Andreas Wirtz*, Julian Lam, and Stefan Wesarg

Automated Cephalometric Landmark Localization using a Coupled Shape Model

Abstract: Cephalometric analysis is an important method in orthodontics for the diagnosis and treatment of patients. It is performed manually in clinical practice, therefore automation of this time consuming task would be of great assistance. In order to provide dentists with such tools, a robust and accurate identification of the necessary landmarks is required. However, poor image quality of lateral cephalograms like low contrast or noise make this task difficult. In this paper, an approach for automatic landmark localization is presented and used to find 19 landmarks in lateral cephalometric images. An initial predicting of the individual landmark locations is done by using a 2-D coupled shape model to utilize the spatial relation between landmarks and other anatomical structures. These predictions are refined with a Hough Forest to determine the final landmark location. The approach achieves competitive performance with a successful detection rate of 70.24% on 250 images for the clinically relevant 2mm accuracy range.

Keywords: cephalometric radiographs, coupled shape model, hough forest, landmark localization

1 Introduction

The analysis of dental radiographic images is an important tool used for clinical diagnosis, treatment procedures and surgery. In orthodontics, the cephalometric analysis is used to extract fundamental information from lateral cephalometric radiographs. To study the soft tissue-, dental- and bony structures of a patient, several different landmark positions have been defined. The type and number of landmarks varies between different classification methods (e.g Steiner [7] or Ricketts [5]). In clinical practice, these landmarks need to be manually traced which is a time consuming and subjective task. An automated landmark detection would be beneficial but is difficult to achieve with high precision and success rate. Fig.

1 shows a cephalometric image with 19 landmark locations marked.

Several methods for the automatic detection of cephalometric landmarks have been proposed in the past. In 2015 Wang et al. [8] organized a Grand Challenge at the International Symposium on Biomedical Imaging (ISBI) on this topic and provided a publicly available data set of 400 images for training and testing. The best results were achieved by approaches employing Random Forests for the classification of the intensity appearance of individual landmarks while constraining the possible locations of landmarks using statistical shape models. In 2017, Arik et al. [1] employed the rapidly developing deep learning techniques for the detection of the landmarks. They trained a convolution neural network to output probabilistic estimations of different landmark locations. Again, better results could be achieved by using a shape model to refine the landmark potions. Recently, Payer et al. [4] proposed a CNN-based framework for anatomical landmark detection for small training sets. Their SpatialConfiguration-Net (SCN) splits the localization task into two sub-problems: ambiguous but accurate candidate prediction and reducing ambiguities by incorporating spatial relations. All approaches achieve similar results on the ISBI 2015 Challenge dataset ranging from 68.13% to 73.33% successful detection rate of the landmark positions in the clinically acceptable 2mm accuracy range. Furthermore, all approaches benefit from utilizing the spatial relations between landmark points to eliminate potential candidate positions which are improbable based on their relative position to other landmarks. This indicates that landmarks cannot be located only based on their appearance, instead the search space needs to be restricted or constrained.

In this paper, we present an approach that combines a 2-D coupled shape model for predicting the individual landmark locations with a Hough Forest to refine each location. Unlike the aforementioned methods, the coupled shape model exploits not only the spatial relation between landmarks but also the relation to other anatomical structures like the skull or the skin. Thereby, landmark locations can be predicted based on the shape and location of these structures and the search space for each landmark can be narrowed down to a small region of interest. The final location of each landmark is then determined based on the local intensity appearance by a landmark-specific Hough Forest.

---

*Corresponding author: Andreas Wirtz, [a]Fraunhofer IGD, Fraunhoferstr. 5, 64283 Darmstadt, Germany, e-mail: andreas.wirtz@igd.fraunhofer.de, [b]Technical University of Darmstadt, Department of Computer Science, Darmstadt, Germany
Julian Lam, Fraunhofer IGD, Darmstadt, Germany
Stefan Wesarg, [a]Fraunhofer IGD, Darmstadt, Germany, [b]Technical University of Darmstadt, Department of Computer Science, Darmstadt, Germany

L1: Sella
L2: Nasion
L3: Orbitale
L4: Porion
L5: Subspinale
L6: Supramentale
L7: Pogonion
L8: Menton
L9: Gnathion
L10: Gonion
L11: Lower incisor
L12: Upper incisor
L13: Libialis superior
L14: Libialis inferior
L15: Subnasale
L16: Soft pogonion
L17: Post. nasal spine
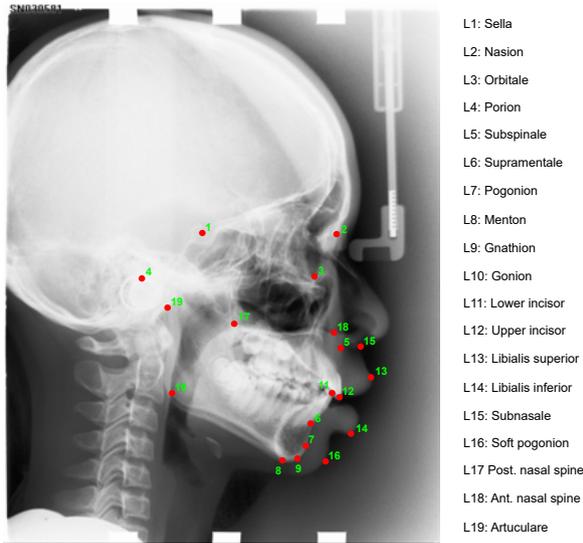L18: Ant. nasal spine
L19: Articulare

Fig. 1: Lateral cephalometric image with 19 landmarks locations.

2 Methods

The presented approach for cephalometric landmark localization can be roughly divided into two steps: location prediction and location refinement. While it is possible to detect the location of each landmark individually, much better results can be achieved by also considering the spatial relation between landmarks [1, 4, 8]. In this approach, we not only use the relation between landmarks but also include their spatial relation to additional structures like skull or skin. The outline of these additional structures is much easier to detect in cephalometric images and can be used to predict landmark locations in the image. After the prediction step, the search space for each landmark can be narrowed down and the final position in a small region around the predicted location can be determined based on its local intensity appearance. This enables the training of more specialized Hough Forests. Additionally, negative samples can be picked from a much smaller area, potentially resulting in a more robust descriptor.

2.1 Location prediction

A 2-D coupled shape model is used in this work to exploit the spatial relation between landmarks as well as the outline of skull and skin. The model combines the statistical knowledge about the shape of each structure with information about their spatial relation and has been used successfully to segment structures in 2-D dental radiographs [9, 10]. The individual structures are represented as statistical shape models and are generated using a point distribution model (PDM) and principal component analysis (PCA). The relative position of each

structure is described by an affine 2-D transformation. Please refer to Wirtz et al. [9] and Steger et al. [6] for more information on the coupled shape model.

While the landmarks themselves cannot be represented by shape models, they can still be included in the coupled model. Therefore, the model has been extended to support so called single point structures (SPS) whose spatial relation to other structures is part of the model framework and learned during model training. However, as single point structures they do no possess any shape information and therefore no shape model is trained for these structures. When adapting the coupled model to an input image, only the structures with statistical shape information are fitted to the image data based on gradient features (bottom-up information). In this case this applies to the outline of the skull and the skin. SPS, which represent landmark locations, are only changed by the learned a priori statistical knowledge about their spatial relation to the other structures (top-down information), but not yet by any image features. Nonetheless, a good prediction of the landmark position is achieved allowing us to restrict the space of potential locations to a small region of interest.

The coupled shape model has been trained on a subset of 50 images from the training set since manual segmentation of the skull and skin in these images is required for generating the shape models for these structures. These shape models consist of 80 contour points each (cf. Fig. 2) and only the principal components describing 95% of the shape variation are kept.

2.2 Location refinement

The final refinement of the landmark positions is performed using a landmark-specific Hough Forest [2]. These has already been previously used for the localization of cephalometric landmarks [3]. Hough Forests are effective for object localization by combining votes from multiple random trees. Each of these trees maps the local appearance of a landmark to its respective leafs and each leaf is attributed a probabilistic vote for an objects position. The individual trees are trained independently of each other with a random subset of the training data. By combining the votes of all individual trees, the Hough Forest can achieve accurate and robust localization results.

Landmark specific forests have been trained on a subset of 120 images from the training set. 120 positive samples were used which are centered at the ground truth locations. To improve the localization accuracy 70 negative samples were added to the training process. These negative samples are chosen from locations which incorrectly received a high probability during a detection process trained with only positive samples. For each landmark 20 forests with a size of 200×200 pixels are trained.
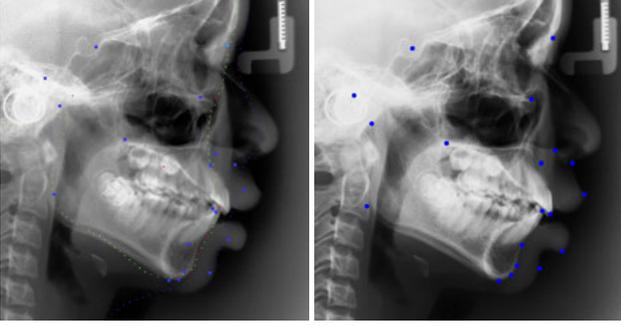
Fig. 2: Left: Result after landmark prediction step with predicted locations (blue) and structure outlines. Right: Final landmark locations (blue) after location refinement step.

3 Experiments & results

3.1 Dataset & metrics

The presented landmark localization approach has been trained and tested on the ISBI 2015 Challenge[1] dataset [8]. It consists of 400 cephalometric images whereby 150 images are designated for training and the remaining 250 image are for testing. For each image, two experts have manually identified 19 individual landmarks which serve as ground truth. The image sizes are 2400 × 1935 pixels. The physical resolution is 0.1 mm per pixel along both directions.

For evaluating the performance of the presented approach, the same metrics as in the original publication [8] have been used, namely mean radial errors with standard deviation and successful detection rate with various ranges of accuracy. The radial error between the detected landmark coordinate d and the reference landmark coordinate r is computed by

$$R(d,r) = \sqrt{(d_x - r_x)^2 + (d_y - r_y)^2},\qquad(1)$$

where x and y are coordinates on the X- and Y-axis. The mean radial error for a single landmark is then defined as

$$MRE = \frac{1}{N} * \sum_{i=1}^{N} R_i,\qquad(2)$$

where N is the number of images. The standard deviation is given by

$$SD = \sqrt{\frac{1}{N} * \sum_{i=1}^{N} (R_i - MRE)^2}.\qquad(3)$$

The successful detection rate for each landmark is then defined as the fraction of correct detections over the total amount of detections (i.e. number of images). A position is considered to be correctly detected if the radial error between the detected

1 http://www-o.ntust.edu.tw/~cweiwang/ISBI2015/challenge1/

Tab. 1: Experimental results on the ISBI 2015 Challenge dataset [8]. Mean Radial Error (MRE) with Standard Deviation (SD) and Successful Detection Rate (SDR) in four categories are show for each landmark as well as the total average (AVG) over all landmarks.

ID	MRE ± SD	SDR (%)			
	(in mm)	2.0mm	2.5mm	3.0mm	4.0mm
1	1.616 ± 3.084	88.80	90.00	90.80	93.20
2	1.632 ± 2.181	73.20	80.00	84.00	90.00
3	1.824 ± 3.815	76.00	81.20	86.00	92.80
4	3.366 ± 5.539	68.00	75.20	77.20	80.80
5	2.426 ± 1.991	50.80	63.20	71.20	84.00
6	2.133 ± 2.406	58.00	73.20	78.80	86.80
7	2.042 ± 3.593	69.20	85.20	90.80	96.00
8	1.099 ± 0.816	89.20	91.20	94.80	100.0
9	1.399 ± 0.983	79.20	90.80	96.00	97.20
10	7.657 ± 8.383	30.80	36.00	40.80	49.20
11	1.782 ± 3.185	78.80	81.20	84.00	88.00
12	1.771 ± 4.127	85.20	86.00	88.80	90.00
13	3.067 ± 4.499	56.80	77.20	90.00	92.80
14	2.376 ± 4.786	80.80	90.80	92.00	95.20
15	1.755 ± 2.736	77.20	86.00	88.00	92.00
16	2.019 ± 2.771	67.20	78.80	85.20	92.00
17	1.564 ± 2.234	80.00	91.20	94.00	96.80
18	2.278 ± 2.598	57.20	70.00	75.20	84.80
19	2.516 ± 3.776	68.00	76.00	78.80	86.00
AVG	2.333 ± 3.342	70.24	79.14	83.49	89.82

landmark coordinate d and the reference landmark coordinate r is not greater then a threshold T. The four thresholds used in the Challenge are 2.0mm, 2.5mm, 3.0mm and 4.0mm.

3.2 Landmark localization results

The evaluation of the landmark localization is performed on the 250 test images which have not been used during model training or development. Table 1 shows the localization results for each individual landmark in terms of mean radial errors with standard deviation and successful detection rate in four categories. Fig. 2 shows exemplary results of the prediction step (left) and final landmark locations (right). The prediction step result includes the predicted locations of landmarks (blue) as well as the detected outline of skull and skin.

4 Discussion

The achieved localization results vary between landmarks. Landmarks 1 and 8 have the highest detection rate of around 89% in the 2.0mm accuracy range, whereas landmark 10 is only at 30.8%. This landmark suffers greatly from the fact that

Tab. 2: Comparison between different methods on the ISBI 2015 Challenge dataset. Successful Detection Rate (SDR) in four categories is listed for each individual approach.

Method	SDR (%)			
	2.0mm	**2.5mm**	**3.0mm**	**4.0mm**
Ibragimov et al.	68.13	74.63	79.77	86.87
Lindner and Cootes	70.65	76.93	82.17	**89.85**
Arik et al.	72.29	78.21	82.27	86.80
Payer et al.	**73.33**	78.76	83.23	89.75
Proposed	70.24	**79.14**	**83.49**	89.82

it is ill defined (cf. Fig. 1). The spatial relation - especially to the skull outline - helps to predict the landmark location very good since the mean radial error after the prediction step is among the lowest of all landmarks. However, no robust Hough Forest for the localization based on its intensity appearance could be trained, emphasizing that it is indeed ill defined. Similar problems occur for landmarks 5,6 and 18. While 5 and 6 are located on the skull contour, their location is not well defined in vertical direction resulting in inaccurate localization. This is specially evident when comparing the detection rates for different accuracy thresholds as they rise by about 30% from the 2.0mm to the 4.0mm threshold, which is significantly more than for other landmarks.

A comparison to the performance of other approaches on the 2015 Challenge dataset is shown in Tab. 2. The presented approach is slightly behind the best results in the 2.0mm accuracy range. However, it achieves the best results for 2.5mm and 3.0mm ranges and is very close for the 4.0mm range. This indicates that the landmarks prediction step correctly estimates the locations as intended and shows the potential of the coupled shape model in this context. However, the local refinement of the locations still leaves room for improvement. Fortunately, the two-step approach offers many possibilities to implement such improvements.

5 Conclusion & future work

The presented approach combines a 2-D coupled shape model for landmark prediction with a Hough Forest for the final refinement. The coupled shape model exploits spatial relation between landmarks as well as outline of skull and skin to narrow the potential search space for individual landmarks locations. The Hough Forest utilizes the local intensity appearance to detect the final landmark location. On a test set of 250 images the approach achieves a competitive performance with a successful detection rate of 70.24% in the clinically relevant 2mm accuracy range.

Future work includes the substitution of the Hough Forest with a more advanced method for landmark refinement to better localize difficult to detect landmarks, improving the robustness of the landmark position prediction to further narrow the local search area and extending the coupled shape model framework such that single point structures (landmarks) can help guide the process of adapting the model to the image. Furthermore, we plan to integrate the teeth segmentation from [10] to enable the detection of additional dental landmarks.

Author Statement

References

[1] Arik, Sercan Ö., Ibragimov, Bulat, and Xing, Lei. "Fully automated quantitative cephalometry using convolutional neural networks." In: *Journal of Medical Imaging* 4.1 (2017).

[2] Gall, Juergen et al. "Hough Forests for Object Detection, Tracking and Action Recognition." In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33.11 (2011).

[3] Lindner, Claudia et al. "Robust and Accurate Shape Model Matching Using Random Forest Regression-Voting." In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 37.9 (2015).

[4] Payer, Christian et al. "Integrating spatial configuration into heatmap regression based CNNs for landmark localization." In: *Medical Image Analysis* 54 (2019).

[5] Ricketts, Robert M. et al. "Orthodontic diagnosis and planning." In: *Denver: Rocky Mountain Data Systems* 1 (1982).

[6] Steger, Sebastian, Jung, Florian, and Wesarg, Stefan. "Personalized articulated atlas with a dynamic adaptation strategy for bone segmentation in CT or CT/MR head and neck images." In: *Medical Imaging 2014: Image Processing.* Vol. 9034. SPIE, 2014.

[7] Steiner, Cecil C. "Cephalometrics for you and me." In: *American journal of orthodontics and dentofacial orthopedics* 39.10 (1953).

[8] Wang, Ching-Wei et al. "A benchmark for comparison of dental radiography analysis algorithms." In: *Medical Image Analysis* 31 (2016).

[9] Wirtz, Andreas, Mirashi, Sudesh Ganapati, and Wesarg, Stefan. "Automatic Teeth Segmentation in Panoramic X-Ray Images Using a Coupled Shape Model in Combination with a Neural Network." In: *Medical Image Computing and Computer Assisted Intervention.* Springer, 2018.

[10] Wirtz, Andreas, Wambach, Johannes, and Wesarg, Stefan. "Automatic Teeth Segmentation in Cephalometric X-Ray Images Using a Coupled Shape Model." In: *OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis.* Springer, 2018.