Localisation of Vertebrae on DXA Images Using Constrained Local Models with Random Forest Regression Voting

P.A. Bromiley, J.E. Adams and T.F. Cootes

Abstract Fractures associated with osteoporosis are a significant public health risk, and one that is likely to increase with an ageing population. However, many osteoporotic vertebral fractures present on images do not come to clinical attention or lead to preventative treatment. Furthermore, vertebral fracture assessment (VFA) typically depends on subjective judgement by a radiologist. The potential utility of computer-aided VFA systems is therefore considerable. Previous work has shown that Active Appearance Models (AAMs) give accurate results when locating landmarks on vertebra in DXA images, but can give poor fits in a substantial subset of examples, particularly the more severe fractures. Here we evaluate Random Forest Regression Voting Constrained Local Models (RFRV-CLMs) for this task and show that, while they lead to slightly poorer median errors than AAMs, they are much more robust, reducing the proportion of fit failures by 68 %. They are thus more suitable for use in computer-aided VFA systems.

1 Introduction

Osteoporosis is a common skeletal disorder defined by a reduction in bone mineral density (BMD) resulting in a T-score of <2.5 (i.e. more than 2.5 standard deviations below the mean in young adults), measured using dual energy X-ray absorptiometry (DXA) images [15]. It significantly increases the risk of fractures, most commonly

P.A. Bromiley (✉) · T.F. Cootes
Imaging Sciences Research Group, University of Manchester,
Manchester, UK
e-mail: paul.bromiley@manchester.ac.uk

T.F. Cootes
e-mail: timothy.f.cootes@manchester.ac.uk

J.E. Adams
Radiology & Manchester Academic Health Science Centre,
Central Manchester University Hospitals NHS Foundation Trust,
Manchester, UK
e-mail: judith.adams@manchester.ac.uk

J. Yao et al. (eds.), *Recent Advances in Computational Methods and Clinical Applications for Spine Imaging*, Lecture Notes in Computational Vision and Biomechanics 20, DOI 10.1007/978-3-319-14148-0_14

occurring in the hips, wrists or vertebrae. Approximately 40 % of postmenopausal caucasian women are affected, increasing their lifetime risk of fragility fractures to as much as 40 % [15]. Osteoporosis therefore presents a significant public health problem for an ageing population.

Accurate identification of vertebral fractures is clinically important in the diagnosis of osteoporosis. Radiological assessment typically uses a semi-quantitative approach [10] requiring subjective judgement by a radiologist. Furthermore, only about one third of vertebral fractures present on images come to clinical attention; they are frequently not noted by radiologists, not entered into medical notes, and do not lead to preventative treatments [6]. Many of these cases involve images acquired for purposes other than VFA. However, a recent multicenter, multinational prospective study [8] has found a false negative rate of 34 % in VFA performed on lateral radiographs of the thoracolumbar spine. The potential utility of computer-aided VFA systems is therefore considerable.

Several authors have investigated the use of methods based on statistical shape models to segment vertebrae in both radiographs and DXA images (e.g. [16]) as a preliminary step for VFA. However, state-of-the-art results achieved using active appearance models (AAMs) [17] exhibit significant numbers of large errors due to fit failures, particularly on the more severely fractured vertebrae. This is the result of two effects. First, osteoporosis patients with vertebral fractures most commonly have only one or two fractures (e.g. [11]). Therefore, models encompassing multiple vertebrae must typically be trained on datasets containing more normal than fractured vertebrae, potentially introducing a bias against the most severe shape changes. Second, work on natural images of faces has shown that holistic methods such as AAMs, which rely on a single model of shape and intensity that covers all landmarks, tend to generalise poorly [7]. An alternative is to use a set of models, each covering an individual landmark. The ambiguity inherent in the use of local image patches may be dealt with by imposing a global shape constraint (e.g. [9]). In particular, regression voting (RV) methods (e.g. [19]), especially those (e.g. [3, 7, 13]) based on Random Forests (RFs) [1] tend to be robust. The RFRV Constrained Local Model (RFRV-CLM) [3, 13], which uses a RF regressor for each point constrained by a global shape model, has been applied successfully to the annotation of landmarks both in facial (e.g. [3]) and clinical (e.g. [13]) images, and shows superior generalisation on facial images compared to the AAM [18].

The hypothesis investigated here is that the superior generalisation capability of the RFRV-CLM will lead to performance improvements, compared to AAMs, in terms of the number of fit failures on DXA spinal images. RFRV-CLMs are applied to annotate vertebral landmarks in a dataset of 320 such images, the first time they have been applied to this task. Extensive experiments were performed to investigate the effect of the free parameters. The results were compared to those from [17] using AAMs on the same dataset, and show that RFRV-CLMs provide a considerable (68 %) reduction in fit failures across all vertebral classifications.

2 Method

The reader is referred to [3, 13] for full details of the RFRV-CLM algorithm; the key points are described below.

Constrained Local Models CLMs [5] build on previous work on Active Shape Models (ASMs) [4] and AAMs [2], providing a method for matching the points of a statistical shape model to an image. They combine global shape constraints with local models of the pattern of intensities. Given a set of training images with manual annotations $\mathbf{x}_l$ of a set of n landmarks $l = 1, \ldots, n$ on each, a statistical shape model is trained by applying principal component analysis (PCA) to the aligned shapes [2]. This yields a linear model of shape variation, which represents the position of each landmark l using

$$\mathbf{x}_l = T_\theta(\bar{\mathbf{x}}_l + \mathbf{P}_l\mathbf{b} + \mathbf{r}_l) \tag{1}$$

where $\bar{\mathbf{x}}_l$ is the mean position of the point in a suitable reference frame, $\mathbf{P}_l$ is a set of modes of variation, $\mathbf{b}$ are the shape parameters, $\mathbf{r}_l$ allows small deviations from the model, and T_θ applies a global transformation (e.g. similarity) with parameters θ.

To match the model to a query image, $\mathbf{I}$, the overall quality of fit Q, of the model to the image is optimised over parameters $\mathbf{p} = \{\mathbf{b}, \theta, \mathbf{r}_l\}$

$$Q(\mathbf{p}) = \Sigma_{l=1}^{n} C_l(T_\theta(\bar{\mathbf{x}}_l + \mathbf{P}_l\mathbf{b} + \mathbf{r}_l)) \quad \text{s.t.} \quad \mathbf{b}^T\mathbf{S}_b^{-1}\mathbf{b} \leq M_t \quad \text{and} \quad |\mathbf{r}_l| < r_t \tag{2}$$

where C_l is a cost image for the fitting of landmark l, $\mathbf{S}_b$ is the covariance matrix of shape model parameters $\mathbf{b}$, M_t is a threshold on the Mahalanobis distance, and r_l is a threshold on the residuals. M_t is chosen using the cumulative distribution function (CDF) of the χ^2 distribution so that 98 % of samples from a multivariate Gaussian of the appropriate dimension would fall within it. This ensures a plausible shape by assuming a flat distribution for model parameters $\mathbf{b}$ constrained within hyper-ellipsoidal bounds [2]. In the original work [2], C_l was provided by normalised correlation with a globally constrained patch model.

RF Regression Voting in the CLM Framework. In RFRV-CLM, C_l in Eq. 2 is provided by voting with a Random-Forest (RF) regressor. To train the RF for a single landmark, the shape model is used to assess the global pose, θ, of the object in each image by minimising $|T_\theta(\bar{\mathbf{x}}) - \mathbf{x}|^2$. Each image is resampled into a standardised reference frame by applying the inverse of the estimated pose. The model is scaled so that the width of the reference frame of the mean shape is a given value, w_{frame}. Sample patches of area w_{patch}^2 are then generated from the resampled images at a set of random displacements from the true point positions. The displacements $\mathbf{d}_j$ are drawn from a flat distribution in the range $[-d_{max}, +d_{max}]$ in x and y. Finally, image features $\mathbf{f}_j$ are extracted from the sample patches. Haar-like features [20] are used, as they have proven effective for a range of applications and can be calculated efficiently from integral images. To allow for inaccurate initial estimates of the pose and to make the detector locally pose-invariant, the process is repeated with random perturbations in scale and orientation of the pose estimate. A RF [1] is then trained,

using a standard, greedy approach, with the feature vectors $\mathbf{f}_j$ as inputs and the displacements $\mathbf{d}_j$ as regression targets. Each tree is trained on a bootstrap sample of N_s pairs $\{(\mathbf{f}_j, \mathbf{d}_j)\}$ from the training data. At each node, a random sub-set of n_{feat} features are chosen from this sample, and a feature f_i and threshold t that best split the data into two compact groups are selected by minimising an entropy measure [3]. Splitting terminates at either a maximum depth, D_{max}, or a minimum number of samples, N_{min}. The process is repeated to generate a forest of size n_{trees}.

RFRV-CLM Fitting Fitting to a query image is initialised via an estimate of the pose of the model e.g. from a small number of manual point annotations or a previous model, providing initial estimates $\mathbf{b}$ and θ (see Sect. 3). Equation 2 is then optimised as follows. The image is resampled in the reference frame using the current pose. Cost images C_l are then computed by evaluating a grid of points in the resampled images over a region of interest around the current estimate of each point; the grid size is defined by a search range $[-d_{search}, +d_{search}]$, and the cost images are calculated for all landmarks independently. At each point $\mathbf{z}_l$ in the grid, the required feature values are extracted and the RF regressor R_l applied. R_l then casts a vote into a cost image C_l using $C_l(\mathbf{z}_l + \delta) \rightarrow C_l(\mathbf{z}_l + \delta) + c$. Each leaf node of the RF contains the mean $\bar{\mathbf{d}}$ and covariance $\mathbf{S}_d$ of the random displacements $\mathbf{d}_i$ from the true point position, in the reference frame, of its training samples. This supports several voting styles (c, δ); a single, unit vote at $\bar{\mathbf{d}}$, or probabilistic voting by weighting with $|\mathbf{S}_d|^{-0.5}$, or by casting a Gaussian spread of votes $N(\bar{\mathbf{d}}, \mathbf{S}_d)$.

The point positions are re-estimated by finding the lowest cost point within a disk of radius r of the current position in each cost image, applying the shape model and moving $\mathbf{b}$ to nearest valid point on the limiting ellipsoid if the shape constraint in Eq. 2 is violated, updating all point positions using $\mathbf{x}_l \rightarrow T_{\theta_r}(\bar{\mathbf{x}}_l + \mathbf{P}_l\mathbf{b} + \mathbf{r}_l)$, and iterating whilst reducing $r \rightarrow k_r r$. The initial disk radius r_{max} was set to the search range d_{search}, the search was terminated at $r_t = 1.5$ pixels (in the reference image), and k_r was set to 0.7. The optimisation is described in full in Algorithm 1.

3 Evaluation

A series of experiments was performed to optimise the various free parameters and options of the RFRV-CLM for application to the task of vertebral localisation in DXA images, and to compare the results to those achieved in [17] using AAMs. To facilitate this comparison, the same dataset and performance metrics were used. The dataset consisted of 320 DXA VFA images scanned on various Hologic (Bedford MA) scanners, obtained from: (a) 44 patients from a previous study [14]; (b) 80 female subjects in an epidemiological study of a UK cohort born in 1946; (c) 196 females attending a local clinic for DXA BMD measurement, and for whom the referring physician had also requested VFA (as approved by the local ethics committee). Manual annotations of 405 landmarks were available for each image, covering the thoracic vertebrae from T7 to T12 and the lumbar vertebrae from L1 to L4. Each of

Algorithm 1 Iterative model matching procedure to estimate the shape and pose parameters in the reference frame, given a set of feature point based cost images C_l.

SHAPE MODEL AND POSE PARAMETER OPTIMISATION

Input: r_{max}, r_t, k_r, $\mathbf{x}_l$ and C_l $\forall 1 \leq l \leq n$

1. Set $r \rightarrow r_{max}$, $\theta_r \rightarrow$ Identity, $\mathbf{x}_l \rightarrow \bar{\mathbf{x}}_l + \mathbf{P}_l\mathbf{b}$, $\mathbf{r}_l = 0$
2. While $r \geq r_t$
   a. For every feature point l, find the best point $\hat{\mathbf{y}}_l$ in a disk of radius r around the current estimate
      $\hat{\mathbf{y}}_l \rightarrow \arg\max_{\mathbf{y}_l:|\mathbf{y}_l-\mathbf{x}_l|<r} C_l(\mathbf{y}_l)$
   b. Fit the shape model to these best points to estimate shape and pose parameters $\{\mathbf{b}, \theta_r\}$ by solving
      $\hat{\mathbf{y}}_l = T_{\theta_r}(\bar{\mathbf{x}}_l + \mathbf{P}_l\mathbf{b})$
   c. If $\mathbf{b}^T\mathbf{S}_b^{-1}\mathbf{b} > M_t$ then move $\mathbf{b}$ to nearest valid point on limiting ellipsoid
   d. Update all feature point positions using $\mathbf{x}_l \rightarrow T_{\theta_r}(\bar{\mathbf{x}}_l + \mathbf{P}_l\mathbf{b} + \mathbf{r}_l)$
   e. Set $r \rightarrow k_r r$ with $0 < k_r < 1$
3. Transform the resulting feature point positions into the image frame using T_θ with $\theta \rightarrow \theta \circ \theta_r$

these vertebrae in each image was also classified by an expert radiologist into one of five groups (normal, deformed but not fractured, and grade 1, 2 and 3 fractures according to the Genant definitions [10]; see Fig. 1).

Following [3, 13], 2-stage, coarse-to-fine RFRV-CLMs were used and, in common with [17], individual models were trained for each vertebra; each covered the target vertebra and its two neighbours, or one neighbour for the top (T7) and bottom (L4) vertebrae. Fitting of the first-stage model was initialised using the approximate centre points of each of the vertebrae covered by the model, calculated from the manual annotations as the centroid of the two central points on the upper and lower vertebral end-plates. This approach was adopted to avoid the significant reduction in random error, compared to individual manual annotations, that would occur if the centre points were calculated as the centroids of all manual annotations on each vertebra. Second-stage fitting was initialised using the results from the first stage. To compensate for the aperture problem present when annotating points on an extended edge, errors on automatic annotations were calculated as the mean, over each vertebra, of the minimum Euclidean distances between the automatic annotations and a Bezier spline through the manual annotations. This was applied to the points on the central vertebra in each model i.e. no use was made of the multiple fits for each point (see Sect. 4).

RFRV-CLM Parameter Optimisation. The free parameters of the RFRV-CLM, as described in Sect. 2, were divided into two sets; RF structure parameters (n_{trees}, n_{feat}, N_{min} and D_{max}), and image parameters (w_{frame}, w_{patch}, d_{max} and d_{search}).[1] These were optimised empirically and, to limit processor time requirements, a sequential approach was applied across both parameters and stages i.e. each first-stage parameter

[1] Parameters w_{patch}, d_{max} and d_{search} were defined in the reference image.

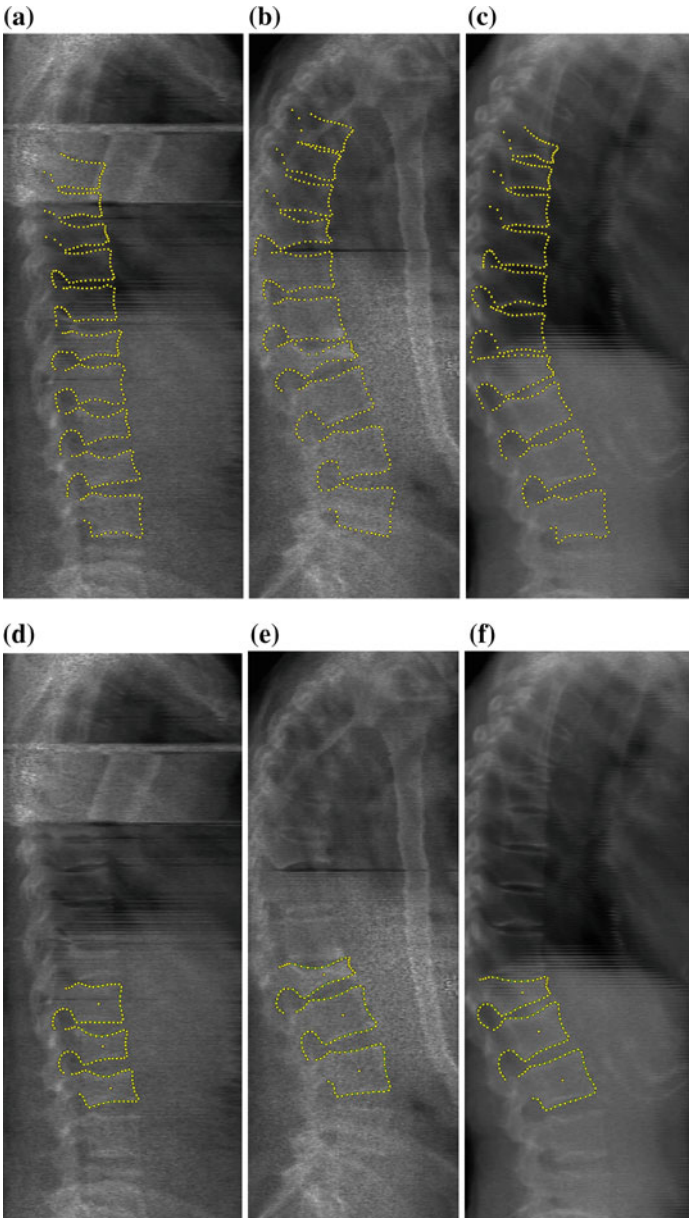

Fig. 1 Example DXA spinal images. **a**–**c** 405-point manual annotations. **d**–**f** Automatic annotation of the L2 vertebra (using the L1-L3 model), using the fully optimised, 2-stage RFRV-CLM. Example (**a**, **d**) shows grade 2 fractures on L2 and L3, (**b**, **e**) show a grade 3 fracture on L1, and (**c**, **f**) show a grade 3 fracture on L1 and a grade 1 fracture on L2

was optimised independently without applying the second-stage model; the optimal values were then fixed, and each second-stage parameter was optimised independently in a two-stage approach. Furthermore, the optimisation experiments were performed only on the L2 vertebra (i.e. the L1-L3 triplet model). L2 was chosen as it was the least obscured by confounding bony structures (ribs, scapulae, the iliac crest etc.) and imaging artefacts, minimising the contamination of the results with fitting failures. The optimisation results were then extended to the other vertebral levels by scaling the optimised w_{frame} using data on mean vertebral heights from [12], such that all image-based parameters were scaled. The data set was divided randomly into halves for training and testing. RF training includes a stochastic element, both in the random selection of data used to train each tree, and the random sub-selection from that data at each node. Therefore, each experiment was repeated five times to evaluate random errors.

For the sake of brevity, complete results are reported only for w_{frame} and w_{patch}, the parameters showing the greatest effect on performance; these are shown in Fig. 2. The graphs show the proportional area under the CDF of mean point-to-curve error on the L2 vertebra across the 160 test images. Performance generally increased with first-stage w_{frame}; however, d_{search} and w_{patch} are defined in the reference frame, and so reducing w_{frame} increases the capture range. Therefore, w_{frame} was set using the point at which the performance increase ceased to be significant, giving 40 and 110 pixels for the first and second stages. Varying w_{patch} had a smaller effect on performance over most of the range tested, but showed more complex behaviour; values of 18 and 21 pixels were selected for the first and second stages, respectively, since these were close to optimal over large portions of the tested ranges of w_{frame}. The remaining parameters were optimised similarly, giving (first stage, second stage): $n_{trees} = 2, 15$; $n_{feat} = 100, 200$; $N_{min} = 1, 1$; $D_{max} = 30, 30$; $w_{frame} = 40, 110$ pixels; $w_{patch} = 18, 21$ pixels; $d_{max} = 15, 15$ pixels; $d_{search} = 15, 10$ pixels (all pixel units except for w_{frame} were defined in the reference frame). In general, the dependence of performance on parameters was weak over large ranges.

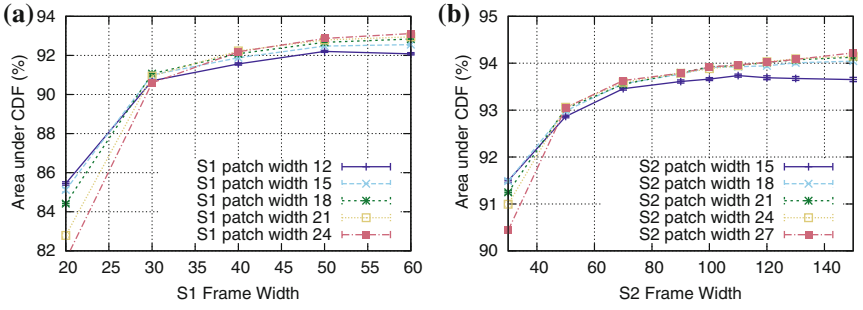

Fig. 2 Optimisation of the patch and frame width for first—(S1, **a**) and second-stage (S2, **b**) RFRV-CLMs on the L2 vertebra. Performance was measured as the percentage area under the CDF of mean point-to-curve error across 160 test images. Error bars are too small to be visible at this scale

Effects of the Shape Model Constraint. A concern with the application of shape models is that the shape constraint may introduce a bias towards the mean of the training data; this is sub-optimal in clinical applications, where the pathological (i.e. outlying) cases are of most interest. To evaluate the effect, experiments were performed using the procedure and optimal parameters described above, both with and without the application of the shape model constraint during the fitting of the second stage of the RFRV-CLM, such that the shape constraint only aided in the approximate location of the global optimum; the final result was based on image information alone. The results are shown in Fig. 3 as the CDF of mean point-to-curve error for the L2 vertebra over the 160 test images. The elimination of the shape model constraint resulted in a small and statistically insignificant change in performance, indicating that any shape model bias had an insignificant effect on the results given the error measure used here.

Effects of the Voting Style. As described in Sect. 2, several methods for voting into the cost images were available. These alternatives were evaluated using the experimental procedure and optimised parameters described above, to determine whether probabilistic voting provided performance enhancements. The results are shown in Fig. 4 as the CDF of mean point-to-curve error for the L2 vertebra over the 160 test images.

The results show that probabilistic voting provided no performance advantage. The performance of Gaussian voting was almost identical to that of single, unit voting. Single, weighted voting resulted in a small decrease in performance; however, these differences were not statistically significant. Similar results have previously been found when applying RFRV-CLMs to facial images [3].

Performance across Multiple Vertebrae. A set of leave-1/4-out experiments was performed to evaluate the RFRV-CLM on all vertebrae between T7 and L4 in all 320 images. The optimised parameters were derived from the L2 vertebra; they were adapted for the other vertebrae by scaling w_{frame} according to the ratio of mean vertebral heights in normal subjects from [12]. Shape model constraints were applied in all stages and single, unit voting was used.

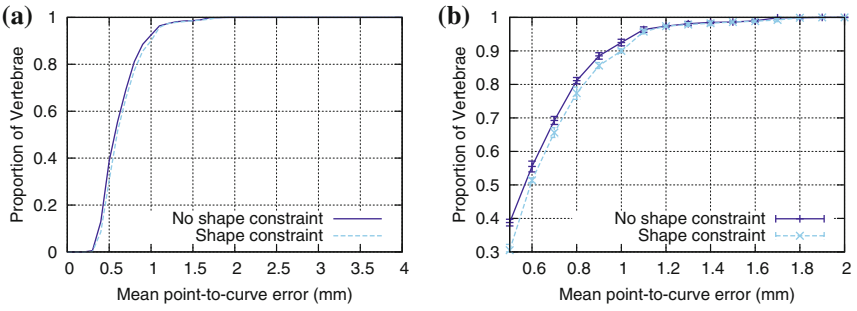

Fig. 3 Evaluation of the effect of the shape constraint in the final fitting stage using the L2 vertebra. Error bars are given as the standard deviation across five repeats; (**b**) is an expanded view of (**a**)

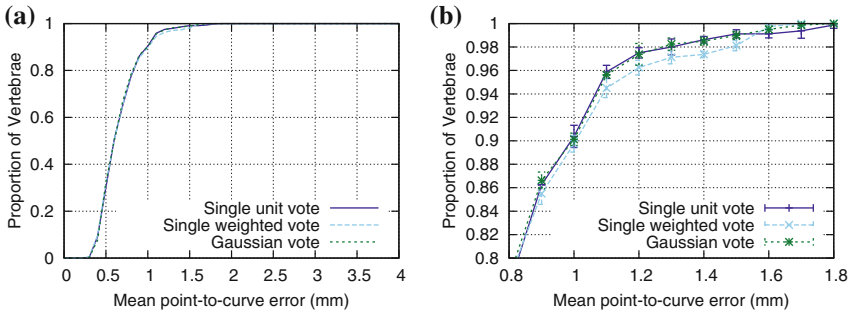

Fig. 4 Evaluation of various voting strategies for the L2 vertebra. Error bars are given as the standard deviation across five repeats; (**b**) is an expanded view of (**a**)

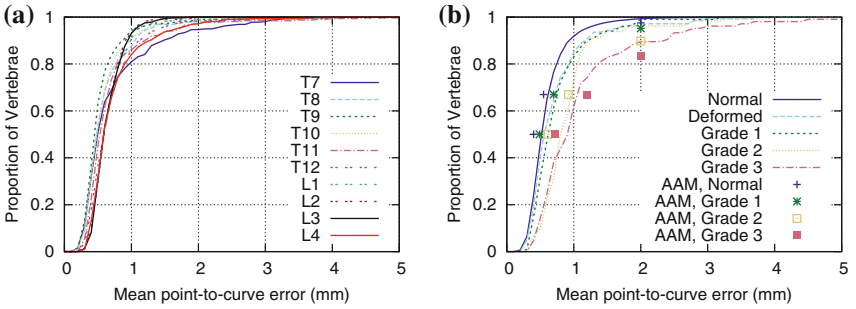

Fig. 5 Evaluation of optimised, two-stage RFRV-CLMs on all vertebrae between T7 and L4. **a** CDF of mean point-to-curve error for each vertebral level. **b** CDF of mean point-to-curve error for each vertebral status. The points show the results achieved by applying an AAM to the same data ([17], Table 1)

The results are shown in Fig. 5, and example fits are shown in Fig. 1. Annotation accuracy for T7 and L4 was lower than that for the other vertebrae, reflecting the fact that the corresponding models covered only two vertebrae, rather than three. The performance also decreased with increasing vertebral level; this may reflect the smaller size of the higher vertebrae and the presence of confounding bony structures (ribs, scapulae; see Fig. 1) in the thoracic region. However, mean errors of less than 2 and 4 mm were achieved for 95 and 99 % respectively of the vertebrae at all levels. The results divided according to vertebral status show that, as expected, performance decreased with increasing severity of fracture i.e. increasing deformation relative to the mean shape. However, mean errors of <4 mm were achieved for 95 % of grade 3 fractures, and 100 % of other classifications.

Table 1 provides numerical performance measures for the RFRV-CLMs, and compares them to the state-of-the-art results reported in [17], which applied AAMs to the same task and dataset. The AAM achieves better performance at the lower end of the CDF, as indicated by lower median errors, indicating smaller random errors on individual points. However, the RFRV-CLM achieves better mean errors for the

Table 1 Statistics of the mean point-to-curve errors on each vertebra for AAM and RFRV-CLM; bold numbers are the best results for each statistic/status

Vertebra status	AAM [17]				RFRV-CLM			
	Percentage of sample (%)	Mean (mm)	Median (mm)	Percentage errors >2 mm (%)	Percentage of sample (%)	Mean (mm)	Median (mm)	Percentage errors >2 mm (%)
Normal	84.9	**0.55**	**0.40**	2.5	83.25	0.59	0.52	**0.68**
Deformed	–	–	–	–	4.38	0.72	0.57	3.57
Grade 1	5.9	**0.70**	**0.49**	4.8	3.16	0.73	0.60	**0.99**
Grade 2	5.1	0.92	**0.61**	10.2	4.06	**0.91**	0.80	**3.84**
Grade 3	4.1	1.19	**0.72**	16.5	3.28	**1.11**	0.90	**10.48**

A total of 60 vertebrae were unclassified by the expert, due to one of the end-plates not being visible, and so 3,140 vertebrae were included in the RFRV-CLM results. Roberts et al. [17] also excluded the deformed class, hence the differences between columns 2 and 6

more severely fractured vertebrae, and substantially lower numbers of vertebrae with mean errors >2 mm regardless of classification, indicating better performance at the higher end of the CDF i.e. a smaller number of fit failures (errors >2 mm on 3.6 % of all vertebrae for AAMs versus 1.2 % for RFRV-CLMs). The median and mean errors across all 3,200 vertebrae were 0.60 and 0.65 mm respectively, compared to 0.43 and 0.60 mm from [17]. Mean search time was 366 ms per triplet per image on a Dell Precision workstation with 2 Intel Xeon 5670 processors and 24 GB RAM, running OpenSuse 11.3 × 64 (Linux kernel 2.6.34), using a single core. Mean search time per image (i.e. for all ten triplets) was 3.7 s.

4 Discussion and Conclusions

This paper has compared the performance of multi-stage RFRV-CLMs to that of AAMs in the task of vertebral landmark annotation on DXA spinal images. Several preliminary experiments were performed to optimise the various free parameters and options of the algorithm. In particular, no significant performance differences were observed, for the error metrics used, either when implementing fully probabilistic regression voting or when eliminating the shape model constraint in the final stage of fitting, such that the result was driven by image information alone.

Comparison of the errors on automatic landmarks from AAMs and RFRV-CLMs can be divided into two components; the random errors on landmarks from successful fits, best represented by the median of the error distribution due to its non-Gaussian shape, and the number of fit failures. Application of fully optimised models to ten vertebral levels in 320 DXA spinal images showed that, whilst the AAM produced smaller median errors, the difference was small at less than 0.2 mm regardless of classification. For comparison, the Genant method for vertebral fracture classification [10] defines grade 1, 2 and 3 fractures as vertebral height reductions of 20–25, 25–40, and >40 %, respectively, and [12] measured mean vertebral heights varying from 22.97 ± 1.52 mm for T7 to 35.62 ± 2.21 mm for L4 in a sample of 108 normal women. Therefore, vertebral fractures are defined via height reductions of $\gtrsim$5 mm regardless of grade or level. The more significant difference between the two techniques is in terms of the number of fit failures, since these represent cases where accurate diagnosis of the vertebral status using the automatic landmarks would not be possible. Fit failures were identified using a threshold of 2 mm, for ease of comparison to the results presented in [17]. The RFRV-CLM produced lower numbers of vertebrae with errors >2 mm for all classifications; a reduction of 68 %. Therefore, in the region of the CDF important for computer-aided VFA, RFRV-CLMs out-perform AAMs.

In this work, no use was made of the multiple fits to each vertebra provided by the overlaps of the models; only the points on the central vertebra in each were used. In future work, we intend to explore the combination of the multiple fits with goodness-of-fit measures both to improve the accuracy of the automatic annotation, in terms of random error, and to detect instances of fit failures i.e. systematic errors in individual fits. Furthermore, we intend to extend the work to include both radiographs

and mid-line sagittal CT images. Finally, we intend to investigate the use of automatic landmarks for vertebral classification, comparing the accuracy of approaches based on the Genant height ratios to classifiers applied both to the point locations themselves, and to the shape parameters generated during fitting.

Acknowledgments This publication presents independent research supported by the Health Innovation Challenge Fund (grant no. HICF-R7-414/WT100936), a parallel funding partnership between the Department of Health and Wellcome Trust. The views expressed in this publication are those of the authors and not necessarily those of the Department of Health or Wellcome Trust.

References

1. Breiman, L.: Random forests. Mach. Learn. **45**, 5–32 (2001)
2. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active appearance models. IEEE Trans. Pattern Anal. Mach. Intell. **23**, 681–685 (2001)
3. Cootes, T.F., Ionita, M.C., Lindner, C., Sauer, P.: Robust and Accurate Shape Model Fitting Using Random Forest Regression Voting. In: Proc. ECCV, pp. 278–291. Springer, Berlin (2012)
4. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active shape models—their training and application. Comput. Vis. Image Underst. **61**(1), 38–59 (1995)
5. Cristinacce, D., Cootes, T.: Automatic feature localisation with constrained local models. J. Pattern Recognit. **41**(10), 3054–3067 (2008)
6. Cummings, S.R., Melton, J.: Epidemiology and outcomes of osteoporotic fractures. Lancet **359**(9319), 1761–1767 (2002)
7. Dantone, M., Gall, J., Fanelli, G., Van Gool, L.: Real-time facial feature detection using conditional regression forests. In: Proc. CVPR, pp. 2578–2585 (2012)
8. Delmas, P.D., van de Langerijt, L., Watts, N.B., Eastell, R., Genant, H.K., Grauer, A., Cahall, D.L.: Underdiagnosis of vertebral fractures is a worldwide problem: the IMPACT study. J. Bone Miner. Res. **20**(4), 557–563 (2005)
9. Donner, R., Menze, B., Bischof, H., Langs, G.: Fast anatomical structure localization using top-down image patch regression. In: Medical Computer Vision, Recognition Techniques and Applications in Medical Imaging, Lecture Notes in Computer Science, vol. 7766, pp. 133–141 (2013)
10. Genant, H.K., Wu, C.Y., Kuijk, C.V., Nevitt, M.C.: Vertebral fracture assessment using a semiquantitative technique. J. Bone Miner. Res. **8**(9), 1137–1148 (1993)
11. Leech, J.A., Dulberg, C., Kellie, S., Pattee, L., Gay, J.: Relationship of lung function to severity of osteoporosis in women. Am. Rev. Respir. Dis. **141**(1), 68–71 (1990)
12. Leidig-Bruckner, G., Minne, H.W.: The spine deformity index (SDI); a new approach to quantifying vertebral crush fractures in patients with osteoporosis. Vertebral Fracture in Osteoporosis, pp. 235–252. Osteoporosis Research Group, University of California, California (1995)
13. Lindner, C., Thiagarajah, S., Wilkinson, J.M., arcOGEN Consortium, T., Wallis, G.A., Cootes, T.F.: Fully automatic segmentation of the proximal femur using random forest regression voting. IEEE Trans. Med. Imaging **32**(8), 1462–1472 (2013)
14. McCloskey, E., Selby, P., de Takats, D., Bernard, J., Davies, M., Robinson, J., Francis, R., Adams, J., Pande, K., Beneton, M., Jalava, T., Loyttyniemi, E., Kanis, J.A.: Effects of clodronate on vertebral fracture risk in osteoporosis: a 1-year interim analysis. Bone **28**(3), 310–315 (2001)
15. Rachner, T.D., Khosla, S., Hofbauer, L.C.: Osteoporosis: now and the future. Lancet **377**(9773), 1276–1287 (2011)
16. Roberts, M.G., Cootes, T.F., Adams, J.E.: Vertebral morphometry: semi-automatic determination of detailed shape from DXA images using active appearance models. Investig. Radiol. **41**(12), 849–859 (2006)

17. Roberts, M.G., Cootes, T.F., Adams, J.E.: Automatic Location of Vertebrae on DXA Images Using Random Forest Regression. In: Proc. MICCAI 2012, LNCS, vol. 7512, pp. 361–368. Springer-Verlag, Berlin (2012)
18. Sauer, P., Cootes, T., Taylor, C.: Accurate Regression Procedures for Active Appearance Models. In: J. Hoey, S. McKenna, E. Trucco (eds.) Proc. BMVC, pp. 30.1–30.11 (2011)
19. Valstar, M.F., Martinez, B., Binefa, X., Pantic, M.: Facial point detection using boosted regression and graph models. In: Proc. CVPR, pp. 2729–2736 (2010)
20. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Proc. CVPR, pp. 511–518. IEEE Computer Society (2001)