# Automatic Cephalometric X-Ray Landmark Detection by Applying Game Theory and Random Forests

Bulat Ibragimov, Boštjan Likar, Franjo Pernuš, and Tomaž Vrtovec

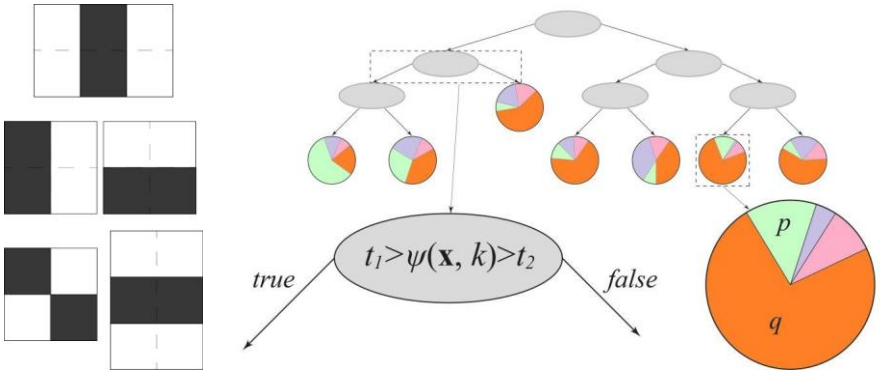
University of Ljubljana, Faculty of Electrical Engineering,
Tržaška cesta 25, SI-1000 Ljubljana, Slovenia

**Abstract.** In this paper, we propose a novel framework for cephalometric X-ray landmark detection. The framework describes the appearance of each landmark by Haar-like features, and uses the theory of random forests (RFs) to combine them into a detector of landmark candidate points in the target image. The framework relies also on spatial relationships among landmarks, modeled by Gaussian kernel density estimation and RFs, and uses concepts from the game theory to determine the optimal landmark candidate points. After training the framework on 100 cephalometric X-ray images, it was validated on a different set of 100 images, and the obtained landmark detection accuracy was 1.82 ± 1.61 mm in terms of mean radial error and 72.7 % in terms of successful detection rate (within 2 mm neighborhood). The results indicate that the proposed framework can be used for computerized cephalometry, moreover, the framework is universal and can be therefore applied to images of various anatomical structures acquired by different modalities.

## 1 Introduction

Cephalometry is a clinical procedure for measuring morphometric parameters of the human head, usually performed from lateral head-and-neck X-ray images. The morphometric parameters are mainly represented by distances and angles among anatomically relevant points, i.e. landmarks. An experienced clinician spends up to 20 minutes to manually place landmarks in one image [1], which makes the procedure time-consuming and subjected to errors. These limitations can be overcome by computerized cephalometric landmarking, which can be used for orthodontic treatment planning or statistical analysis of large image data. At early stages of computerized cephalometry, intensity- and gradient-based methods were used to detect landmarks [2, 3], or anchor structures and lines around landmarks [4]. However, these methods often generated landmark outliers, as they relied strongly on the quality of cephalometric X-ray images, and the results were therefore affected by image artifacts, shadows and differences in the projective displacement of imaged structures. The detection robustness was increased by developing more sophisticated appearance features, restricting search areas for each landmark and/or considering spatial relationships among landmarks (i.e. the shape of the observed structures). For example, anchor boundaries [5], auxiliary landmarks with well-defined intensity appearance [6], and

adfa, p. 1, 2014.




(a) (b)

**Figure 1.** (a) An illustration of different types of Haar-like features. (b) An example of a decision tree with the maximal depth of four. Each split node (gray ellipse) checks if the value $\psi(k, \mathbf{x})$ of the $k$-th feature from set **x** is between thresholds $t_1$ and $t_2$. Each terminal node (pie diagram) is a predictor that stores the posterior probabilities for landmarks $p, q, \ldots \in \mathcal{P}$ by taking into account feature set **x.**

manually or automatically identified regions of interest [7, 8] proved to considerably increase the overall detection robustness. On the other hand, the shape information was incorporated through rigid template registration [1] or active shape and active appearance models [8, 9].

In this paper, we propose a novel framework for landmark detection in cephalometric X-ray images. The framework describes the intensity appearance of each landmark by Haar-like features, applies the theory of random forests (RFs) to combine these features into a landmark candidate point detector, and models the spatial relationships among landmark candidate points by Gaussian kernel density estimation and RFs. The landmark intensity appearance and spatial relationships are finally used to optimize the position of landmark candidate points by applying concepts from game theory.

## 2 Methodology

### 2.1 Appearance likelihood maps

Let the training set $T$ of two-dimensional cephalometric X-ray images be annotated by corresponding landmarks $p \in \mathcal{P}$. To detect the same landmarks in an unknown target image, the intensity appearance information that describes landmarks has to be efficiently extracted. However, local appearance information in the form of individual pixel intensities is not descriptive enough, as images can be of different quality, subjected to local intensity variability and/or contain projective shadows. A more robust description can be obtained by Haar-like features, which describe the appearance information by intensities within rectangular areas, and can be rapidly computed from the integral image representation [10] (Fig. 1a). By extracting Haar-like features from the neighborhood of landmark $p$ in images from the training set $T$, consistent patterns

among the extracted features can be learned and used to detect the same landmark $p$ in the target image.

Recent studies [11, 12] showed that RFs efficiently capture consistent image patterns, and can be therefore used to accurately and robustly detect landmarks for a large variety of objects of interest. A RF is a collection of decision trees, i.e. connected hierarchical graphs with no loops, where split nodes perform simple classification tests and terminal nodes serve as predictors. A data point, represented by a set of features, arrives at the initial split node of a decision tree and passes to a terminal node according to the results of binary classification tests at internal split nodes. Each terminal node stores the posterior probabilities of the data point belonging to specific classes (Fig. 1b). The resulting classification is therefore the sum of the posterior probabilities for all trees in the forest, which forms the appearance likelihood map $f_p$ for each landmark $p \in \mathcal{P}$:

$$f_p(v) = F_p(\mathbf{x}_v) = T_p^1(\mathbf{x}_v) + T_p^2(\mathbf{x}_v) + \cdots + T_p^n(\mathbf{x}_v), \tag{1}$$

where $\mathbf{x}_v$ is a set of Haar-like features, extracted around the observed pixel $v$ in the target image, and $F_p$ is a RF of $n$ decision trees $T_p^i$; $i = 1, 2, \ldots, n$, defined for landmark $p$ in images from the training set. If a pixel $v$ in the target image is assigned a high appearance likelihood $f_p(v)$, then it is similar in appearance to landmark $p$ in images from the training set. Candidate points $\mathcal{S}_p = \{ s_{p1}, s_{p2}, \ldots \}$ for landmark $p$ are therefore located at the local maxima of the appearance likelihood map $f_p(v)$.

### 2.2 Shape likelihood maps

Spatial relationships among landmarks are important for object description and, when combined with appearance information, increase the accuracy of landmark detection [13]. We model the spatial relationships between any two landmarks $p, q \in \mathcal{P}$ by the Gaussian kernel density estimation of the distance $d$ and angle $\varphi$ between the two landmarks, resulting in the distance feature map $D_{p,q}$ and angle feature map $\Phi_{p,q}$:

$$D_{p,q}(d) = \sum_{i \in T} \frac{1}{\sigma_D} exp\left( - \frac{\left( d_{p,q}(i) - d \right)^2}{2{\sigma_D}^2} \right), \tag{2}$$

$$\Phi_{p,q}(\varphi) = \sum_{i \in T} \frac{1}{\sigma_\Phi} exp\left( - \frac{\left( \varphi_{p,q}(i) - \varphi \right)^2}{2{\sigma_\Phi}^2} \right), \tag{3}$$

where $d$ and $\varphi$ are, respectively, arbitrary distances and angles, $d_{p,q}(i)$ and $\varphi_{p,q}(i)$ are, respectively, the distance and angle between landmarks $p$ and $q$ in the $i$-th image from the training set $T$, and $\sigma_D$ and $\sigma_\Phi$ are constants that tune the sensitivity of feature maps to changes in the distance or angle, respectively, between landmarks. The obtained feature maps are used to compute the shape likelihood map $g_{p,q}(d, \varphi, \Delta, \Theta)$:

$$g_{p,q}(d, \varphi, \Delta, \Theta) = \lambda D_{p,q}(\Delta d) \ + (1 - \lambda) \Phi_{p,q}(\varphi + \Theta), \tag{4}$$

where $\lambda$ is a weighting factor; $0 \leq \lambda \leq 1$, while $\Delta$ and $\Theta$ additionally scale and rotate, respectively, the system of landmarks $p$ and $q$ in order to compensate for the slight scaling or rotation of the observed structures among images. If a distance $d$ and an angle $\varphi$ in the target image are assigned a high shape likelihood $g_{p,q}(d, \varphi, \Delta, \Theta)$, then they are similar to the distance and angle between landmarks $p$ and $q$ in images from the training set.

### 2.3 Game-theoretic landmark detection

The game-theoretic framework (GTF) for the determination of optimal landmark candidate points [13] considers landmarks $p \in \mathcal{P}$ as players, landmark candidate points $s_p \in \mathcal{S}_p$ (located at the local maxima of appearance likelihood maps $f_p$) as player strategies, and likelihoods that landmarks are represented by specific candidate points as player payoffs, which depend on appearance likelihood maps $f_p$ (Eq. 1) and shape likelihood maps $g_{p,q}$ (Eq. 4). For each landmark $p$ and its set of candidate points $\mathcal{S}_p$, partial payoffs $W_{p,q}$ are evaluated against each other landmark $q \in \mathcal{P} \setminus \{p\}$ and its set of candidate points $\mathcal{S}_q$:

$$W_{p,q}(s_p, s_q, \Delta, \Theta) = (1 - \tau)\, f_p\left(v_{s_p}\right) + \tau\, g_{p,q}\left(d_{s_p,s_q}, \varphi_{s_p,s_q}, \Delta, \Theta\right), \tag{5}$$

where $v_{s_p}$ is the position of candidate point $s_p \in \mathcal{S}_p$ for landmark $p$ in the target image, $d_{s_p,s_q}$ and $\varphi_{s_p,s_q}$ are the distance and angle, respectively, between candidate points $s_p \in \mathcal{S}_p$ and $s_q \in \mathcal{S}_q$, and $\tau$ is a weighting factor; $0 \leq \tau \leq 1$. The optimal landmark candidate points are found by maximizing the total payoff for all landmarks, which, in terms of game theory, corresponds to the solution of the cooperative game with the grand coalition [13]:

$$\vartheta^*(\mathcal{P}, \mathcal{S}, \mathcal{W}) = \arg\max_{\omega}\left(\sum_{p \in \mathcal{P}} \sum_{q \in \mathcal{P} \setminus \{p\}} 1^*_{p,q} \cdot W_{p,q}(s_p, s_q)\right), \tag{6}$$

where $\mathcal{S} = \mathcal{S}_p \cup \mathcal{S}_q \cup \ldots$ is the set of candidate points for all landmarks $p \in \mathcal{P}$, and $\mathcal{W}$ is the set of partial payoffs for every pair of landmarks; $\mathcal{W} = \{W_{p,q}; \forall\, p, q \in \mathcal{P}, p \neq q\}$. The indicator function $1^*_{p,q}$ is related to the shape representation and defines whether the connection between landmarks $p$ and $q$ is established ($1^*_{p,q} = 1$) or not ($1^*_{p,q} = 0$). It was showed that by replacing the complete graph of connections (i.e. $1^*_{p,q} = 1$ for every pair of landmarks) with an optimal sparse graph of connections (i.e. $1^*_{p,q} = 0$ for selected pairs of landmarks, such as the graphical lasso-based graph [14] or optimal transportation-based graph [15]), not only the computational complexity and memory requirements can be reduced (i.e. if $1^*_{p,q} = 0$, the corresponding $W_{p,q}$ and $g_{p,q}$ do not need to be computed), but also the performance of landmark detection can be improved. Among existing graph-based shape representations we selected the optimal assignment-based graph with landmark clustering, which demonstrated the best performance for segmentation of vertebrae and femoral heads from computed tomography images [15].

Although Equation 4 represents a non-deterministic polynomial-time hard (NP-hard) problem, we apply an optimization algorithm that is able to find a set of locally optimal candidate points in polynomial time [13]. We additionally re-evaluate each locally optimal candidate point for landmark $p$ by considering the spatial domain $\mathcal{Q}_p$ that encompasses all candidate points $s_p \in \mathcal{S}_p$ in the target image. The re-evaluated optimal candidate points $\sigma^* = \{s_p^*, s_q^*, ...\}$ represent the optimal position of landmarks $\mathcal{P} = \{p, q, ...\}$ in the target image, as defined by GTF.

### 2.4 Shape-based refinement of landmark candidate points

To determine the optimal landmark candidate points, GTF combines Haar-like appearance features of individual landmarks with spatial relationships of pairs of landmarks. Although such approach may be computationally efficient, it does not fully exploit the multi-landmark environment, where landmarks in a subset can jointly predict the position of landmarks that are not included in the subset. To incorporate such a multi-landmark environment, the position of the optimal candidate point $s_p^*$ for landmark $p$ can be approximated from positions of optimal candidate points for landmarks from a subset $U_p \subset \sigma^* \backslash \{s_p^*\}$:

$$R_{p,U_p}(v) = \bar{F}_p\left(v, d_{p,q_1}^x, d_{p,q_1}^y, \dots d_{p,q_i}^x, d_{p,q_i}^y, \dots\right), \tag{7}$$

where $v$ is the observed pixel in the target image, $\bar{F}_p$ is a RF that models the spatial dependence between point $s_p^*$ and points from subset $U_p$ in images from the training set, and $d_{p,q_i}^x$ and $d_{p,q_i}^y$ are, respectively, horizontal and vertical distances between points $s_p^*$ and $s_{q_i}^* \in U_p$. The resulting RF-based shape likelihood map $R_{p,U_p}$ is generated within the spatial domain $\mathcal{Q}_p$ that encompasses all candidate points $s_p \in \mathcal{S}_p$ in the target image, and is used to refine the position of the optimal candidate point $s_p^* \in \sigma^*$ defined by GTF:

$$s_p^{**} = \arg\max_{v \in \mathcal{Q}_p} \left( R_{p,U_p}(v) \cdot f_p(v) \right). \tag{8}$$

In an iterative procedure, the refined point $s_p^{**}$ is included into subset $U_p$; $U_p \leftarrow U_p \cup \{s_p^{**}\}$, and used to refine the next optimal candidate point from $\sigma^*$. The order of the refinement is defined from the detection performance of GTF for each landmark $p \in \mathcal{P}$, i.e. the landmarks with low detection errors are analyzed first and used to support the detection of landmarks with high detection errors. At the first steps of the refinement, when the size of subset $U_p$ is small and therefore not representative enough, five candidate points that correspond to landmarks with the lowest predicted detection error are iteratively refined according to the leave-one-out principle. The refined optimal candidate points $\sigma^{**} = \{s_p^{**}, s_q^{**}, ...\}$ represent the final position of landmarks $\mathcal{P} = \{p, q, ...\}$ in the target image, as defined by GTF and shape-based refinement.

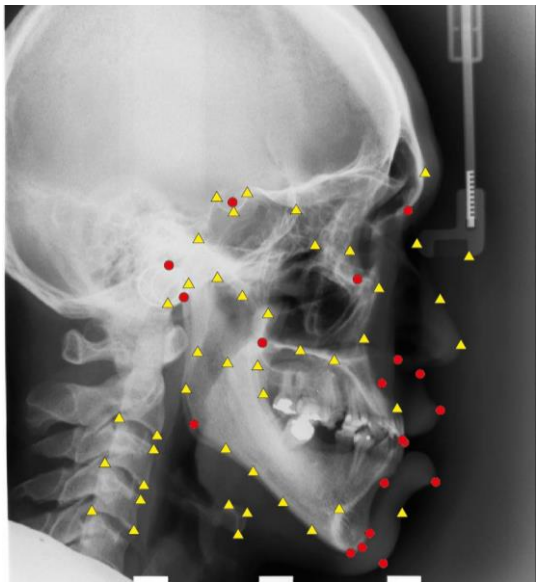


**Figure 2**. A cephalometric X-ray image with superimposed anatomical landmarks (red circles) and auxiliary landmarks (yellow triangles).

## 3 Results

The image dataset of 200 cephalometric X-ray images (size: 1935 × 2400 pixels, pixel size: 0.1 × 0.1 mm) was divided into training and test sets, each containing 100 images. In each image, a clinical expert manually placed 19 anatomical landmarks, while we manually placed an additional of 44 auxiliary landmarks at visually distinctive locations such as the nose tip, corners of cervical vertebrae, etc. (Fig. 2). Auxiliary landmarks do not necessarily mark anatomically relevant points but serve to assist the detection of anatomical landmarks.

All images were first resampled to a resolution of 0.3 × 0.3 mm to keep the set of appearance features more compact and representative, and to speed up the landmark detection process. To generate appearance likelihood maps, five types of Haar-like features at 4-, 8- and 16-pixel large scales (Fig. 1a) were computed at 121 pixels of the $21^2$-pixels large sparse neighborhood of landmark $p$ in the 100 images from the training set. For each appearance likelihood map and each RF-based shape likelihood map, RFs consisted of 40 trees with the maximal depth of 10, i.e. a path from any initial split node to any terminal node contained up to nine internal split nodes. The remaining parameters of the framework can be found in our previous publications [13, 15].

The obtained landmark detection results for the 100 images in the test set are summarized in Table 1. The results are given in terms of mean radial error (MRE), i.e. the mean Euclidean distance between the obtained landmarks $\sigma^{**} = \{s_p^{**}, s_q^{**}, \dots\}$ and the corresponding manually placed anatomical landmarks $\mathcal{P} = \{p, q, \dots\}$, and in terms of

**Table 1.** Landmark detection results in terms of mean radial error (MRE, ± standard deviation) and successful detection rate (SDR) within 2.0, 2.5, 3.0, and 4.0 mm neighborhoods.

| Landmark name | MRE (mm) | SDR (%) | | | |
|---|---|---|---|---|---|
| | | 2.0 mm | 2.5 mm | 3.0 mm | 4.0 mm |
| Sella turcica | 0.95 ± 1.36 | 96.0 | 96.0 | 96.0 | 96.0 |
| Nasion | 1.55 ± 1.49 | 74.0 | 79.0 | 82.0 | 92.0 |
| Orbitale | 1.85 ± 1.68 | 70.0 | 81.0 | 83.0 | 89.0 |
| Porion | 4.98 ± 2.76 | 20.0 | 25.0 | 30.0 | 37.0 |
| Subspinale | 2.63 ± 2.17 | 54.0 | 62.0 | 69.0 | 82.0 |
| Supramentale | 2.28 ± 1.99 | 58.0 | 71.0 | 78.0 | 85.0 |
| Pogonion | 1.27 ± 0.98 | 75.0 | 90.0 | 95.0 | 98.0 |
| Menton | 1.08 ± 0.70 | 88.0 | 96.0 | 99.0 | 100.0 |
| Gnathion | 0.95 ± 0.73 | 92.0 | 95.0 | 98.0 | 100.0 |
| Gonion | 3.59 ± 2.75 | 33.0 | 42.0 | 49.0 | 61.0 |
| Lower incisal incision | 1.27 ± 1.58 | 84.0 | 87.0 | 88.0 | 93.0 |
| Upper incisal incision | 0.94 ± 1.80 | 91.0 | 91.0 | 93.0 | 96.0 |
| Upper lip | 0.95 ± 0.61 | 92.0 | 98.0 | 100.0 | 100.0 |
| Lower lip | 0.88 ± 0.65 | 94.0 | 97.0 | 99.0 | 100.0 |
| Point Pm or MN | 1.27 ± 1.37 | 85.0 | 87.0 | 92.0 | 96.0 |
| Soft tissue pogonion | 2.19 ± 1.93 | 67.0 | 72.0 | 73.0 | 79.0 |
| Posterior nasal spine | 1.15 ± 0.86 | 83.0 | 88.0 | 94.0 | 100.0 |
| Anterior nasal spine | 2.15 ± 2.00 | 62.0 | 72.0 | 81.0 | 86.0 |
| Articulate | 2.62 ± 3.23 | 64.0 | 68.0 | 72.0 | 76.0 |
| **Total** | **1.82 ± 1.61** | **72.7** | **78.8** | **82.7** | **87.7** |

successful detection rate (SDR), i.e. the percentage of the obtained landmarks $\sigma^{**}$ located within predefined neighborhoods of the corresponding manually placed anatomical landmarks $\mathcal{P}$.

## 4 Conclusion

We presented a novel framework for landmark detection, which combines landmark appearance information, described by Haar-like features and modeled by RFs, and landmark shape information, described by both Gaussian kernel density estimation and a novel RF-based shape model. The framework was applied to automatically detect 19 anatomical landmarks from cephalometric X-ray images, and the obtained MRE of 1.82 ± 1.61 mm and SDR of 72.7% (within 2 mm neighborhood) indicate that the proposed framework may be used for computerized cephalometry. Moreover, the described landmark detection framework is universal and can be therefore straightforwardly applied to images of various anatomical structures acquired by different imaging modalities.

## Acknowledgements

This work was supported by the Slovenian Research Agency, under grants P2-0232 and L2-4072.

## References

1. El-Fegh, I., Galhood, M., Sid-Ahmed, M., Ahmadi, M.: Automated 2-D cephalometric analysis of X-ray by image registration approach based on least square approximator. In: 30th International Conference of the IEEE Engineering in Medicine and Biology Society - EMBS 2008. 3949–3952 (2008).
2. Cardillo, J., Sid-Ahmed, M.A.: An image processing system for locating craniofacial landmarks. IEEE Trans. Med. Imaging 13, 275–289 (1994).
3. Forsyth, D.B., Davis, D.N.: Assessment of an automated cephalometric analysis system. Eur. J. Orthod. 18, 471–478 (1996).
4. Lévy-Mandel, A.D., Venetsanopoulos, A.N., Tsotsos, J.K.: Knowledge-based landmarking of cephalograms. Comput. Biomed. Res. Int. J. 19, 282–309 (1986).
5. Grau, V., Alcañiz, M., Juan, M.C., Monserrat, C., Knoll, C.: Automatic localization of cephalometric landmarks. J. Biomed. Inform. 34, 146–156 (2001).
6. Mohseni, H., Kasaei, S.: Automatic localization of cephalometric landmarks. In: 7th IEEE International Symposium on Signal Processing and Information Technology - ISSPIT 2007. 396–401 (2007).
7. Ciesielski, V., Innes, A., John, S., Mamutil, J.: Genetic programming for landmark detection in cephalometric radiology images. Int. J. Knowl.-Based Intell. Eng. Syst. 7, 164–171 (2003).
8. Kaur, A., Singh, C.: Automatic cephalometric landmark detection using Zernike moments and template matching. Signal Image Video Process., in press (2014).
9. 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 Trans. Biomed. Eng. 53, 1615–1623 (2006).
10. Viola, P., Jones, M.: Robust real-time face detection. Int. J. Comput. Vis. 57, 137–154 (2004).
11. Donner, R., Menze, B.H., Bischof, H., Langs, G.: Global localization of 3D anatomical structures by pre-filtered Hough forests and discrete optimization. Med. Image Anal. 17, 1304–1314 (2013).
12. Criminisi, A., Shotton, J.: Decision forests for computer vision and medical image analysis. Springer (2013).
13. Ibragimov, B., Likar, B., Pernuš, F., Vrtovec, T.: A game-theoretic framework for landmark-based image segmentation. IEEE Trans. Med. Imaging 31, 1761 –1776 (2012).
14. Friedman, J., Hastie, T., Tibshirani, R.: Sparse inverse covariance estimation with the graphical lasso. Biostat. 9, 432–441 (2008).
15. Ibragimov, B., Likar, B., Pernus, F., Vrtovec, T.: Shape representation for efficient landmark-based segmentation in 3D. IEEE Trans. Med. Imaging, in press (2014).