Improving an Active Shape Model with Random Classification Forest for Segmentation of Cervical Vertebrae

S M Masudur Rahman Al Arif[1], Michael Gundry[2], Karen Knapp[2] and Greg Slabaugh[1]

[1]Department of Computer Science, City University London, UK
[2]University of Exeter Medical School, UK

Abstract. X-ray is a common modality for diagnosing cervical vertebrae injuries. Many injuries are missed by emergency physicians which later causes life threatening complications. Computer aided analysis of X-ray images has the potential to detect missed injuries. Segmentation of the vertebrae is a crucial step towards automatic injury detection system. Active shape model (ASM) is one of the most successful and popular method for vertebrae segmentation. In this work, we propose a new ASM search method based on random classification forest and a kernel density estimation-based prediction technique. The proposed method have been tested on a dataset of 90 emergency room X-ray images containing 450 vertebrae and outperformed the classical Mahalanobis distance-based ASM search and also the regression forest-based method.

Keywords: ASM, Classification forest, Cervical, Vertebrae, X-ray.

1 Introduction

The cervical spine or the neck region is vulnerable to high-impact accidents like road collisions, sports mishaps and falls. Cervical radiographs is usually the first choice for emergency physicians to diagnose cervical spine injuries due to the required scanning time, cost, and the position of the spine in the human body. However, about 20% of cervical vertebrae related injuries remain undetected by emergency physicians and roughly 67% of these missing injuries result in tragic consequences, neurological deteriorations and even death [1, 2]. Computer aided diagnosis of cervical X-ray images has a great potential to help the emergency physicians to detect miss-able injuries and thus reducing the risk of missing injury related consequences.

Segmentation of the cervical vertebra in X-ray images is a major part of any computer aided injury detection system. Due to the clinical importance of vertebrae segmentation, there is a large body of research in the literature [3–11]. Based on this literature, arguably the most successful segmentation method is the statistical shape model (SSM). Active shape model (ASM) is one version of the SSMs that has been performing with success in various fields including medical and facial images. Since its inception, the algorithm has been studied and modified by many researchers [12–16]. In [12], a simple gradient maxima

search has been introduced for this task. However, this method is limited to edge like object boundaries. An improved Mahalanobis distance-based search method has been introduced in [13]. This method involves a training phase and an optimization step to find the amount of displacement needed to converge the mean shape on the actual object boundary. The method has been shown to work well on cervical vertebra X-ray images in [3, 4]. In [15], a conventional binary classifier and a boosted regression predictor has been compared and used to improve the performance of ASM segmentation during image search phase. While these methods detect the displacement of the shape towards the possible local minima, [16] have proposed a method to directly predict some of the shape parameters using a classification method. In the state-of-the-art work on vertebra segmentation [17], a random regression forest has used to predict the displacement during image search of constrained local model (CLM), another version of SSM.

In this paper, we propose a one-shot random classification forest-based displacement predictor for ASM segmentation of cervical vertebrae. Unlike the Mahalanobis distance-based method used in [3, 4, 13, 14], this method predicts the displacement directly without a need of a sliding window-based search technique. Our method uses a multi-class forest in contrast with the binary classification method used in [15]. A kernel density estimation (KDE)-based classification label prediction method has been introduced which performed better than traditional classification label prediction method. The proposed algorithm has been tested on a dataset of 90 emergency room X-ray images and achieved 16.2% lower error than the Mahalanobis distance-based method and 3.3% lower fit-failure compared with a regression-based framework.

2 Methodology

Active shape model (ASM) has been used in many vertebrae segmentation frameworks. In this work, we have proposed an improvement in the image search phase of ASM segmentation using a one-shot multi-class random classification forest algorithm. The proposed method is compared with a Mahalanobis distance-based method and a random regression forest-based method. The ASM is briefly described in Sec. 2.1, followed by the Mahalanobis distance-based search method in Sec. 2.3, regression forest-based search method in Sec. 2.4 and finally the proposed search method is explained in Sec. 2.5.

2.1 Active Shape Model

Let $\mathbf{x}_i$, a vector of length $2n$ describing n 2D points of the i-th registered training vertebra, is given by:

$$\boldsymbol{x}_i = [x_{i1}, y_{i1}, x_{i2}, y_{i2}, x_{i3}, y_{i3}, ..., x_{in}, y_{in}] \tag{1}$$

where (x_{ij}, y_{ij}) is the Cartesian coordinate of the j-th point of the i-th training vertebra. A mean shape, $\bar{\mathbf{x}}$, can be calculated by averaging all the shapes:

$$\bar{\boldsymbol{x}} = \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{x}_i \tag{2}$$

where N is the number of vertebrae available in the training set. Now, the covariance, $\boldsymbol{\Lambda}$, is given by

$$\boldsymbol{\Lambda} = \frac{1}{N-1}\sum_{i=1}^{N}(\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})^T \tag{3}$$

Principal component analysis (PCA) is performed by calculating $2n$ eigenvectors $\boldsymbol{p}_k$ $(k = 1, 2, ..., 2n)$ of $\boldsymbol{\Lambda}$. The eigenvectors with smaller eigenvalues (λ_k) are often result from noise and/or high frequency variation. Thus any shape, $\boldsymbol{x}_i$, can be approximated fairly accurately only by considering first m eigenvectors with largest eigenvalues.

$$\hat{\boldsymbol{x}}_i \approx \bar{\boldsymbol{x}} + \boldsymbol{P}_s\boldsymbol{b}_i; \qquad \boldsymbol{P}_s = [\boldsymbol{p}_1, \boldsymbol{p}_2, ..., \boldsymbol{p}_m] \tag{4}$$

where $\boldsymbol{b}_i$ is a set of weights known as shape parameters. The standard practice to select m is to find the first few eigenvalues, λ_k's, that represent a certain percentage of the total variance of the training data. For any known shape, $\boldsymbol{x}_i$, shape parameter $\boldsymbol{b}_i$ can be computed as:

$$\boldsymbol{b}_i = \boldsymbol{P}_s^T(\boldsymbol{x}_i - \bar{\boldsymbol{x}}) \tag{5}$$

2.2 ASM Search

When segmenting a vertebra in a new image, the mean shape is approximately initialized near the vertebra using manually clicked vertebra centers [18,19]. The model then looks for displacement of the mean shape towards the actual vertebra based on the extracted profiles perpendicular to the mean shape in the image (see Fig. 1). In [12], a simple gradient maxima search have been introduced for this task but however this method is limited to edge like object boundaries. An improved Mahalanobis distance-based search method has been introduced in [13]. This method has been used for vertebrae segmentation in [3,4].

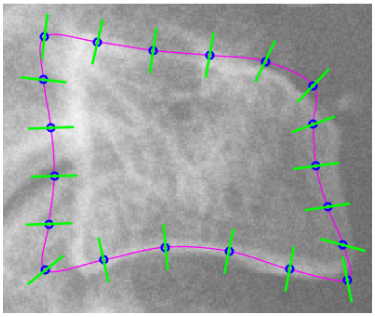

Fig. 1. ASM search: extraction of normal profiles. Initialized mean shape (magenta), extracted profiles (green) and shape describing points (blue).

2.3 Mahalanobis distance-based ASM search (ASM-M)

The Mahalanobis distance-based ASM search involves a training phase and an optimization step to find the amount of displacement needed to converge the mean shape on actual object boundary. During training for each landmark point, intensity profiles of length $2l+1$ are collected from all the objects. The normalized first derivatives these profiles ($\boldsymbol{g}$) are then used to create a mean profile ($\bar{\boldsymbol{g}}$) and a covariance matrix ($\boldsymbol{\Lambda_g}$). When a new profile, $\boldsymbol{g}_k$, is given, then the Mahalanobis distance can be calculated as:

$$M(\boldsymbol{g}_k) = (\boldsymbol{g}_k - \bar{\boldsymbol{g}})\boldsymbol{\Lambda}_{\boldsymbol{g}}^{-1}(\boldsymbol{g}_k - \bar{\boldsymbol{g}}) \tag{6}$$

The profile $\boldsymbol{g}_k$ is then shifted from the mean shape inwards and outwards by l pixels and Mahalanobis distance is computed at each position. The desired amount of displacement ($\hat{k}$) is then computed by minimizing $M(\boldsymbol{g}_k)$, which is equivalent to maximizing the probability that $\boldsymbol{g}_k$ originates from a multidimensional Gaussian distribution learned from the training data. The one-dimensional displacements ($\hat{k}$) for all the points are then mapped into 2D displacement vector $d\boldsymbol{x}$. This $d\boldsymbol{x}$ reconfigures the mean shape towards the actual object boundary.

$$d\boldsymbol{b} = \boldsymbol{P}_s^T d\boldsymbol{x}; \qquad \boldsymbol{b}_t = \boldsymbol{b}_{t-1} + d\boldsymbol{b}; \qquad \hat{\boldsymbol{x}}_t = \hat{\boldsymbol{x}}_{t-1} + \boldsymbol{P}_s\boldsymbol{b}_t \tag{7}$$

where $\hat{\boldsymbol{x}}_0 = \bar{\boldsymbol{x}}$ and $\boldsymbol{b}_0$ is an all zero vector. The process is iterative. The reconfiguration stops if number of iterations, t, crosses a maximum threshold or $d\boldsymbol{x}$ is negligible.

2.4 Random Regression Forest-based ASM search (ASM-RRF)

Regression-based method has been used for ASM search in [15]. Random forest (RF) is a powerful machine learning algorithm [20]. It can be applied to achieve classification and/or regression [21]. Recent state-of-the-art work on vertebrae segmentation [17], proposed a random forest regression voting (RFRV) method for this purpose in the CLM framework. The regressor predicts a 2D displacement for the shape to move towards a local minimum. In order to compare the performance of our proposed one-shot multi-class random classification forest-based ASM search, a random regression forest-based ASM search has also been implemented. This forest trains on the gradient profiles collected during a training phase and predicts a 1D displacement during ASM search.

ASM-RRF Training: The ASM-M produces a displacement $\hat{k}$ by minimizing Eqn. 6 over a range of displacements. The predicted displacement ($\hat{k}$) can take any value from $-l$ to $+l$ representing the amount of shift needed. The ASM-RRF is also designed to produce the same, only by looking at the gradient profile ($\boldsymbol{g}$). To achieve this, the profile vectors ($\boldsymbol{g}$) for each landmark point are collected based on the manual segmentation curves of the object. The profiles with manually segmented annotations at the center pixel are assigned shift label 0, then the profile extraction window is shifted inwards or outwards to create positive and negative shift labels representing the position of the manually annotated

segmentation pixel with respect to the center. To make the process equivalent to the previous ASM-M method, the amount of shifting was limited to $\pm l$ pixels, giving us a total of $2l + 1$ regression target values to train the forest. The forest is trained using standard information gain and regression entropy i.e. variance of the target values.

$$IG = H(S) - \sum_{i\epsilon\{L,R\}} \frac{|S^l|}{S} H(S^i) \tag{8}$$

$$H(S) = H_{reg}(S) = Var(L_S) \tag{9}$$

where S is a set of examples arriving at a node and S^L, S^R are the data that travel left or right respectively and L_S is the set of target value available at the node considered. In our case, $L_S \subset \{-l, -l+1, ..., 0, ..., l-1, l\}$. The node splitting stops when the tree reaches a maximum depth (D_{max}) or number of elements at node falls below a threshold ($nMin$). The leaf node records the statistics of the node elements by saving the mean displacement, $\bar{k}_{ln}$ and the standard deviation $\sigma_{k_{ln}}$ of the node target values.

ASM-RRF Prediction: At test time new profiles are fed into the forest and they regress down to the leaf nodes of different trees. A number of voting strategies for regression framework are compared in [17]: a single vote at $\bar{k}_{ln}$, a probabilistic voting weighted by $\sigma_{k_{ln}}$ or a Gaussian spread of votes $N(\bar{k}_{ln}, \sigma_{k_{ln}})$. They reported the best performance using the single vote method. Following this, in this paper, the displacement $\hat{k}$ is determined by Eqn. 10. This $\hat{k}$ is then returned to the ASM search process to reconfigure the shape for next iteration.

$$\hat{k} = \frac{1}{T} \sum_{t=1}^{T} (\bar{k}_{ln_t}) \tag{10}$$

where T is the number of trees in the forest.

2.5 Random Classification Forest-based ASM search (ASM-RCF)

The main contribution of this work is to provide an alternative to the already proposed ASM-M and ASM-RRF methods with the help of random classification forest (ASM-RCF) algorithm. Classification-based ASM search methods have previously been investigated in [15, 16]. A binary classification-based method is proposed in [15] to predict the displacement while [16] proposed another classification-based to determine first few shape parameters directly. Like [15] our also ASM-RCF method determines the displacement but instead of a binary classification, the problem is designed as a multi-class classification problem. Thus like regression, it can predict the displacement in one-shot without the need of sliding window search like ASM-M method and [15].

ASM-RCF Training: The training data is the same as the ASM-RRF method. Gradient profiles ($\boldsymbol{g}$) are collected using manual segmentations and shifted inwards and outwards to create $2l + 1$ shift labels. But, instead of considering

the shift labels as continuous regression target values, here we consider them as discrete classification labels. The classification forest is then trained to predict $2l+1$ class labels. The same information gain of Eqn. 8 is used but the entropy $H(S)$ is replaced by the classification entropy of Eqn. 11.

$$H(S) = H_{class}(S) = -\sum_{c \epsilon C} p(c) log(p(c)) \tag{11}$$

where C is the set of classes available at the node considered. Here, $C \subset \{-l, -l+1, ..., 0, ..., l-1, l\}$. Both ASM-RRF and ASM-RCF are parametrized by maximum depth (D_{max}), minimum node element ($nMin$), number of trees (T), numbers of random variables ($nVar$) and numbers of random threshold values ($nThresh$) to considers for node optimization.

ASM-RCF Prediction: The leaf nodes of our classification forest are associated with a set of labels, $C_{ln_{xy}}$, that contains all the target classification labels present at that leaf node.

$$C_{ln_{xy}} = \{c_1, c_2, ..., c_{n_{Leaf_{xy}}}\} \tag{12}$$

where $n_{Leaf_{xy}}$ is the number of elements at the x-th leaf node of y-th tree of the forest. At test time, a new profile is fed into the forest and it reaches different leaf nodes of different trees. We have experimented with two different prediction methods. First, a classical $ArgMax$ based classification label predictors and other is a Gaussian kernel-based classification label predictor.

ArgMax based label prediction (RCF-AM): The set of leaf node labels from each tree is collected in C_{forest}. The collection of labels is then converted into a probabilistic distribution over the $2l+1$ classification labels as $p(C_{forest})$. Finally, the predicted label $\hat{c}$ (or the displacement $\hat{k}$) is determined by finding the label that maximizes $p(C_{forest})$.

$$\hat{k} = \hat{c}_{argmax} = \arg\max_c p(C_{forest}) \tag{13}$$

where

$$C_{forest} = \{C_{tree_1} \cup C_{tree_2} \cup .... \cup C_{tree_T}\} \tag{14}$$

$$C_{tree_y} = C_{ln_{xy}} = \{c_1, c_2, ..., c_{n_{Leaf_{xy}}}\} \tag{15}$$

Kernel based label prediction (RCF-KDE): Apart from the RCF-AM method, we propose a new kernel density estimator (KDE) based method to determine the classification label $\hat{c}$ or the displacement $\hat{k}$ of the test profile. KDE-based predictors are most commonly used in regression forests. But here, we demonstrate its usefulness as a multi-class label predictor for classification forest. Like RCF-AM, here we also collect the all the leaf node labels as C_{forest}. Then at each element c_{in} of C_{forest} a zero mean Gaussian distribution with variance σ^2_{kde} is added. The predicted class $\hat{c}$ is determined by the label that maximizes the resultant distribution.

$$\hat{k} = \hat{c} = \arg\max_c \left(\frac{1}{s} \sum_{i=1}^{s} \left(\frac{1}{\sigma_{kde}\sqrt{2\pi}} \exp^{-\frac{(x - c_{in_i})^2}{2\sigma^2_{kde}}} \right) \right) \tag{16}$$

where s is number of class labels in C_{forest}. Fig. 2 shows an toy example of different prediction methods for a forest with a single tree.

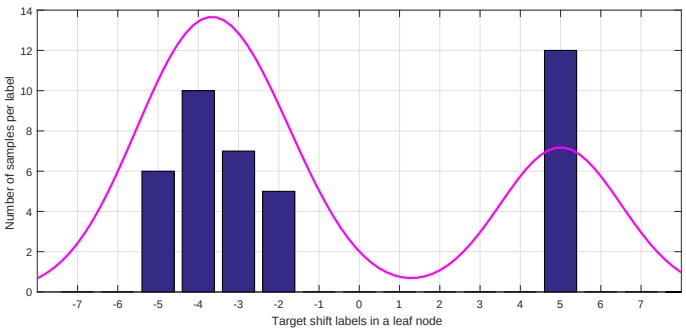

Fig. 2. An example leaf node and different prediction methods: In this particular leaf node there are in total 40 samples: the shift labels are -5, -4, -3, -2 and 4; number of samples per label are 6, 10, 7, 5 and 12 respectively. The regression mean of ASM-RRF method for this leaf node is zero, class label prediction with ArgMax (ASM-RCF-AM) method is label 5 and with ASM-RCF-KDE method is label -4. The magenta curve represents the summed kernel densities.

3 Experiments

3.1 Data

A total of 90 X-ray images have been used for this work. Different (Philips, Agfa, Kodak, GE) radiographic systems were used for the scans. Pixel spacing varied from 0.1 to 0.194 pixel per millimetre. The dataset is very challenging and contains natural variations, deformations, injuries and implants. A few images from the dataset are shown in Fig. 3. Each of these images was then manually

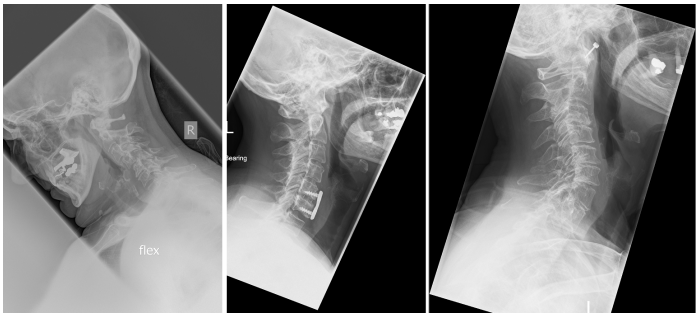

Fig. 3. Example of images in the dataset.

demarcated by expert radiographers. Radiographers clicked on 20 points along the vertebra boundary. Some manual demarcation points for C3-C7 are shown in Fig. 4. The axis (C1) and atlas (C2) of the cervical vertebrae have not been

studied in this work due to their ambiguity in lateral X-ray images similar to other work in the literature [3, 4].

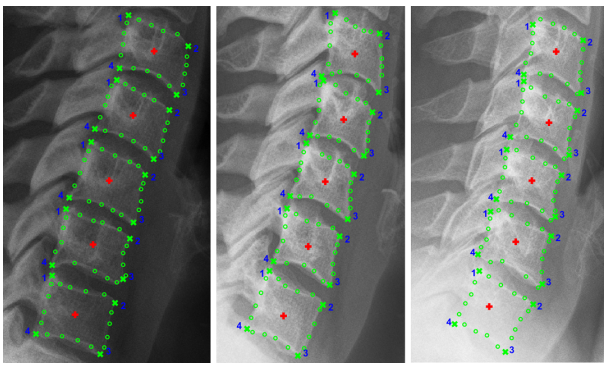

Fig. 4. Manual segmentations: centers (+), corners (×) and other points (o).

3.2 Training

An ASM is trained for each vertebra separately. Five models are created for five vertebrae. The means and covariance matrices for ASM-M are computed separately for each landmark point of each vertebra. The forests (ASM-RRF and ASM-RCF) are trained separately for each side of the vertebrae (anterior, posterior, superior and inferior). To increase the forest training samples, the 20 point segmentation shape is converted into a 200 point shape using Catmull-Rom spline. The training profiles are collected from these points. The training profiles for all ASM search methods are of length 27 i.e. $l = 13$, giving us a total of 27 shift labels: $\{-13, -12, ..., 0, ..., 12, 13\}$.

3.3 Segmentation evaluation

A 10-fold cross-validation scheme has been followed. For each fold, the ASM, ASM-M, ASM-RRF and ASM-RCF training has been done on 81 images and tested on 9 images. After training, each fold consists of 5 vertebrae ASM models (mean shape, Eigen vectors and Eigen values), (5 vertebra × 20 points =) 100 mean gradient profiles and covariance matrices for ASM-M method and (5 × 4 sides =) 20 forests each for ASM-RRF and ASM-RCF methods. Each forest is trained on (200 points × 27 labels × 81 training images ÷ 4 sides =) 109350 samples. The experiment is repeated 10 times so that each image in our dataset of 90 images are considered as test image once. At the end of the experiments, the Euclidean distance between predicted vertebra shape points and manual segmentation curves are computed in millimetres as the error metric. The distance errors are calculated for each segmentation point and averaged over all the vertebrae as a single metric.

3.4 Parameter optimization

There are five free parameters in the random forest training: the number of trees (T), maximum allowed depth of a tree (D_{max}), minimum number of elements at a node ($nMin$), number of variables to look at in each split nodes ($nVar$) and number of thresholds ($nTresh$) to consider per variable. Apart from these, the kernel density estimation function requires a bandwidth (BW) which is the variance σ^2_{kde} of Eqn. 16. A greedy sequential approach is employed to optimize each parameter due to time constraint. The sequence followed is: BW, T, D and $nMin$ in a 2D fashion, and $nVar$ and $nThresh$ in a 2D fashion. The cost function for the optimization is the average absolute difference between predicted and actual class labels. Fig. 5 shows an example parameter search for BW. BW is chosen based on the minimum error found on the graph i.e. 1.5. Similarly, all the parameters are optimized and reported in Table 1.

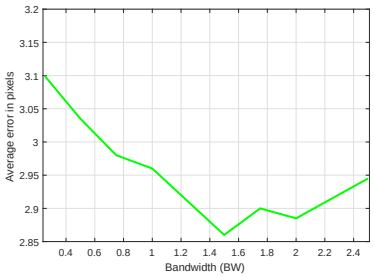

Fig. 5. Bandwidth Optimization.

Table 1. Optimized parameters.

Parameters	**Value**
BW	1.5
T	100
D_{max}	10
$nMin$	50
$nVar$	6
$nThresh$	5

4 Results

The mean, median and standard deviation of the average errors in millimeters have been reported in Table 2. Both random forest-based methods perform better than the ASM-M method. Among two options of ASM-RCF methods, KDE method outperforms the ArgMax (AM) method. ASM-RCF-KDE shows an improvement of 16.1% in terms of median error over ASM-M method. ASM-RCF-KDE also outperforms regression-based ASM-RRF slightly in terms of mean and median. The algorithms are also compared using fit-failures. In this work, fit-failure is defined as the percentage vertebra having an average error of 1 mm or higher. The last row of Table 2 report the fit-failures. In terms of this metric, both classification (ASM-RCF) based methods outperform other methods and ASM-RCF-KDE performs the best with the lowest failure rate of 16.67%. The algorithms are also compared in Fig. 6 where the proportion of the vertebrae is shown as a cumulative distribution function over the errors. More area under the curve indicates better performance. It can be seen that ASM-RCF-KDE and ASM-RRF are the two best algorithms for ASM search. The cropped and zoomed version, Fig. 6 (right), indicates that our proposed method ASM-RCF-KDE slightly outperforms the current state-of-the-art regression-based method. Some qualitative segmentation results for ASM-RCF-KDE method with the manual segmentation has been shown in Fig. 7. The method successfully segments most

of the vertebrae. The performance is satisfactory for the vertebrae with low contrast (4a) and implants (5a) too. But the algorithm still requires future work. Row (b) of Fig. 7 shows challenging segmentation cases where the segmentation was unsuccessful. It can be seen that the segmentation sometimes suffers from low contrast and bad initialization (1b), deformity (2b) and implants (3b,4b and 5b).

Table 2. Performance comparison: average error in MM.

	ASM-M	**ASM-RRF**	**ASM-RCF-AM**	**ASM-RCF-KDE**
Median	0.8019	0.6933	0.7054	**0.6896**
Mean	0.8582	0.7704	0.8060	**0.7688**
Standard deviation	**0.3437**	0.3766	0.3998	0.3965
Fit failures (%Errors >1 mm)	24.00	21.78	20.00	**16.67**

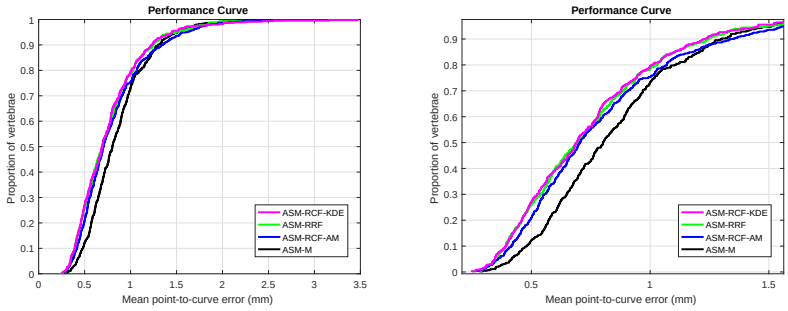

Fig. 6. Comparison of performance.

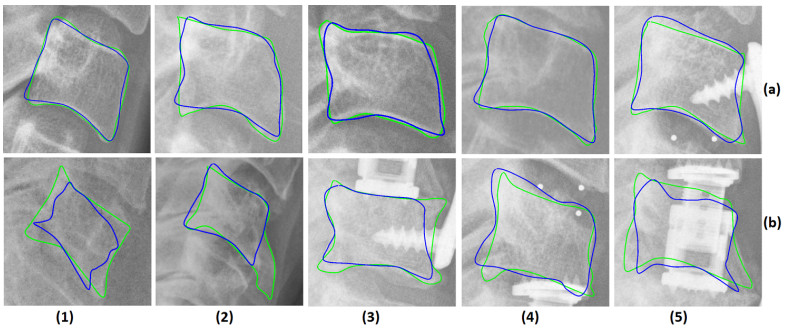

Fig. 7. Segmentation results: Manual segmentation (green), ASM-RCF-KDE (blue).

5 Conclusion

In this paper, we have provided an alternative to standard active shape model algorithm by introducing a one-shot multi-class random forest in the ASM

search process. The new algorithm has been formulated as a classification problem which eliminates the sliding window search of Mahalanobis distances-based method. The improved algorithm provides better segmentation results and an improvement of 16.1% in point to line segmentation errors has been achieved for a challenging dataset over ASM-M. The proposed classification forest-based framework with kernel density-based prediction (ASM-RCF-KDE) outperformed regression-based methods by 3.3% in terms of fit-failures. The proposed KDE-based prediction helps to predict the displacement with better accuracy because it can nullify the effect of false tree predictions by using information from neighbouring classes (Fig 2).

Our algorithm has been tested on a challenging dataset of 90 emergency room X-ray images containing 450 cervical vertebrae. We have achieved a lowest average error of 0.7688 mm. In comparison, current state-of-the-art work [17], reports an average error of 0.59 mm for a different dataset of DXA images on healthy thoraco-lumbar spine. Their method uses the latest version of SSM, constrained local model (CLM) with random forest regression voting based search method. In the near future, we plan to work on CLM with a classification forest-based search method. As our overarching goal is to develop an injury detection system, the segmentation still requires further work to apply morphometric analysis to detect injuries especially for the vertebrae with conditions like osteoporosis, fractures and implants. Our work is currently focused on improving the segmentation accuracy by experimenting with better and automatic initialization using vertebrae corners [18, 19] and/or endplates [5].

References

1. P. Platzer, N. Hauswirth, M. Jaindl, S. Chatwani, V. Vecsei, and C. Gaebler, "Delayed or missed diagnosis of cervical spine injuries," *Journal of Trauma and Acute Care Surgery*, vol. 61, no. 1, pp. 150–155, 2006.
2. J. W. Davis, D. L. Phreaner, D. B. Hoyt, and R. C. Mackersie, "The etiology of missed cervical spine injuries.," *Journal of Trauma and Acute Care Surgery*, vol. 34, no. 3, pp. 342–346, 1993.
3. M. Benjelloun, S. Mahmoudi, and F. Lecron, "A framework of vertebra segmentation using the active shape model-based approach," *Journal of Biomedical Imaging*, vol. 2011, p. 9, 2011.
4. S. A. Mahmoudi, F. Lecron, P. Manneback, M. Benjelloun, and S. Mahmoudi, "GPU-based segmentation of cervical vertebra in X-ray images," in *Cluster Computing Workshops and Posters (CLUSTER WORKSHOPS), 2010 IEEE International Conference on*, pp. 1–8, IEEE, 2010.
5. M. G. Roberts, T. F. Cootes, and J. E. Adams, "Automatic location of vertebrae on DXA images using random forest regression," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012*, pp. 361–368, Springer, 2012.
6. M. G. Roberts, T. F. Cootes, and J. E. Adams, "Vertebral shape: automatic measurement with dynamically sequenced active appearance models," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2005*, pp. 733–740, Springer, 2005.
7. M. G. Roberts, T. F. Cootes, E. Pacheco, T. Oh, and J. E. Adams, "Segmentation of lumbar vertebrae using part-based graphs and active appearance models,"

in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2009*, pp. 1017–1024, Springer, 2009.
8. M. Roberts, E. Pacheco, R. Mohankumar, T. Cootes, and J. Adams, "Detection of vertebral fractures in DXA VFA images using statistical models of appearance and a semi-automatic segmentation," *Osteoporosis international*, vol. 21, no. 12, pp. 2037–2046, 2010.
9. S. Casciaro and L. Massoptier, "Automatic vertebral morphometry assessment," in *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, pp. 5571–5574, IEEE, 2007.
10. M. A. Larhmam, S. Mahmoudi, and M. Benjelloun, "Semi-automatic detection of cervical vertebrae in X-ray images using generalized hough transform," in *Image Processing Theory, Tools and Applications (IPTA), 2012 3rd International Conference on*, pp. 396–401, IEEE, 2012.
11. M. A. Larhmam, M. Benjelloun, and S. Mahmoudi, "Vertebra identification using template matching modelmp and K-means clustering," *International journal of computer assisted radiology and surgery*, vol. 9, no. 2, pp. 177–187, 2014.
12. T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models-their training and application," *Computer vision and image understanding*, vol. 61, no. 1, pp. 38–59, 1995.
13. T. Cootes and C. Taylor, "Statistical models of appearance for computer vision, wolfson image anal. unit, univ. manchester, manchester," tech. rep., UK, Tech. Rep, 1999.
14. B. Van Ginneken, A. F. Frangi, J. J. Staal, B. M. Romeny, and M. A. Viergever, "Active shape model segmentation with optimal features," *medical Imaging, IEEE Transactions on*, vol. 21, no. 8, pp. 924–933, 2002.
15. D. Cristinacce and T. F. Cootes, "Boosted regression active shape models.," in *BMVC*, vol. 1, p. 7, 2007.
16. Y. Zheng, A. Barbu, B. Georgescu, M. Scheuering, and D. Comaniciu, "Four-chamber heart modeling and automatic segmentation for 3-D cardiac CT volumes using marginal space learning and steerable features," *IEEE transactions on medical imaging*, vol. 27, no. 11, pp. 1668–1681, 2008.
17. P. Bromiley, J. Adams, and T. Cootes, "Localisation of vertebrae on DXA images using constrained local models with random forest regression voting," in *Recent Advances in Computational Methods and Clinical Applications for Spine Imaging*, pp. 159–171, Springer, 2015.
18. S. M. M. R. Al-Arif, M. Asad, K. Knapp, M. Gundry, and G. Slabaugh, "Hough forest-based corner detection for cervical spine radiographs," in *Medical Image Understanding and Analysis (MIUA), Proceedings of the 19th Conference on*, pp. 183–188, 2015.
19. S. M. M. R. Al-Arif, M. Asad, K. Knapp, M. Gundry, and G. Slabaugh, "Cervical vertebral corner detection using Haar-like features and modified Hough forest," in *Image Processing Theory, Tools and Applications (IPTA), 2015 5th International Conference on*, IEEE, 2015.
20. L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.
21. J. Gall, A. Yao, N. Razavi, L. Van Gool, and V. Lempitsky, "Hough forests for object detection, tracking, and action recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 33, no. 11, pp. 2188–2202, 2011.