European Journal of Orthodontics 32 (2010) 233–241
doi:10.1093/ejo/cjp099
Advance Access Publication 4 March 2010

Automatic landmarking of cephalograms using active appearance models

Predrag Vučinić*, Željen Trpovski** and Ivana Šćepan***
Departments of *Orthodontics and **Telecommunications and Signal Processing, University of Novi Sad and ***Clinic of Orthodontics, University of Belgrade, Serbia

SUMMARY There have been many attempts to further improve and automate cephalometric analysis in order to increase accuracy, reduce errors due to subjectivity, and to provide more efficient use of clinicians' time. The aim of this research was to evaluate an automated system for landmarking of cephalograms based on the use of an active appearance model (AAM) that contains a statistical model of shape and grey-level appearance of an object of interest and represents both shape and texture variations of the region covered by the model.

Multi-resolution implementation was used, in which the AAM iterate to convergence at each level before projecting the current solution to the next level of the model. The AAM system was trained using 60 randomly selected, hand-annotated digital cephalograms of subjects between 7.2 and 25.6 years of age, and tested with a leave-five-out method that enabled testing not only of the accuracy of the AAM system but also the accuracy of each AAM. Differences between methods were examined using the non-parametric Wilcoxon signed rank test.

An average accuracy of 1.68 mm was obtained, with 61 per cent of landmarks detected within 2 mm and 95 per cent of landmarks detected within 5 mm precision. A noticeable increase in overall precision and detection of low-contrast cephalometric landmarks was achieved compared with other automated systems. These results suggest that the AAM approach can adequately represent the average shape and texture variations of craniofacial structures on digital radiographs. As such it can successfully be implemented for automatic localization of cephalometric landmarks.

Introduction

It was not until introduction of the cephalostat (Broadbent, 1931; Hofrath, 1931) that cephalometric analysis revolutionized diagnostics and treatment planning in orthodontics. For the first time, it was possible to analyze not only the dentoalveolar but also the underlying skeletal characteristics of the viscero- and neurocranium. Since then, it has become a standardized diagnostic method in everyday orthodontic practice and research.

Two approaches may be used for tracing lateral cephalograms: a manual approach and a computer-aided approach. The manual approach is the oldest and still most widely used and is carried out by placing a sheet of acetate paper over the cephalometric radiograph and manually tracing skeletal and soft tissue features, identifying landmarks, and measuring distances and angles between landmark locations. Computer-aided cephalometric analysis uses manually identified landmarks, based either on transferring landmarks from cephalometric radiographs with a digitizing pad connected to a computer, or direct landmark identification with a mouse cursor on the computer monitor (Baumrind and Miller, 1980; Richardson, 1981; Turner and Weerakone, 2001). The computer software then completes the cephalometric analysis by automatically measuring distances and angles.

It is widely acknowledged that both approaches are time-consuming and prone to errors, both systematic and random, due to problems and inconsistencies in features and landmark identification, drawing lines between landmarks and measuring with a ruler and protractor (Baumrind and Frantz, 1971; Kamoen *et al.*, 2001). The latter is successfully eliminated in computer-aided cephalometric analysis, but the greatest error still lies in landmark identification (Houston *et al.*, 1986; Nimkarn and Miles, 1995; Kamoen *et al.*, 2001). Variability in landmark identification has been determined to be five times greater than measurement variability, with both methods (Miller *et al.*, 1971; Savage *et al.*, 1987). In addition, the process is open to considerable subjectivity, since landmarks currently are defined using subjective criteria rather than strict mathematical specifications.

It is generally accepted that accuracy in landmark identification ideally should be less than 0.5 mm. However, measurements with errors within 2 mm are considered acceptable, and are often used as a reference to evaluate the recognition success rate, but the former level of precision is considered desirable (Forsyth and Davis, 1996). Research studies on the accuracy of landmark identification have also shown that manual landmark identification errors vary significantly depending on the landmark, observer, and quality of the radiograph (Cohen and Linney, 1986; Parthasarathy *et al.*, 1989). In some studies, the mean estimating error of expert landmarking identification has been reported to be 1.26 mm (Parthasarathy *et al.*, 1989).

There have been many attempts to further improve and automate cephalometric analysis. Computer systems, which automatically identify relevant skeletal and soft tissue structures and landmarks have the potential to increase accuracy, reduce errors due to subjectivity, provide more efficient use of clinician time, and improve the ability to correctly diagnose orthodontic cases.

The first attempt at automated landmarking of cephalograms was made by Cohen *et al.* (1984). Since then, automatic cephalometric analysis has been the subject of a number of studies, and the automatic identification of landmarks has been attempted by more than 20 independent researchers using different approaches and computer systems, with varying degrees of success. All these methods can be divided into three categories. The first, pure edge tracking, follows a strategy similar to that employed by clinicians and uses a combination of image processing techniques to detect and extract the important edges, subsequently used to locate landmarks on line crossings (Cohen *et al.*, 1984; Cohen and Linney, 1986; Lévy-Mandel *et al.*, 1986; Parthasarathy *et al.*, 1989; Davis and Taylor, 1991; Davis, 1994; Liu *et al.*, 2000). The second category, knowledge-based template matching methods, implements a grey-level model around each landmark to reduce the search area (Cardillo and Sid-Ahmed, 1994; Ren *et al.*, 1998; Rudolph *et al.*, 1998; Desvignes *et al.*, 2000; Hutton *et al.*, 2000; Grau *et al.*, 2001; Romaniuk *et al.*, 2002). The third category employs neural networks and fuzzy inference systems to locate the landmarks (Uchino and Yamakawa, 1995; Sanei *et al.*, 1999; Innes *et al.*, 2002; Ciesielski *et al.*, 2003; El-Feghi *et al.*, 2004).

The pure edge-based approach identifies pixels near the object boundaries. Boundaries are detected as areas with a high gradient value. Landmarks are then found in relation to these boundaries. Attempts to use this approach for automatic landmark identification have both experimental design flaws and very limited results. In most cases, the method was tested on the same set of radiographs used to develop the algorithm (Lévy-Mandel *et al.*, 1986; Parthasarathy *et al.*, 1989; Davis and Taylor, 1991; Liu *et al.*, 2000). Some of the studies used a very small number of radiographs (Lévy-Mandel *et al.*, 1986; Parthasarathy *et al.*, 1989). It was also observed that tested methods only worked on high quality images (Parthasarathy *et al.*, 1989). These heuristic methods are based on *ad hoc* rules for finding each specific landmark. The main problem is that the rules become increasingly difficult as more complex landmarks, structures, and variations in image quality and contrast are introduced. This may explain the limited accuracy of these algorithms and their inability to produce a potentially clinically applicable approach.

The knowledge-based template matching approach, also known as the learning approach, uses mathematical models to narrow down the search area for each landmark, subsequently applying various pattern-matching algorithms to pinpoint the exact location of the landmark (Cardillo and Sid-Ahmed, 1994; Uchino and Yamakawa, 1995; Ren *et al.*, 1998; Rudolph *et al.*, 1998; Desvignes *et al.*, 2000; Hutton *et al.*, 2000; Grau *et al.*, 2001; Romaniuk *et al.*, 2002). These methods have proved more accurate, especially in detecting complex landmarks with less contrast characteristics. They also produce consistently better results with radiographic images of varying quality. Nevertheless, overall accuracy is still far beyond applicability in everyday clinical practice and research.

The newer generation of knowledge-based systems make use of an additional statistical model that takes into account the variation of characteristics in the images. The first attempt was undertaken by Hutton *et al.* (2000), who applied active shape models (ASMs) to detect cephalometric landmarks. Those authors concluded that even though ASMs were not sufficiently accurate for clinical application, they should provide a model for future studies and a framework for further improvements.

Active appearance models (AAMs), recently proposed by Cootes *et al.* (2001), Cootes and Taylor (2001), and Stegmann (2004), modelling both shape and texture variability seen in a training set, should make the search more precise and robust.

The aim of this study was to evaluate the accuracy of a computerized automatic landmark identification system, based on the AAM approach.

Materials and methods

Experimental design

Sixty cephalograms were randomly selected from the records of patients who had attended for orthodontic assessment and treatment at the Orthodontic Department of the Clinic of Dentistry, Medical Faculty, University of Novi Sad, Serbia. The subjects were aged between 7.2 and 25.6 years (mean age 14.7 years; Table 1).

All the radiographs were taken on Soredex Cranex Tome Ceph digital X-ray machine (Soredex, Tuusula, Finland) using a phosphorus IP-plate (24 × 30 cm). The image plate was processed by a PCT-Digora medical image laser scanner (Soredex). This yielded images that were 2400 × 3000 pixels, giving a pixel size of 0.1 mm, with 256 grey levels in Bitmap format. According to visual assessment, the radiographs varied in quality from average

Table 1 Characteristics of image sample.

Characteristics	Number
Males	37
Females	23
Skeletal Class I	24
Skeletal Class II	27
Skeletal Class III	9
Normal face height	27
Short face height	16
Long face height	17

to high, and were overall considered of good rather than exceptional quality, and as such represented typical lateral cephalograms taken on a modern radiographic machine. By applying this non-selective method of sample collecting, it was hoped to obtain a wide range of variations of both morphological characteristics of skeletal and soft tissue structures and quality of the radiographs.

In order to compare the proposed system for automatic landmarking with previous studies, the images were reduced to 945 × 1181 pixels by pixel averaging. The pixel size in the resultant image was increased from 0.1 to 0.22 mm. The loss in accuracy due to this resolution reduction was considered negligible. The real impact of the full resolution AAMs on performance of the proposed method will be analyzed in future studies.

For building training sets and testing the accuracy of the algorithm, a modified drop-one-out scheme was used (Rudolph *et al.*, 1998; Hutton *et al.*, 2000). Instead of removing just one radiograph from the initial set of cephalograms, five radiographs that were later used for performance testing were removed. Thus, not only accuracy of the algorithm could be tested but also the accuracy of each AAM.

Building statistical models of shape and texture

The first step in building statistical models of appearance is data acquisition. The training set consists of annotated images, where key landmark points are marked on each example object. Suitable normalization is then undertaken after which the data are ready for analysis and can be described in terms of statistical models. The process is divided into three steps: capture, normalization, and analysis (Stegmann, 2000).

In total, 17 standard cephalometric landmarks and 114 pseudo-landmarks were used to define statistical models of shape and texture (Figure 1). For this purpose, the open C++ source code set of the AAM tools were partially modified and used in this study. Each of the 131 landmarks was manually identified by one observer on five occasions. The 'gold standard' (the closest assessment of a landmark position that can be achieved with existing technology and science) was defined as the mean of the five recordings. This gold standard was also used to assess and compare landmarking errors in both the automatic and manual approaches.

Given such a set, a statistical model of shape and texture variation can be generated by applying principal component analysis to the set of vectors describing the shapes and textures in the training set (Cootes *et al.*, 2001; Cootes and Taylor, 2001; Vucinic, 2006). The shape of an object can be represented as a vector x and the texture (or grey levels) as vector g. The appearance model has parameters c controlling the shape and texture according to:

$$x = \overline{x} + Q_s \cdot c,$$
$$g = \overline{g} + Q_g \cdot c,$$

where $\overline{x}$ is the mean shape, $\overline{g}$ the mean texture, and Q_s and Q_g the matrices describing the modes of variation derived from the training set.

Subsequently, a full synthetic image of modelled objects can be synthesized for a given c by generating a texture image from the vector g and warping it using the control points described by x (Figure 2).

AAM matching

To identify the landmarks on a new cephalogram, an initial template was placed over the image. The method described by Cootes *et al.* (2001) and Cootes and Taylor (2001) was used. AAM treat interpretation as an optimization problem in which it seeks to minimize the difference between the new image and the one synthesized by the appearance model. A multi-resolution implementation was used, in which AAMs iterate to convergence at each level before projecting the current solution to the next level of the model (Figure 3). This is more efficient and can converge to the correct solution

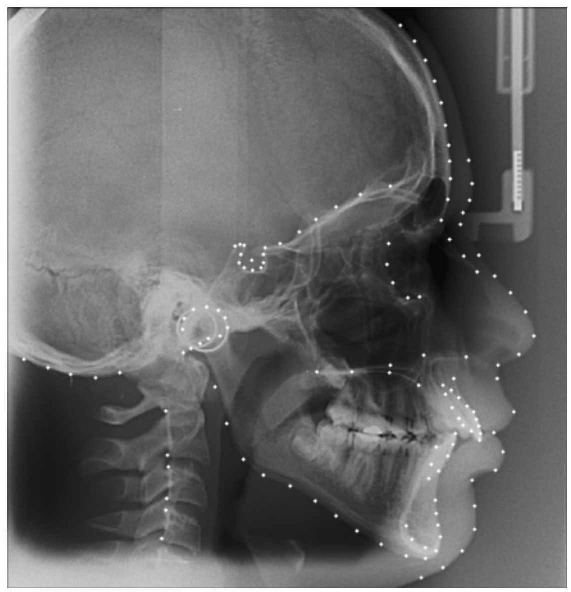

Figure 1 Example of cephalometric image annotated with 131 landmark points.

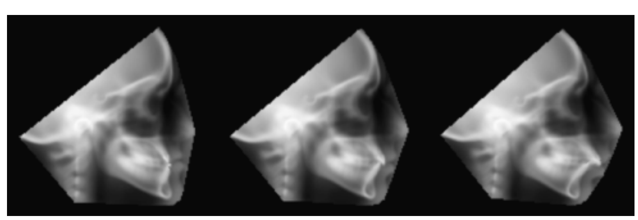

Figure 2 First mode of variation of an appearance model, describing some possible variations in both the shape and the texture component of the synthesized image, seen across 60 training images (Left, –28D; centre, mean; right, +28D).

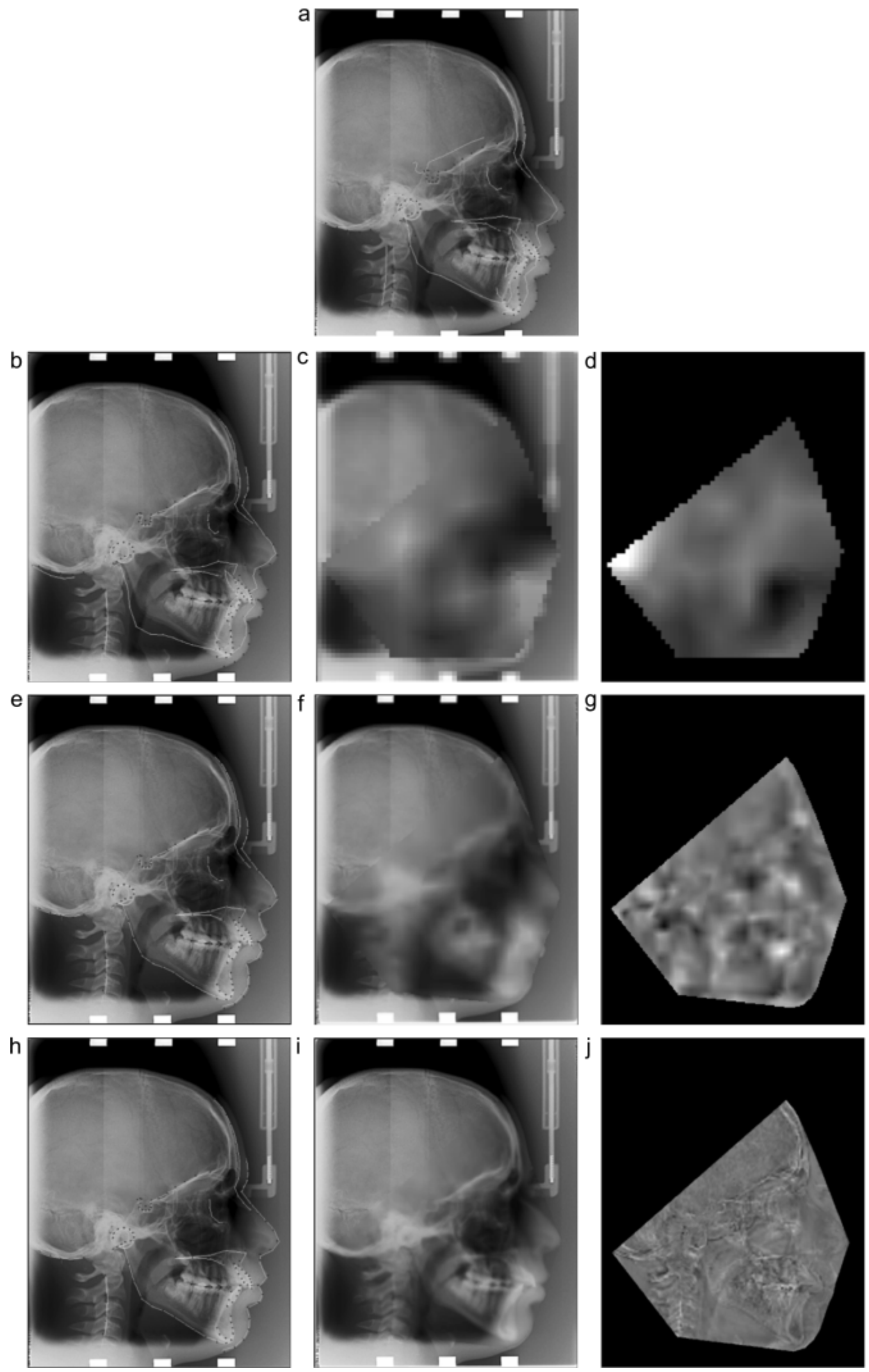

Figure 3 Multi-resolution active appearance model (AAM) search. (a) Initial positioning of the AAM; (b–g) AAM search through different resolution levels; (h–j) final convergence of the AAM and image.

faster and from further away than searching at a single resolution (Cootes *et al.*, 2001; Stegmann, 2004).

Model and method evaluation

The accuracy of each of the 60 templates was tested on five cephalograms that had not been used in the training phase, according to the leave-five-out methodology. For each template, training was undertaken on 55 cephalograms and testing on the five remaining cephalograms. In this way, performance of the AAM method was tested 300 times and its accuracy then estimated as the average error of all landmark detections.

Error in automatic landmark identification was calculated as the Euclidean, *x*- and *y*-axis distance from their manually determined position (gold standard). The bisecting line from the image of the cephalostat through the centre of the machine ear rod was defined as the *y*-axis, and the line perpendicular to the y-axis through the centre of the machine ear rod as the x-axis. Thus, it was possible to evaluate if automatic landmark identification followed a certain envelope pattern, similar to manual detection. The identification of points subspinale (A) or supramentale (B), for example, is prone to error in the perpendicular rather than in the horizontal plane (Liu *et al.*, 2000).

Statistical analysis

Statistical evaluation using the Shapiro–Wilk test, frequency histogram, and normal probability plot confirmed that the collected data were not normally (Gaussian) distributed (Stevens and D'Agostino, 1986). Therefore, non-parametric statistical analysis was applied to the data: the average landmarking errors of repeat measurements for each method were determined using the median value and 80th percentile as a measure of spread. Differences between methods were examined using the non-parametric Wilcoxon signed rank test (Conover, 1980; Turner and Weerakone, 2001). Statistical analyses were performed using the Analyse-it for Microsoft Excel, version 1.62 (Analyse-it Software, Ltd, Leeds, UK) and the Statistical Package for Social Sciences, version 13.00 (SPSS, Inc., Chicago, Illinois, USA).

Error of the method

Ten cephalograms were randomly selected and the gold standard for all the landmarks reassessed 1 month after the original recordings. As recommended by Houston (1983) and Battagel (1993), the error of the method was assessed for random error using Dahlberg's formula. A paired t-test was also performed to assess systematic error. Dahlberg's values demonstrated that random error ranged from 0.11 to 0.35 mm for manual landmark identification. A paired t-test of the repeated measures showed no systematic errors.

Results

Average Euclidean, x- and y-axis landmarking errors for manual, and automatic landmark identification are shown in Tables 2 and 3. Best recognition performances were for Apex inferior (Ap1l) and Supramental (B) and the lowest were for Porion (Po) and Articulare (Ar). These results were compared with the manual method (Table 4) and with those obtained in previous studies (Tables 5–7). The overall success rate for all landmark detection attempts for automatic recognition was 28 per cent within 1 mm, 61 per cent within 2 mm, and 95 per cent within 5 mm precision (radii).

Discussion

Unlike previously tested methods (edge tracking and the ASM approach) where the search is made around the current position of each point using models of the image texture in small regions around each landmark, the AAM manipulates a full model of appearance, which represents both shape variation and the texture of the region covered by the model. This was used to generate full synthetic images of the modelled objects. AAM then uses the difference between the current synthesized image and the target image to update its parameters (Cootes *et al.*, 1999). After initial testing of the system, the shape-based modification of the AAM algorithm was used, in which linear shape update is predicted from current texture error, since it was able to locate the points slightly more accurately than the original formulation.

The manual method for landmark detection is still considered the most accurate. Differences in accuracy between the manual and proposed method varied from 0.17 to 3.06 mm. It was negligible for three of the 17 landmarks

Table 2 Median and 80th percentile values of landmarking errors for manual landmark identification (in mm).

Cephalometric landmarks	Median (Euclidean)	Minimum	Maximum	80th percentile	Median (x-axis)	Median (y-axis)
Sella (S)	0.18	0.02	0.45	0.29	0.07	0.12
Nasion (N)	0.38	0.06	3.83	1.36	0.25	0.29
Porion (Po)	0.21	0.03	2.87	0.51	0.20	0.05
Orbitale (Or)	0.24	0.07	2.10	0.70	0.18	0.10
Subspinale (A)	0.92	0.09	16.62	3.30	0.16	0.55
Anterior nasal spine (ANS)	0.47	0.06	8.56	1.12	0.34	0.13
Posterior nasal spine (PNS)	0.40	0.07	2.52	0.78	0.29	0.15
Supramentale (B)	0.80	0.06	4.25	1.52	0.09	0.79
Pogonion (Pg)	0.23	0.00	1.16	0.51	0.04	0.22
Gnathion (Gn)	0.34	0.02	1.28	0.58	0.19	0.17
Menton (Me)	0.66	0.09	2.57	1.02	0.59	0.15
Gonion (Go)	0.83	0.09	4.67	1.62	0.45	0.65
Articulare (Ar)	0.36	0.06	4.93	1.49	0.21	0.28
Incision superior (Is1u)	0.14	0.03	1.48	0.22	0.10	0.07
Apex superior (Ap1u)	1.07	0.09	3.19	1.64	0.38	0.85
Incision inferior (Is1l)	0.19	0.02	2.94	0.30	0.10	0.12
Apex inferior (Ap1l)	0.90	0.22	4.88	2.37	0.44	0.70
X (average)	0.49	0.06	4.02	1.14	0.23	0.32

Table 3 Median and 80th percentile values of landmarking errors for automatic landmark identification (in mm).

Cephalometric landmarks	Median (Euclidean)	Minimum	Maximum	80th percentile	Median (*x*-axis)	Median (*y*-axis)
Sella (S)	1.87	0.12	9.13	3.27	1.24	0.91
Nasion (N)	1.42	0.04	13.34	2.66	0.94	0.99
Porion (Po)	3.27	0.14	9.46	4.69	2.12	1.73
Orbitale (Or)	2.22	0.22	6.44	3.81	1.59	1.05
Subspinale (A)	1.41	0.03	12.27	3.09	0.62	1.01
Anterior nasal spine (ANS)	1.99	0.14	11.75	3.49	1.31	1.00
Posterior nasal spine (PNS)	1.56	0.02	8.76	2.46	0.81	0.92
Supramentale (B)	1.20	0.04	9.77	2.15	0.59	0.68
Pogonion (Pg)	1.33	0.17	9.54	2.41	0.71	0.84
Gnathion (Gn)	1.31	0.04	9.39	2.34	0.74	0.76
Menton (Me)	1.23	0.09	10.49	2.33	1.01	0.45
Gonion (Go)	2.13	0.13	9.94	3.58	1.15	1.15
Articulare (Ar)	2.31	0.36	6.96	3.56	1.31	1.28
Incision superior (Is1u)	1.34	0.09	15.19	2.22	0.84	0.85
Apex superior (Ap1u)	1.72	0.03	12.90	2.58	0.66	1.23
Incision inferior (Is1l)	1.24	0.06	10.45	2.56	0.93	0.72
Apex inferior (Ap1l)	1.07	0.07	12.48	1.93	0.62	0.70
X (average)	1.68	0.11	10.49	2.89	1.01	0.96

Table 4 Wilcoxon signed rank test comparing manual and automatic landmark identification.

Cephalometric landmarks	Difference (mm)	Wilcoxon test	*P*	
Sella (S)	1.70	1829	<0.0001	***
Nasion (N)	1.04	1506	<0.0001	***
Porion (Po)	3.06	1829	<0.0001	***
Orbitale (Or)	1.98	1816	<0.0001	***
Subspinale (A)	0.49	1074	0.2418	NS
Anterior nasal spine (ANS)	1.53	1435	0.0001	***
Posterior nasal spine (PNS)	1.16	1674	<0.0001	***
Supramentale (B)	0.40	1139	0.0991	NS
Pogonion (Pg)	1.10	1774	<0.0001	***
Gnathion (Gn)	0.97	1702	<0.0001	***
Menton (Me)	0.57	1547	<0.0001	***
Gonion (Go)	1.31	1645	<0.0001	***
Articulare (Ar)	1.94	1679	<0.0001	***
Incision superior (Is1u)	1.19	1819	<0.0001	***
Apex superior (Ap1u)	0.65	1266	0.0098	**
Incision inferior (Is1l)	1.05	1693	<0.0001	***
Apex inferior (Ap1l)	0.17	893	0.8713	NS

$^{**}P < 0.01$; $^{***}P < 0.001$; NS, not significant.

(subspinale, supramentale, and apex inferior; Table 4). On average, for all the cephalometric landmarks, the proposed system had a precision of 1.68 mm, which is a considerable improvement with regard to other complementary automatic systems (Table 5; Rudolph *et al.*, 1998; Hutton *et al.*, 2000; Liu *et al.*, 2000; Grau *et al.*, 2001). However, care must be taken when comparing and interpreting such results, since the training data and validation data differs for all published studies.

Taking into account the nature of multivariate linear regression, it is anticipated that using a larger number of sample images for model training and building will lead to a better prediction model and consequently more accurate results. Therefore, the real impact of increasing the number of training images (potentially capturing more variations of morphological structures of the human skull and quality of cephalograms) on the performance of the proposed method requires further investigation. Considering that the accepted normal range of error for most cephalometric measurements is approximately ±2 mm and that inter-expert variability can vary up to 5 mm (Liu *et al.*, 2000), these results justify the potential of the studied method for clinical application. The AAM approach provides the opportunity for building more robust and precise systems for automatic landmark detection, as suggested by Hutton *et al.* (2000).

The magnitude of error in landmark identification depends on the position of the landmark. If the landmark is in a clear border of the craniofacial structure, such as sella (S) or pogonion (Pg), the error will be smaller. On the other hand, if the landmark is located on poorly defined structures which have a low signal to noise ratio, with many craniofacial structures overlying each other, such as porion (Po) and orbitale (Or), the error will be larger (Baumrind and Frantz, 1971). Experienced clinicians may be able to infer the position of landmarks from their background knowledge of cephalometry, even poorly defined ones, whereas the automatic systems are still unable to compete in this capacity. However, the results of the present research showed a significant improvement in locating low-contrast landmarks over other methods (Table 7), except the approaches presented by Grau *et al.* (2001) and Liu *et al.* (2000), but their systems were tested on a very small number of low resolution radiographs, making the results statistically unreliable. This can be attributed

Table 5 Comparison of the average automatic landmarking errors (mm) in different studies.

Cephalometric landmarks	Rudolph *et al.* (1998)	Hutton *et al.* (2000)	Liu *et al.* (2000)	Grau *et al.* (2001)	Present study
Sella (S)	5.06	5.5	0.94	1.92	1.87
Nasion (N)	2.57	5.6	2.32	1.40	1.42
Porion (Po)	5.67	7.3	2.43	—	3.27
Orbitale (Or)	2.46	5.5	5.28	1.92	2.22
Subspinale (A)	2.33	3.3	4.29	0.90	1.41
Anterior nasal spine (ANS)	2.64	3.8	2.90	0.75	1.99
Posterior nasal spine (PNS)	—	5.0	—	1.13	1.56
Supramentale (B)	1.85	2.6	3.69	—	1.20
Pogonion (Pg)	1.85	2.7	2.53	0.95	1.33
Gnathion (Gn)	—	2.7	1.74	1.44	1.31
Menton (Me)	3.09	2.7	1.90	0.48	1.23
Gonion (Go)	—	5.8	4.53	1.10	2.13
Articulare (Ar)	—	—	—	—	2.31
Incision superior (Is1u)	2.02	2.9	2.36	0.84	1.34
Apex superior (Ap1u)	2.17	2.9	—	0.89	1.72
Incision inferior (Is1l)	2.46	3.1	2.86	0.90	1.24
Apex inferior (Ap1l)	2.67	3.9	—	0.54	1.07
X (average)	2.83	4.08	2.91	1.08	1.68

Table 6 Comparison of the overall success rate of all landmark detection attempts, within 1, 2, and 5 mm precision (radii) of the active shape model (ASM) and active appearance model (AAM) approach.

	Present study		Hutton *et al.* (2000) automatic ASM %
	Manual %	Automatic AAM %	
<1 mm	72	28	13
<2 mm	87	61	35
<5 mm	98	95	74

to the fact that cephalometric images are very rich in subtle grey-level variations and in such cases an appearance model can represent both the shape and texture variability seen in a training set, better than edge-tracking methods.

The pattern of errors for most landmarks was similar to that found with manual tracing. Distribution of detection attempts for landmarks located on edges follows the shape of the border of the craniofacial structures. Landmarks lying on vertical borders, such as nasion (Na), are more accurately located in the horizontal dimension as opposed to the vertical dimension (Figure 4). Similarly, landmarks lying on horizontal edges are more accurately located in the vertical dimension. This is in accordance with the findings of Forsyth and Davis (1996).

Example of failure

Figure 5 shows an example where the AAM failed to locate boundaries correctly on images used for testing. In this case, the subject showed a more extreme shape variation from the mean. Due to this variation, the model was not able to locate the outer boundaries. This is because the model only samples the image under its current location and within variation limits. There is not always sufficient information to drive the model outward to the correct outer boundary. This can be overcome by modelling the whole surface of the radiograph or by using a larger number of sample images for model training that will capture even these extreme variations. Alternatively, it may be possible to combine different methods and include explicit searching outside the current patch, for instance by searching along normal to current boundaries as in the ASM (Cootes and Taylor, 2001).

Conclusions

Based on the obtained results, the following conclusions were made:

1. The AAM approach can adequately represent the average shape and texture variations of craniofacial structures on digital radiographs. As such it can successfully be implemented for automatic localization of cephalometric landmarks.
2. An increase in overall precision and detection of low-contrast cephalometric landmarks was achieved in comparison with other automatic systems.

Considering the established potentials and advantages of the AAM, it is expected that by creating more precise statistical models of shape and texture (based on a larger number of radiographs), and refining the AAM algorithm in the final adaptation phase, it will be possible to use it as a completely automatic system for automatic detection of cephalometric landmarks.

Table 7 Comparison of mean landmarking errors (mm) of low (L)- and high (H)-contrast cephalometric landmarks.

	Rudolph *et al.* (1997)		Rudolph *et al.* (1998)		Liu *et al.* (2000)		Hutton *et al.* (2000)		Grau *et al.* (2001)		Present study	
	L	H	L	H	L	H	L	H	L	H	L	H
Mean error	4.30	2.65	3.85	2.56	2.88	2.91	5.02	3.52	1.32	0.91	2.03	1.47
Difference	1.65		1.29		−0.03		1.50		0.41		0.56	

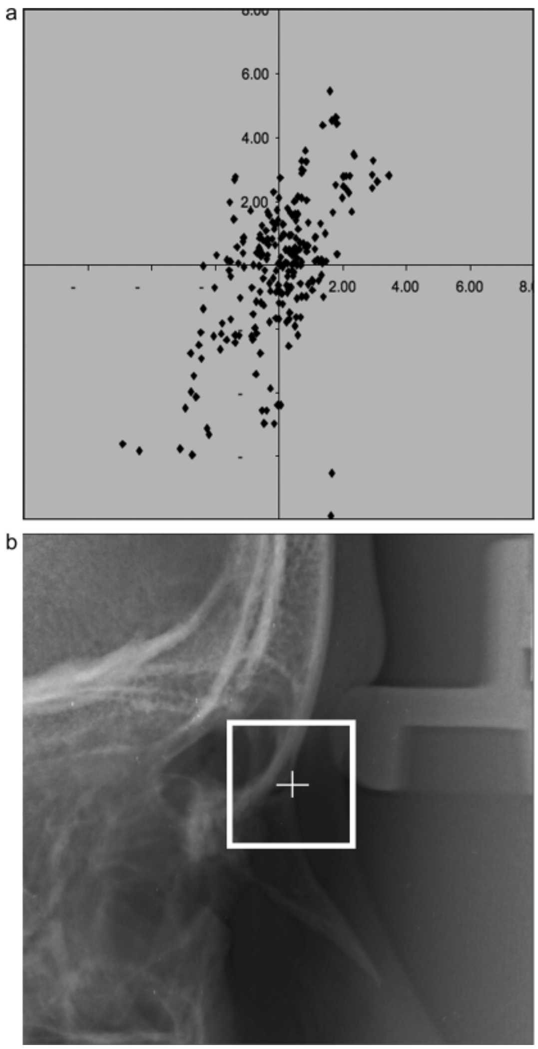

Figure 4 Distribution of automatic detection attempts (a) for the landmark nasion (N) (b).

Figure 5 An example of a search failure, where the active appearance model did not locate the correct outer boundaries of the lower jaw.

Address for correspondence

Predrag Vucinic
Clinic of Dentistry
Faculty of Medicine
Hajduk Veljkova 12
21000 Novi Sad
Serbia
E-mail: pvucinic@neobee.net

Acknowledgement

The authors are grateful to Professor Tim Cootes, Division of Imaging Science and Biomedical Engineering, University of Manchester, UK, for providing the open C++ source code set of AAM tools for this research and for his kind help.

References

Battagel J M 1993 A comparative assessment of cephalometric errors. European Journal of Orthodontics 15: 305–314

Baumrind S, Frantz R C 1971 The reliability of head film measurements. 1. Landmark identification. American Journal of Orthodontics 60: 111–127

Baumrind S, Miller D M 1980 Computer-aided head film analysis: the University of California San Francisco method. American Journal of Orthodontics 78: 41–65

Broadbent H B 1931 A new X-ray technique and its application to orthodontia. Angle Orthodontist 1: 45–66

Cardillo J, Sid-Ahmed M A 1994 An image processing system for locating craniofacial landmarks. IEEE Transactions on Medical Imaging 13: 275–289

Ciesielski V, Vic A, Innes A, John S, Mamutil J 2003 Genetic programming for landmark detection in cephalometric radiology images. International Journal of Knowledge-Based and Intelligent Engineering Systems 7: 164–171

Cohen A M, Linney A D 1986 A low cost system for computer-based cephalometric analysis. British Journal of Orthodontics 13: 105–108

Cohen A M, Ip H H, Linney A D 1984 A preliminary study of computer recognition and identification of skeletal landmarks as a new method of cephalometric analysis. British Journal of Orthodontics 11: 143–154

Conover W J 1980 Practical non-parametric statistics, 2nd edn. Wiley, New York

Cootes T F, Taylor C J 2001 Statistical models of appearance for medical image analysis and computer vision. In: Proceedings of the Conference on Medical Imaging. SPIE Proceedings Series, San Diego, 4322: 236–248

Cootes T F, Edwards G J, Taylor C J 1999 Comparing active shape models with active appearance models. In: Proceedings of the 10th British Machine Vision Conference 1: BMVA Press, Nottingham, pp. 173–182

Cootes T F, Edwards G J, Taylor C J 2001 Active appearance models. IEEE Transactions on Pattern Analysis and Machine Intelligence 23: 681–685

Davis D 1994 Knowledge-based cephalometric analysis: a comparison with clinicians using interactive computer methods. Computers and Biomedical Research 27: 10–28

Davis D N, Taylor C J 1991 A blackboard architecture for automating cephalometric analysis. Journal of Medical Informatics 16: 137–149

Desvignes M, Romaniuk B, Demoment R, Revenu M, Deshayes M J 2000 Computer assisted landmarking of cephalometric radiographs. In: Proceedings of the Fourth IEEE Southwest Symposium on Image Analysis and Interpretation. IEEE Computer Society, SSIAI, Austin, pp. 296–300

El-Feghi I, Sid-Ahmed M A, Ahmadi A 2004 Automatic localization of craniofacial landmarks for assisted cephalometry. Pattern Recognition 37: 609–621

Forsyth D B, Davis D N 1996 Assessment of an automated cephalometric analysis system. European Journal of Orthodontics 18: 471–478

Grau V, Alcañiz M, Juan M C, Monserrat C, Knoll C 2001 Automatic localization of cephalometric landmarks. Journal of Biomedical Informatics 34: 146–156

Hofrath H 1931 Die Bedeutung der Röntgenfern und Abstandsaufnahme für die Diagnostik der Kieferanomalien. Fortschritte der Kieferorthopdiäe 1: 232–258

Houston W J B 1983 The analysis of errors in orthodontic measurements. American Journal of Orthodontics 83: 382–390

Houston W J B, Mahler R E, McElroy D, Sherriff M 1986 Sources of error in measurements from cephalometric radiographs. European Journal of Orthodontics 8: 149–151

Hutton T J, Cunningham S, Hammond P 2000 An evaluation of active shape models for the automatic identification of cephalometric landmarks. European Journal of Orthodontics 22: 499–508

Innes A, Ciesielski V, Mamutil J, Sabu J 2002 Landmark detection for cephalometric radiology images using pulse coupled neural networks. In: Proceedings of the International Conference on Artificial Intelligence. CSREA Press, Las Vegas, 2: 511–517

Kamoen A, Dermaut L, Verbeeck R 2001 The clinical significance of error measurement in the interpretation of treatment results. European Journal of Orthodontics 23: 569–578

Lévy-Mandel A D, Venetsanopoulos A N, Tsotsos J K 1986 Knowledge-based landmarking of cephalograms. Computers and Biomedical Research 19: 282–309

Liu J K, Chen Y T, Cheng K S 2000 Accuracy of computerized automatic identification of cephalometric landmarks. American Journal of Orthodontics and Dentofacial Orthopedics 118: 535–540

Miller R L, Dijkman D J, Riolo M L, Moyers R E 1971 Graphic computerization of cephalometric data. Journal of Dental Research 50: 1363–1369

Nimkarn Y, Miles P G 1995 Reliability of computer-generated cephalometrics. International Journal of Adult Orthodontics and Orthognathic Surgery 10: 43–52

Parthasarathy S, Nugent S T, Gregson P G, Fay D F 1989 Automatic landmarking of cephalograms. Computers and Biomedical Research 22: 248–269

Ren J *et al.* 1998 A knowledge-based automatic cephalometric analysis method. Proceedings of the 20th International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Computer Society, Hong Kong, 20: 723–727

Richardson A 1981 A comparison of traditional and computerized methods of cephalometric analysis. European Journal of Orthodontics 3: 15–20

Romaniuk B, Desvignes D, Revenu M, Deshayes M J 2002 Linear and non-linear model for statistical localization of landmarks. In: Proceedings of the 16th International Conference on Pattern Recognition 4: IEEE Computer Society, Washington, p. 40393

Rudolph D J, Coggins J M, Moon H 1997 Investigation of filter sets for supervised pixel classification of cephalometric landmarks by spatial spectroscopy. International Journal of Medical Informatics 47: 183–191

Rudolph D J, Sinclair P M, Coggins J M 1998 Automatic computerized radiographic identification of cephalometric landmarks. American Journal of Orthodontics and Dentofacial Orthopedics 113: 173–179

Sanei S, Sanaei P, Zahabsaniei M 1999 Cephalogram analysis applying template matching and fuzzy logic. Image and Vision Computing 18: 39–48

Savage A W, Showfety K J, Yancey J 1987 Repeated measures analysis of geometrically constructed and directly determined cephalometric points. American Journal of Orthodontics and Dentofacial Orthopedics 91: 295–299

Stegmann M 2000 Active appearance models: theory, extensions and cases. Thesis, University of Denmark, Lingby, pp. 37–39

Stegmann M 2004 Generative interpretation of medical images. Thesis, University of Denmark, Lingby

Stevens M A, D'Agostino R B 1986 Goodness of fit techniques. Marcel Dekker, Inc., New York

Turner P J, Weerakone S 2001 An evaluation of the reproducibility of landmark identification using scanned cephalometric images. Journal of Orthodontics 28: 221–229

Uchino E, Yamakawa T 1995 High speed fuzzy learning machine with guarantee of global minimum and its application to chaotic system identification and medical image processing. In: Proceedings of the Seventh International Conference on Tools with Artificial Intelligence. IEEE Computer Society, Herndon, pp. 242–249

Vučinić P 2006 Primena metode aktivnog modela oblikovanja u oblasti automatske identifikacije rendgenkefalometrijskih tacaka (Application of active appearance models for the automatic cephalometric landmarks recognition). Thesis, University of Novi Sad, Serbia