

Landmark Detection for Cephalometric Radiology Images using Genetic Programming

Andrew Innes*, Vic Ciesielski†, John Mamutil‡, Sabu John and Alan Harvey§

31st October 2002

Abstract

The focus of this paper is to determine whether Genetic Programming can be used to find key craniofacial features in digital X-rays. The evolved programs use a feature set of pixel statistics of regions customised to the shape of the landmark idiosyncrasies. The feature set is extracted from an input field large enough to contain the object of interest and traversed over the images. The fitness function is based on detection rate and false alarm rate. The method was tested on two landmark detection problems of increasing difficulty in terms of background noise and contrast. The image set consisted of 109 images of which 81 were used for training and the remaining 28 for testing using 4-fold cross-validation. The fittest genetic program from each cross-validation was used to locate the nose point on test images with a detection rate of 97.2% and no false alarms. The detection rate for the incisal upper incisor was 88.4% with a false alarm rate of 11.6%. The landmarks were all located within the required 2mm tolerance. Further work needs to be done on other landmarks and on increasing the detection rate and decreasing the false alarm rate, however, it appears that the genetic programming approach could be used to find many of the landmarks.

Cephalometric analysis, Genetic Programming , Object detection.

1 Introduction

Advances and affordability in digital radiographic imaging have recently seen a demand for medical professions to automate analysis and diagnosis tasks that were once performed manually. Currently a cephalometric analysis is manually intensive, and it can take an experienced orthodontist up to thirty minutes to analyse one X-ray. The analysis involves finding a number of craniofacial landmarks that are used to determine distances and angles between them. Treatment planning is based on the result of this analysis. Figure 1 shows the landmarks specific to Braces Pty Ltd cephalometric analysis, and are located in both bony structure and soft tissue.

The aim of this paper is to establish whether genetic programs can locate cephalometric landmarks in digital X-rays by using pixel level features. The two regions shown in Figure 2 each contain a cephalometric landmark, where each region increases in difficulty in terms of variability of background noise and contrast. The ‘easy’ and ‘difficult’ regions contain the nose and incisal upper incisor respectively (refer to Figure 2). These regions have been delimited heuristically by using prior knowledge of the geometry of the human face and the easy to find corner point of the ruler. For example, the tip of the nose must be around a nose length below the corner of the ruler.

Previous attempts at locating landmarks have had limited success [1, 2], but accurately locating landmarks for a large database has been unsuccessful. Realistically any method will probably not locate landmarks on all grades of digital images, as the method will be refined to work on digital X-rays for a pre-defined set of conditions such as signal-noise ratio or specific to digital imaging equipment. For example, if the method locates landmarks on digital radiographic imaging equipment, the method will most likely not locate landmarks on digitised film X-rays unless the method is refined. This paves the way for an improvement on previous attempts by developing

*Department of Mechanical and Manufacturing Engineering, RMIT University, Melbourne, Australia

†School of Computer Science and Information Technology, RMIT University, Melbourne, Australia

‡Braces Pty Ltd, 404 Windsor Road, NSW 2153, Australia

§School of Electrical and Computer Engineering, RMIT University, Melbourne, Australia

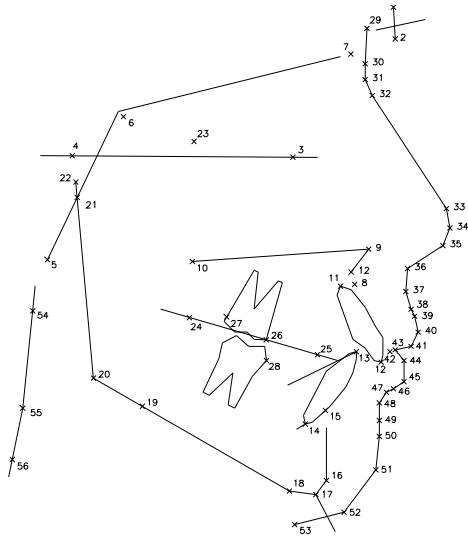


Figure 1: Line tracing of the fifty six landmarks required for a cephalometric analysis. Soft tissue landmarks are represented by points 29 -53.

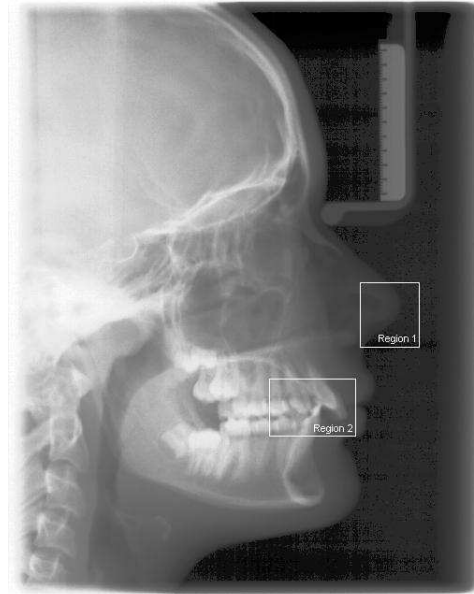


Figure 2: A digital cephalogram depicting two regions containing the mid nose and incisal upper incisor landmark.

a robust method that can locate landmarks on a large database of images within a tolerance acceptable for a cephalometric analysis. Rakosi [3] suggests when locating landmarks an error of $\pm 2\text{mm}$ is acceptable. Error is defined by the difference between the automatic landmark, and the location as specified by an experienced orthodontist.

2 Previous Work

Traditionally the location of landmarks for a cephalometric analysis was performed manually using a tracing from X-ray film. More recently the film was digitised and semi-automatic systems allowed the orthodontist to plot landmarks directly to the screen. As a natural progression and whilst not a new concept, an automated cephalometric analysis was proposed by Hussain et al. [4] in 1985.

Automatic cephalometric analysis has been attempted by no fewer than 20 independent researchers with varying degrees of success. The research can be categorised by two methodologies, i.e. prior knowledge and a learning approach. Prior to 1989, the focus was to locate landmarks using image processing techniques in conjunction with prior knowledge of the cranial structure [5, 6]. Only during the past decade has research focused on artificial intelligent approaches to assist in detecting landmarks [1, 7, 8, 2]. Although some of the recent work has produced promising results [1, 2], to date an automatic cephalometric analysis is still not forthcoming.

Earlier methods revealed handcrafted techniques could be improved using better image processing techniques, but the underlying flaw is that they are limited to a small population of images unless the code is written to handle variations in biological shapes. Some of the other shortcomings may have been a result of detail lost when digitising film X-rays, or working with low resolution images. In any case, the handcrafted methods are limited to a finite population of images defined by the code. Later attempts using learning approaches achieved better results due to a more flexible approach to the problem. The work by Cardillo et al. [1] has produced the most promising results locating 85% of landmarks within 2mm, although the results are based on a small test set. Recent research by Hutton, T.J. et al. [2] while not producing results as promising as Cardillo et al. uses a larger test set. The general consensus seems to be the underlying problems in accurately locating landmarks are the large variations in biological features, abnormalities, areas of soft tissue, and areas with subtle changes

in greyscale.

Recent methods of classifying landmarks show a greater success when using a learning approach [1, 2]. This is due to a greater ability to cope with a large variety of biological shapes. Although some of the recent methods have had limited success with small test sets [1], the size of the test set may suggest an inability to work on larger test sets. The most likely solution to improve automatic landmarking is a learning approach using prior knowledge of facial geometry with an emphasis on pre-processing. Our overall goal is to develop an automated system that will be successful on a large number of test images. The system will use prior knowledge of face geometry to determine the region in which a landmark should be located and then genetic programming to find the landmark in this region. This paper describes our work on two landmarks - the mid nose and incisal upper incisor.

3 Methodology

The method developed here is to locate landmarks in regions categorised in terms of difficulty as easy and hard (refer to Figure 2). The original image is scaled down to 20% of the original image. This reduces the number of genetic program evaluations during training and reduces the effect of the Gaussian noise on the image. The feature set is based on dividing the area surrounding a landmark into specific shapes (refer to Figure 3) individual to the landmark characteristics. The features consist of the mean and standard deviation that are calculated for each shape from pixel intensities. The genetic programming software used in these experiments is based on code to locate multiple class objects such as coins and haemorrhages and micro-haemorrhages in retina images[9].

A brief outline of the method is as follows:

1. Assemble a database of images with the known positions of landmarks to be located.
2. Use domain knowledge to extract regions in which the landmarks are expected to be. For example, we know the approximate position of the nose relative to the bottom left corner of the ruler, so we can use the ruler as a datum point and extract a sub-image of the region encompassing the nose landmark (region 1 in Figure 2).
3. Reserve some sub-images as 'unknowns' for measuring detection performance as the test set.
4. Determine the appropriate regions that are expected to generalise the training images and discriminate the object from the background (refer to Figure 3).
5. Invoke an evolutionary process to generate a program which can determine the class of an object in its input field.
6. Apply the generated program as a moving template to the reserved test images from step 3, and obtain the position of the landmark. Calculate the detection rate and the false alarm rate on the test set as the measure of performance.

3.1 Genetic Programming

3.1.1 The Terminal Set

In the context of genetic programming in object detection problems, terminals correspond to image features. Because each landmark is distinct in shape, grayscale and contrast, a different set of features or shapes is required to define each landmark. The features presented in this paper correspond to different shapes with the resulting mean and standard deviation (refer to figure 3). In addition to these features, a terminal that generates a random number in the range of [0,255] was included.

3.1.2 The Function Set

The function set $\{+, -, *, /\}$ represents four arithmetic operations. The $+$, $-$ and $*$ have their usual meanings, while $/$ represents a protected division which is the usual division operator except that a divide by zero produces zero. Additional arithmetic operators have been considered, but previous work by Zhang [10] has shown that additional operators (dabs, sin and exp) in the function set did not improve detection rate. Convergence was

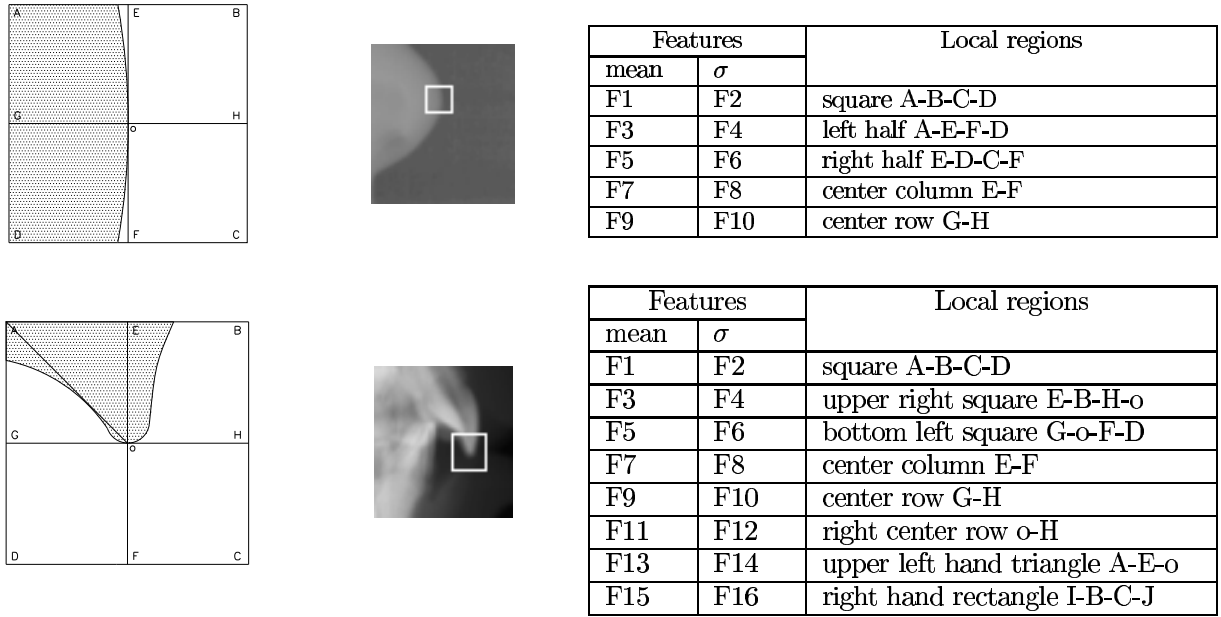


Figure 3: The two diagrams in the left column depict the map used to extract the features for the nose and incisal upper incisor landmark. The features consist of the mean and standard deviation calculated for each shape from grey level intensities. The corresponding pictures in the middle show the map size relative to the sub-image. The map dimensions (square size) used on the nose and incisal upper incisor was 10 pixels and 14 pixels respectively.

also shown to be slightly slower when training to detect objects in easy pictures, however additional operators were shown to improve rate of convergence when training to detect objects in more difficult images.

The output of the genetic program, *ProgOut*, is a floating point number indicating if the position in the image is an object of interest or background. If *ProgOut* < 0 the position is classified as background, otherwise the position is an object. During training an object can include both true positives and false positives, while during testing the object is found using the highest value of *ProgOut*.

3.1.3 The Fitness Function

The fitness of a program during training is calculated by using detection rate, false alarm rate and distance error (found location-actual location) on training images. The fitness is calculated as follows:

1. Apply the program as a moving window to each training image and obtain the output value (*ProgOut*) of the program at each pixel location. Label each position as per the object classifier defined in Section 3.1.2.
2. Match the detected object with the known location of the landmark. A match (true positive) occurs for the highest output from *ProgOut* within a set TOLERANCE of the known location. Pixels marked as objects within SQUARE_SIZE/2 of the true positive are disregarded as objects. SQUARE_SIZE is a constant corresponding to the size of the square used to define the feature set (refer to figure 3)
3. The remaining matches are marked as false positives.
4. Compute the fitness as per equation 1.

$$fitness(Fr, Dr) = A \times Fr + B \times (1 - Dr(1 - DP)) \quad (1)$$

where FR is the false alarm rate, DR is the detection rate, A and B are constants that reflect the importance of the false alarm rate and detection rate and DP is a distance penalty.

The fitness function defined in equation 1 calculates the detection accuracy of the program. The fitness function is constructed so as program fitness increases the fitness function approaches zero. The fitness

graph in figure 4 reflects five runs of the genetic program for the nose landmark over 100 generations reflecting optimal training was not achieved. It is worth noting that after 60 generations an improvement in program fitness was marginal. The fittest program at the end of 100 generations resulted in a detection rate of 97.5% when tested on training images.

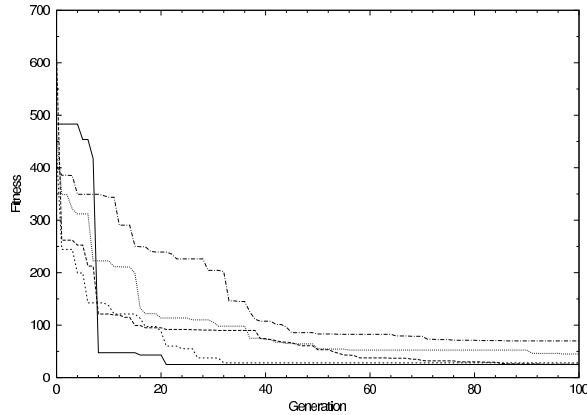


Figure 4: Fitness graph of genetic programs for 5 runs over 100 generations.

3.1.4 Genetic Programming Parameters

The values used in the genetic programming training are defined in Table 1. Population size is the number of individuals in the population, Elitism (%) is the percentage of best individuals in the current population copied to the next generation, Cross rate (%) is the percentage of individuals to be produced by cross over, Mutation rate (%) is the percentage of individuals to be created by mutation, Cross chance term (%) is the probability that in a crossover operation two terminals will be swapped, Cross chance func (%) is the probability that in crossover operation random subtrees will be swapped, Maximum depth is the maximum depth allowed for programs, Maximum generations is the termination of the evolutionary process. For definitions of A and B refer to section 3.1.3.

Parameters	Region 1 (Easy) Nose	Region 2 (Hard) Incisal upper incisor
Population size	100	100
Elitism (%)	10	10
Cross rate (%)	70	70
Mutation rate (%)	20	20
Cross chance term (%)	15	15
Cross chance func (%)	85	85
Maximum depth	6	6
Maximum generations	100	100
A	50	50
B	1000	1000
Tolerance (pixels)	2	2
Square size (pixels)	10	14

Table 1: Parameters used for GP training for the nose and incisal upper incisor landmark.

4 Results

To determine whether genetic programming can locate landmarks, two landmarks were selected. An easy landmark (nose) located on the boundary of soft tissue and background, and one landmark of medium difficulty

(incisal upper incisor) located on the edge of bone and soft tissue. A landmark is classified as found if the genetic program locates the landmark position within 2mm (5pixels) of the actual position as found by an orthodontist. If the landmark position is not within 2mm of the actual position the landmark is recorded as a false alarm. When a landmark has not been found by the genetic program it is unclassified (neither found or false alarm). Training and testing results are based on 4-fold cross-validation.

4.1 Region 1 (Nose mid)

The genetic program in Figure 6 is indicative of the program used for testing during cross-validation. Training results achieved a detection rate of 98.7% and false alarm rate of 0.93%. The detection rate of test images produced a detection rate of 97.2% with no false alarms. The unclassified images are a result of the nose located on the edge of the image (refer to images in Figure 5). Images in Figure 6 are indicative of the type of images from the test set containing the nose landmark. The white dot corresponds to the location of the nose as found by the genetic program.

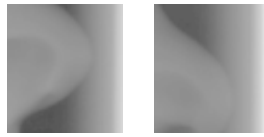


Figure 5: Two images from the training set showing the nose landmark located on the right most edge of the image. The genetic program failed to locate the nose landmark for both of these images .

(+ (* (+ (- (* (+ f1 f6) f1) (* 148.222 f5)) (* (- f6 (+ f6 f4)) (+ 11.904 f1))) (* (/ (/ f4 (* f8 f2)) (* f6 f2)) f3))
 (* (+ (- (* (+ f7 f6) f1) (* (+ f7 f6) f5)) (* (- (/ f10 f3) (+ f6 f4)) (+ 11.904 f1))) (* (/ (* f3 f1) (* f8 f2)) f3)))

Figure 6: A generated program for the nose landmark.

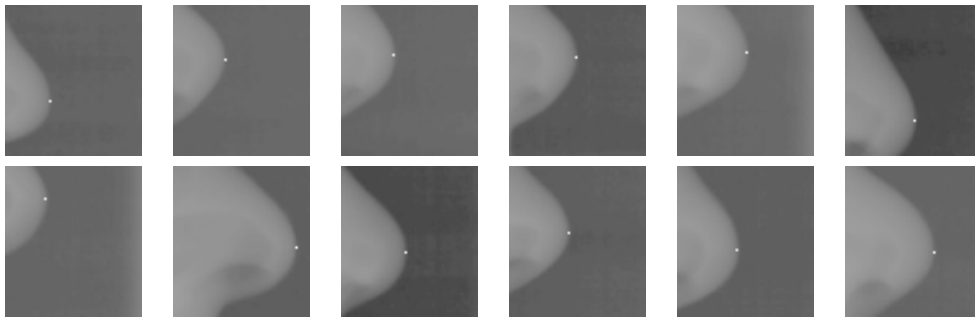


Figure 7: The 12 images are indicative of the variations in shape surrounding the nose. The white point indicates the position found using the program shown in Figure 6. The genetic program was successful for all of these images.

	Training	Testing
No. of objects	81	27
No. of objects found	80	26.25
No. of false alarms	0.75	0
No. unclassified	0.25	0.75
Detection rate (%)	98.7	97.2
False alarm rate (%)	0.93	0

Table 2: Training and testing results for the nose landmark using 4-fold cross-validation.

4.2 Region 2 (Incisal upper incisor)

The genetic program in Figure 8 is a program used for cross-validation. Training results from 4-fold cross-validation produced a detection rate and false alarm rate of 92.3% and 7.7% respectively. Results from test images unseen during training produced a detection rate of 88.4% and false alarm rate of 11.6%. The images in Figure 9 is indicative of the types of images containing the incisal upper incisor from the test set. The white dot corresponds to the location of the incisal upper incisor landmark as found by the genetic program. The detection performance was better on images that contained overbite as opposed to images that contained underbite. The results for the incisal upper incisor are not as good as for the nose landmark. This is expected as there is clearly more variation and complexity in grey levels surrounding this point.

```
(- (- f8 (- (- (* f15 f11) f12) (- f8 (* (- 195.687 f5) 195.687)))) (* (- f13 (+ (* (+ 126.29 f6) (- f16 f4)) (+ (* f10 f7) (+ f14 f16)))) (- (- (/ (+ 126.29 f6) f15) (* f8 f12)) f9)))
```

Figure 8: A generated program for the incisal upper incisor.

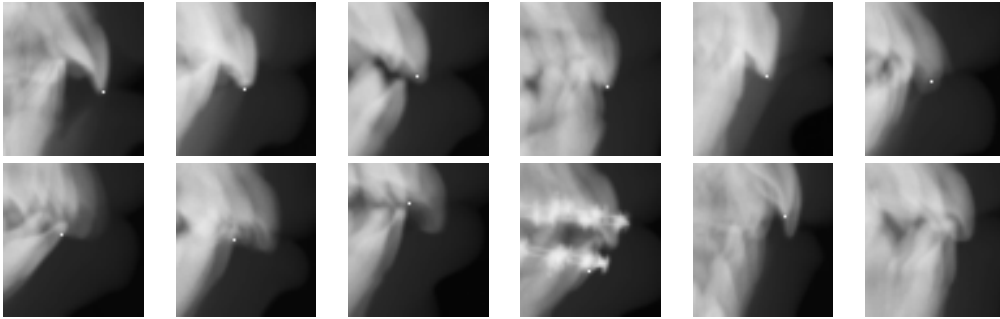


Figure 9: The 12 images are indicative of the variations in shape surrounding the incisal upper incisor. The white point in the images indicates the position found by the program. The top row are images where the program found the landmark and the bottom row are false alarms with one image unclassified.

	Training	Testing
No. of objects	81	28
No. of objects found	74.75	24.75
No. of false alarms	6.25	3.25
Detection rate (%)	92.3	88.4
False alarm rate (%)	7.7	11.6

Table 3: Training and testing results for incisal upper incisor landmark using 4-fold cross-validation.

5 Discussion

An evolved program can be viewed as a classifier which discriminates a landmark of interest (class 1) from the other points (class 2) in the images. Thus there will be a decision boundary in feature space between these two classes. The boundary is the multi-variate polynomial represented by the program. It is reasonable to ask whether the same detection performance could be achieved with a simpler boundary. Earlier work [10] suggests that a complex boundary is needed to avoid a high false positive rate. In this work, neural network classifiers trained by backpropagation were used in place of genetic programs. The networks were able to find objects of interest but gave a significant number of false positives. The false positive rate could be reduced by subsequent refinement of the weights using a genetic algorithm with a similar fitness function to equation 1. It appears that a complex decision boundary found by penalising false positives during training is necessary for accurate detectors.

6 Conclusion

Preliminary results have shown that genetic programming was able to successfully evolve a program for the nose landmark and with limited success for the incisal upper incisor. The detection rate for the nose and incisal upper incisor landmark was 97.2% and 88.4% respectively. False alarms for the incisal upper incisor landmark was 11.6% and the genetic program performed better on images that contain overbite. Improving the detection rate of the incisal upper incisor is the objective, but it is equally important to eliminate false alarms. To reduce the number of false alarms and improve detection rate more features are required, but more importantly having quality features that can distinguish between background and landmark. Currently the shapes of these features are manually created to accommodate individual characteristics of the respective landmarks. To improve this process, a method is required to evolve these shapes automatically.

Acknowledgments

This work is funded by the Australian Research Council (ARC) SPIRT grant scheme in partnership with Braces Pty Ltd from grant no. C00107119.

References

- [1] J. Cardillo and M.A. Sid-Ahmed, "An image processing system for locating craniofacial landmarks," *IEEE transaction on Medical Imaging June 1994*, vol. 13, no. 2, pp. 275–289, 1994.
- [2] T.J. Hutton, S. Cunningham, and P. Hammond, "An evaluation of active shape models for the automatic identification of cephalometric landmarks," *European Journal of Orthodontic*, vol. 22, no. 5, pp. 499–508, October 2000.
- [3] T. Rakosi, *An atlas of cephalometric radiography*, Wolfe Medical Publications, London, 1982.
- [4] Z. Hussain and H.H.S. Ip, "Automatic identification of cephalometric features on skull radiographs," *ACTA polytechnica Scandinavica-Applied physics series*, , no. 150, pp. 194–197, 1985.
- [5] A.D. Levymandel, A.N. Venetsanopoulos, and J.K. Tsutos, "Knowledge-based landmarking of cephalograms," *Computers and Biomedical Research*, vol. 19, no. 3, pp. 282–309, June 1986.
- [6] W. Tong, S.T. Nugen, P.H. Gregson, G.M. Jensen, and D.F. Fay, "Landmarking of cephalograms using a microcomputer system," *Computers and Biomedical research*, vol. 23, no. 4, pp. 358–379, August 1990.
- [7] Y. Chen, K. Cheng, and J. Liu, "Improving cephalogram analysis through feature subimage extraction," *IEEE engineering in Medicine and Biology*, pp. 25–31, 1999.
- [8] M. Desvignes, B. Romaniuk, R. Demoment, M. Revenu, and M.J. Deshayes, "Computer assisted landmarking of cephalometric radiographs," .
- [9] M. Zhang and V. Ciesielski, "Genetic programming for multiple class object detection," in *Proceedings of the 12th Australian Joint Conference on Artificial Intelligence*. 1999, pp. 180–191, Springer-Verlag.
- [10] M Zhang, *A domain independent approach to 2D object detection based on the neural and genetic paradigms*, Ph.D. thesis, RMIT University, 2000.