Automatic computerized radiographic identification of cephalometric landmarks

D.J. Rudolph, DDS, PhD, a P.M. Sinclair, DDS, MSD, b and J.M. Coggins, PhD c

Los Angeles, Calif., and Chapel Hill, N.C.

Computerized cephalometric analysis currently requires manual identification of landmark locations. This process is time-consuming and limited in accuracy. The purpose of this study was to develop and test a novel method for automatic computer identification of cephalometric landmarks. Spatial spectroscopy (SS) is a computerized method that identifies image structure on the basis of a convolution of the image with a set of filters followed by a decision method using statistical pattern recognition techniques. By this method, characteristic features are used to recognize anatomic structures. This study compared manual identification on a computer monitor and the SS automatic method for landmark identification on minimum resolution images (0.16 cm² per pixel). Minimum resolution (defined as the lowest resolution at which a cephalometric structure could be identified) was used to reduce computational time and memory requirements during this development stage of the SS method. Fifteen landmarks were selected on a set of 14 test images. The results showed no statistical difference (p > 0.05) in mean landmark identification errors between manual identification on the computer display and automatic identification using SS. We conclude that SS shows potential for the automatic detection of landmarks, which is an important step in the development of a completely automatic cephalometric analysis. (Am J Orthod Dentofacial Orthop 1998;113:173-9.)

Three possible approaches may be used to perform a cephalometric analysis. The most common method is by manually placing a sheet of acetate over the cephalometric radiograph, tracing salient features, identifying landmarks, and measuring distances and angles between landmark locations. Another approach is computer aided. Landmarks are located manually while these locations are digitized into a computer system. The computer then completes the cephalometric analysis. The third approach is completely automated. The cephalometric radiograph is scanned into the computer. The computer automatically locates landmarks and performs the cephalometric analysis.¹

The current technique of manually analyzing a cephalometric radiograph by measuring spatial relationships between cephalometric landmarks is both time-consuming and somewhat inaccurate. Time pressures in the clinical environment can contribute to decreased reliability.

Currently, errors in manual cephalometric analysis can be attributed to errors in reproducibility and validity. Reproducibility errors are caused by variations in image acquisition, landmark identification, and measurement errors. Variations in image acquisition can result from beam alignment and repositioning, exposure parameter variations, and film processing variations. Limitations of human visual performance may result in cephalometric landmark identification errors. These errors may be expressed as either interobserver or intraobserver variation. Interobserver variation also may be caused by variations in training and experience, or by the subjective nature of landmark identification. Furthermore, intraobserver variation may result from variations in lighting and image orientation. It is generally accepted that landmarking errors should be less than 0.5 mm. This rule of thumb, however, has not been proven in research studies. The mean estimating error of landmark identification has been reported to be 1.26 mm in one study. This study also showed that manual landmark identification errors varied significantly depending on the landmark and the observer.

Landmarks currently are defined using subjec-
tive criteria rather than mathematically rigorous specifications. Landmarks, for the most part, are described as extrema of a tissue component with respect to a spatial coordinate. For example, pogonion is the most anterior point of the chin. Thus the orientation of the patient's head affects the location of pogonion. Other landmarks that are affected by image orientation include a point, b point, menton, and gonion.

Measurement of distances and angles between landmark locations is defined by the limitations of the measurement devices (rulers and protractors), as well as the limitations of human visual performance. Human errors can occur both in recording of the measurement or in the use of the measurement devices. In most cases, landmarks are not labeled before measurement because of time constraints, causing additional errors.

The time required for manual analysis depends on how comprehensive the measurements are. Some clinicians make no measurements at all; they hold the radiograph up to the light and get a “clinical impression” of the skeletal and dental pattern of the patient. Other clinicians make extensive, time-consuming measurements, and some clinicians direct a staff member (nonorthodontist) to prepare the analysis. In each case the technique is inefficient, time-consuming, or prone to error at several steps.

**Computer-Aided Analysis**

Recently a computer-assisted method of cephalometric analysis has been developed. This process requires manual landmark identification and digitization along with computerized measurement of the landmark relationships. This approach provides for comprehensive measurements in less time, but it is still time-consuming and error-prone. Currently, several commercially available systems can perform basic cephalometric analysis tasks. The user locates landmarks manually with a mouse cursor on the display monitor on some systems. Other systems digitize landmark locations on a digitizing pad. In either case a computer algorithm performs a cephalometric analysis by calculating distances and angles between landmark locations. In addition, the algorithm connects these landmarks with line segments to produce a tracing. Some systems are capable of moving the tissues to simulate treatment effects, growth effects, and surgical prediction. Finally, some of these systems also are able to produce a time series of images using landmark locations, not superimposition contours, to register images.

Generally, these systems do not save time, are expensive, and require technical training. The accuracy of these computer-aided programs has been demonstrated to be similar to that of manual digitization, and because manual landmark identification programs require subjective user point identification, they are limited in scope. In addition, the number of landmarks required are high; this tends to negate any time saved using this method. Although the analysis uses a computer, the process of manual point digitization can be time-consuming and error-prone.

**Automatic Landmark Identification**

A third approach to cephalometric analysis is completely automated. The cephalometric image is scanned into a computer and both landmark identification and cephalometric analysis are automated. The process has the potential to increase accuracy, provide more efficient use of clinicians' time, and improve our ability to correctly diagnose orthodontic problems. Additionally, this process may provide mathematical descriptions of landmark locations that could be applied to new ways of evaluating cephalometric radiographs to derive clinically important information. Also, computerization itself may provide for alternative methods of form description besides distances and angles of landmark locations and may make the work of Bookstein as well as techniques such as finite...
element analysis, allometric models, symmetric axis transform, mesh diagram, and angular invariants more clinically applicable.

The previous published attempts to automatically identify landmarks and produce a cephalometric analysis have used a similar approach. Image pixels that are in regions of high intensity gradient are identified as edges, and these edges are assumed to be object boundaries. Landmarks are then found in relation to these labeled boundaries. This edge-based approach (Fig. 1) involves four steps.

1. Remove noise.
2. Label pixels according to their edginess.
3. Connect pixels and label edges.
4. Find landmarks based on position or relationship to a labeled edge.

In these studies, noise is removed by using a median filter and pixel averaging. Edginess of pixels is found by thresholding on the intensity or thresholding on the image convolved with gradient operators such as Hueckel and Kirsch operators. Edges are labeled by position in the image. For example, a vertical edge closest to the right border of the image is labeled soft tissue profile. Landmarks are then labeled according to their (x,y) position along these labeled edges.

Only four algorithms reported in the literature have attempted to locate image landmarks. These attempts have had both experimental design flaws and limited results. Landmarking was evaluated for correctness either by comparison with experts or by inspection. The number of landmarks attempted varied from 10 to 27. One study did not report the accuracy of landmarks directly but rather reported the angles and distances between landmarks.

These previous results have several pitfalls. The most glaring is testing on the training set. In these cases, the method was tested on the same set of radiographs used to develop the algorithm. Some of the studies used a very small sample size, including two and five radiographs. Another problem is that some studies only reported the images that resulted in relatively successful landmarking. They did not report the total number of radiographs attempted. The results reported in these studies were inadequate for clinical use. Also, it was often noted that the results reported only worked on ideal images. There are two reasons why the results of these previous studies may have been limited. First of all, they are based on sets of rules (heuristic), and second, they are edged based.

Heuristic methods are based on ad hoc rules for finding each specific landmark. An example would be to find a point defined by being on a vertical edge in the lower right portion of the image that is most anterior and call it pogonion. Using this approach, locating teeth and other interior structures such as temporomandibular joint, orbitale, and posterior nasal spine is more difficult. The major works showing application of heuristic methods to landmarking were presented more than 5 years ago. No new insights have been reported since. The problem is that the rules become increasingly difficult as more landmarks, internal structures, and variations in image contrast are introduced. Thus some landmarks require more complex rules than others. In addition, these methods are not invariant to rotation of the radiograph, magnification, and contrast. This may explain why these algorithms currently are limited to only ideal radiographs and may not represent a potentially clinically applicable future approach.
The use of edges as a basis for image analysis is replete with problems. First, edges are notoriously unreliable indices of object location. Second, describing an image by edges ignores rich structural information from the gray scale geometry of the image. Third, edge representations fail to capture essential elements of object geometry such as “center” and “width.” Finally, edge representations are made on the basis of a single small scale, but image structure can be more fully described by measurements of image geometry across multiple scales.

MATERIALS AND METHODS

Spatial spectroscopy (SS) is an automatic approach to image interpretation and was used in this study to automatically locate landmarks on cephalometric radiographs. The details of this method are reported elsewhere.\(^{18,19}\) In short, a cephalometric image is broken down into 75 feature values \(F_n\) for each picture element. The process involves a mathematical operation called a convolution of the image with a Gaussian filter and taking the derivatives of image intensity in both the \(x\) and \(y\) direction (Equation 1).

\[
F_n(x_0, y_0, p, q) = \frac{\partial}{\partial x} \left( \frac{\partial}{\partial y} [G_x \otimes I] (x_0, y_0) \right)
\]  

(1)

Where \(p = 0,1,2,3\), \(q = 0,1,2,3\); \(G_x\) is a two-dimensional Gaussian distribution of variance \(\sigma\); \(I(x_0, y_0)\) is the image intensity distribution. Applying this equation to an image results in 75 different feature values at each image pixel location that describe the multiscale geometry at and around each pixel.

In a given pixel location, the probability that the pixel portrays a specific landmark, \(P(C_i | F_n)\), is determined by Equation 2:

\[
P(C_i | F_n) = \text{Nexp}\left( -0.5 |F_n - \mu_i| \Sigma_i^{-1} \right)
\]

\[
F_n - \mu_i \right)^T (2)
\]

where \(N\) is a normalization constant and \(\mu_i\) and \(\Sigma_i\) (the means and covariance) are determined by sampling the feature values that result from known landmark locations.

Experimental Design

The algorithm was tested on 14 radiographs in a leave-one-out design where 13 radiographs were used to estimate the class-conditional densities for each landmark \(\mu_i\) and \(\Sigma_i\), and the remaining one was tested by the algorithm. This was repeated 14 times, leaving a different radiograph out to be tested each time. This method maximizes the use of the data set and yet does not test on the training set. The same parameter used to define the mathematical equations were not derived from the radiograph being tested.

Image Sample

Sixteen radiographs were randomly selected from the retention file of the Department of Orthodontics at the University of North Carolina at Chapel Hill, of which 14 were used in this study. One radiograph was rejected because the patient wore appliances, and one was rejected because the patient had postsurgical rigid fixation. Table I shows the composition of the sample.

Because these were retention cases, the radiographs were taken over a long span of time. Radiographs were not all taken on the same x-ray unit. They represented a wide range of exposure parameters and techniques. They also included several patients with a cervical collar. Thus
this sample represents a wide variation of cephalometric radiographs.

Image Acquisition

The cephalometric radiographs were scanned for the 14 test images with a Sony camera (Sony, Inc.) and captured on a Data Translation board. All parameters on the digitizing camera were kept constant, including F-stop, magnification, and room lighting. The image orientation was maintained by orienting all images in natural head position. A 10 square-inch area was sampled to \(256 \times 256\) pixels maintaining rotational, scale, and contrast invariance in the digitization process. This yielded a resolution of \(1\) mm\(^2\) per pixel. Then the images were reduced to \(64 \times 64\) pixels by pixel averaging. The resolution of an image pixel of the resultant image was \(0.16\) cm\(^2\).

Landmarks

The 15 landmarks studied are shown in Table II. They comprised four cranial base, two maxillary, five mandibular, and four dental landmarks. Each of the 15 landmarks was manually identified by one observer on 5 separate days on the set of 14 \(64 \times 64\) cephalometric images. The gold standard (the closest assessment of a position that can be achieved with the existing technology and science) was defined as the mean of these five attempts. This gold standard was used to identify mean errors in both the automatic and manual approaches. Mean errors were determined for each of the five attempts identified manually and for the automatic approach. Mean errors in this study were defined as mean magnitude in distance between the gold standard and selected landmarks for all the 14 radiographs.

RESULTS

Table III shows that the automatic landmark identification was more accurate than manual identification for 9 of the 15 landmarks. No significant difference \((p > 0.05)\) was noted, however, between the total mean error for manual and automatic landmark identification. The total mean error for both approaches was almost identical and was less than 1 pixel (0.77 pixel, 3.1 mm). Furthermore, when individual landmarks were compared, only sella showed differences between automatic and manual identification \((p < 0.05)\). Pogonion and b point \((0.457\) pixels, 1.848 mm) were the most accurately identified landmarks using the automatic method, whereas upper incisor tip \((0.497\) pixels, 1.988 mm) was the most accurate landmark identified manually. Porion was the least accurate for both manual and automatic identification.

DISCUSSION

Minimum resolution \((0.16\) cm\(^2/\)pixel\)) was used in this preliminary study to limit computational time and memory requirements in developing the use of SS for automatic landmark identification. The spa-
tial errors of landmark identification in terms of pixels (pixel element) may give a theoretical indication of the limits of automatic landmark identification as higher image resolutions are achieved. At higher resolutions, the area represented by a pixel becomes reduced. Because optical systems are available that have higher resolution than the human optical system, greater resolution and less landmark identification errors with automatic analysis may be achievable. The cost, however, is computation time and memory usage.

The results in this study for automatic landmark identification and manual landmark identification on the computer monitor were similar. No significant difference was found in the mean total landmark identification. These results support the hypothesis that SS may be useful in automatic identification of cephalometric landmarks. Automatic identification errors for sella were significantly higher than for manual identification. This may have occurred because sella is a unique landmark. Unlike other landmarks, it is not defined as a directional extrema of a tissue boundary or the intersection of tissue boundaries. Sella is defined by its relationship to another structure (geometric center of the pituitary fossae).

The results showed that pogonion and b point (0.457 pixels) were the most accurate automatic landmarks identified by the automatic method, whereas upper incisor tip (0.497 pixels) was the most accurate landmark identified manually on the monitor. Porion was the least accurate for both manual monitoring and automatic identification. Sella, mandibular notch, TMJ, and porion all had high automatic landmarking errors, whereas the other 11 landmarks tested had relatively low automatic landmarking errors. Porion and TMJ had high landmarking errors when manually found on the computer monitor.

The mean error (3.1 mm) of this study at minimum resolution is somewhat greater than the landmarking error of 1.26 mm reported for manual identification on cephalometric radiographs at high resolution. Furthermore, (except for sella) landmarking errors tended to be qualitatively similar regardless of the method used (automatic, manual on monitor, or manual on radiograph). Landmarks that were reported in the literature to have high errors when identified manually on cephalometric radiographs also had high errors in this study. In addition, those that had low errors manually had low errors in this study.

Two potential sources of errors existed in the automatic approach: errors in defining the gold standard, and limitations of the SS model. Small spatial errors in the selection of the gold standard may have caused larger errors in the values of the features selected. Thus slight spatial inaccuracies in the determination of the true landmarks may have caused greater effects on the accuracy of the algorithm than the original error in identification of the gold standard. Furthermore, the true landmark may be located in the center, corners, or edges of a pixel. At minimum resolution, the distances between these indistinguishable positions may have contributed to the errors demonstrated.

One limitation of the SS model is that the features used from Equation 1 and the decision process from Equation 2 have not yet been optimized. Another limitation is the limited number of images in the training set in the face of a large number of features used to make decision as to the landmark location. Only 13 images were used to determine the values of $\mu_{ni}$ and $\Sigma_i^{-1}$ from Equation 2.

Future Studies

Future developments will include the following points: 1) optimization of SS by optimizing the parameters in Equations 1 and 2; 2) use of a larger number of images to more accurately identify ideal feature values; 3) use of more landmarks to enable a complete cephalometric analysis; 4) higher image resolution to reduce the area represented by one

Fig. 3. Four typical examples of automatic landmark identification.
image pixel and thus potentially increase accuracy; 5) improvement of the accuracy of identification of sella by locating landmarks on the clinoid processes and locating sella on the basis of the relationship to these landmarks; 6) refinement of the determination of the gold standard by examining the feature value produced at neighboring locations in the selection of the gold standard and considering the use of these as more accurate landmark locations; and 7) use of larger scales to capture larger regions of image context in the descriptions of the landmarks.

CONCLUSION

This preliminary study presented a novel method (SS) for locating landmarks on cephalometric radiographs. This study tested the approach at minimum resolution to stay within initial computational and memory constraints during the development stage of SS. The mean total landmarking error for 15 landmarks at this resolution was not significantly different from that of manual identification on the computer display. The preliminary results presented here indicate that an automatic approach to landmark identification will be found. Automatic landmarking is the first step in the development of a completely automatic cephalometric analysis.

REFERENCES