Automatic Identification and Localization of Craniofacial Landmarks Using Multi Layer Neural Network

I. El-Feghi, M.A. Sid-Ahmed, and M. Ahmadi

Department of Electrical and Computer Engineering, University of Windsor,
Windsor, ON, Canada N9B 3P4
{idrise,ahmed,ahmadi}@uwindsor.ca

Abstract. Cephalometric evaluation of lateral x-rays of the skull, used mainly by orthodontists, is usually carried-out manually to locate certain craniofacial landmarks. This process is time consuming, which is both tedious and subject to human error. In this paper we propose a novel algorithm based on the use of the Multi-layer Perceptron (MLP) to locate landmarks on the digitized x-ray of the skull. The main feature of this proposed algorithm is that its performance is independent of the quality of radiographs. Preprocessing techniques are used to enhance the quality of the image and to extract the outer edges of the skull. Four points are selected to form the basis for additional features representing the size, rotation and offset of the skull. The extracted features are then used as inputs to the MLP. The corresponding outputs represent the horizontal and vertical coordinates of the selected landmark. MLP's are efficient function approximators and in this work are trained to locate landmarks by using a number of manually labeled data as a training set. After training, the MLP is used to locate landmarks on target digitized images of radiographs. The MLP is trained using 55 manually labeled images and tested on a separate set consisting of 134 images, which are not used for training. Results obtained show an improvement over template-matching and line-following techniques. This is apparently evident when the search encounters a lost tooth, cavity filling or when the image is of a low quality.

Keywords: Cephalometry, MLP, Craniofacial, Landmarks, Template-matching.

1 Introduction

Cephalometry is defined as the scientific measurement of the head usually on radiographic x-ray of a skull. This measurement is performed by orthodontists based on location of a set of agreed-upon points known as craniofacial landmarks. There are 20-30 landmarks visible on the x-ray of a human skull, which are used by orthodontists in what is known as cephalometric evaluations. Location and definition of the most commonly used landmarks as defined in [1] are shown in Fig.1. Once the landmarks are located several linear and angular measurements are performed to assess the treatment, evaluate treated cases, plan the treatment or compare the measurements with norms of populations.

R.E. Ellis and T.M. Peters (Eds.): MICCAI 2003, LNCS 2878, pp. 643–654, 2003.

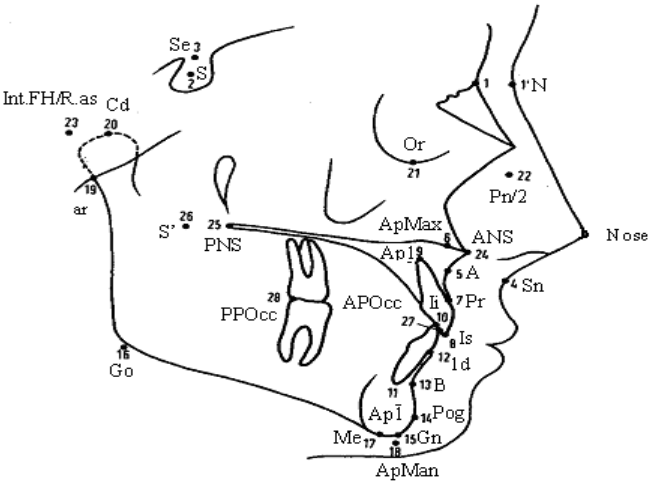

Fig. 1. The most commonly used landmarks for cephalometric evaluation as depicted on the cephalometric tracing.

Currently orthodontists locate landmarks manually by following two steps: first the orthodontist traces the x-ray then locate landmarks on line intersections or geometrical line shapes. This process requires an experienced orthodontist to spend 15-20 minutes in assessing each patient. This is a time consuming, tedious process which is subject to human error.

A computerized system that will carry out this tedious task is obviously needed and can be useful not only for cephalometric evaluation but also for keeping and organizing dental records. Automatic location of landmarks is a difficult task due to large variability in the morphology of the human head and the variation of the head position in radiographs. No two x-rays are the identical even if they are taken one after the other for the same patient. A small shift in the distance of the head from the cephalostat will result in a large change in the scale of the head while shifts in the position of the head will result in shift of the landmarks position. In this paper we present a new algorithm for the automatic location of craniofacial landmarks on digitized radiographs of the skull. The proposed algorithm is based on the use of a MLP as function approximators. The weights of the MLP are adjusted based on the location of the landmarks. The inputs to the MLP are features extracted from several radiographs. The features represent the size; rotation and shift of the skull while the outputs are the approximate location of the target landmark. The MLP is trained using manually located data. Locations of landmarks on target images (images that were not used in training) are determined by the MLP with variable weights which are set depending on the landmark to be located. The advantage of this approach it can deal with images having irregularities such as a missing tooth or a cavity filling.

Automated cephalometry has been subject to research for many years and has been attempted by several independent researchers with varying degree of success. Levy-Mandel *et al* (1986) [2] introduced the first step toward an automatic extraction of landmarks. In their study they used image enhancing techniques such as median filter

and histogram equalization to enhance the contrast and remove the noise from the image. Next, they used Mero-Vassy Operator [3] for edge detection to extract the relevant edges. Algorithm with prior knowledge is used to track lines in a predetermined order of detected lines. Positions of the landmarks are determined based on a set of predefined geometrical properties of lines, lines intersections and exterior boundaries. As an example the tip of the nose can be selected as the most anterior point on the x-ray. This system was tested on two high quality x-rays and it located 23 out of 36 landmarks. Parthasarathy *et al*(1989) [4] improved the system used in [2] by including a resolution pyramid to reduce the processing time. The resolution of the x-ray images is reduced to 64x60 pixels, once the landmarks are located this resolution is scaled back to the original size and the locations are fine-tuned. Their system was tested on five x-ray of different quality. The system was able to locate 58 % of the nine landmarks with accuracy of ±2mm. Tong *et al* (1990) [5], Davis and Forsyth (1994) [6] and Forsyth and Davis (1996)[7] presented similar algorithms for x-ray landmarking.

Algorithms presented on [2][4-7] located landmarks by edge detection, which makes their performance highly correlated to the quality of the images; moreover, not all landmarks are located on significant edges. Another problem with knowledge-base systems is their rigidity, which makes it difficult to add new rules to the system.

Cardillo and Sid-Ahmed (1994) [8] used mathematical modeling to reduce the search area for the landmark then applied a template matching techniques based on mathematical morphology to pin point the exact location of the landmarks. The algorithm was tested on 20 x-ray images and it located 76% of 20 landmarks with accuracy of less than or equal to ±2mm. The sizes of the search windows were obtained through a training algorithm.

Rudolph *et al* (1998) [9] used special spectroscopy to characterize the gray-level around landmarks from a training set located by hand. To facilitate testing they used a 'drop-one-out' scheme to enable testing one image and use the rest of the images as training set. To reduce the computation burden, they used images of size 64x64 pixel. They reported that 100% of the landmarks are located within ±4mm.

Chen *et al* (1999) [10] used neural network and genetic algorithm to search for sub images representing landmarks. Nothing was mentioned about the accuracy of the algorithm.

Hutton et *al* (2000)[11] used active shape models for cephalometric landmarking. Permissible deformations of a template were established from a training set of hand-annotated images and the resulting model was used to fit unseen images. The algorithm was tested on 63 randomly selected cephalograms. On average, 55% of 16 landmarks were within an acceptable range of ±2 mm. It was concluded that the current implementation did not give sufficient accuracy for complete automated landmarking, but could be used as a time saving tool to provide a first-estimate location of the landmarks.

Grau *et al* (2001) [12] improved the work of Cardillo *et al* [8] by using a line detection module to search for the most significant lines, such as the jaw line or nasal spine, then utilized mathematical morphology approach similar to that used by [8] for shape recognition. Twenty images were used for training and another 20 used for testing the algorithm. They reported that 90% of the 17 landmarks were located within ±2mm.

Recently, Innes, *et al* (2002)[13] used pulse coupled Neural Networks (PCNN) to highlight regions containing key craniofacial features from digital x-rays. They applied different size averaging filters prior to using the PCNN to minimize the noise in the different regions of the image. In this study a bigger set of images (109 x-rays) was used and tested on locating three landmarks with a success rate of 36.7%, 88.1% and 93.6%. Although PCNN's have shown to be capable of image smoothing and segmentation, they require a large amount of manual intervention to set the required parameters.

2 Problem Formulation

Despite the effort to automate the problem, location of landmarks is still done by hand. No automatic landmarking routine is reliable enough to be used. Most of the existing algorithms suffer one or more of the following problems

1- Low recognition rate.
2- Small number of test sets.
3- Performance is dependent on the quality of x-ray images.

Most of the proposed algorithms are based on template matching or line crossings in lines detected in the x-ray image using edge detection techniques. Therefore, results are dependent on the quality of the x-ray image. Since each landmark differs in shape and in size from one x-ray to the other, no one template can closely represent all variations of the shape of any landmark. Examples of different shapes of the landmark "PPocc" which is entry 28 in fig.1, located on the posterior point of the occlusal plane are shown in the x-ray images Fig. 2. This landmark is very difficult to locate using template matching specially when one of the most posterior molars is missing. The same shape variation applies to other landmarks such as Sella 'S' point 1 and "Is" point on fig 1.

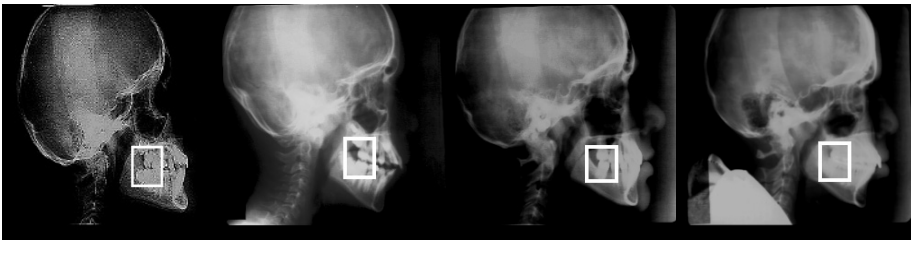

Fig. 2. Example of different shape of the same landmark on different x-rays.

Cardillo *et al* [8] solved the problem partially by dividing the shape into several smaller shapes which were then used to locate the landmark. In spite of the fact that it becomes more reliable to locate a complex shape if it was divided into several smaller shapes, the algorithm will fail completely if the shape happens not to be present in the x-ray due to an anomaly such a missing tooth. In this paper we present automatic landmark detection by utilizing an MLP neural network as function approximators. The localization of landmarks is formulated as a function approximation problem. Locations of landmarks will not depend on the quality of the x-ray but on the size,

rotation and shifts of the outer contour of the skull. We start by extracting features for the image then use these features as inputs to the MLP. The MLP is trained by a number of labeled images and then used to predict the location of the landmarks on new images (not used for training) based on the knowledge obtained during the training.

3 Features Extraction

The method developed here is based on first highlighting the features of the image representing the soft and bonny structures. Histogram equalization is applied to enhance the appearance of the images, Sobel operators is used to extract the most important edges of the skull then a 5×5 median filter is used to enhance the edges. Images are then converted to binary to show the outer edge of the skull

Four points are located on the binary image as in Fig.3. These points are located by tracing the image in the directions shown in the figure until the first or last non-zero point is found. Points (P1) and (P4) are located by tracing the image vertically from top to bottom starting from the left side for P1 and right side for P4. Point (P2) is found by extending a line from P1 until the last non-zero value is found. Point (P3) is found by scanning diagonally from right to left starting from the right hand corner. A shape similar to that shown in fig 4 is constructed and more features are made available using the extracted four points as listed in Table 1. Lines are drawn (see Fig.4) to connect P1, P2, P3, and P4 and the center of gravity of the shape (C_x, C_y) calculated from [14]:

$$C_x = \frac{1}{6A}\sum_{i=0}^{N-1}(x_i + x_{i+1}).(x_i y_{i+1} - x_{i+1} y_i) \tag{1}$$

$$C_y = \frac{1}{6A}\sum_{i=0}^{N-1}(y_i + y_{i+1}).(x_i y_{i+1} - x_{i+1} y_i) \tag{2}$$

where N the Number of Points and A is the area of the shape measured in pixels.

Table 1 gives the names and descriptions of all the measurements constructed using the mentioned five points.

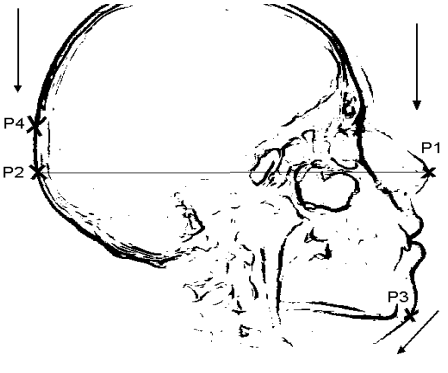

Fig. 3. Extractions of four points

4 Selection of Training Set

Since it is not feasible to train the network to map all cases in its given domain, a small set of training data is to be determined such that the network would be generalized enough and capable of learning from the training data to perform on any case with acceptable tolerance. How to select the training set to accomplish near-optimal performance plays an important role in any pattern recognition system. There is not yet an optimal algorithm for the selection of a training set available [15]. An important property of the training set is that it must cover all the expected variations in the data. A method is needed to cluster images in such a way that similar images are grouped together to form an exemplar of the human skull. To come up with the prescribed training set, images had to be clustered into several groups based on their feature vector using K-means [16-17]. In order for the K-means algorithm to converge, initial estimates of cluster centers are needed.

After extracting the features listed in Table 1, each x-ray images is represented by a vector of 21 elements. Because of the sensitivity of the clustering algorithms, all feature vectors are normalized beforehand so that parameters in the training set are uniformly distributed in the scaled range between 0 and 1.

The length of the image vector, which is the distance between the features vector and the origin of the hypercube, is computed as follows:

$$\|L\| = \sqrt{\sum_{i=1}^{i=NF} (x_i)^2} \tag{3}$$

where $\{x_1, x_2, \ldots, x_{NF}\}$ is the feature vector and *NF* is the number of features.

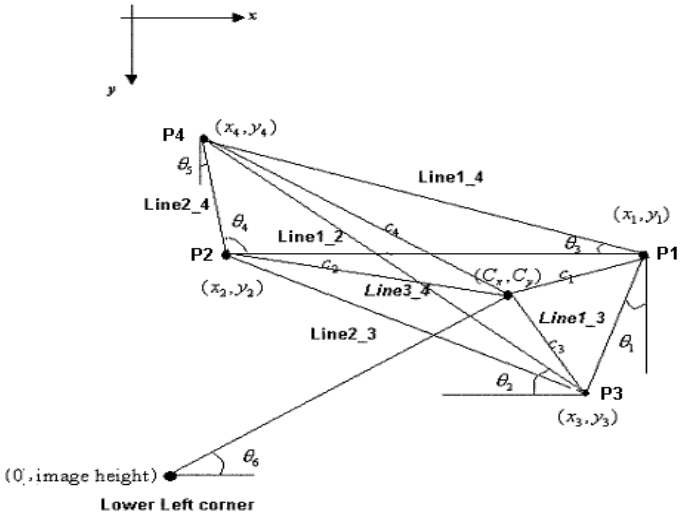

Fig. 4. The feature made from the extracted four points

Table 1. Name and description of the made features

No	Name	Description
1	Line1_2	Distance between P1 and P2
2	Line1_3	Distance between P1 and P3
3	Line1_4	Distance between P1 and P4
4	Line2_4	Distance between P2 and P4
5	Line2_3	Distance between P2 and P3
6	Line3_4	Distance between P3 and P4
7	c_x	Horizontal coordinate of the center of gravity
8	c_y	Vertical coordinate of the center of gravity
9	c_1	Distance between center of gravity and P1
10	c_2	Distance between center of gravity and P2
11	c_3	Distance between center of gravity and P3
12	c_4	Distance between center of gravity and P4
13	θ_1	Angle between Line1_3 and the vertical axis
14	θ_2	Angle between Line3_4 and the horizontal axis
15	θ_3	Angle between Line1_4 and the horizontal
16	θ_4	Angle between Line2_4 and Line1_2
17	θ_5	Angle between Line2_4 and the vertical axis.
18	Pr	Perimeter of the shape=Line1_3+Line3_2+Line2_4+Line4_1
19	A	Area of the shape.
20	θ_6	Angle between the center of gravity and the horizontal
21	Dc	Distance between center of gravity and the lower left corner of the image

The feature vectors are normalized and sorted in ascending order according to their length. These are then represented in a diagram where the horizontal axis represents the normalized length of the vectors. At each location a vertical line equal in length to the value of the normalized vector is placed to represent each image. The horizontal distance between 0 and 1 is divided into equal intervals of length 1/(number of samples). This will generate an estimate of the distribution of the images with respect to the length of their vectors. The number of samples was found through empirical trials to be 55. After several trials in which intervals with zero number of vectors were discarded and any interval with relatively high number of vectors is divided into two, it was found that our set can be initially represented by 55 groups.

Starting with 189 images of sizes ranging between 480×480 to 564×564 pixels, a smaller training set is formed using the K-means algorithm with initial estimates of centers obtained from the previous step.

The objective of clustering algorithms is to group the data set into several groups such that each group contains data with high similarity. One of the most widely used clustering algorithms is the K-means algorithm. The K-means algorithm partitions the given data set into K classes and calculates class centers so that the centers minimize an objective function which is the sum of the distances between data in a group and the group center. After applying the K-means, images are clustered into 55 groups and one image is selected from each group as a representative of that group in the training set. Landmarks in training set is located by hand.

5 MLP Training

MLP's have been widely used in function approximation problems [18-20]. Estimating the location of landmarks can be formulated as a function approximation problem, for which MLP's are well suited due to their universal approximation property. For this approach every landmark will have its unique set of weights.

The MLP is set up as a general approximator to `learn' the given input/output relations or the mapping function between the input and the output by updating the connection weights in the approximator according to the back propagation algorithm [21]. The MLP is presented with the different training item until it learns the mapping of the inputs to the proper output. The learning process is terminated when some error criterion are met. Based on the Stone-Weierstrass theorem, Cotter [22] suggested that a two-layer feedforward network could be used to approximate any continuous function by mapping variables from $\Re^n$ to $\Re^m$. The task of approximating an unknown function from input-output pairs can be formulated as follows:

The input-output pairs say $T = \{(X_p, Y_p); p = 1,..., P\}$ P is the number of training pairs.

The input vector $X_p = \{x_{p1},...,x_{pn}\}^T$ and the output vector $Y_p = \{y_{p1},..., y_{pn}\}^T$ are related by unknown function f such that $Y_p = f(X_p) + e_p$. The e_p is an error vector. The task of the MLP is to find an estimator f' of f such that some metric for the error is minimized.

In this research f is the horizontal location of the landmark as obtained by a human expert. We are seeking f' such that the predicted values obtained from f' are close to the values of f.

The proposed MLP is a three-layer network. An input layer, a hidden layer and an output layer. MLP approximates function by ensembles of simpler function as follows.

Let $f_x(X)$ and $f_y(X)$ be the functions for approximating the horizontal and vertical coordinates of a landmark, where $X = [x_1, x_2,..., x_n]^T$ is the feature vector, the goal of the function approximation is to describe $f_x(X)$ and $f_y(X)$ by combinations of simpler functions $\varphi(X)$ and $\lambda(X)$:

$$\hat{f}_x(\mathrm{X},\ \mathrm{W}) = \sum_{i=1}^{N} w_i \phi_i(\mathrm{X}) \tag{4}$$

$$\hat{f}_y(\mathrm{X},\ \mathrm{W}) = \sum_{i=1}^{N} w_i \lambda_i(\mathrm{X}) \tag{5}$$

where N is the number of ensembles and w_i are real-valued entries representing the connection weights coefficient vector $W = \{w_1, w_2, ..., w_m\}$ such that

$$\left| f_x(X) - \hat{f}_x(X,W) \right| < \varepsilon_x \tag{6}$$

$$\left| f_y(X) - \hat{f}_y(X,W) \right| < \varepsilon_y \tag{7}$$

ε_x and ε_y are minimizing a least mean squared sense to obtain an approximation for the desired output vector.

The basis function can be wavelets, sinc function or polynomial. Since $f_x(X)$ and $\lambda(X)$ are both nonlinear functions, there is no natural choice of the proper number of basis functions (for our application we used the sinc function). If too many terms are used we will be faced with an over-fitting problem and an under-fitting problem if the number is too small. The best approximation is obtained with minimal number of basis function which we can obtain by experimentation. The basis will change with input data, which means the weights in the input layer changes the orientation of the basis while the weights in the output layer find the proper amplitude of the units [23].

6 Experimental Results

After the MLP is trained using the training set, it is used to approximate the location of the landmarks based on the knowledge obtained during the training phase. The algorithm was tested on 55 x-ray images which were not used for training. Results obtained by the algorithm are compared to those obtained by a human expert. If the difference between the approximated result (obtained form the MLP) and the expected result (obtained from human expert) is less than or equal ± 2mm, then the approximation is considered successful and acceptable otherwise it is considered a failure. The algorithm was tested to locate 20 landmarks on a set of 134 images not used in training. We compare the obtained results with those reported by previous work of Cardillo [8] and Grau [12]. Cardillo *et al* tested their algorithm on 20 x-ray used a set of 20 landmarks. They obtained a recognition rate of 75 %. Grau et al [12] tested their algorithm on 20 images using a set of 20 landmarks, of which 14 are also presented in this study. In Table 2, we present a comparison of results obtained by [8] and [14]. Considering the fact that our system was tested on a larger test set, we can see that , our system outperforms the two systems.

Table 2. Comparisons of experimental results

No.	Landmark	Car-dillo [8] 1994	Grau [12] 2001	Proposed System 2003	Improve-ments over [8]	Improve-ments Over[12]
1	N	1994	2001	100	8%	6%
2	S	53%	65%	77%	24%	12%
3	Nose	94%		100%	6%	
4	Point A	77%	95%	94%	17%	-1%
5	Is	76%	100%	100%	24%	0%
6	AP $\bar{1}$	89%	100%	100%	11%	0%
7	AP 1	79%	100%	100%	21%	0%
8	Li	64%	90%	88%	24%	-2%
9	Point B	71%		85%	14%	
10	Pog	97%	95%	100%	3%	5%
11	Gn	100%	90%	100%	0%	10%
12	Go	61%	85%	87%	26%	2%
13	Me	78%	100%	84%	6%	-16%
14	Ar	89%		97%	8%	
15	Or	40%	65%	74%	34%	9%
16	SoftPog	91%		100%	9%	
17	ANS	68%	90%	92%	24%	2%
18	PNS	71%	80%	100%	29%	20%
19	APOcc	48%		68%	20%	
20	PPOcc	71%		93%	22%	
	Average	**75%**	**88.6**	**91.6**	**16%**	**3%**
	Test Set	**20**	**20**	**134**		

7 Conclusions

In this paper, we have shown that localizing the craniofacial landmarks can be formulated to a function approximation problem. MLP is used to approximate the location of the landmarks based on the knowledge obtained by using a set of labeled data for training. The method is tested to locate 20 landmarks on a set of 55 x-ray it was found that it is possible to locate most of the landmarks with average accuracy higher than 91%.

References

[1] T. Rakosi, "An Atlas and Manual of Cephalometric Radiology, London: Wolfe Medical Publications, 1982.

[2] Lévy-Mandel A, Venetsanopoulos A, Tsotsos J. "Knowledge-based Landmarking of Cephalograms," Computer *and biomedical Research*, vol. 19, pp. 282–309, June1986.

[3] Mero, L., and Z Vassy, " A Simplified and Fast Version of the Hueckel Operator for Finding Optimal Edges in Pictures," *in Proc. 14th International Joint Conference on Artificial Intelligence,* 1975, pp. 650–655.

[4] Parthasaraty S., Nugent S., Gregson P., Georgen, P. G. and Fay, D.F. "Automatic Landmarking of Cephalograms, " *Computer and biomedical research,* vol. 22(3), pp. 248–269, June 1989.

[5] W. Tong, S. T. Nugent, P. H. Gregson, G. M. Jensen, and D. F. Fay "Landmarking of Cephalograms Using a Microcomputer System," *Computer and Biomedical Research*, vol. 23(4), pp. 358–379, Aug. 1990.

[6] Davis D. and Douglas Forsyth "Knowledge-Based Cephalometric analysis: A Comparison with Clinicians using interactive Computer methods," *Computer biomedical Research,* vol. 27(3), pp. 210–228, June 1994.

[7] D. B. Forsyth and D. N. Davis "Assessment of an Automated Cephalometric Analysis System*," European Journal of Orthodontists*, vol. 18(5), pp. 471–478, Oct. 1996.

[8] J. Cardillo and M. A. Sid-Ahmed "An Image Processing System for the Automatic Extraction of Craniofacial Landmarks," *IEEE Trans. In Medical Imaging*, vol. 13(2), pp.275–289, June 1994.

[9] Rudolph. D. J, Sinclair P. M. Coggins. J. M, "Automatic Computerized Radiographic Identification of Cephalometric Landmarks,*" American Journal of Orthodontics and Dentfacial Orthopedics,* vol. 113, pp. 173–179, Feb. 1998.

[10] Yen-Ting Chen, Kuo-Sheng Cheng, and Jia-Kuang Liu "Improving Cephalogram Analysis through Feature Subimages Extraction," *IEEE Engineering in Medicine and Biology*, vol.18, pp. 25–31, Feb.1999.

[11] T.J. Hutton, S. Cunningham and P. Hammond "An evaluation of active shape models for the automatic identification of cephalometric landmarks," *The European Journal of Orthodontics*, vol. 22(5), pp. 499–508, Oct. 2000.

[12] V. Grau, M. C. Juan, C. Monserrat and C. Knoll "Automatic Localization of Cephalometric Landmarks, " *Journal of biomedical information* vol.34 pp. 146–156, Sept. 2001.

[13] Andrew Innes, Vic Ciesielski, John Mamutil and Sabu John "Landmark Detection for Cephalometric Radiology Images using Pulse Coupled Neural Networks," *International conference in computing in Communications,* pp. 391–96, June 2002.

[14] Paul Bourke (July 1988) Calculating the Area and Centroid of Polygon. [Online]. Available:http://astronomy.swin.edu.au/~pbourke/geometry/polyarea/

[15] MacQueen, J.B. "Some Methods for Classification and Analysis of Multivariate Observations*," Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, pp. 281–297. 1967.

[16] Singh, M, Patel, P., Khosla, D., Kim, T.*"* Segmentation of Functional MRI by K-means Clustering**,"** *IEEE Transactions on Nuclear Science*, vol.43 (3), pp. 2030 –2036, June 1996.

[17] M. Wann, T. Hediger, N.N. Greenbaun, "The Influence of Training Sets on Generalization in Feed-Forward Neural Networks," IJCNN, 1990, 17–21, 1990.

[18] K. Hornik, M. Stinchcombe, H. White, "Multi Layer Feedforward Networks are universal approximators, " Neural Networks, vol. 2, pp. 559–366, 1989.

[19] Thananchai Leephakpreeda, "Novel Determination of Differential-Equation Solutions: Universal Approximation Method," *Journal of Computational and Applied Mathematics*, vol.146, pp. 443–457, 2002.

[20] Franco Scarselli and Ah Chung Tsoi, "Universal Approximation Using Feedforward Neural Networks: A Survey of Some Existing Methods, and Some New Results," Neural Networks, vol. 11, No.1, pp.15–37, 1998.
[21] Rumelhart, D[.E., McClelland, J. L and the PDP Research Group *" Parallel Distributed Processing*," vol. 1 &2. MIT Press, 1986.
[22] N.E. Cotter, "The Stone–Weierstrass theorem and its Application to Neural Nets," *IEEE Trans. Neural Networks*, vol. 1(4), pp. 290–295, Dec. 1990.
[23] José. C. Principe, Neil R. Euliano and W. Curt Lefebvre, Neural and Adaptive Systems, New York: Wiley & Sons, 2000, pp. 238–242.