the same three points as Smyth
et al., they achieved a point-to-line error of 0.79mm.
15 Better results have been achieved by Roberts et al
(Proceedings of Medical Image Understanding and Analysis,
2006, Vol.1, pp 120-124), who recently reported a 0.64 mm
mean point-to-line error (95% of the points 2mm within the
contour) in radiographs of healthy vertebrae. The user is
20 required to approximately mark the center of each
vertebra. The algorithm was extended to DXA images by
Roberts et al (Proceeding of MICCAI Conference Workshop on
joint and bone disease, 2006, Vol. 1, pp 1-8). They
achieved 0.69mm mean point-to-line errors in normal
25 vertebrae and 0.96 in fractures.

According to a first aspect of the present invention,
there is provided a method of locating a contour of a
structure in an image by processing said image including
30 the structure, comprising the steps of:

taking a starting set of digital data
representative of the image including the structure, the
structure in said image having annotated on it from three
to ten landmark positions;

fitting a statistical model of said structure to the landmark positions annotated on the image, and making an initial estimate of the contour of the structure; and using grey level information derived from points adjacent the estimated contour iteratively to move the contour boundary to produce a final estimate of the contour of the structure.

The invention is generally applicable to any structure in an image, for example, an x-ray, computer tomography or magnetic resonance image, which has a generally predictable shape and where there is some difference in light intensity between the structure and the background of the image. Previous images that include similar structures can then be used in the training of a statistical model that is then used to locate the contour of a structure in a new image.

The method may be used for locating body parts and specifically bones. In this respect, bones generally have a predictable shape such that a statistical model may be used with assistance from manual annotations on an image of a bone, to locate the contour of a bone.

In a preferred embodiment, the structure is a vertebra and the image is of part of a spine. The method is described, hereinafter, with reference to its application in locating the contour of a vertebra. However, it will be appreciated that the method is equally applicable to extracting a contour of any structure in an image for which a general expectation of its shape is deducible by statistical analysis of the contours of similar structures in a set of such images.

For example, the described method may also be used to locate the contour of bones in x-ray images of a hand or wrist, or to locate the contours of chambers of a heart or heart walls in MR or CT scans.

5

Providing multiple landmark positions on a vertebra to which a statistical model may then be fitted provides a good starting point for the iterative process of locating the edge of the vertebra. Preferably, the vertebra has
10 annotated on it four to eight, for example six, landmark positions, one on each of the corners and one in the middle point of each vertebra endplate. Annotating vertebrae with six landmarks as described above is current standard practice. As a result, compared to standard
15 practice, no additional manual work is required.

In an embodiment, the method comprises training the statistical model of the vertebra using information from points approximating respective contours of a set of other
20 vertebrae. It will be appreciated that any number of points may be used that provides an approximation of the contour of the vertebrae. However, in a preferred embodiment, 20 or more points are used to approximate the contour, for example, more than 40, more than 50 or more
25 than 60 points.

Preferably, the method further comprises training the statistical model using information from three to ten, for example, four to eight, landmark positions annotated on
30 vertebrae in said set of other vertebrae. In a preferred embodiment six landmark positions are annotated on the set of vertebrae.

In a preferred embodiment, the set of vertebrae used
35 in training the statistical model includes unfractured and

fractured vertebrae. By including fractured vertebrae in the training, the statistical model is enabled to locate the edges of fractured vertebrae aswell as unfractured vertebrae. Vertebrae with osteophytes may also be
5 included in the training stage to increase the likelihood of finding osteophytes in new images with this method.

In an embodiment, the statistical model is a conditional point distribution model. However, it will be
10 appreciated that if the mathematical process is appropriately adapted, other statistical shape models may alternatively be used, for example, statistical models based on spherical harmonics, Fourier descriptors, distance maps, image warps and volumetric or medial
15 representations.

Preferably, the conditional point distribution model is constructed from information approximating the respective contours of a set of vertebrae and information
20 of three to ten, four to eight, or six landmarks annotated on each member of said set of vertebrae.

Additionally, and/or alternatively, the conditional point distribution model is constructed from a first point
25 distribution model constructed from information approximating the respective contours of said set of vertebrae and a second point distribution model constructed from information of three to ten, four to eight, or six landmarks annotated on each member of said
30 set of vertebrae.

In a preferred embodiment, the initial estimate of the contour is the mean of the conditional point distribution model fitted to the landmark positions.

Preferably, the iterative movement of the estimated contour is constrained by its proximity to the current estimate of the contour. Additionally and/or alternatively, the iterative movement is constrained by the conditional covariance. These constraints reduce the search space around the initial estimate and result in plausible shapes.

Additionally and/or alternatively, the movement of the contour boundary is constrained by restricting divergence of grey level information derived from points adjacent the estimated contour with equivalent information derived from said statistical model.

Preferably, the iterative movement of the contour boundary is continued until the difference between the estimated contours at two consecutive iterations is smaller than a preset limit. For example, when the distance between consecutive contour estimations is less than 2 pixels or less than 1 pixel, the iterative process stops.

In an embodiment, a grey level profile is built by sampling grey level information in the image along the normal to the contour across each contour point.

The invention has principally been defined as a method of deriving information from a digital image. However, it is of course equally applicable as an instruction set for a computer carrying out a said method or as a suitably programmed computer.

Embodiments of the present invention will hereinafter be described, with reference to the accompanying diagrams, in which:

Figure 1 shows an example of a vertebra with six initial landmark positions and the contour annotated;

Figure 2 shows an example of an initial estimate of the contour of a vertebra of an embodiment of the present invention;

Figure 3 illustrates the influence of the maximum allowed Mahalanobis distance on a result;

Figure 4 shows the distance of a contour as determined by the present invention from the real contour in the form of a histogram and cumulative distribution function;

Figure 5 shows some examples from leave-one-out experiments;

Figure 6 shows a graph of mean point-to-line error against point index that illustrates the error depending on the point number;

Figure 7 shows a graph of sum of squared errors against α , where a) shows the dependence of the mean sum of squared errors with α , b) shows the dependence of the maximum sum of squared errors with α and c) shows the dependence of the mean sum of squared errors with the number of fractures present in the training set for $\alpha=1.75$; and

Figure 8 shows a comparison of segmentation according to the present invention, illustrating the difference between the use of standard PCA and α -PCA.

The present invention will hereinafter be described with particular reference to the analysis of x-ray images of vertebrae of a spine. It will, however, be appreciated that the described method could be applied to other medical images of a spine for example, DXA, Computer Tomography (CT) or Magnetic Resonance (MR).

All of the steps described below are equally

applicable to extracting a contour of any structure in an image for which a general expectation of its shape is deducible by statistical analysis of the contours of similar structures in a set of such images. For example,
5 the method may be applied to extracting the contour of bones in x-ray images of a hand or wrist, or in extracting the contour of chambers of the heart and the heart walls in MR or CT scans.

10 The preferred method of locating a contour of a vertebra of the present invention consists of two main steps.

The first step is to construct a conditional point
15 distribution model (PDM). For this, initially two point distribution models (PDM) are constructed, a first derived from a training set of vertebrae annotated with the traditional six landmark points, and a second derived from the same training set of vertebrae annotated with a large
20 number of points, for example, 20 or over, that approximate the actual contour outline. The relationship between the two PDMs is then modelled to make it possible, for a new case, to construct a conditional PDM for the full contour depending on the position of six points that
25 are manually annotated by a clinician.

The second step is then to apply this conditional PDM to a new image of a vertebra with the traditional six landmark points annotated and to approximate an initial
30 contour of the vertebra. Active shape modelling is then used to manipulate the initial contour to find the actual contour of the vertebra subject to the constraints of the conditional PDM covariance.

35 The two steps above will now be described in more

detail with reference to a specific example.

The training set used in this specific example consisted of information of full contours of vertebrae and sets of six landmark positions of vertebrae from 142 patients. Where, as a result of imaging and projection of a vertebra in the image, a vertebra was shown to have two contours, the lower one was always chosen. Vertebrae L1, L2, L3 and L4 were analysed, so 568 vertebrae (including 64 fractured vertebrae) were included in the study.

The images were 12-bit deep and their resolution was 570DPI. All images were stored in DICOM format. As the application did not require such a high resolution, the images were smoothed with a Gaussian kernel and down-sampled by a factor of five.

The six landmark positions and the contours were marked for training purposes by three different radiologists. The radiologist who marked the landmark positions on an image would always annotate the contour too. In the annotation of the six landmarks, the corners are marked first and then the perpendicular bisector of the segment joining the upper corners is displayed. It serves as a guide for the radiologist, who is supposed to place a landmark on the point of minimum height, and if it is unclear, as close to the bisector as possible. The process is then repeated for the lower plate. The displayed bisectors help the radiologists be consistent throughout the annotation process, minimizing the impact of interobserver variability in the PDM.

In order to annotate the full contour, the radiologists drew a polygonal line with as many vertices as they wanted. This contour was used as the ground truth

for the study. As can be seen in Figure 1, the six landmarks and the contours were annotated in different passes without showing the earlier annotation, so they do not necessarily overlap.

5

A method of a preferred embodiment is based on the following steps:

Training:

10 * six landmarks and an approximation of the contours of vertebrae are annotated on the training images by an expert radiologist;

 * the vertebrae in the data set are aligned

 * two PDMs are constructed, a first using
15 information of an approximation of the respective contours of a set of vertebrae, and a second using information of landmark positions annotated on each of the same set of vertebrae.

20 Vertebral edge location:

 * six landmarks are marked on the actual image of a vertebra by a radiologist

 * a conditional model based on the position of the six landmarks on the image and on the two PDMs derived in
25 the training stage is built for analysis

 * an initial solution is estimated, using the mean of the conditional model

 * active shape modelling is used to fit the contour to the vertebra in the image, using the
30 conditional covariance to constrain the solution.

These steps will now be described in further detail.

An embodiment of the method is described with
35 reference to use of point distribution models. However,

it will be appreciated that the mathematic processes may be adapted to enable use of other statistical models, for example, statistical shape models based on spherical harmonics, Fourier descriptors, distance maps, image warps
5 and volumetric or medial representations.

As a point distribution model is to be built, the contour points must be placed on the vertebrae of the training image at equivalent points of the vertebrae. The
10 number of points can be arbitrarily chosen by the radiologist, however, it has been shown that 20 or more points is sufficient to mark the outline of the contour. When building the point distribution model, it is necessary for the number of contour points to be the same
15 for every case. Accordingly, resampling is necessary in the contours, as the number of points marked by radiologists is likely to be inconsistent. In the training set described in this example, the maximum number of points annotated by a radiologist was 53.

20

For convenience, in the described embodiment, it was decided that the contour model would consist of 67 points. This allows for an equal number of points between landmarks. It will, however, be appreciated that a
25 different, fixed, number of points could be used, provided that the number of points allows a sufficient degree of accuracy for the outline of the contour. Accordingly, to arrive at 67 points, the contour was completed, using the points marked by the radiologist, and then the 67 points
30 were assigned to the contour. In the described embodiment, the points of the contour closest to the initial six landmark positions were chosen to be points 1, 13, 25, 43, 55 and 67. The rest of the contour points were equidistantly placed between these six. The third
35 segment has 50% more points because it is on average

(approximately) 50% longer than the other four. A sample image is shown in Figure 1. Here the six initial landmark positions are marked as stars and the contour with selected points are marked as asterisks.

5

This specific example has been described with reference to the use of six landmark positions. However, it will be appreciated that a range of between three and ten landmarks may be positioned that are still, in general, representative of the shape of the vertebra.

10

In the following step, the 6-point representations of the vertebrae were aligned in order to eliminate translation, scale and rotation deviations between point sets. The Procrustes method was used, using only the information from the corners. The complex representation of an annotation shape can be defined as:

15

$$w_i = (x_1 + i * y_1, x_2 + i * y_2, \dots, x_n + i * y_n)^T$$

20 and generalised Procrustes alignment can be expressed as

$$w_i^p = w_i^* \hat{\mu} w_i^* / w_i^* w_i, \quad i = 1, 2, \dots, n,$$

where each w_i^p is the full Procrustes fit of w_i onto $\hat{\mu}$. The full Procrustes estimate of the mean shape $[\mu]$ can be found as the eigenvector corresponding to the largest eigenvalue of the complex sum of squares and products matrix:

25

$$S = \sum_{i=1}^n w_i w_i^* / w_i^* w_i \quad i = 1, 2, \dots, n.$$

The transform parameters for each vertebra were then applied on the full contours, so that both representations

30

were aligned. Each shape was then represented by a vector, where $N = 6$ or $N = 67$, depending on the model:

$$\mathbf{x} = [x_1, x_2, \dots, x_N, y_1, y_2, \dots, y_N]^t$$

5

Once the training shapes were aligned, a point distribution model was built for both the six landmark positions and the full contour representations, mixing vertebrae L1, L2, L3 and L4 into a single model. In order
 10 to build the model, a Principal Component Analysis (PCA) was performed on the aligned data vectors. PCA is a technique that can be used to reduce the dimensionality of a dataset. It is a linear transformation towards a new coordinate system such that the greatest variance by any
 15 projection of the data lies on the first coordinate (known as the first principal component), the second greatest variance on the second coordinate, and so on. It is hence possible to simplify the dataset by keeping only the first principal components in the new representation. It can be
 20 shown that an orthonormal base of the new space is given by the eigenvectors of the covariance matrix of the dataset.

The shape model is then characterised by a mean shape
 25 $\bar{\mathbf{x}}$, the kept eigenvalues $\lambda_i, i=1, \dots, k$, and the corresponding eigenvectors, which are grouped columnwise in the $\mathbf{P}$ matrix, which accomplishes that $\mathbf{P}^t \mathbf{P} = \mathbf{I}$. Each shape can then be approximated by

$$30 \quad \mathbf{x} \approx \bar{\mathbf{x}} + \mathbf{P}\mathbf{b}$$

where $\mathbf{b}$ can be calculated for a certain shape:

$$\mathbf{b} = \mathbf{P}^t (\mathbf{x} - \bar{\mathbf{x}})$$

35

and the error vector is:

$$= x - \bar{x} - \mathbf{Pb} = x - \bar{x} - P(P'(x - \bar{x}))$$

5 where b is a column vector of k components, representing the projection of the shape onto the space of the model. Along the training set, the mean of this vector will be zero, and the covariance C will be a diagonal matrix including the k eigenvalues.

10

In order to verify whether a certain vector b corresponds to a plausible shape, it must be checked that it is not too far away from the mean of the model, that is, the zero vector. At the same time, plausible shapes
15 can be generated just by taking b vectors close to the zero vector. The valid region is defined by limiting the Mahalanobis distance of b . The limit $d_{\max}$ can be chosen using the chi-square distribution.

$$20 \quad d = \sqrt{b^t C^{-1} b} = \sqrt{\sum_{i=1}^k \left(\frac{b_i^2}{\lambda_i} \right)}$$

If the condition $d < d_{\max}$ is not true, the b vector can be modified:

25

$$b' = b(d_{\max} / d)$$

As the number of fractures in the data set is low, compared to the number of healthy ones, their influence on
30 the model was increased by giving them a higher weight when building it. Two different weights were given to normal and fractured vertebrae when calculating the mean and the variance of the shapes, so that their total

contributions were equal. As 504 healthy and 64 fractured vertebrae were available, the weights were $\omega_h = \frac{0.5}{504}$ and $\omega_f = \frac{0.5}{64}$, respectively.

5 As an alternative to this method of weighting, a variation on PCA (described above) can be implemented that deals with the importance of outliers. There are cases in which a small subset of samples may appear as outliers and still correspond to plausible data. This is true where
 10 some of the data obtained of vertebrae include fractured vertebrae. These may appear as outliers but the information they provide is still important. In this case, the modelling of outliers is important and rather than disregarding them, they may be used to enhance the
 15 results.

PCA is an orthogonal linear transformation that spans a subspace which approximates the data optimally in a least-squares sense. This is accomplished by maximising
 20 the variance of the transformed coordinates.

If the dimensionality of the data is to be reduced to N , an equivalent formulation of PCA is to find the N -set of orthonormal vectors, grouped in the $\mathbf{P}$ matrix, which
 25 minimises the error made when reconstructing the original data points in the data set. The error is measured in a L_2 norm fashion:

$$C = \sum \|\mathbf{P}\mathbf{P}'\mathbf{x}_i - \mathbf{x}_i\|^2$$

30 where C is the cost, N is the number of training cases and $\mathbf{x}_i$ are the centred data vectors to approximate.

Least-squares is not robust when outliers are present

in the dataset, as they can skew the result from the desired solution, leading to inflated error rates and distortions in statistical estimates (Hampel et al. : "Robust statistics: the approach based on influence functions" Wiley 1986). It has been attempted to reduce the impact of outliers on PCA by modifying the cost in the above equation, for example, by introducing a binary variable that is zero when a data sample is considered to be an outlier:

$$C_{Xu} = \sum_{i=1}^N [V_i \|\mathbf{P}\mathbf{P}^t \mathbf{x}_i - \mathbf{x}_i\|^2 + \eta(1 - V_i)]$$

where V_i is the set of binary variables. The term $\eta(1 - V_i)$ prevents the optimisation from converging to the trivial solution $V_i = 0, \forall i$.

In contrast to this and in contrast to directly minimising the squared data reconstruction error as in normal PCA, the presented Φ -PCA algorithm minimises:

$$C = \sum \Phi[\|\mathbf{P}\mathbf{P}^t \mathbf{x}_i - \mathbf{x}_i\|^2]$$

where Φ is a twice differentiable function such that $\Phi(x^2)$ is convex. The fact that Φ is twice differentiable makes it possible to use Hessian-based methods in the optimisation, providing quadratic convergence. The convexity requirement ensures the existence of just one minimum for C .

A simple and at the same time powerful form of the function is $\Phi(x) = x^\alpha$, with $\alpha > 0.5$ in order to accomplish the convexity condition. This special case will be called

α -PCA. Large values for α ($\alpha > 1$, in general), will enhance the outliers, as they become more expensive compared to normal cases. In particular, $\alpha = \infty$ would lead to minimising the L_∞ norm, and hence the maximum reconstruction error over shapes measured in a L_2 norm fashion. On the other hand, smaller values ($0.5 < \alpha < 1$) will have the opposite effect, leading to a more robust PCA. The case $\alpha = 0.5$ minimises the L_1 norm. Finally $\alpha = 1$ amounts to standard PCA.

10

The data points $\mathbf{x}_i$ must be centred, which means that their means must be subtracted from them: $\mathbf{x}_i = \mathbf{s}_i - \boldsymbol{\mu}$, where $\mathbf{s}_i$ represents the original, non-zero mean data samples. In the proposed algorithm, the "mean" is no longer the component-wise arithmetic mean of the data points as in the standard PCA, but the vector which minimises (assuming M dimensions for the data points):

15

20

$$C_{\mu^\Phi} = \sum_{i=1}^N \Phi[\|\mu^\Phi - \mathbf{s}_i\|^2] = \sum_{i=1}^N \Phi \left[\sum_{t=1}^M (\mu_t^\Phi - s_{it})^2 \right]$$

Once the $\mathbf{x}_i$ vectors have been calculated the ϕ -PCA which consists of searching the basis vectors $\mathbf{P}$ that minimise the cost function in the original equation (??), can be performed. Numerical methods will be required in both minimising C and C_μ , as there is no closed-form expression for μ^Φ or $\mathbf{P}$.

30

An important difference between standard and ϕ -PCA is that, in the latter, the principal components have to be recalculated if the desired dimensionality changes. In standard PCA, on the other hand, the first N_1 principal components are common for analyses with N_1 and N_2 assuming

35

that $N_2 > N_1$.

The expressions for the gradient and the Hessian of the cost C_{μ^Φ} are quite simple and fast to calculate.

5 Using the component-wise arithmetic mean as initialisation, Newton's method converges rapidly to the solution:

$$10 \quad \mu_{n+1}^\Phi = \mu_n^\Phi - [HC_{\mu^\Phi}(\mu_n^\Phi)]^{-1} \nabla C_{\mu^\Phi}(\mu_n^\Phi),$$

where the gradient ∇C_{μ^Φ} is a column vector consisting of the first-order derivatives:

$$15 \quad \frac{\partial C_{\mu^\Phi}}{\partial \mu_k^\Phi} = 2 \sum_{i=1}^N \Phi' \left[\sum_{l=1}^M (\mu_l^\Phi - s_{i_l})^2 \right] (\mu_k^\Phi - s_{i_k})$$

and the Hessian matrix H consists of the second-order
20 derivatives:

$$H_{kk} = \frac{\partial^2 C_{\mu^\Phi}}{\partial \mu_k^{\Phi^2}} = 2 \sum_{i=1}^N \left\{ 2\Phi'' \left[\sum_{l=1}^M (\mu_l^\Phi - s_{i_l})^2 \right] (\mu_k^\Phi - s_{i_k})^2 + \Phi' \left[\sum_{m=1}^M (\mu_m^\Phi - s_{i_m})^2 \right] \right\}$$

$$25 \quad H_{ku} = H_{uk} = \frac{\partial^2 C_{\mu^\Phi}}{\partial \mu_k^\Phi \partial \mu_u^\Phi} = 4 \sum_{i=1}^N \Phi'' \left[\sum_{l=1}^M (\mu_l^\Phi - s_{i_l})^2 \right] (\mu_k^\Phi - s_{i_k})(\mu_u^\Phi - s_{i_u})$$

Once the mean has been subtracted from the data points, the original cost C must be minimised. The
30 function has the interesting property that it reaches its global minimum for an orthonormal $\mathbf{P}$ matrix such that $\mathbf{P}'\mathbf{P}=\mathbf{I}$. This makes it possible not to have to constrain $\mathbf{P}$ to accomplish this condition during the optimisation, even it that implies that in general $\mathbf{PP}'$ will not represent a
35 projection matrix, and hence $\mathbf{PP}'\mathbf{x}_i - \mathbf{x}_i$ does not express the

reconstruction error any longer.

In the minimisation problem, only the expression for the gradient is implemented, as the one for the Hessian matrix is too complex and its computation too expensive. Using matrix calculus, all the partial derivatives can be calculated simultaneously:

10

$$\begin{aligned} \frac{dC}{d\mathbf{P}} &= \frac{d}{d\mathbf{P}} \sum_{i=1}^N \Phi [\|\mathbf{P}\mathbf{P}^t \mathbf{x}_i - \mathbf{x}_i\|^2] = \sum_{i=1}^N \frac{d}{d\mathbf{P}} \Phi [(\mathbf{P}\mathbf{P}^t \mathbf{x}_i - \mathbf{x}_i)^t (\mathbf{P}\mathbf{P}^t \mathbf{x}_i - \mathbf{x}_i)] = \dots \\ &\sum_{i=1}^N \Phi' [\mathbf{x}_i^t \mathbf{P} \mathbf{P}^t \mathbf{P} \mathbf{P}^t \mathbf{x}_i - \mathbf{x}_i^t \mathbf{x}_i - 2\mathbf{x}_i^t \mathbf{P} \mathbf{P}^t \mathbf{x}_i] [-4\mathbf{x}_i \mathbf{x}_i^t \mathbf{P} + 2 [\mathbf{x}_i \mathbf{x}_i^t \mathbf{P} \mathbf{P}^t + \mathbf{P} \mathbf{P}^t \mathbf{x}_i \mathbf{x}_i^t] \mathbf{P}] \end{aligned}$$

15

Once the gradient is known, different standard techniques can be used to update $\mathbf{P}$. In a simple gradient descent scheme, for example:

20

$$\mathbf{P}_{n+1} = \mathbf{P}_n - k \frac{dC}{d\mathbf{P}}$$

where k is the step size.

25

Line search can then be used with a normal PCA as initialisation in order to quickly keep the optimal $\mathbf{P}$. In this algorithm, different step sizes are probed at each iteration, keeping the one that leads to the minimum value of the cost function C . It is important to mention that the orthonormality condition, which would simplify the expressions of the cost and the gradient, cannot be assumed throughout the process, as the $\mathbf{P}$ matrix is being modified unconstrainedly (even though it converges to an orthonormal matrix).

35

In shape models (Cootes et al. 1995), a set of landmarks is defined on a set of previously aligned shapes. One data vector $\mathbf{s}_i$ is built per shape by stacking
 5 of the x and y coordinates of the landmarks. Next, the mean is subtracted from them and PCA performed on the resulting $\mathbf{x}_i$ data vectors, aiming at representing the shapes with a lower dimensionality and with a higher specificity than the explicit Cartesian coordinates, at
 10 the expense of a certain approximation error. The differences between shape models based on standard and ϕ -PCA will be described.

First, the shapes are aligned with the Procrustes
 15 method and their mean calculated. Rotation, translation and scaling are allowed for aligning the shapes. The alignment parameters and the mean are optimised simultaneously, minimising:

$$20 \quad C_{align} = \sum_{i=1}^N \Phi[\|\mathbf{T}_i(\mathbf{z}_i, \theta_i) - \mu^\Phi\|^2] = \sum_{i=1}^N \Phi[\|\mathbf{s}_i - \mu^\Phi\|^2] = \sum_{i=1}^N \Phi[\|\mathbf{x}_i\|^2]$$

where $\mathbf{T}_i(\mathbf{z}_i, \theta_i)$ represents the aligned $\mathbf{s}_i$ shape according to
 25 the set of parameters θ_i . The constraint $\mu^\top \mu = 1$ prevents the shapes from shrinking towards zero. The iterative algorithm described in Cootes et al. was used for solving the problem:

- 30 1. Normalise the size of the first shape and use it as a first estimate of the mean.
2. Align all the shapes to the current estimate of the mean.
3. Update the estimate of the mean by finding the mean
 35 of the aligned shapes.

4. Normalise the size of the new estimate of the mean.
5. Go to step 2 until convergence.

The mean in the third step must be found by
 5 numerically minimising the cost. However, as minimising $\phi(t^2)$ is equivalent to minimising $t^2 = \|\mathbf{P}\mathbf{P}'\mathbf{x} - \mathbf{x}\|^2$, the alignments in the second step can be easily calculated by minimising the sum of squared distances in the standard way. Another consequence of this property is that the PCA
 10 coordinates of $\mathbf{b}_i$ of a shape can still be calculated in the same way as in the normal PCA:

$$\mathbf{s}_i \approx \mu^\phi + \mathbf{P}\mathbf{b}_i$$

$$15 \quad \mathbf{b}_i = \mathbf{P}'(\mathbf{s}_i - \mu^\phi)$$

The distribution of the principal components of the full-contour model F can be modelled as a conditional
 20 Gaussian, dependent on the principal components of the landmark positions L . If $\mu_F, \mu_L, \Sigma_F, \Sigma_L$ are the means and covariances of the principal components for the two models across the training data, and Σ_{FL}, Σ_{LF} represent their cross-covariances, it is then possible to write:

25

$$P(F|L) = N(\mu, \Sigma)$$

30

$$\mu = \mu_F + \Sigma_{FL}\Sigma_{LL}^{-1}(L - \mu_L) = \Sigma_{FL}\Sigma_{LL}^{-1}L$$

$$\Sigma = \Sigma_{FF} - \Sigma_{FL}\Sigma_{LL}^{-1}\Sigma_{LF}$$

35

where μ is the conditional mean and Σ the conditional covariance matrix for the principal component coordinates of the full contour given the principal component coordinates L of the six points.

5

It is also possible to model the position of the points and landmarks instead of the principal components. The latter has an advantage as the size of the covariance matrices is smaller. If the coordinates of the points are
10 used directly, $\sum LL$ might become non-invertible due to multi-collinearity in the positions, and regularisation would be required.

As shown in Figure 2, the mean of the conditional
15 model can be used as initialisation and the covariance is useful when fitting the model to the images. The conditional covariance is in general much "smaller" than the unconditional covariance of **C**: the differential entropy of the distribution decreases almost 10
20 logarithmic units from the unconditional to the conditional model (from -13.88 to -23.85). It is thus possible to look for the solution around the conditional mean, in a region limited by a certain value of Mahalanobis distance defined according to the new
25 conditional covariance. The search space will hence be reduced, making it easier to fit the model and making shapes relatively far away from the six landmark positions unlikely.

30 In an embodiment having fewer or more than six landmarks, these landmark positions will be used in the same way as in the described embodiment when calculating the conditional covariance.

35 In the described example, the six landmarks in the

training data are not constrained to stay on their corresponding points on the contour. The lack of this constraint allows the 11-D full contour conditional shape model to represent exactly the same shapes as the non-conditional one, although with higher Mahalanobis distances and hence smaller likelihood. This is important because it allows the conditional covariance matrix to remain full-rank, and thus invertible.

10 The active shape model (ASM) is an iterative algorithm that tries to fit the shape model to the contours of the vertebrae in the image. The first step is to find the translation(t_x, t_y), rotation (θ) and scale (s) parameters that best fit the corners of the six given
15 landmark positions to the mean of the shape model. These pose parameters define the transform that allows to switch between the positions of the points in the image X (in "physical" coordinates), and their positions in the shape model "normalised" coordinates x . The pose parameters will
20 be kept constant throughout the process.

Starting from an initial solution, which is calculated in the "normalised" coordinates with the six landmark positions, a translation along the normal to the
25 contour is proposed for every contour point in the model at each iteration. In order to calculate it, a set of candidate positions t are selected along the normal to the contour at each contour point. A grey value profile is then built for each t by sampling the grey levels in the
30 image along the normal to the contour and around the point t . The derivative is calculated for the points in the profile, and then scaled so that the sum of the absolute values of the derivative profile is one. This makes the algorithm robust against contrast variations. The
35 resulting profiles $p_i(t)$ are then compared to the ones of

the training cases on the contour at the same contour position.

If $p_i(t)$ represents the vector of normalised derivatives around point t in the profile around contour point i (let us say in the interval $[t-T_i, t+T_i]$), a fitness function $f_i(t)$ can be calculated for each t by comparing $p_i(t)$ to the model built from the training examples (with semilength T_i):

$$f_i(t) = (p_i(t) - \bar{p}_i)' S_{pi}^{-1} (p_i(t) - \bar{p}_i)$$

where $\bar{p}_i$ is the average of the profiles of length $2T_i+1$ around point i in the training cases, and S_{pi} is a diagonal matrix including the (independent) variances of each element in the profile. The function $f_i(t)$ will just be defined in the interval $[-T_p, T_p]$. Hence, T_p defines how far from the current point location the search for the contour is performed. If t_m minimises $f(t)$, the shift $\frac{2}{3}t_m$ is chosen to make the algorithm evolve in a smoother way.

This new desired shape $X + dX$ is translated into the normalised coordinates, becoming $x + dx$. The shape model parameters b are then updated to fit $x + dx$ as well as possible:

$$(P'W_s)dx = (P'W_sP)db$$

where W_s is a diagonal matrix with weights that measure the importance of each point in the fitting. The weights depend on the magnitude of the displacement and on the goodness of the fit:

$$\omega'_i = \frac{1}{2 + |dX_i|^2} \frac{1}{f_i(t_m)}$$

$$\omega_i = \frac{\omega'_i}{\sum \omega'_j}$$

5 Before updating the contour, it is important to check that **db** leads to a plausible shape. It is at this point that the information in the covariance of the conditional model becomes useful.

$$\begin{aligned} 10 \quad & \mathbf{b}'_{i+1} = \mathbf{b}_i + \mathbf{db} \\ & d = \sqrt{(\mathbf{b}'_{i+1} - \mu)^t \Sigma^{-1} (\mathbf{b}'_{i+1} - \mu)} \end{aligned}$$

and then

$$15 \quad \mathbf{b}_{i+1} = \begin{cases} \mathbf{b}'_{i+1}, & \text{if } d \leq d_{\max} \\ \mu + (\mathbf{b}'_{i+1} - \mu)(d_{\max}/d), & \text{if } d > d_{\max} \end{cases}$$

20

Hence, $d_{\max}$ is the parameter that controls how free the algorithm is to fit the contour to the edges in the image. A large value allows the result to move around the principal component space, which can lead to implausible solutions if the edges are not clear in the image. A small value makes the algorithm rely mostly on the model, leading to more conservative solutions, closer to the mean of the distribution. This can prevent the algorithm from finding the correct solution, especially in abnormal cases with fractures or osteophytes, in which the solution is relatively far from the initialisation in the principal component space, as shown in Figure 3. Specifically, Figure 3 shows the influence of maximum allowed by Mahalanobis distance on the result. In Figure 3(a) the

25
30
35

shape model is unable to fit the contour to the osteophyte. In figure 3(b) the threshold has been increased by 1.5 and the contour approximates the osteophyte better.

5

The new coordinates are easily calculated using:

$$x \approx \bar{x} + \mathbf{Pb}$$

10 and then transformed back to physical coordinates by the transform defined by $-t_x, -t_y, -\Theta, s^{-1}$. A new translation is then proposed for each point once more, starting a new iteration. When the difference between the shapes at two consecutive iterations is smaller than a certain limit,
15 the process stops. For example, when the distance between consecutive estimated contours is less than 2 pixels or less than 1 pixel, the iterative process stops.

First of all, several preliminary experiments were
20 run in order to find suitable values for the different parameters in the algorithm. Regarding the selection of the number of principal components to be kept, it depends on the proportion of the variance to be preserved. Seven and eleven components, for the six landmarks and the full
25 contour respectively, were enough to keep approximately 97% of the total variance. From that point on, raising the proportion becomes very expensive in terms of the number of principal components.

30 Regarding the profile lengths, it was found that using a grey value model profile semilength $T_i = 12$ pixels (2.7mm) and making $T_p = 8$ pixels (1.8mm) leads to good results. T_p represents how far from the current solution one tries to find the contour. Making this parameter too
35 large would make the search profile too long and hence

make it more likely that the algorithm captures a wrong edge, especially if this edge does not represent an implausible shape. This typically happens in the type of "double contour" cases shown in Figure 5(b).

5

Finally, the choice of the maximum Mahalanobis distance $d_{\max}$ was made with the help of the χ^2 distribution for eleven degrees of freedom. Arbitrarily fixing a 10% tail probability, a compromise limit equal to 10 4.1 was picked. All the parameters were kept fixed for all the further experiments, unless additional description is provided.

A leave one out experiment was carried out. For each 15 vertebra, a complete model (six landmarks, full contour, profiles) was built from the other images and used for the contour location. The distance between each point of the physician-annotated contour and the closest point in the detected contour (point-to-line distance) was calculated 20 for each vertebra. The distribution of the error is represented in Figure 4. Specifically, Figure 4 shows the distance to real contour in the form of a histogram and cumulative distribution function.

25 The RMS error was equal to 0.68 mm, while the mean error was 0.48mm and its standard deviation was 0.48mm. 89% of the points were located within 1mm of the manually annotated contour, 96% within 1.5mm and 98% within 2mm. The average of the maximum errors in each vertebra was 30 1.53mm.

The fact that fractures were given a higher weight when building the model prevents the performance of the model from decreasing too much when fitting shapes 35 corresponding to fractured vertebrae. The mean error for

the fractures was 0.54mm. 86% of the points were 1mm within the real contour in fractures, and 94% within 1.5mm.

5 The results were also judged visually, especially in cases with osteophytes and fractures, in which the contour may be oversmoothed. Some sample images are shown in Figure 5. As can be seen, the model has difficulties approximating the ground truth when the ground truth takes
10 an abrupt turn (like around osteophytes), but otherwise works quite well, as the quantitative results suggest.

It is also important to note that the results are not very sensitive to changes in the different tuned
15 parameters. The increments in the mean error when changing the main parameters are:

- * For the profile length used for building the model: 4µm in a 4 pixels wide region around the chosen value;
- * For the profile length used in the contour search: 6
20 µm in a 4 pixels wide region around the chosen value; and
- * For the cut-off Mahalanobis distance: 25 µm in a 1 unit wide region around the chosen value.

These values suggest that the method is robust
25 against non-optimal choices of the parameters. The cut-off Mahalanobis distance is the one that affects the results the most, representing a trade-off between freedom (better approximation to outliers) and safety (lower likelihood of implausible shapes). The profile lengths
30 affect mostly the convergence speed.

In the preferred embodiment, higher weights are given when building the model to fractures, lowering the overall mean performance but increasing the accuracy in the most
35 difficult cases and hence lowering the largest errors,

too. In spite of that, an average point-to-line error lower than 0.5mm was achieved. 95% of the errors were lower than 1.36mm. The performance is higher than those of the most recent systems described.

5

By annotating the six points and the full contour in different passes, the solution is not constrained to pass through the six landmarks. This adds flexibility to the shape model, making it possible for it to fit a wider
10 range of different shapes. The fact that different radiologists made the annotations also adds some interesting variability to the model.

Figure 6 shows an illustration of the error depending
15 on the point number. The points corresponding to the six landmarks are marked with a star and the distances from the manually placed landmarks to the true contour are marked with crosses. It is clear here that the mean position error peaks clearly just after the third landmark
20 and just before the fourth, which are the typical locations of osteophytes. The curve has local minima around the points corresponding to the landmarks, except for the middle point of the lower endplate. This is possible because the six landmarks are not constrained to
25 be on the contour. The mean distances from these points to the contour are marked with crosses in the same figure.

In an embodiment, further training cases including osteophytes could be used, or a higher weighting could be
30 given to such cases, as with the fractures. Alternatively, more flexibility could be allowed for the ASM around points 27 and 41 (the typical locations for osteophytes), for example by modifying the contour with some kind of smooth curve to fit the points. A more
35 complex solution could be to improve the correspondence of

the points around the osteophytes. Minimum description length (MDL), in which the best hypothesis for a given set of data is the one that leads to a largest compression, can be used for choosing corresponding points.

5

Results are also provided of an alternative dataset based on the alternative weighting applied to outliers using α -PCA. The study is based on a dataset which consists of lateral X-rays from the spine of 141 patients. 10 Vertebrae L1 through L4 were outlined by three different expert radiologists, providing the ground truth of the study. 65 landmarks were extracted for each vertebra using the MDL algorithm, described in Thodberg "Description Length Shape and Appearance Models", 15 proceedings of Information Processing in Medical Imaging (2003) Springer. The same radiologists also provided information regarding the fracture type (wedge, biconcave, crush) and grade (mild, medium, severe) for the vertebrae. In addition, they also annotated the six landmarks used in 20 the standard six-point morphometry, located on the corners and in the middle point of both vertebra endplates. These points define the anterior, middle and posterior heights, which are used to estimate the fracture grade and type.

25 Both normal PCA and α -PCA (for different values of α) were applied on the dataset keeping 7 (α -) PCA coordinates, capable of preserving approximately 95% of the total variance in the data in all the cases. For both algorithms the mean and maximum squared reconstruction 30 errors were calculated. The dependence of the error on the number of fractures in the training set was also studied. It should be noted that a higher number of components would achieve better precision and still provide a good trade-off with respect to the specificity 35 of the model, but a smaller amount was kept in this

experiment in order to better illustrate the difference between PCA and α -PCA.

Finally, PCA and α -PCA were tested in an active
 5 shape model for segmenting the L1-L4 vertebrae in the images. Two shape models were built, one for the six landmarks and the other for the full contour, and the relationship between the (α -) PCA coordinates of both models fitted to a conditional Gaussian distribution. In
 10 order to allow for more flexibility in the model, a higher number of principle components was utilised, seven for the six landmarks and eleven for the complete contour, keeping approximately 98% of the total variance in both cases.

15 The mean of the conditional distribution was used as initialisation for the segmentation of the full contour. At each iteration, the gray level information along a profile perpendicular to the contour was used to calculate a desired position for each point at the following
 20 looping. The new contour can then be calculated by fitting the model to the new points. The conditional covariance was used to measure the Mahalanobis distance from the new (α -)PCA coordinates $\mathbf{b}$ to the conditional mean. In case of it being larger than a certain threshold
 25 $D_{\max}$, the vector is scaled down $\mathbf{b}' = \mathbf{b}(D_{\max}/D(\mathbf{b}))$ to ensure $D(\mathbf{b}) \leq D_{\max}$. This way, the solution is constrained to stay close to the six landmarks. The process is repeated until convergence.

30 Figures 7(a) and 7(b) show the dependence on α of the sum of squared errors when fitting the model to labelled points. The maximum error decreases with α as expected doing it faster for the fractures. The mean error shows how values of α lower than one tend to
 35 increase the error in fractures, as they are no longer

important in the model, and decrease it in unfractured vertebrae, even if not by much. Unfractured vertebrae are in general quite well modelled already. Values larger than one initially improve the results in fractures, at the expense of making them slightly worse in unfractured vertebrae. Finally, if α increases too much, the model tends to fit merely the most unlikely cases, making the average results worse both for unfractured vertebrae and mild fractures.

10

In this experiment, the model was built with all the unfractured vertebrae and different fractions of the total amount of available fractures: from 12.5 % to 100 % in 12.5 % increments. α was set to 1.75, providing a good trade-off between the maximum and mean errors in fractured and unfractured vertebrae, according to the results presented above. Figure 7 (c) shows that α -PCA is especially useful, clearly outperforming the normal PCA, when the number of fractures in the training set is relatively small.

20

Vertebrae L1 through L4 were segmented from the available images using a shape model conditioned on the six landmarks annotated by the radiologists, using both standard and α -PCA ($\alpha=1.75$). The experiments were performed in a leave-one-out fashion: the model used for segmenting a certain image is built upon all the other ones.

25

The point-to-line errors from the true contour to the output of the algorithm are displayed in the table below, along with the p-values resulting from a paired, double-sided t-test and a paired, double-sided Wilcoxon signed rank test. The results show that the standard PCA leads to a lower mean error in unfractured vertebrae, but α -PCA

30

35

provides more uniform results along the different grades of fractures severity, at the expense of a slight increase in the total mean error. Moreover, α -PCA significantly outperforms the standard PCA in fractures, especially in the severe ones. It also has the property of assigning different importance to each case in a continuous manner without requiring fracture information for the training data. If this information was available, it would be possible to build two different models, but then a large number of training fractures would be required. Regarding the p-values, both tests indicate that the difference in the means between the two setups is significant.

	No. shapes	Error PCA (mm)	Error α -PCA (mm)	p t-test	p signed rank
Unfractured	500	0.44	0.47	$1.79 \cdot 10^{-4}$	$6.05 \cdot 10^{-4}$
Mild fractures	15	0.62	0.56	$1.14 \cdot 10^{-2}$	$1.25 \cdot 10^{-2}$
Medium fractures	38	0.66	0.57	$7.87 \cdot 10^{-3}$	$1.40 \cdot 10^{-2}$
Severe fractures	11	0.97	0.60	$4.78 \cdot 10^{-3}$	$9.77 \cdot 10^{-3}$
All fractures	64	0.70	0.57	$3.10 \cdot 10^{-11}$	$3.53 \cdot 10^{-12}$

Finally, two radiographs which have been segmented with a standard and α -PCA ($\alpha=1.75$) are displayed along with the contour provided by the radiologists in Figure 8. They both correspond to severe fractures. α -PCA provides a better approximation of the real shape, especially around the points in which it changes its direction rapidly.

It will be appreciated that modifications of, and alterations to, the embodiments as described and illustrated may be made within the scope of this application.

CLAIMS

1. A method of locating a contour of a structure in an
5 image by processing said image including the structure,
comprising the steps of:

taking a starting set of digital data
representative of the image including the structure, the
structure in said image having annotated on it from three
10 to ten landmark positions;

fitting a statistical model of said structure to
the landmark positions annotated on the image, and making
an initial estimate of the contour of the structure; and

using grey level information derived from points
15 adjacent the estimated contour iteratively to move the
contour boundary to produce a final estimate of the
contour of the structure.

2. A method as claimed in claim 1, wherein the structure
20 is a bone.

3. A method as claimed in claim 1 or claim 2, wherein
the structure is a vertebra and the image is of part of a
spine including said vertebra.

25 4. A method as claimed in claim 3, further comprising
training the statistical model of the vertebra using
information from points approximating respective contours
of a set of other vertebrae.

30 5. A method as claimed in claim 4, further comprising
training the statistical model using information from
three to ten landmark positions annotated on vertebrae in
said set of other vertebrae.

35

6. A method as claimed in claim 4 or claim 5, wherein the set of vertebrae used in training the statistical model includes unfractured and fractured vertebrae.

5 7. A method as claimed in any preceding claim wherein the statistical model is a conditional point distribution model.

8. A method as claimed in claim 7, wherein the
10 conditional point distribution model is constructed from information approximating the respective contours of a set of vertebrae and from information of three to ten landmarks annotated on said set of vertebrae.

15 9. A method as claimed in claim 7 or claim 8, wherein the conditional point distribution model is constructed from a first point distribution model constructed from information approximating the respective contours of a set of vertebrae and a second point distribution model
20 constructed from information of three to ten landmarks annotated on said set of vertebrae.

10. A method as claimed in any one of claims 7 to 9, wherein the conditional point distribution model is a
25 conditional Gaussian dependent on the positions of the landmark positions in the image being processed.

11. A method as claimed in claim 10, wherein the conditional point distribution model is a conditional
30 Gaussian modelling the principal components of the point distribution model constructed from information approximating the respective contours of a set of vertebrae, dependent on the principal component coordinates of the six landmark positions in the image
35 being processed.

12. A method as claimed in any of claims 7 to 11, wherein
the initial estimate of the contour is the mean of the
conditional point distribution model fitted to the
5 landmark positions.

13. A method as claimed in any preceding claim, wherein
the iterative movement of the estimated contour is
constrained by the conditional covariance and the
10 proximity of the conditional mean to the current estimate
of the contour.

14. A method as claimed in any of claims 1 to 12, wherein
the movement of the contour boundary is constrained by
15 restricting divergence of grey level information derived
from points adjacent the estimated contour with equivalent
information derived from said statistical model.

15. A method as claimed in any preceding claim, wherein
20 the iterative movement of the contour boundary is
continued until the difference between the estimated
contours at two consecutive iterations is smaller than a
preset limit.

25 16. A method as claimed in any preceding claim, wherein a
grey level profile is built by sampling grey level
information in the image along the normal to the contour
across each contour point.

30

35

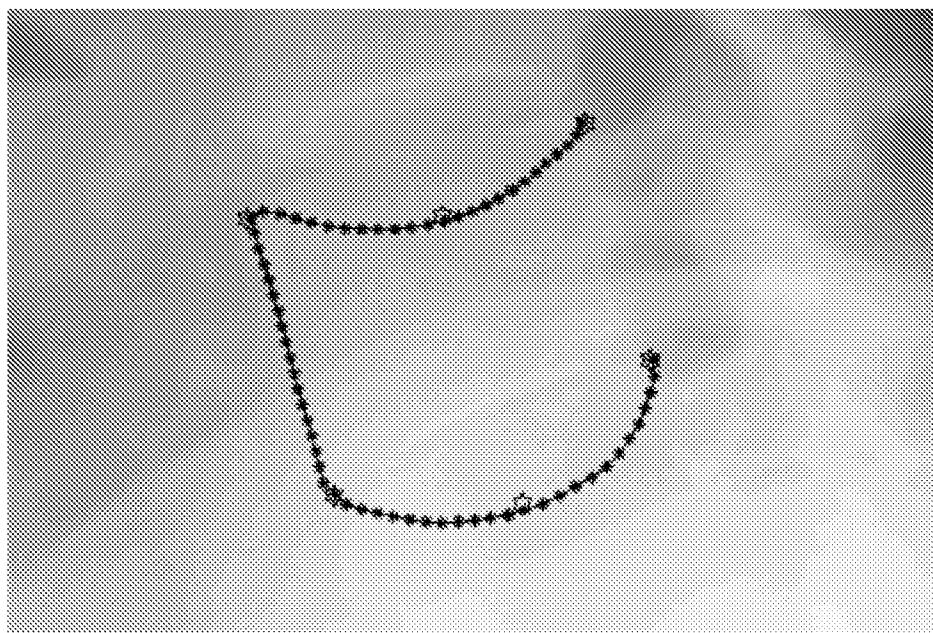


Figure 1

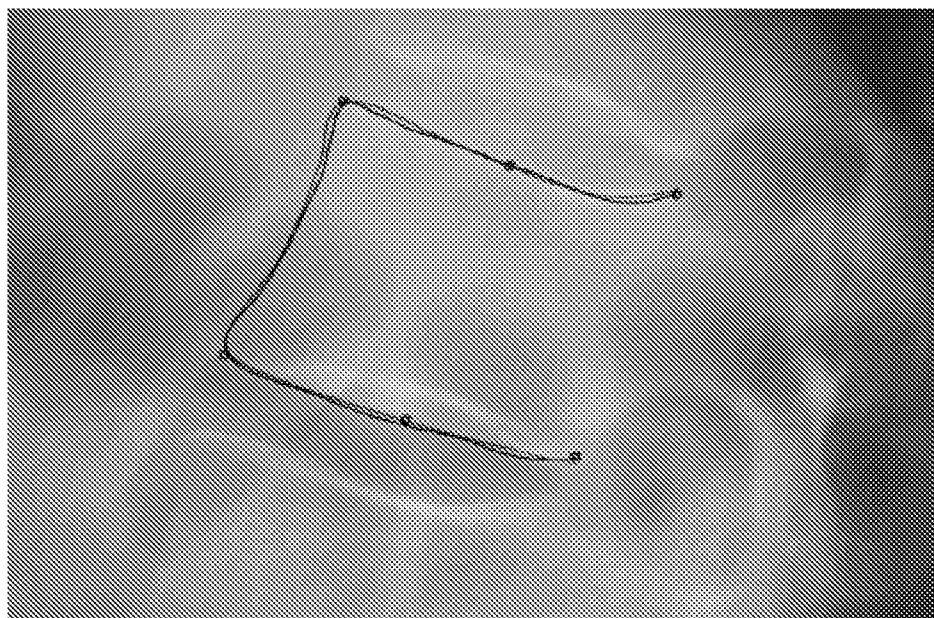


Figure 2

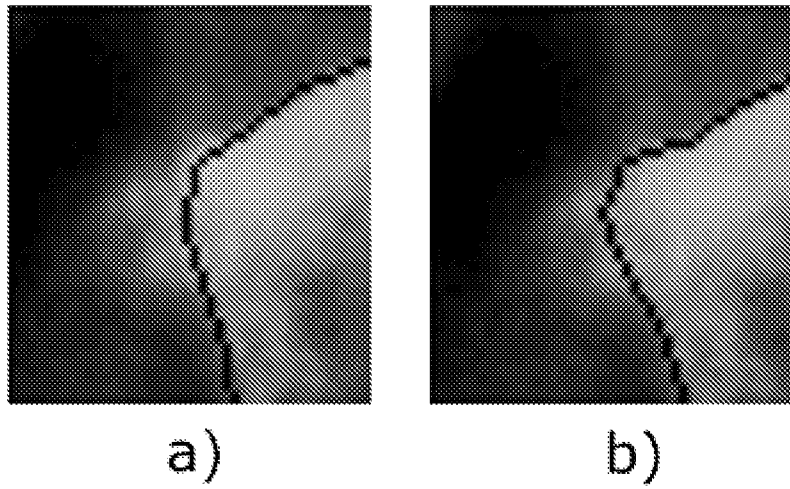


Figure 3

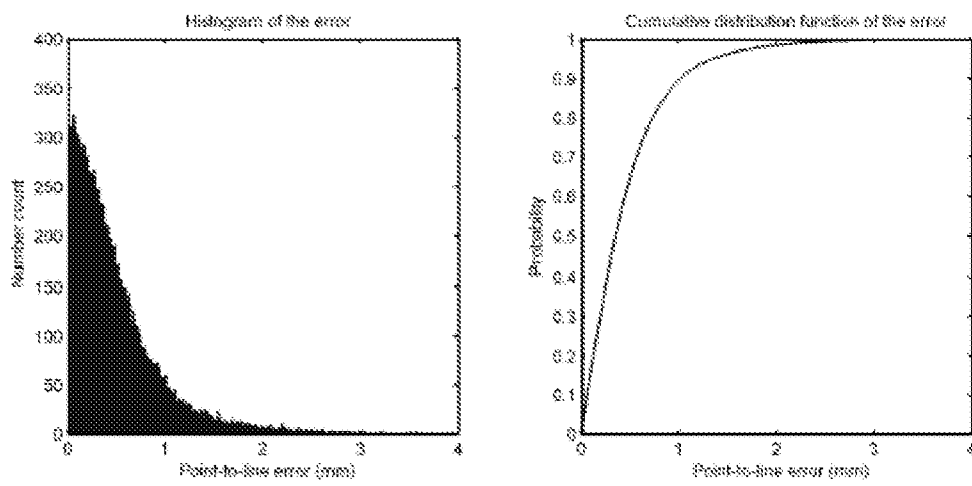


Figure 4

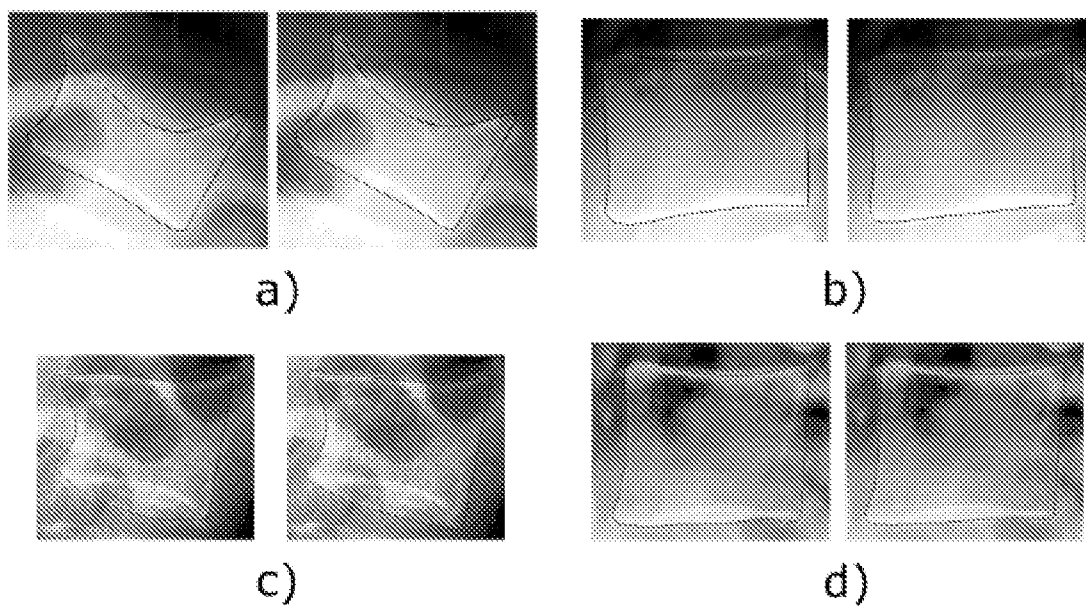


Figure 5

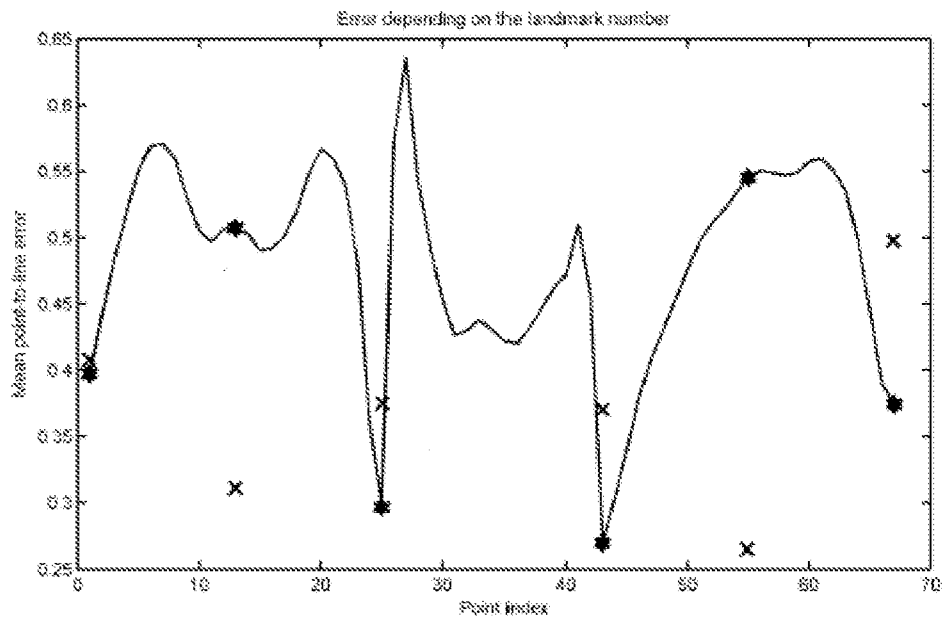


Figure 6

6/7

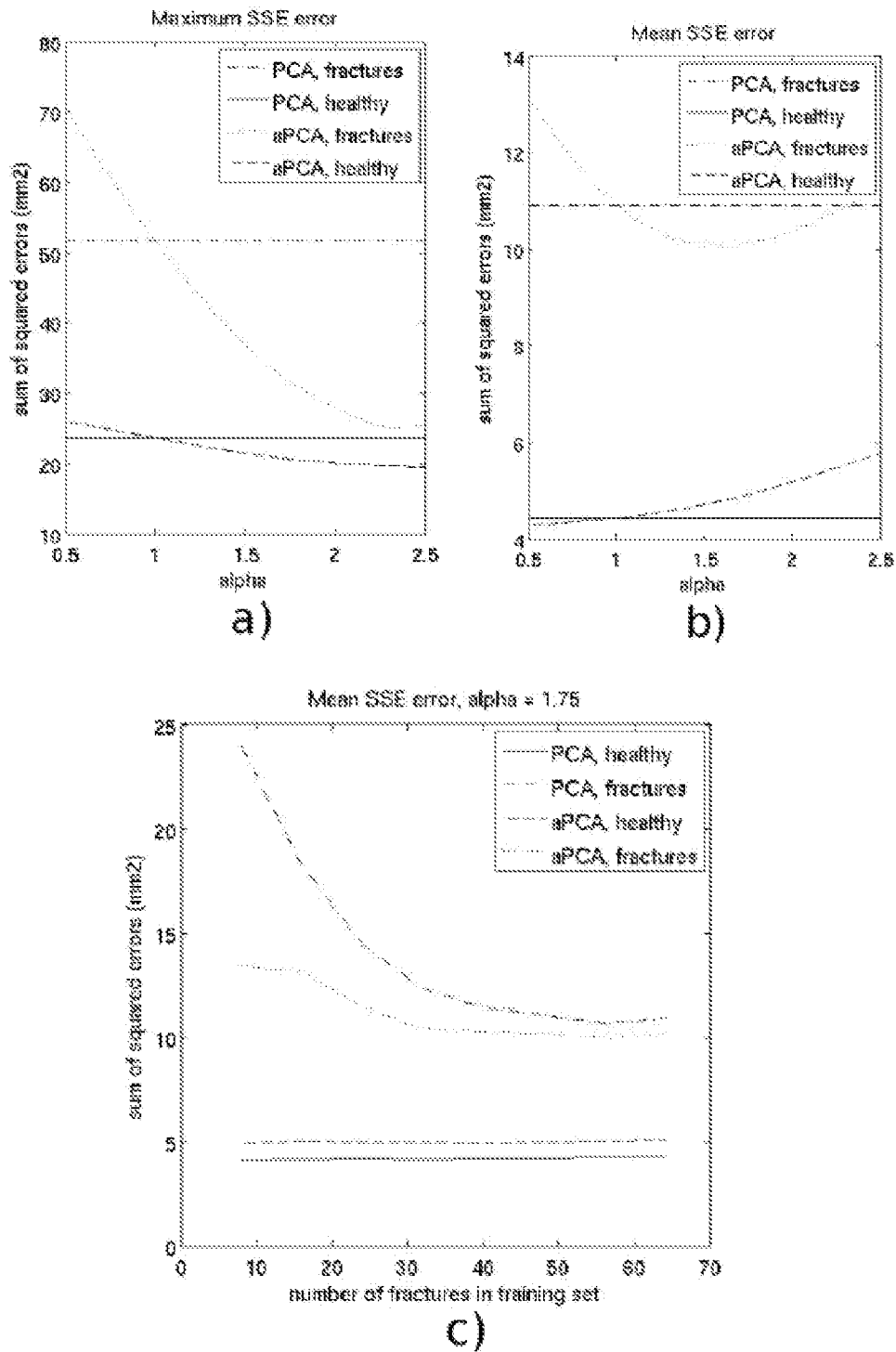


Figure 7

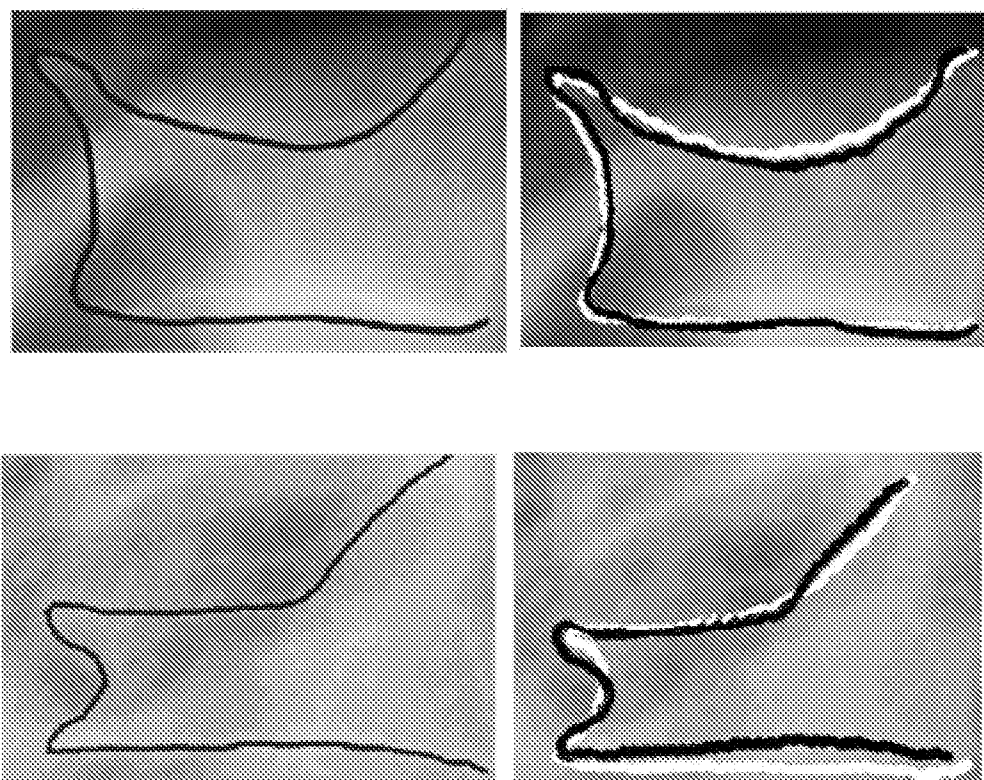


Figure 8