Fully Automatic Cephalometric X-Ray Landmark Detection Using Random Forest Regression and Sparse shape composition

Chengwen Chu[1], Cheng Chen[1], Ching-Wei Wang[2,3], Cheng-Ta Huang[2], Chung-Hsing Li[3,4], Lutz-P.Nolte[1], and Guoyan Zheng[1]

[1] Institute for Surgical Technology and Biomechanics, University of Bern, Switzerland chengwen.chu@istb.unibe.ch, guoyan.zheng@ieee.org
[2] Graduate Institute of Biomedical Engineering, National Taiwan University of Science and Technology, Taiwan
[3] Dental Department, National Defense Medical Center, Taiwan
[4] Orthodontics and Pediatric Dentistry Division, Dental Department, Tri-Service General Hospital, Taiwan

Abstract. Landmark detection from 2D cephalometric X-ray images is a basic and important function for cephalometric analysis which widely used in oral surgery planning and evaluation. We present a fully automatic landmark detection method for cephalometric analysis. Our method works by combining random forest (RF) regression based landmark detection with sparse shape composition model based landmark correction. Validation on 100 cephalometric X-ray images show that 77.79% landmarks can be detected by our method with an error less than 4.0 mm.

1 Introduction

Recently, cephalometric analysis becomes more and more important and widely used in treatment planning and evaluation of orthodontics, orthopedics, and other areas of oral and maxillofacial surgery [1]. In cephalometric analysis on 2D X-ray images, the dentists aim to study relationships between the dental and skeletal in the head. To reach this purpose, a set of critical craniofacial landmark positions [2] are necessary to be identified as these landmarks providing the interpretation of patients' bony structures. Since manual identification of landmark position is very time consuming and hard to achieve reproducible results, developing an automatic landmark detection algorithm for cephalometry is highly demanded.

However, fully automatic landmark detection on 2D X-ray projections is still a challenge due to the following three reasons: 1) The quality of radiographs may vary considerably in terms of contrast, resolution and the region of the skull bone. 2) Plain film radiographs do not provide homogeneous values for the same structures due to overlapping body parts. 3) Deformities of the dental and skeletal bone may cause the loss of distinguishable radiographic key features. Therefore, the existing techniques using edge detection embedded with

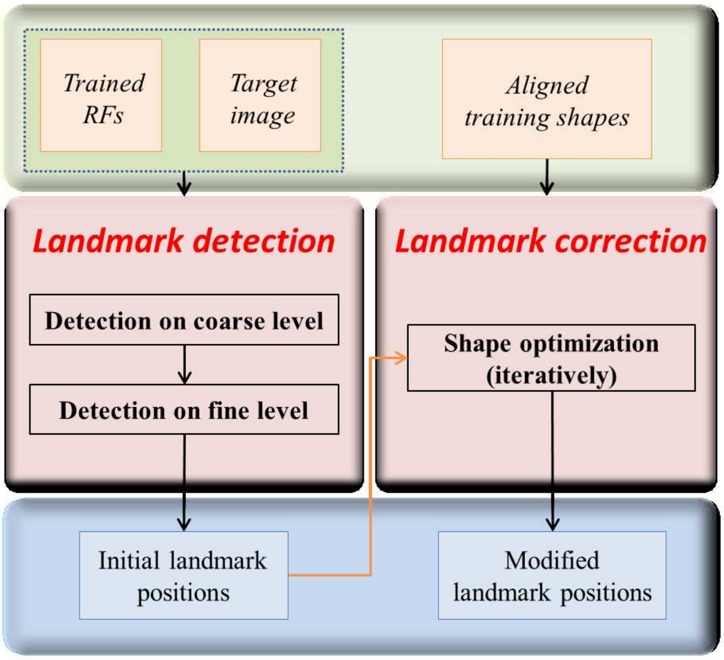

Fig. 1. The flowchart of our proposed segmentation method

prior knowledge [3, 4] are difficult to provide satisfactory results since the performance of these methods are highly related to the quality of input images. Instead, methods using statistical knowledge are proposed [5] [6, 7]. However, model based methods suffer from the requirement of proper initialization, which is typically done manually, and the limited converging region, leading to unsatisfactory results.

In order to address problems mentioned above, machine learning based methods have been introduced for landmark detections from medical images. In [8], EI-Feghi applied neural network to identify 25 most commonly used landmarks. In [9], Chakrabartty implemented support vector machines (SVM) to detect 16 landmarks and more than 95% accuracy was obtained. In [10], Zheng et al. proposed marginal space learning to automatically localize the heart chamber from 3D CT. More recently, random forest (RF) regression has been proved to be effective for automatically organ localization from 2D Xray images [11].

In this paper, we address the landmark detection problem on 2D cephalometric X-ray images by a fully automatic landmark detection method using RF regression. Since there might be some outliers which are not correctly detected, we further add shape constraints to modify the detected landmark positions. This is addressed by using a sparse shape composition model for landmark correction. Fig. 1 shows an overview of our proposed method.

2 Method

2.1 Landmark detection by random forest regression

Basic algorithm. We have a separate RF landmark detector for each landmark. An overview of our landmark detection framework is shown in Fig. 2. Please note that both this paper and [12] use the same framework but different landmark detection algorithms. During training, in each training image, we sample a set of rectangular image patches around the ground-truth landmark position which is known. Each sampled patch is represented by its visual feature $\mathbf{f}_i \in \mathbb{R}^{d_f}$ and the displacement $\mathbf{d}_i \in \mathbb{R}^2$ from its center to the landmark (Fig. 2(a)). Let us denote all the sampled patches in all training images as $\{P_i = (\mathbf{f}_i, \mathbf{d}_i)\}_{i=1...N}$ (Fig. 2(b)). The goal is then to learn a mapping function $\phi : \mathbb{R}^{d_f} \to \mathbb{R}^2$ from the feature space to the displacement space. Principally, any regression method can be used. In this paper, similar to [11, 13], we utilize the random forest regressors [14].

Once the regressor is trained, given a new image (Fig. 2(c)), we randomly sample another set of patches $\{P'_k = (\mathbf{f'}_k, \mathbf{c'}_k)\}_{k=1...N'}$ all over the image (or an region of interest if an initial guess of the landmark position is known), where $\mathbf{f'}_k$ and $\mathbf{c'}_k$ are the visual feature and center coordinate of the kth patch, respectively (Fig. 2(d)). Through the trained mapping ϕ, we can calculate the predicted displacement $\mathbf{d'}_k = \phi(\mathbf{f'}_k)$, and then $\mathbf{d'}_k + \mathbf{c'}_k$ becomes the prediction of the landmark position by a single patch P'_k (Fig. 2(e)). Note that each tree in the random forest will return a prediction. Therefore, supposing that there are t trees in the forest, we will get $N' \times t$ predictions. These individual predictions are very noisy, but when combined, they approach an accurate prediction. To this end, we consider each single vote as a small Gaussian distribution. We use a general probability aggregating algorithm to add these distributions to get a soft probability map called *response image* which gives, for every position of the image, its probability of being the landmark (Fig. 2(f)).

Visual feature. As for the visual feature for the image patches, we use the multi-level HoG (Histogram of Oriented Gradient) [15] as our feature for image patches. To keep the randomness of our trained RFS, we use a feature selection algorithm propose in [16] to efficiently select only the most relevant feature components. In our method, only one third of the features are randomly selected from all the dimensions of extracted features when training a new node.

Multi-resolution processing. We process each landmark detection in a two-step coarse-to-fine strategy. In the first step, we down-sample the images to a coarse resolution, on which we performed the training and test as described above. In the test stage, we sample image patches throughout the entire image. In the second step, training and test are done in the original image resolution. The detection result from the first step is used as initialization and therefore in the test stage we only sample patches in a local region around the initial landmark locations. A more detailed parameter values are introduced in the following.

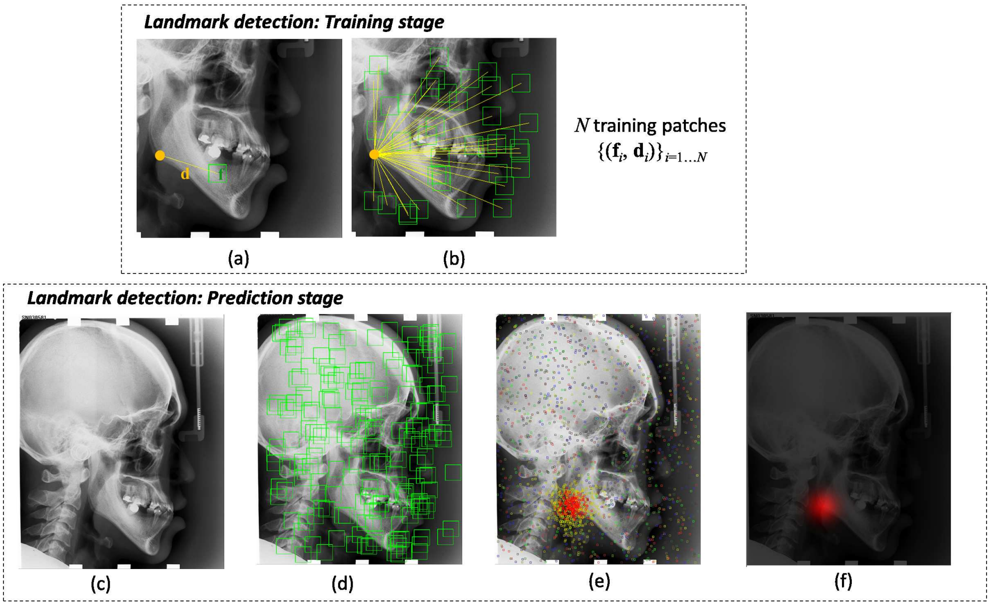

Fig. 2. The RF training and landmark detection. Illustration on coronal slice for easy understanding. (a) A patch sampled around the true landmark position. (b) Multiple sampled training patches from one atlas. (c) A target image. (d) Multiple sampled test patches over target image. (e) Each patch gives a single vote for landmark position. (f) Response image calculated using Gaussian transform.

2.2 Landmark Correction using Sparse Shape Model

In this section we present our method which searches for the optimal landmark positions under shape constraints using a spare shape composition model. Prior to the searching, we have the initial shape $y_0 \in \mathbb{R}^{2K}$ for all the K landmarks, and also a shape model $\mathcal{M}$ includes a set of pre-aligned training shapes y_i, where $i \in \{0, 1, ..., n-1\}$. Each shape y_i is described by a vector containing $2K$ vertices: $y_i = \{x_0, y_0, x_1, y_1, ..., x_{2K-1}, y_{2K-1}\}$. The task is then to find the optimal shape $y^* \in \mathbb{R}^{2K}$, starting from the initial shape y_0, constrained by the prior information encoded in the shape model.

The procedure is shown in Table 1. Basically, starting from the initial shape, we update the shape iteratively. In each iteration, we perform two actions: regularize the shape by the shape model, and update the shape pose. These steps are straightforward except the shape regularization, which regularizes the locally updated shape by the shape model to remove noise (step 2 in Table 1). Traditionally, this can be done by the ASM [5] based on PCA (Principal Component Analysis). In this paper, we instead employ the recently proposed shape composition model based on sparse representation [17]. Here we briefly explains this method.

Table 1. The algorithm of shape optimization.

Input: Initial shape y_0, shape model $\mathcal{M}$ with aligned training shapes $\{y_i\}_{i=1,...,N}$
Output: Optimal shape y^*, pose transform matrix T
Procedure:
1. Initialize $y = y0$, $T =$ the optimal similarity transform from y to shape model $\mathcal{M}$ 2. Regularize shape y by the shape model $\mathcal{M}$. 3. Update pose T by the optimal similarity transform from y to shape model $\mathcal{M}$ 4. Repeat steps 2 to 4 until convergence. Then $y^* = y$

The shape model consists of a set of pre-aligned training shapes $\{y_i\}_{i=1,...,N}$. For each new shape y' to be regularized, after computing a transformation T (which is evaluated separately in step 3 of Table 1), it should be approximated by a linear combination involving only a small subset of the training shapes, plus a sparse error:

$$T(y') \approx Yx + e = [YI] \begin{bmatrix} x \\ e \end{bmatrix} = Y'x' \tag{1}$$

where $Y' = [Y, I]$, and $x' = [x^\top, e^\top]^\top$. In Eq. 1, both the linear coefficients x and the error e are sparse. Therefore, the composite coefficient x' is also sparse. Our goal becomes to solve the following L_1-regularized least squares problem:

$$x'_{opt} = \arg\min_{x'} \left(\parallel T(y') - Y'x' \parallel_2^2 + \lambda \parallel x' \parallel_1 \right) \tag{2}$$

where λ is a parameter controlling the importance of the sparsity constraint. There are a number of solves for Eq. 2, and we employ the method using truncated newton interior-point method.

The interpretation of Eq. 1 is clear: the shape y' should be approximated (with a transformation T) as a linear combination of only a small number of "basis, which can either be the training shapes, or standard basis of the $\mathbb{R}^{2K}$ space. The contribution from the training shapes represents the "true part of shape y' that is consistent to the shape model, and the contribution from the standard basis accomodates large but sparse errors (noises). Therefore, after we get the optimal x'_{opt} by Eq. 2, we decompose x'_{opt} by $x'_{opt} = [x^\top_{opt}, e^\top_{opt}]^\top$ as in Eq. 1, discard the e_{opt} which corresponds to the noises, and the regularized shape is given by back-projecting the"true part of the shape:

$$y'_{regularized} = T^{-1}(Yx_{opt}) \tag{3}$$

Thus we complete the shape regularization step in Table 1.

3 Experiments and Results

Implementation Details. Totally 100 labeled training images are used for RF training. In coarse landmark detection stage over down-sampled image space,

$N = 20000$ patches are separately sampled for training each RF. The size of down-sampled image is 1/3 of the original image both in x-, y- directions. Each trained RF contains 10 separated random trees, and each tree has a maximum of decision depth of 15. Same training parameters are used for training RFs in original image space. In landmark prediction step, $N' = 2000$ patches are sampled in down-sampled image as same as in original image for each landmark position.

Dataset and Results. For RF training, all the provided 100 training images and its ground-truth of landmark positions are used. For testing, the given 100 test images are used. Both training images and test images are 2D cephalometric X-ray images, and these images have the same value of 1935×2400.

We first conducted a leave-twenty-out study based on the 100 training images. More specifically, we subdivided the 100 training images into two groups. The first group include 80 images which we used to train our RF regressor and the rest 20 images were used for test. From this study, taking the given label data as the ground truth, we found that 84.21% landmarks detected with our method are smaller than 4.0 mm.

We also evaluated our presented method on the given 100 test images. We found that the average distance between the automatic detection and ground truth of 77.79% landmarks is smaller than 4.0 mm, which is slightly lower than the results of our leave-twenty-out study. This is properly due to the over-fitting for leave-twenty-out study since training images are similar with each other in image features. Fig. 3 shows a couple of test image with automatically detected landmarks.

The computation time to detect all 19 landmarks cost on average $1 \sim 2$ minutes. All test on a computer with 3.0 GHz CPU and 12G RAM.

4 Conclusion

This paper proposed a fully automatic method for landmark detection from 2D cephalometric X-ray images. Our method works by combining random forest (RF) regression based landmark detection with sparse shape composition model based landmark correction.Results from experiments conducted on 100 cephalometric X-ray images show that our method detected 77.79% landmarks with an error less than 4.0 mm. The present method can be extended to other kind of medical images such as CT and MRI.

References

1. Yue, W., Yin, D., Li, C., Wang, G., Xu, T.: Automated 2-D Cephalometric Analysis on X-ray Images by a Model-based Approach. IEEE T-BME, 53(8), 1615-1623 (2006)
2. Rakosi, T.: An Atlas and Manual of Cephalometric Radiology. Wolfe Medical, 1982.

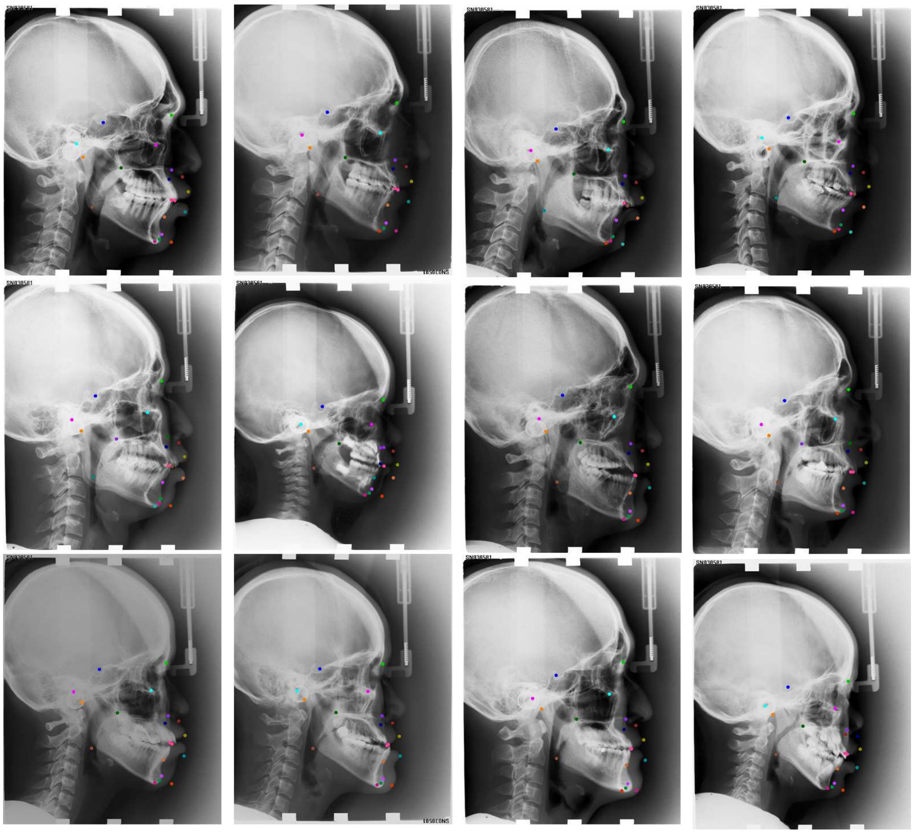

Fig. 3. Results of landmark detection.

3. Ren, J., C., Liu, D., Shao, J., L.: A knowledge-based automatic cephalometric analysis method. In: IEEE EMBC 1998, vol. 20, no. 2, pp. 723-727.
4. Wei, M., Zhou, Z., Ding, W.: Prelimary study of computerized automatic identification of landmarks in X-ray cephalometry Chin. J. Oral Maxillofac. Surg., 10(4), 283-287 (2000)
5. Cootes, T., F., Taylor, C., J., Cooper, D., H., Graham, J.: Active shape models: their training and application. Comput. Vis. Image Understanding, 61(1), 38-59 (1995)
6. Dong, X., Zheng, G.: Automatic extraction of proximal femur contours from calibrated x-ray images using 3D statistical models: an in vitro study. IJMRCAS 5(2), 213222 (2009)
7. Cristinacce, D., Cootes, T.: Automatic feature localization with constrained local models. Pattern Recognition 41(19), 30543067 (2008)
8. El-Feghi, I., Sid-Ahmed, M., A., Ahmadi, M.: Automatic localization of craniofacial landmarks for assisted cephalometry. Pattern Recognit., 37(3), 609-621 (2004)
9. Chakrabartty, S., Yagi, M., Shibata, T., Gauwenberghs, G.: Robust cephalometric landmark identification using support vector machines. In Int. Conf. Multimedia

and Expo 2003, vol. 3, pp. 429432.
10. Zheng, Y., Barbu, A., Georgescu, B., Scheuering, M., Comaniciu, D.: Four-chamber heart modeling and automatic segmentation of 3-D cardiac CT volumes using marginal space learning and steerable features. IEEE T. Med. Imaging 27(11), 16681681 (2008)
11. Lindner, C., Thiagarajah, S., Wilkinson, J. M., arcOGEN Consortium, Wallis, G., Cootes, T. F.: Fully automatic segmentation of the proximal femur using random forest regression voting. IEEE TMI 32(8),14621472 (2013)
12. Chen, C., Wang, C-W., Huang, C-T., Li, C-H, Zheng, G.: Fully-Automatic Landmark detection in Skull X-Ray images. submitted to Automatic Cephalometric X-ray Landmark Detection Challenge 2014.
13. Criminisi, A., Shotton, J., Robertson, D., Konukoglu, E.: Regression forests for efficient anatomy detection and localization in CT studies. In: Menze, B., Langs,G., Tu, Z., Criminisi, A. (eds.) MICCAI 2010. LNCS, vol. 6533, pp. 106117. Springer, Heidelberg (2011)
14. Breiman, L.: Random forests. Machine Learning 45(1), 5–32 (2001)
15. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: CVPR, vol. I, pp. 886893 (2005)
16. Chen, C., Yang, Y., Nie, F., Odobez, J.M.: 3D human pose recovery from image by efficient visual feature selection. CVIU 115(3), 290299 (2011)
17. Zhang, S., Zhan, Y., Dewan, M., Huang, J., Metaxas, D.N., Zhou, X.S.: Sparse shape composition: a new framework for shape prior modeling. In: CVPR (2011)