

# An Approach for the Automatic Cephalometric Landmark Detection Using Mathematical Morphology and Active Appearance Models\*

Sylvia Rueda and Mariano Alcañiz

Medical Image Computing Laboratory (MedICLab), Universidad Politécnica de Valencia, UPV/ETSIA, Camino de Vera s/n, 46022 Valencia, Spain  
silruelo@degi.upv.es, malcaniz@degi.upv.es

**Abstract.** Cephalometric analysis of lateral radiographs of the head is an important diagnosis tool in orthodontics. Based on manually locating specific landmarks, it is a tedious, time-consuming and error prone task. In this paper, we propose an automated system based on the use of Active Appearance Models (AAMs). Special attention has been paid to clinical validation of our method since previous work in this field used few images, was tested in the training set and/or did not take into account the variability of the images. In this research, a top-hat transformation was used to correct the intensity inhomogeneity of the radiographs generating a consistent training set that overcomes the above described drawbacks. The AAM was trained using 96 hand-annotated images and tested with a leave-one-out scheme obtaining an average accuracy of 2.48mm. Results show that AAM combined with mathematical morphology is the suitable method for clinical cephalometric applications.

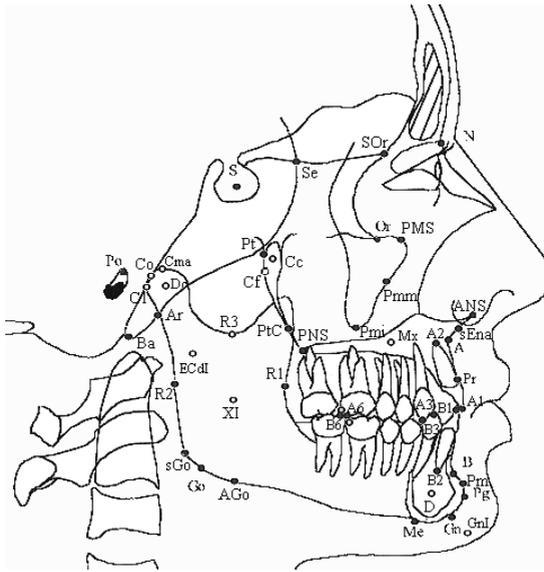
## 1 Introduction

Cephalometry means measurement of the head. Accordingly, a cephalometric analysis consists of characterizing distances and angles between significant structures in a x-ray image of the head. The whole process is based on the localization of cephalometric landmarks. These points correspond to specific locations on the radiograph and are accurately defined for hard, external soft and pharyngeal tissues [1]. Figure 1 shows the location of commonly chosen cephalometric points for several methods. Once located, lines are traced between the points to measure angles and distances, which are the cephalometric indices. After a comparison of these measurements with the norms stated for the same age, ethnicity and sex, a diagnosis can be established and a treatment forecast.

Manual positioning of landmarks is a tedious, time-consuming and error prone task. Depending on the quality of the radiograph and the experience of the user, the location will be slightly different. According to Dean et al. [2] and Geelem et al. [3], inter-expert variability during the positioning task varies from 3.3mm to 5mm and intra-expert variability is of 1mm.

---

\* Radiographs provided by Y. Perrin, A. Sanpietro and the Eastman Dental Institute.



**Fig. 1.** Lateral cephalometric points of bony tissue for different methods of analysis

The orthodontist spends on average 15-20 minutes per analysis depending on the quality of the cephalogram, his/her experience and the number of points considered in the method selected. A fully automated system would detect landmarks, thus reducing the time required to carry out an analysis, improving the accuracy of landmark localization and decreasing errors due to expert subjectivity.

## 2 Previous Work

Several approaches have been proposed for the automatic detection of cephalometric landmarks. They can be classified into 3 groups:

**Knowledge-based Systems.** The first attempt was undertaken by Levy et al. [4], who applied enhancement filters and a knowledge-based line tracker to extract edges. Landmarks were located according to their geometric definitions. The method was tested on two high-quality x-ray images. 23 of 36 landmarks were located on good quality images. Parthasaraty et al. [5] improved the previous work by including a four level resolution pyramid to reduce time-processing. The testing was done in five images. Nine of the landmarks considered in that study, 58% were within  $\pm 2\text{mm}$ , 18% within  $\pm 1\text{mm}$  and 100% within  $\pm 5\text{mm}$ . These studies used few images for training and tested the method on the training set. In the presence of artifacts and bad quality images, these methods will not work and, therefore, cannot be used for clinical cephalometric systems.

**Template Matching.** Template matching methods implement a grey-level model around each landmark on each image in the training set. Then, they search

for a correspondence between the model and new images. Cardillo et al. [6] applied a pattern-matching algorithm based on mathematical morphology to detect landmarks. They considered 40 images in the training set. The testing was done on 20 images. Of 20 landmarks, 66% were within  $\pm 2\text{mm}$ . Grau et al. [7] improved on the work of Cardillo et al. by using a line detection module to search for the most significant lines. They used 20 images for training and 20 for testing and reported that 90% of the 17 landmarks tested were repeatedly located within  $\pm 2\text{mm}$ . Chen et al. [8] developed a method of neural networks and genetic algorithms to search for sub-images containing each of the cephalometric points. These methods consider the texture around each point. However, this appearance can vary from one image to another making local detection worse. Moreover, not enough images were used to represent the variability existing in cephalograms. In our experience, these methods, when used on images slightly different from the ones of the training set, are not sufficiently accurate for clinical applications.

**Statistical Models.** Statistical methods are one of the most suitable options for considering the huge amount of variability in cephalograms, as they take into account the variation of characteristics in the images. Hutton et al. [9] applied Active Shape Models (ASMs) to cephalometric point detection. They used 63 randomly selected cephalograms and tested the training set with a drop-one-out method. 13% of 16 landmarks were within  $\pm 1\text{mm}$ , 35% within  $\pm 2\text{mm}$  and 74% within  $\pm 5\text{mm}$ . The authors concluded that ASM did not give sufficient accuracy for landmark detection in clinical applications and could only be used as a good starting point for global landmark identification. ASM mainly considers the variation in shape present on the images. The intensity variations are only modeled in the profiles normal to the contour. Accordingly, to improve accuracy it is necessary to add more information to the system. The appearance of the radiographs contains useful information that should be taken into account.

### 3 Method

Active Appearance Models (AAMs), recently proposed by Cootes et al. [10][11], can model both shape and texture variability existing in a training set. To create a model we need a set of labeled images representative of the real variability of the object we want to segment. AAMs have proved to be powerful tools for medical image segmentation and understanding [12]. The more homogeneous the images of the training set are, the more robust the model will be. The problem of the images we work with in this research is that they can be greatly different and present huge luminosity inhomogeneity. Mathematical morphology can solve this; in particular, we will apply a top-hat transformation [13] on the images to extract light objects from an uneven background.

#### 3.1 Selection of the Training Set

The images forming the training set must be representative of the variability existing in clinical cases. The data and the source should reflect the target

population for the algorithm that is being developed. In particular, it is necessary to consider: the anatomical and morphological variability of the human head, the variability of texture depending on image quality, the variability of structures present in a cephalogram (not all the radiographs have the same size and include the same structures), the variability of capturing the x-ray (i.e. double structures appearing because of the non-orthogonal position of the patient during capture of a lateral head film) and the variability of the source.

### 3.2 Luminosity Inhomogeneity Correction

Depending on the source, the cephalogram can present varied textures and thus be completely different from one another. Background cannot be separated from the rest due to its non-uniformity. Therefore, a purely preprocessing approach for intensity inhomogeneity correction is required to improve results when using AAM. Mathematical morphology [14], which is based on set theory, provides powerful tools for image analysis. Fundamental operations are erosion, dilation, opening and closing. An opening consists of an erosion followed by a dilation. A structuring element defines the size and the shape of the transformation to be done. In our case we will use a circular structuring element with a radius of 75 pixels. We use this size to extract the structures of interest correctly. The white top-hat  $\rho(f)$  extracts bright structures and is defined as the difference between the image  $f$  and its opening  $\gamma(f)$ :

$$\rho(f) = f - \gamma(f) \quad (1)$$

Subtracting the open image from the original provides an image where removed objects stand out clearly. This technique is used to extract contrasted components with respect to the background. The result of applying this transformation can be seen in Fig. 2 (right column).

### 3.3 Training the Model

To generate a statistical model of shape and texture variation it is necessary to describe the shape and the texture of each training example. Therefore, we represent  $n$  landmark points,  $(x_i, y_i)$ , for each image as a  $2n$  vector,  $\mathbf{x}$ , where  $\mathbf{x} = (x_1, \dots, x_n, y_1, \dots, y_n)^T$  will describe the shape of an object. The annotated training set is aligned into a common co-ordinate frame using a Procrustes Analysis [15]. Hence, we obtain the Point Distribution Model (PDM) for all the images of the training set. The mean shape is extracted and the appearance variation collected by establishing a piece-wise affine warp (based on the Delaunay triangulation) between each image of the training set and the mean shape. Next, the intensity is sampled from the shape-normalized images over the region covered by the mean shape. The resulting samples are then normalized to minimize the effect of global lighting variation and the texture (grey-levels) vector  $\mathbf{g}$  is obtained. Finally, for both shape and texture, we perform a Principal Component Analysis (PCA) on the aligned training set to describe the shape

and appearance variations of the object. Often, shape and texture are correlated. Therefore, we can deduce a combined model of shape and texture with parameters  $\mathbf{c}$ , controlling the shape and texture at the same time. We get:

$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{Q}_s \mathbf{c}, \mathbf{g} = \bar{\mathbf{g}} + \mathbf{Q}_g \mathbf{c} \quad (2)$$

where  $\bar{\mathbf{x}}$  and  $\bar{\mathbf{g}}$  are the mean shape vector and the mean normalized grey-level vector,  $\mathbf{Q}_s$  and  $\mathbf{Q}_g$  are matrices describing the modes of variation derived from the training set and  $\mathbf{c}$  the combined PCA parameters. New images of the object can be synthetically constructed from  $\mathbf{c}$ . More details about AAMs can be found in [10][11].

### 3.4 Segmenting New Images

We first place an initial template model over the unseen image. We then use a principal component multivariate linear regression model to generate new images to fit the unseen image in the best way. Once the process converges, a match can be declared. Finally, an optimization scheme accomplishes further refinement of the match.

### 3.5 Model Evaluation

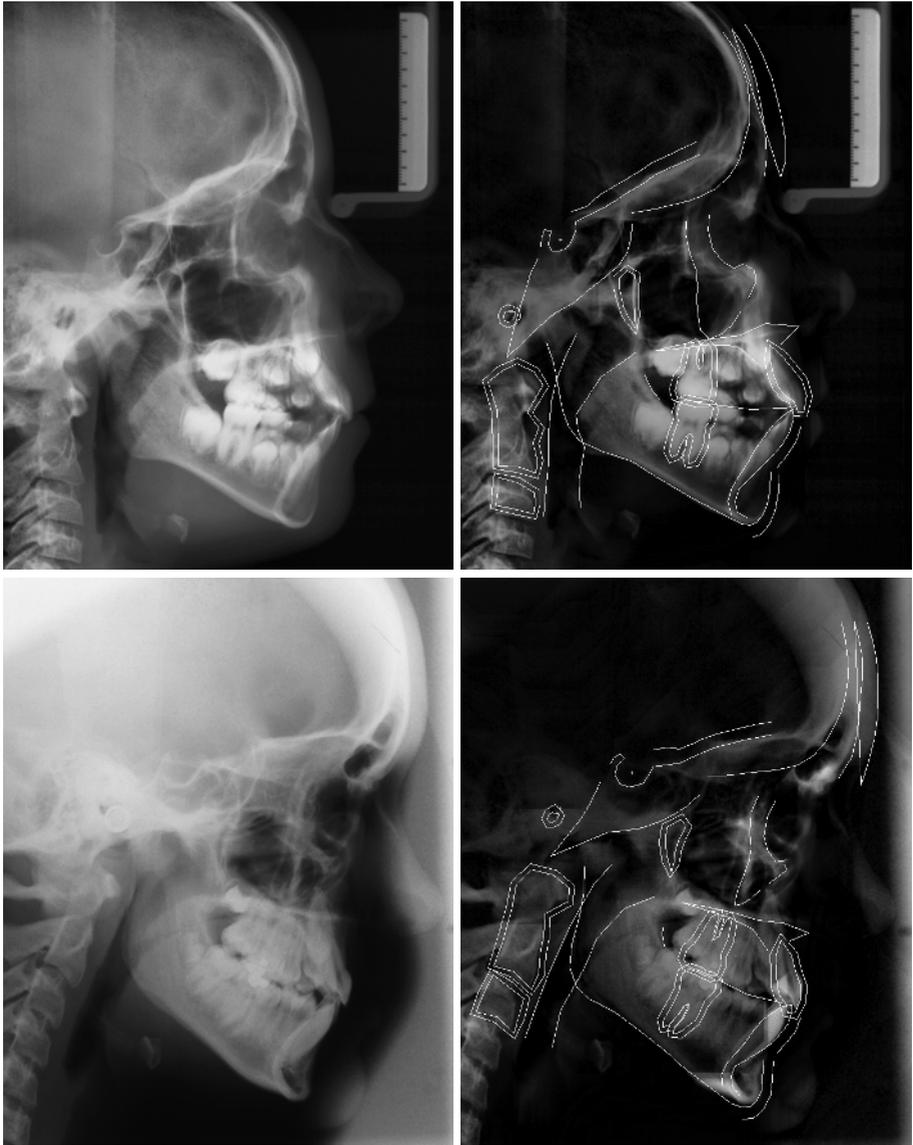
The final step of the whole process consists of testing the performance of the algorithm. Care should be taken to test the algorithm in data that has not been used in the training.

In this case, we will use a common methodology called leave-one-out. In this approach, a set of  $M$  images (ground truth) is split in  $M$  different ways into a training set of size  $M-1$  and a test set of size 1. For each of the  $M$  splits, training is done on the  $M-1$  images and then testing is done on the remaining 1 item. Performance is then estimated as the average of the  $M$  tests.

## 4 Experimental Results

We applied the method described to automatically detect cephalometric landmarks. The open C++ source code AAM-API [17] was partially used in this study. We implemented the leave-one-out algorithm in Matlab to evaluate the model.

96 images, annotated by an expert, form the training set used in this study. Image selection fulfils the requirements stated in Sect. 3.1. This number of images is sufficient to include all the existing variability. Each image has a size of  $780 \times 1000$  pixels and 256 grey-levels. 43 cephalometric landmarks were considered to assess most of the cephalometric analysis. We used a combination of scanned and digital x-ray images, coming from different sources, to consider different image qualities in the training set. Anatomical variability is assessed with good proportion between ages, different norm groups (Caucasian, black and oriental), incisor relationship (I,II or III) and dentition (mixed or permanent).



**Fig. 2.** Results of the search in a digital image (first row) and a scanned x-ray (second row) after luminosity inhomogeneity correction. We can appreciate that the original x-rays are very different and the intensity normalized images are more homogeneous.

We have created a template with 289 points. Points were divided into cephalometric, mathematical and pseudo-landmarks [15]. Double contours were added to some structures to improve the search giving more specificity to the model [16] (Fig. 2). The evaluation of the model was assessed using a leave-one-out method. We considered the initialization of the system to have failed if more than 60% of

the points have an error larger than 5mm. In these conditions, for 96 images used in this study, the automatic initialization failed in 9.37% of the cases. Figure 2 shows the results of different quality cephalograms. The first row assesses the result in a digital x-ray whereas the second row shows the result in a scanned x-ray with high luminosity variation, black regions and rotation of the head. For both, the segmentation works well and with high precision. Table 1 assesses the accuracy of the system for the main landmarks considered. The names of the landmarks correspond to locations in Fig. 1.

**Table 1.** Results and comparisons of the main landmarks

	mean (mm)	std (mm)	≤ 2mm (%)	≤ 3mm (%)	≤ 5mm (%)		mean (mm)	std (mm)	≤ 2mm (%)	≤ 3mm (%)	≤ 5mm (%)
<b>Na</b>	2.31	1.78	56.32	75.86	91.95	<b>SOr</b>	2.16	1.33	49.42	80.46	96.55
<b>Se</b>	2.04	1.12	52.87	80.46	100	<b>S</b>	2.29	1.26	39.08	70.11	97.70
<b>Or</b>	2.05	1.23	57.47	73.56	97.70	<b>Po</b>	3.66	2.12	18.39	42.53	78.16
<b>PMS</b>	2.24	1.25	48.28	75.86	96.55	<b>Co</b>	2.47	1.35	37.93	64.37	94.25
<b>Ba</b>	2.70	1.79	37.93	67.82	91.95	<b>Ar</b>	2.27	1.41	44.83	74.71	96.55
<b>Pt</b>	2.19	1.22	47.13	78.16	98.85	<b>PtC</b>	2.14	1.37	55.17	78.16	97.70
<b>Cf</b>	2.02	1.18	54.02	85.06	98.85	<b>R1</b>	1.95	1.02	60.92	85.06	100
<b>B</b>	2.19	1.26	43.68	80.46	97.70	<b>Pm</b>	1.97	1	56.32	85.06	98.85
<b>Pg</b>	1.83	1.06	57.47	89.66	98.85	<b>Gn</b>	1.58	1.12	73.56	86.21	98.85
<b>Me</b>	1.59	1.07	70.11	86.21	100	<b>Go</b>	3.88	2.41	26.44	43.68	66.67
<b>R2</b>	2.46	1.43	41.38	71.26	91.95	<b>LIT</b>	1.52	0.90	77.01	94.25	100
<b>LIR</b>	1.55	0.92	67.82	94.25	100	<b>PNS</b>	2.67	1.42	36.78	68.97	91.95
<b>ANS</b>	2.12	1.28	55.17	78.16	96.55	<b>A</b>	2	1.41	67.82	83.91	94.25
<b>UIT</b>	1.82	1.85	75.86	87.36	94.25	<b>UIR</b>	1.98	1.43	65.52	82.76	95.40

On average, for all the cephalometric landmarks considered, the system has a precision of 2.48mm and an average standard deviation of 1.66mm. Considering that inter-expert variability varies from 3.3mm to 5mm, these results are to date the best option for clinical applications. In this study, 50.04% of the landmarks were located within 2mm, 72.62% within 3mm and 91.44% within 5mm.

## 5 Conclusions

In this paper, we have shown that AAM achieves precise, automatic detection of cephalometric landmarks that takes into account the variability existing in shape and texture. This leads to the clinical applicability of automatic landmark identification. Previous work in this field used few images, tested their methods in the training set or did not select a consistent training set that well characterized the variability of this type of images. None of the previous studies undertaken had achieved clinical cephalometric analysis. Our method has proved to be useful to locate a considerable number of the landmarks with a high precision rate for clinical cephalometric applications. Image homogeneization and the use of double contours for some structures improve AAM results when applied to cephalometric radiographs.

## References

1. Rakosi, T.: An Atlas and Manual of Cephalometric Radiology, London Wolfe Medical Publications (1982)
2. Dean, D., Palomo, M., Subramanya, K., et al.: Accuracy and precision of 3D cephalometric landmarks from biorthogonal plain-film x rays. *SPIE Med. Imag.* 3335, (1998) 50-58
3. Geelem, W., Wenzel, A., Gotfredsen, E., Kruger, M. and Hansson, L.G.: Reproducibility of cephalometric landmarks in conventional film, and hardcopy and monitor-displayed images obtained by the storage phosphor technique. *Eur. J. Orthod.* 20, (1998) 331-340
4. Lévy-Mandel, A., Venetsanopoulos, A. and Tsotsos, J.: Knowledge-based landmarking of cephalograms. *Comput. Biomed. Res.* 19, (1986) 282-309
5. Parthasaraty, S., Nugent, S., Gregson, P. et al.: Automatic landmarking of cephalograms. *Comput. Biomed. Res.* 22, (1989) 248-269
6. Cardillo, J. and Sid-Ahmed, M.A.: An image processing system for locating craniofacial landmarks. *IEEE Trans. Med. Imag.* 13, (1994) 275-289
7. Grau, V., Alcañiz, M., Juan, M.C., Monserrat, C. and Knoll, C.: Automatic Localization of Cephalometric Landmarks. *J. biomed. inform.* 34, (2001) 146-156
8. Chen, Y.T., Cheng, K.S. and Liu, J.K.: Improving Cephalogram Analysis through Feature Subimages Extraction. *IEEE Eng. in Med. Biol. Mag.* 18, (1999) 25-31
9. Hutton, T.J., Cunningham, S. and Hammond, P.: An Evaluation of Active Shape Models for the Automatic Identification of Cephalometric Landmarks. *Eur. J. Orthod.* 22, (2000) 499-508
10. Cootes, T.F., Edwards, G.J. and Taylor, C.J.: Active Appearance Models. *Proc. European Conference on Computer Vision*, vol.2 (1998) 484-498
11. Cootes, T.F. and Taylor, C.J.: Statistical Models of Appearance for Computer Vision, Report (2004)
12. Cootes, T.F. and Taylor, C.J.: Statistical models of appearance for medical image analysis and computer vision, *Proc. SPIE Med. Imag.* (2001)
13. Sonka, M. and Fitzpatrick, J.M.: Handbook of Medical Imaging, vol.2. *Medical Image Processing and Analysis*, SPIE Press (2000)
14. Sonka, M., Hlavac, V. and Boyle, R.: *Image Processing, Analysis, and Machine Vision*, Second Edition (1999) Brooks/Cole Publishing Company
15. Dryden, I.L. and Mardia, K.V.: *Statistical Shape Analysis*, Wiley (1998)
16. Stegmann, M.B.: Active Appearance Models - Theory, Extensions & Cases, Master Thesis IMM-EKS-2000-25 (2000)
17. Stegmann, M.B., Ersboll, B.K. and Larsen, R.: FAME - a flexible appearance modelling environment. *IEEE Trans. Med. Imag.* (2003)