

Global Localization and Orientation of the Cervical Spine in X-ray Images

S M Masudur Rahman Al Arif¹, Michael Gundry², Karen Knapp² and Greg Slabaugh¹

¹Department of Computer Science, City University London, UK

²University of Exeter Medical School, UK

Abstract. Injuries in cervical spine X-ray images are often missed by emergency physicians. Many of these missing injuries cause further complications. Automated analysis of the images has the potential to reduce the chance of missing injuries. Towards this goal, this paper proposes an automatic localization of the spinal column in cervical spine X-ray images. The framework employs a random classification forest algorithm with a kernel density estimation-based voting accumulation method to localize the spinal column and to detect the orientation. The algorithm has been evaluated with 90 emergency room X-ray images and has achieved an average detection accuracy of 91% and an orientation error of 3.6°. The framework can be used to narrow the search area for other advanced injury detection systems.

Keywords: Random forest, Classification, Cervical, Vertebra, Localization, Orientation.

1 Introduction

The cervical spine is vulnerable to high-impact accidents like automobile collision, sports mishaps and falls. Due to the scanning time required, cost, and the position of the spine in the human body, X-ray is the first mode of investigation for cervical spine injuries. Unfortunately, roughly 20% of cervical vertebrae related injuries remain undetected by emergency physicians and about 67% of these missing injuries result in tragic consequences like loss of motor control, disability to move the neck and other neurological deteriorations [1, 2]. Providing emergency physicians with an automated analysis of the cervical X-ray images has a great potential to reduce the chances of missing injuries. Towards that goal, this paper takes the first step to localise the cervical spine in an arbitrary X-ray image. Our method involves a machine learning process which employs a patch based framework to localize the vertebrae column. It is also able to predict the orientation of the spinal curve.

Some limited work has been presented in the literature for global localization of the cervical spine on X-ray images. Most of the methods revolve around the generalized Hough transform (GHT). Tezmol et al. [3] used a GHT based

framework using mean vertebra templates and an innovative voting accumulator structure. A more recent work [4], proposed another template matching based approach relying on GHT which involves a training phase. In contrast, our work is designed as a machine learning classification problem and votes are accumulated, then refined in a novel fashion to generate a bounding box.

Random forest is a popular machine learning algorithm [5]. It has been used in recent vertebra related literature [6–11]. Glocker et. al. presented a random regression forest based localization and identification framework for vertebrae in arbitrary CT scans [10]. They proposed another framework using random classification forest which have shown better performance in localizing and identifying vertebrae with pathological cases [11]. Our work also uses random classification forest. But instead of localizing and identifying each vertebrae, it finds the global position and orientation of the vertebral column in cervical X-ray images.

The recent work by Bromiley et al. [6], demonstrated a segmentation method based on constrained local model (CLM) and random forest regression voting (RFRV). Like other statistical shape model (SSM)-based approaches [7, 12], this work also requires initialization of the mean shape near the actual vertebra. The initialization is usually done with help of manual click points [6, 12] or other automatic methods [7]. Random regression forest-based initialization method described in [7] requires a bounding box from where the input features are collected. In their work, the bounding box around the vertebrae curve is generated using hard parameters which are empirically found based on the training images. In our work, we propose an automatic way to locate the vertebrae column in X-ray images.

In this work, 90 cervical X-ray images of emergency room patients were evaluated. The images contain a total of 450 cervical vertebrae (C3–C7). A random forest is trained to distinguish between vertebra and non-vertebra image patches from the images. The task is designed as a binary classification problem: vertebra and non-vertebra. The framework employs a two-stage coarse-to-fine approach. In the first coarse localization stage, a sliding window sparsely scans a test image to vote for vertebrae patches. After this sparse voting, an accumulation phase converts the votes into a bounding box which indicates the position of the spinal column inside the image. The fine localization stage scans the resultant bounding box of the first stage densely with different patch sizes and orientations. The same voting accumulation phase is applied again and a refined bounding box is generated. The angle of this bounding box determines the predicted orientation of the vertebrae column. Even on a dataset of emergency room X-ray images, 91% of the vertebrae area has been detected under the first stage bounding box and an average error of 3.6° has been achieved for orientation prediction with the second stage bounding box.

2 Data

Our dataset of 90 lateral view emergency room X-ray images was collected from the Royal Devon and Exeter Hospital, and consists of patients exhibiting symptoms, ranging from pain to serious trauma. Different radiography systems were used. The resolution of the images were in the range from 0.1 to 0.194 mm per pixel and the exposure time varied from 16 to 345 milliseconds. The ages of the patients were in the range from 18 to 91. All the scans were digital and taken in 2014-15. These images were anonymized and collected through appropriate procedures to be used for research.

Along with the data, our partners at University of Exeter have also provided manual segmentations of the vertebrae. A set of 20 landmark (LM) points per vertebra was annotated by experts in the field and these annotations were used in training and to evaluate the performance of our algorithm quantitatively. Fig. 1a shows example images from our dataset and Fig. 1b shows manual segmentation points on a spine. For this work, vertebra C3 to C7 are considered. C1 and C2 are not studied as their appearance is ambiguous in lateral cervical X-ray images.

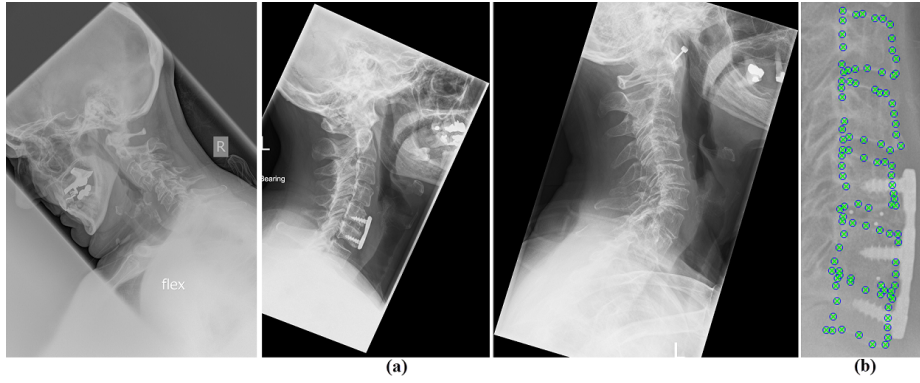


Fig. 1: (a) X-ray images in the dataset. (b) Manual segmentation points.

3 Methodology

The localization framework is based on the detection of vertebrae patches in the images. The detection is done by image patches where a machine learning algorithm decides whether the patch belongs to a vertebra or not. To learn this, a random classification forest [5] has been used. Image patches are generated from the image datasets and labelled into vertebra class and non-vertebra class. The patches are considered with different patch sizes and patch orientations. To generate positive patches, the manual segmentation of the vertebra points is used. The center of the vertebra is used as an anchor point on which different sizes and orientations are considered for training. In order to generate patches for

the non-vertebra class, 50% of the patches are considered from both sides of the vertebral column and the rest are collected from other areas of the image. Fig. 2a shows the areas from which the positive and negative patches are collected; positive patches are collected from the green box, 50% of the negatives patches are collected from the blue boxes and other negative patches are collected from the remainder of the image randomly. More importance is provided in the areas adjacent to the vertebral column for negative patch creation so that the forest has a better opportunity to distinguish these areas. These image patches are then converted to structured forest (SF) feature vectors [13, 14]. This feature vector collects gradient magnitude and orientation information at different scales and angles. This feature vector recently has shown outstanding performance on the edge detection problem [14]. As vertebrae patches are mostly filled with edge-like structures, this feature vector is chosen. Once the feature vectors and corresponding binary output labels are ready, a random classification forest is trained on the data.

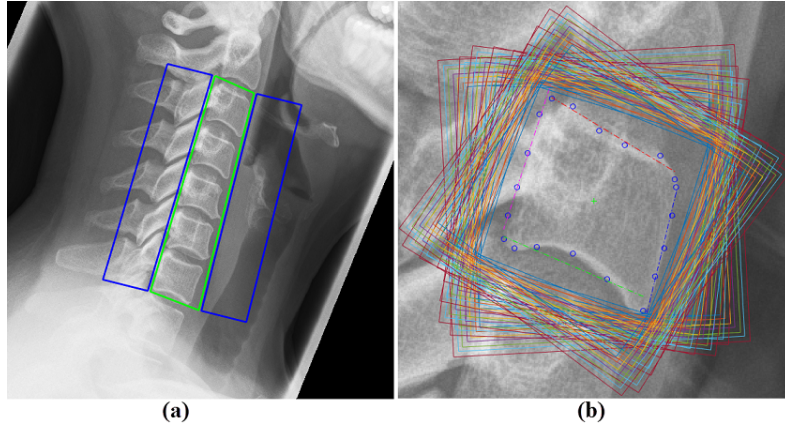


Fig. 2: (a) Area of positive patches (green box) and area of 50% of the negative patches (blue boxes). (b) Positive patch boundaries around a vertebra with different orientations and sizes.

3.1 Stage 1: Coarse Localization

At test time, a new image is fed into the framework for localization. A set of test points is generated on the image at fixed step size (S_1). A single orientation 0° (O_1) and a fixed patch size, P_1 , is considered to generate image patches, one at each of the test points. The generated image patches overlap neighbouring image patches. The amount of overlapping is controlled by the parameters S_1 and P_1 . These patches are fed into the forest. The forest determines which test points belong to vertebrae. These positive predicted points, x_i s, are then passed to the vote accumulation phase to generate a bounding box.

Vote accumulator: The vote accumulator adds a Gaussian kernel at each of the positive votes. The bandwidth, t , of these kernels are automatically estimated using a diffusion-based technique proposed by Botev et al. [15]. This method allows the bandwidth (t) to change dynamically based on the vote distribution from image to image. The resultant distributions are then added together to form a single distribution, F , over the image space.

$$F(\mathbf{x}) = \frac{1}{N} \sum_{i=1}^N \frac{1}{\sqrt{2\pi t}} e^{-\frac{(\mathbf{x}-\mathbf{x}_i)^2}{2t}} \quad (1)$$

where N is the number of total positive votes coming to the accumulator.

This distribution over the image space is converted to a binary image, B , by dynamic thresholding (Eqn. 2). The resulting binary image may be divided into a number of parts, B_j s (Fig. 3c). The area of these parts are measured (A_j) and weighted (w_j) based on the distance from the image center (C_{image}) to the centroid of the concerned image part (C_{B_j}). As the images are taken to diagnose cervical vertebrae related injuries, the assumption is that the spine should be located near the image center, not at any extreme corner of the image. Then some of these areas are eliminated if they are small enough or located far from any adjacent areas (Eqn. 6). This process reduces the chance of misdetection, for example, the area in the skull region of Fig. 3c. Finally, a minimal bounding parallelogram is generated to enclose the rest of areas [16]. This parallelogram is the output of the coarse localization stage. The process is summarized in Fig. 3.

$$B(\mathbf{x}) = \begin{cases} 1 & \text{if } F(\mathbf{x}) > F_t, \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

where $F_t = K \times \max(F)$ and K is an empirically chosen constant. As $\max(F)$ is different for different images, F_t dynamically changes accordingly.

$$A_j = \text{area}(B_j) \quad (3)$$

$$w_j = \frac{1}{\text{distance}(C_{image}, C_{B_j})} \quad (4)$$

$$wA_j = A_j \times w_j \quad (5)$$

where $j = 1, 2, \dots, M$; M is the number of disconnected areas in B and C_a denotes the centroid of the area a . In Fig. 3c $M = 3$.

$$\hat{B}_j = B_j = \begin{cases} \text{valid (kept)} & \text{if } wA_j > A_t \text{ \& } d_{B_j} < d_t \\ \text{invalid (eliminated)} & \text{otherwise.} \end{cases} \quad (6)$$

$$d_{B_j} = \text{minimum} \left(\left\{ \text{distance}(C_{B_k}, C_{B_j}) : k \in \{1, 2, \dots, M\} \text{ and } k \neq j \right\} \right) \quad (7)$$

where A_t and d_j are the empirical area and distance threshold respectively.

$$\text{BoundingBox}_{coarse} = mBP \left(\left\{ \hat{B}_j : j \in \{1, 2, \dots, O\} \right\} \right) \quad (8)$$

where mBP computes the minimum bound parallelogram enclosing the valid B_j s [16] and O is the number of valid disconnected areas. In Fig. 3d $O = 2$.

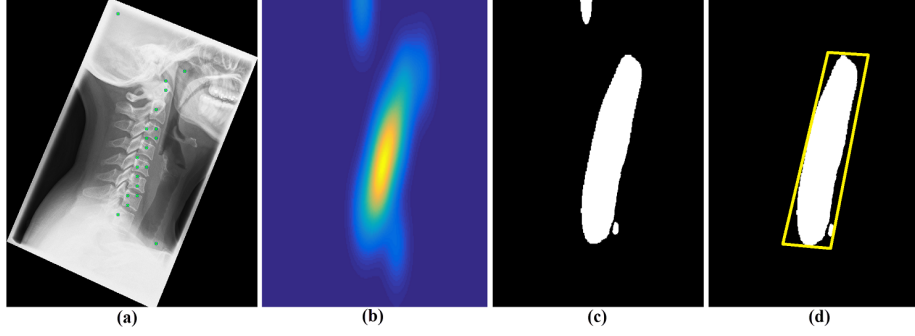


Fig. 3: (a) Positive votes on the image. (b) Resultant distribution F . (c) F after binarization. (d) F after elimination of invalid areas with the minimum bound parallelogram.

3.2 Stage 2: Fine Localization

The previous stage is a single resolution single orientation phase, thus less probable to find vertebra with uncommon orientation or size. As the bounding box of the previous stage is only meant to find the approximate area covered by the vertebra, coarse localization is enough. But in order to find the orientation of the vertebrae curve, a finer localization with multiple patch resolutions and orientations is necessary. In this stage, a new set of test points is created within the coarse localization bounding box, with varying step sizes, S_2 . At each test point, multiple patches are generated with different patch sizes (P_2) and angles (O_2). Then the same random forest patch classification and vote accumulation phase are conducted. This creates a refined bounding box within the first stage bounding box. The orientation angle of this smaller bounding box is computed as the orientation of the vertebrae column.

4 Experiment & Results

To train the random classification forest, different sizes and orientations of the image patches have been considered. The orientation of the patch is defined as the rotation of the angle from the mean vertebral axis. To train the forest, 7 different patch sizes with a step of 0.5 mm (starting from the vertebra size) and 19 orientations of -45° to $+45^\circ$ with a step of 5° have been used. From the 450 cervical vertebrae of our dataset, a total of $450 \times 7 \times 19 = 59,850$ vertebra (positive) images patches were generated. To balance the data, equal numbers of non-vertebra patches were generated from the rest part of each image. Each these image patches was converted to a SF feature vector of length 6116.

The random forest has a number free parameters: maximum allowable tree depth (nD), minimum number of sample at a node ($nMin$), number of trees ($nTree$) and number of variables to test at node split ($nVar$) and number of thresholds to choose from ($nThresh$). To find optimum parameters, a sequential parameter search has been applied to a fixed set of training and test images from the dataset. Final parameters are reported in Table 1. To measure the performance of the trained forest, a ten-fold cross-validation scheme is followed. For each fold, 10% of the images are considered as test images and others are used for forest training. Table 2 reports the patch classification accuracies of each forest.

Table 1: Optimized parameters and values. Table 2: Patch classification accuracy of the forests.

Parameters	Values	Forest	Accuracy
P_1	24 mm	Fold 1	97.90%
S_1	10 mm	Fold 2	98.48%
O_1	0	Fold 3	95.34%
K	0.5	Fold 4	97.91%
A_t	10 pixel	Fold 5	98.21%
d_t	15 mm	Fold 6	95.92%
P_2	20, 30, 40 mm	Fold 7	97.63%
S_2	$P_2/2$	Fold 8	98.25%
O_2	-45, 0, 45	Fold 9	97.62%
nD	10	Fold 10	98.36%
$nMin$	50		
$nTree$	10		
$nVar$	85		
$nThresh$	5		

The localization framework also has a set of free parameters mentioned in Sec 3.1 and 3.2 which are empirically chosen and reported in Table 1. The localization algorithm has been applied on all the images and for each image, the forest was chosen from the ten forests such that the test image is not used in training. We have reported two metrics for the coarse localization bounding box: 1) Average percentage of vertebrae area covered inside the bounding box and 2) Average percentage of landmark points falling outside the bounding box. The orientation of the second stage bounding box is calculated based on the angle of the longer axis of the parallelogram with the horizontal axis. The ground truth orientation is measured by a smallest possible parallelogram that covers the manual annotations (Fig. 4 (1a)). The error is calculated by the absolute different between the ground truth orientation and predicted orientation in degrees ($^\circ$). The results are reported in Table 4 and 3. Overall 91% of the vertebra area fell inside the predicted bounding box. Only 12% of the landmark points were outside the box. The best performance is achieved by the vertebra C4 at 99%, followed by C3 and C5 both at 97%. The performance is worse as we go down the spine, C6 reports 92% and C7 69%. In terms of percentage of landmark points falling outside the bounding box, from C3 to C7, the numbers are

Table 3: Performance of the coarse localization bounding box.

Vertebra	Percentage of area inside the bounding box			Percentage of landmark points outside the bounding box		
	Median	Mean	Std	Median	Mean	Std
C3	100%	97%	14%	0%	7%	16%
C4	100%	99%	4%	0%	2%	7%
C5	100%	97%	12%	0%	4%	14%
C6	100%	92%	22%	0%	11%	24%
C7	87%	69%	37%	33%	37%	36%
Overall	100%	91%	24%	0%	12%	25%

Table 4: Orientation error in degree($^{\circ}$): GTO: Ground truth orientation, ALMP: All landmark points, FLMP: Landmark points inside the first stage bounding box.

GTO type	Coarse Localization		Fine Localization	
	ALMP	FLMP	ALMP	FLMP
Median	5.73	3.89	3.07	2.37
Mean	8.16	6.26	4.59	3.60
Std	8.21	7.18	5.27	4.55

7%, 2%, 4%, 11%, and 37%. Fig. 5 demonstrates the metrics graphically. Almost 80% of the vertebrae have no parts of it outside the bounding box. In terms of landmark points, 70% the vertebrae have no LM points outside the bounding box and about 80% have less than three points out of 20 LM points (15%).

The orientation error metric can be computed in two ways. One with all the vertebrae (ALMP), C3-C7, the other with only the landmark points that fall inside the bounding box of the first stage (FLMP). As the second stage can only use the information what's inside the first stage bounding box, the later seems more fair to judge its ability. When considering all the vertebra the average error is 4.59° while the other results in an average of 3.6° . For the coarse localization bounding box the average errors are larger: 8.16° and 6.26° respectively.

Table 5 reports the average Dice coefficient, sensitivity (true positive rate) and specificity (true negative rate) of the coarse and fine localization bounding boxes. These metrics are computed by comparing the ground truth bounding box (Fig. 4 (1a)) with predicted bounding boxes. The Dice coefficient for coarse localization bounding box averages at 0.62 where it stands at 0.69 for the fine localization bounding box. However in terms of sensitivity, the first stage bounding box scores 0.88 while the second stage bounding box scores only 0.62. Specificity is high for both bounding boxes: 0.97 for coarse localization and 0.99 for fine localization.

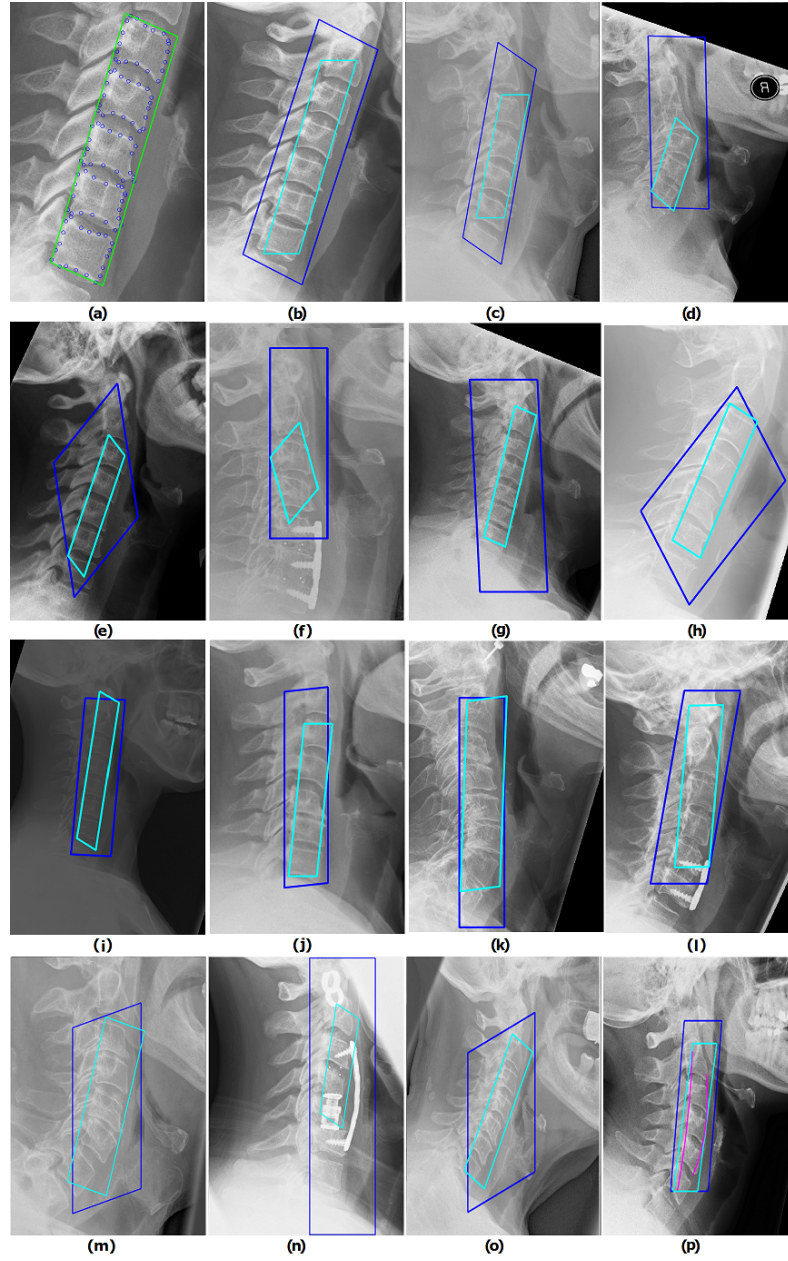


Fig. 4: (a) Manual annotation points and ground truth bounding box (green). (b)-(p) Coarse (blue) and fine (cyan) localization bounding boxes. (p) An example of the ongoing vertebral curve detection method (magenta).

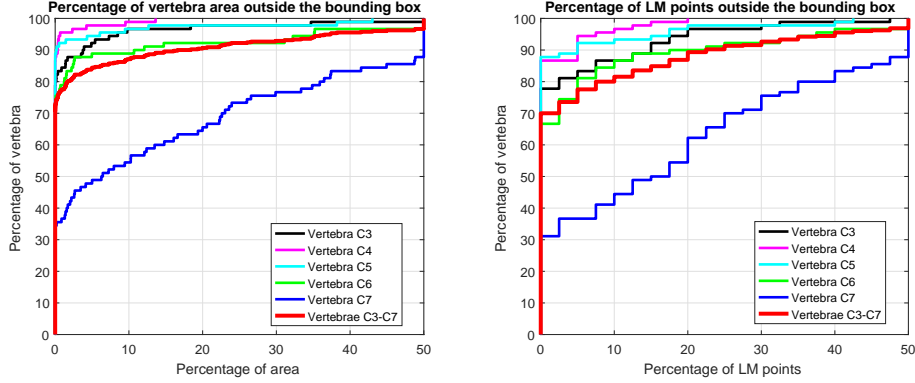


Fig. 5: Percentage of area and landmark points outside the coarse localization bounding box.

Table 5: Localization performance.

	Coarse Localization			Fine Localization		
	Dice coeff.	Sensitivity	Specificity	Dice coeff.	Sensitivity	Specificity
Median	0.65	0.93	0.97	0.71	0.61	1.00
Mean	0.62	0.88	0.97	0.69	0.62	0.99
Std	0.14	0.14	0.03	0.11	0.15	0.01

5 Discussion and Conclusion

In this work, a coarse to fine cervical spine localization algorithm has been evaluated on a set of 90 emergency room X-ray images. The algorithm is based on a random forest patch classifier which distinguishes between the vertebra and the non-vertebra image patches. Based on the centers of vertebra patches on a test image, a novel vote accumulator converts the votes into a bounding box. A second multi-resolution multi-orientation patch classification is applied inside the initial bounding box to determine the orientation of the vertebral column. The resultant coarse localization bounding box covers 91% of the all vertebral area on an average with a maximum of 99% for vertebra C4. C4's location on the spine is key to the increased accuracy. On average only 12% of the landmark points fell outside the bounding box, most of which are from the lowest vertebra, C7, where the image quality is often reduced.

While coarse localization creates a larger bounding box, the fine localization creates a smaller and refined bounding box. This bounding box predicts the orientation of the spinal column better. The average orientation error of the fine localization bounding box is 3.6° only while for the coarse bounding box, the error is 6.26° . The fine localization scans the coarse localization box with more

variation and thus it can find the spinal orientation with better accuracy.

To measure the compactness of the both bounding boxes, Dice coefficients and sensitivity metrics are computed. The Dice coefficient of the fine localization bounding box is 9% higher than the Dice coefficient of the coarse localization box. However, in terms of sensitivity, coarse localization outperforms fine localization bounding box by 30%. Based on the application in which the bounding boxes will be used, the user may choose between the two options.

Our algorithms outperformed the performances of [3,4]. [3] reported an average orientation error of 4.16 degree and [4] reports a vertebra detection 89%. However, [3] report only 10% landmark points to be outside the bounding box which is lower than our 12%. But their landmark points did not consider the posterior points. It is also important to mention that both of these works, has been performed on a small (40 and 50) images from NHANES-II dataset of scanned X-ray images, where the images are collected from healthy patients for the purpose of developing automatic algorithms thus contains less variation, injuries and exposure differences. In our case, the dataset represents X-ray images collected from real life emergency room images where resolution, patient age, injury, orientation, X-ray exposure all vary widely. Fig. 4 shows examples of images with low contrast (h, i), bone implants (f, l, n), displacements (j, m) and osteoporosis (d, k). Our algorithm works well in all these conditions.

The algorithm is written in MATLAB2014b on a Intel Core-i5 3GHz machine with 8GB RAM and have not been optimized for execution time. The unoptimized code takes on average around 2.5 seconds to run the whole localization procedure (both coarse and fine). The execution time varies based on the image size, resolution and number of positive votes at each stage.

The performance of our algorithm can be attributed to the training of the forests and to the novel voting accumulation process. The patch classification accuracy of forests is in the range of 95 to 98% (Table 2) which eliminates the majority of the false detections. The novel voting accumulation method which utilises dynamic diffusion based kernel density estimation and weighted area filtering eliminates the rest of the false detection and thus the final results are good. We are currently working on a vertebral curve detection method (Fig. 4(p)), which can detect the anterior and posterior vertebral curves. A single orientation angle is not capable of describing the spinal column accurately. In many cases, the spinal column is a curve than a straight line (Fig. 4(d, m)). Thus, these curves will tell us more about the global orientation of the spine. Our next target is to detect the vertebrae centers or other landmarks automatically like [6–9]. The output of this work will be helpful in order to limit our search over the image. It can also help algorithms [7–9] where the search area was manually reduced with hard coded parameters.

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. C. Morris and E. McCoy, "Clearing the cervical spine in unconscious polytrauma victims, balancing risks and effective screening," *Anaesthesia*, vol. 59, no. 5, pp. 464–482, 2004.
3. A. Tezmol, H. Sari-Sarraf, S. Mitra, R. Long, and A. Gururajan, "Customized Hough transform for robust segmentation of cervical vertebrae from X-ray images," in *Image Analysis and Interpretation, 2002. Proceedings. Fifth IEEE Southwest Symposium on*, pp. 224–228, IEEE, 2002.
4. 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.
5. L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.
6. 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.
7. 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.
8. 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.
9. 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.
10. B. Glocker, J. Feulner, A. Criminisi, D. R. Haynor, and E. Konukoglu, "Automatic localization and identification of vertebrae in arbitrary field-of-view CT scans," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012*, pp. 590–598, Springer, 2012.
11. B. Glocker, D. Zikic, E. Konukoglu, D. R. Haynor, and A. Criminisi, "Vertebrae localization in pathological spine CT via dense classification from sparse annotations," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 262–270, Springer, 2013.
12. 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.
13. P. Dollár and C. L. Zitnick, "Structured forests for fast edge detection," in *ICCV*, 2013.
14. P. Dollár and C. L. Zitnick, "Fast edge detection using structured forests," *PAMI*, 2015.
15. Z. I. Botev, J. F. Grotowski, D. P. Kroese, *et al.*, "Kernel density estimation via diffusion," *The Annals of Statistics*, vol. 38, no. 5, pp. 2916–2957, 2010.
16. C. Schwarz, J. Teich, E. Welzl, and B. Evans, *On finding a minimal enclosing parallelogram*. Citeseer, 1994.