CHAPTER 7

CEPHALOMETRIC LANDMARK DETECTION USING ANGULAR RADIAL TRANSFORM AND HYBRID LOCAL TEMPLATE MATCHING

7.1. Introduction

The most difficult step in image pattern recognition is the extraction of desired patterns from a background of the large amount of irrelevant information. This is particularly true with respect to cephalometric image analysis owing to the following reasons.

i. Different three dimensional objects from the skull can yield identical two dimensional projections on cephalograms.
ii. The objects of interest can have several different visual appearances.
iii. Edge blurring, shadows and noise make visualization and extraction of the delicate anatomical structures more difficult.

In this Chapter, we propose a three stage algorithm to improve the accuracy of the landmark localization, speed of computation and robustness to image noise. The method can abstract from the background the most meaningful information about each landmark, in presence of image clutter, occlusion, transformations (affine, scale, rotation), nonuniform illumination, and variability in shape. In the first stage, the search window for each landmark location is identified using a global region based shape descriptor utilizing ART coefficients and Procrustes analysis for shape alignment. ART has the same characteristics as the ZMs, minimum information redundancy, and robustness to image noise and minor nonrigid deformations and invariance to rotation. In addition, it has low order of computation complexities both for extraction and matching. In the second stage, the system applies a DoG (Difference of Gaussian) filter, which acts as a band pass filter to each search window to suppress the high-frequency noise and low frequency nonuniform illumination while retaining the desired spatial information.

In the last stage, local template matching is applied to each search window corresponding to each landmark to find its exact location. No single template matching technique based on either the spatial information or some feature is robust for complex cephalometric analysis. We have made our descriptor more accurate and

robust by combining spatial and feature based local template matching, robust in presence of data perturbations. The spatial techniques used are NCC, SSD and a measure based on their fusion (FSM).

For feature based template matching angular radial features are used with a robust distance measure based on both magnitude and phase. The concept of multiple templates is used to accommodate the variability exhibited by the appearance of complex cephalometric structures and instead of using the template of the entire object, hierarchical templates are used to reduce the effect of distortions. The major disadvantage of the template matching techniques and using multiple templates is high computational cost, which is reduced by using local template matching implemented using faster techniques. Normalized correlation and SSD are computed using fast Fourier transform (FFT). A coarse to find template matching strategy is used to further reduce the computational complexity.

7.2. Methodology

To improve the landmark localization a three stage process is followed. If the search for the best match location is done exhaustively over the entire search space, the process is computationally expensive and the rate of false detection will rise due to the presence of similar structures at different locations of the image. Due to this, a small search window for each landmark is identified automatically using ART based image registration. Region based preprocessing is then applied to each search to suppress noise and nonuniform illumination effects. Finally, a hybrid local template matching based on spatial information and ART coefficients is applied to each search window for landmark location detection.

7.2.1. Estimation of search window for each landmark using angular radial transform and shape alignment

Cephalometric landmark detection is a real time application so fast computation is an important issue. Thus in this Chapter we use ART for cephalometric registration process. The ART features have the ability to describe the complex objects effectively and are invariant under rotation, scaling, translation and noise [113-115].

Step for estimating initial landmark locations

Step I: Feature Detection

This step works in two modes (offline and online). In offline mode, the region based ART features of the training set images are computed and saved for future reference. In online mode the ART features of the query image are computed.

Step II: Matching of the ART Descriptors

The correspondence between the ART features in the training and test images is established using a rigorously founded approach, which takes both magnitude and phase information into consideration. The distance measure used is similar to the one proposed in [116] for Zernike moments. This distance measure provides a more accurate rotation invariant similarity measure than considering only the Euclidean distance between magnitude part while keeping the computational complexity same. Using the phase information with the magnitude improves the distance measure in terms of robustness against geometric deformation. Distance between two images I and J is calculated as follows

$$D_{I,J}^2(\phi) = \sum_{(p,q)\epsilon D} \sum \frac{\pi}{p+1}\left[\left|A_{pq}^I\right|^2 + \left|A_{pq}^J\right|^2 - 2\left|A_{pq}^I A_{pq}^J\right| cos\left(q\phi + \left[A_{pq}^J\right] - \left[A_{pq}^I\right]\right)\right] \quad (7.1)$$

Where A_{pq}^I and A_{pq}^J represents the ART coefficients with order p and repetition q of Image I and J.

Step III: Image Transformation

The 5% most similar training images to the query image are selected.

Step IV: Shape Alignment

The selected images from step 3 are aligned with the input image by Procrustes algorithm that iteratively computes the mean shape and aligns all the examples to it until convergence. The mean shape is used to extract the initial approximate location of landmarks.

The shapes from each of the five training set images are aligned, and their average shape is used as the approximate location of landmarks on query image. These locations are used to find search windows for landmarks in the query image.

i. Initially, one shape is considered the mean shape and rest of the shapes are aligned with this mean shape.
ii. Next the average of the aligned shapes is taken as the mean shape and all the shapes are again aligned with the mean shape till the mean does not change.

Following steps are considered to align two shapes:

Consider z' with coordinates(x', y') as the mean shape and shapes z with coordinates(x, y) as the shape to be aligned with shape z'.

Calculate the value of $a, b, t_x,$ and t_y, as given in the following equations

$$t_x = \frac{1}{n}\sum x_i' \tag{7.2}$$

$$t_y = \frac{1}{n}\sum y_i' \tag{7.3}$$

$$a = \left(\sum x_i\, x_i' + \sum y_i\, y_i'\right)/\left(\sum x_i^2 + \sum y_i^2\right) \tag{7.4}$$

$$b = \left(\sum x_i\, y_i' + \sum y_i\, x_i'\right)/\left(\sum x_i^2 + \sum y_i^2\right) \tag{7.5}$$

From these values of a, b , $t_x,$ and t_y, compute values of s and θ as follows

$$s^2 = a^2 + b^2 \tag{7.6}$$

$$\theta = tan^{-1}(b/a) \tag{7.7}$$

Align shape z using the following equation

$$T(Z) = sA(z) + tr \tag{7.8}$$

Where s is scale, A is rotation matrix and tr is the translation.

$$A = \begin{bmatrix} cos(\theta) & -sin(\theta) \\ sin(\theta) & cos(\theta) \end{bmatrix} \tag{7.9}$$

$$tr = \begin{pmatrix} t_x \\ t_y \end{pmatrix} \tag{7.10}$$

7.2.2. Noise and nonuniform illumination suppression

In most template matching techniques, the homogeneity of the image in the training set improves the robustness of these methods. However, cephalograms X-rays lack this property and may suffer from noise and luminance variance. The performance of most existing algorithms is sensitive to illumination variation present on the lower side of cephalograms, which affect the extraction of landmarks like Pog, Me and Rgn. Certain landmarks like Ans are cluttered with unwanted information around them. Illumination is usually supposed to change smoothly across the object so that illumination is more responsible for low frequency property of real image while reflectance is more responsible for high-frequency property of real image. Since reflectance is independent of illumination, reflectance image is illumination normalized. A common approach to address the effects of varying lighting conditions is to preprocess to normalize illumination [117]. Widely used normalization techniques include HE, histogram matching, gamma intensity correction, etc. These

normalization techniques can be applied to an image either globally or regionally. The region based normalization leads to higher identification accuracy than the traditional global normalization.

In the present case, we need a filter that suppresses low frequency component corresponding to illumination and at the same time high frequencies corresponding to noise while still preserving the necessary information from the input radiograph for successful landmark position extraction. DoG is a grayscale image enhancement algorithm that involves the subtraction of one blurred version $f_2(x, y)$ of an original grayscale image $I(x, y)$ from another, less blurred version $f_1(x, y)$ of the original image. The blurred images (f_1, f_2) are obtained by convolving the original grayscale image with Gaussian kernels having differing SDs (σ_1, σ_2) subtracting one image from the other preserves spatial information that lies between the range of frequencies that are preserved in the two blurred images. Bandpass filter suppresses low frequencies ("lighting variations") and high frequencies ("noise") as can be seen in Fig. 7.1.

2-D Gaussian distribution is represented as

$$G_\sigma(x, y) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left[\frac{-x^2 + y^2}{2\sigma^2}\right] \tag{7.11}$$

DoG filter is given by

$$DoG = G_{\sigma 1}(x, y) - G_{\sigma 2}(x, y) \tag{7.12}$$

The enhanced output file $f_o(x, y)$ is obtained by convolving the input image $f(x, y)$ with the DoG kernel.

$$f_o(x, y) = DoG^* f(x, y) \tag{7.13}$$

Except one landmark Co, the location accuracy of all landmarks improves after applying DoG.

Figure 7.1: (a) Region around landmark Me and Pog with illumination effects (b) Low contrast and noisy region near landmark Prn (c) Result of applying DoG to Pog region (d) and Result of applying DoG to Prn region.

7.2.3. Robust local template matching based on fused NCC and SSD

Template matching is a fundamental problem in pattern recognition, which has a wide application in object detection and object tracking. The traditional template matching method uses an object image as a template image and moves the template pixel by pixel over the search image. Similarity measure like normalized correlation value between the template image and the overlapped search image is calculated, to find the position that has the highest normalized correlation value. NCC, SSD is widely used in a pixel by pixel template matching.

A feature based pattern matching helps in the reduction of computational burden in the matching process besides different features having different properties like scale and rotation invariance. However, the performance of a feature based techniques degrades significantly under changes of contrast, presence of false edges, shadowing, nonuniform illumination, etc. Thus in the proposed work, we use spatial information. SSD can be viewed as the squared Euclidean distance and is calculated as follows

$$ssd[i,j] = \sum_{x=-N/2}^{x<N/2} \sum_{y=-N/2}^{y<N/2} (I_{(i+x,j+y)} - T_{(x+N/2,y+N/2)})^2 \quad) \tag{7.14}$$

Where I is the test is image and T is the template image

NCC is computed as follows:

$$\text{ncc[i, j]} = \frac{\sum_{x=-N/2}^{x<N/2} \sum_{y=-N/2}^{y<N/2} \left(T_{(x+N/2,y+N/2)} - \bar{T}\right)(I_{(i+x,j+y)} - \bar{I})}{\sqrt{\sum_{x=-N/2}^{x<N/2} \sum_{y=-N/2}^{y<N/2} \left(T_{(x+N/2,y+N/2)} - \bar{T}\right)^2 (I_{(i+x,j+y)} - \bar{I})^2}} \tag{7.15}$$

where $\bar{I}$ is the mean of test image and $\bar{T}$ is the mean of template image.

NCC captures the statistical difference effectively. However, it fails, when either of the windows contains constant image intensity. This results in division by zero [35]. NCC is invariant to linear brightness changes, but is very sensitive image clutter, occlusion and nonlinear contrast changes. SSD is a quick solution for comparing templates. SSD gives more stable solutions than NCC in presence of clutter and occlusion [119]. SSD is very sensitive to few pixels with large intensity differences between the two compared images. Thus, a single similarity measure may result in suboptimal solution. In case of cephalometric radiographs, the shape variability is high hence instead of using the complete shape a sub-part of the shape is used for template matching. The hierarchical template matching is less affected by shape distortions.

In our proposed algorithm, information from SSD and NCC based template matching is fused to improve the recognition rate of the algorithm. The FSM is as defined below

$$FSM = \frac{[ncc + (1 - ssd)]}{2} \tag{7.16}$$

FSM gives the best match at a location of maximum correlation and minimum distance. However, depending on the radiographic complexity of the region around each landmark, one of the two measures (NCC, SSD) may fail owing to the variability in shape, contrast, illumination effects in the template and test image. However, in such a case the best match location given by FSM is closer to the true location, and its value will be close to either NCC or SSD. In case of very complex structures, these three measures may point to three different locations. One of these locations is the actual landmark position.

Multiple templates are used to accommodate the variability exhibited by the appearance of complex cephalometric structures. In the proposed algorithm, four templates are used for each landmark. Thus, the template matching gives four best match locations corresponding to each landmark. In case of only one best match the values corresponding to different templates and measures may overlap. Else they can be 12 (4×3) unique locations by considering four templates and three measures. To find the best match from these probable locations ranging from 1 to 12 ART feature based pattern matching method is used. This process of using multiple measures with ART features makes the technique robust and improves the recognition rate even for poor-quality cephalograms.

For template matching in search windows, a coarse to fine strategy is used. In step I the template is moved row and column wise in steps of ten pixels and in Step II the template is moved in single pixel steps. The procedure to extract a landmark from the search window is given below:

Step I: Coarse template matching

FOR each template

i. Save the location of maximum FSM and NCC and minimum SSD by moving the template in steps of ten pixels.
ii. If the search window does not suffer from any distortion, these three locations overlap, else they may lie at different location.
iii. Save each unique location.

END

Step II: Refining the best match by ART features

i. Compute ART features for each template in the set and take their average value.

ii. Around each unique landmark location extracted in step, I select a sub-window of the size of the template and compute ART features.

iii. Compare the ART features of each location with template features using the following distance measure using both magnitude and phase

$$D^2_{I,J}(\phi) = \sum_{(p,q)\in D} \sum \frac{\pi}{p+1}\left[\left|A^I_{pq}\right|^2 + \left|A^J_{pq}\right|^2 - 2\left|A^I_{pq}A^J_{pq}\right| cos\left(q\phi + \left[A^J_{pq}\right] - \left[A^I_{pq}\right]\right)\right] \quad (7.17)$$

The location with smallest distance is considered the best match for landmark position. The process is repeated for each landmark.

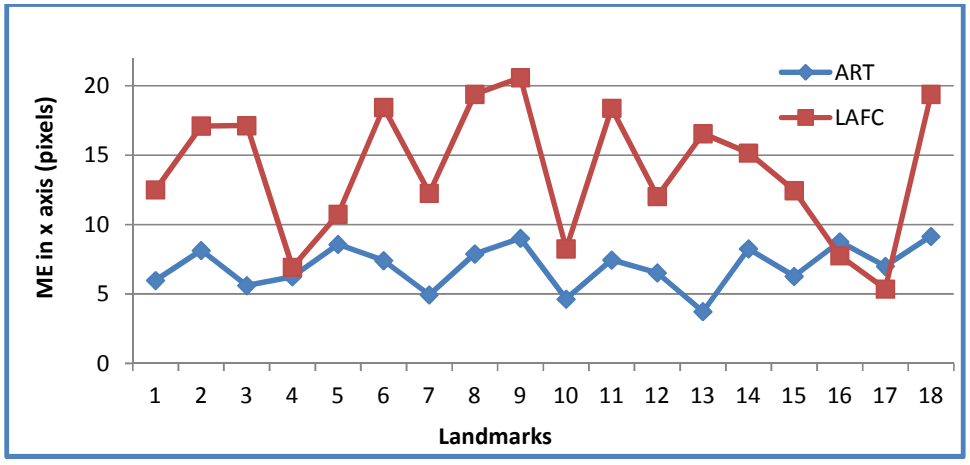
Step III: Fine local template matching

i. Consider a smaller sub-window around each landmark position obtained in step I.

ii. Repeated the process of step I by considering a smaller and more localized template for each landmark and moving the template in steps of single pixel to obtain the exact landmark location.

7.3. Results and Discussions

Eighteen landmarks are selected for this experiment as shown in Table 6.2. If the difference between the two is less than ±2 mm, it is considered successful. Total 85 training images were collected randomly from the data set without judgment of their quality, sex, and age. The system was tested using the drop-one-out algorithm. Each time 84 images were used for training, and one image was excluded for testing. Most existing methods use line and angular features (LAF) for pattern matching. These features are computed using few reference landmarks, found using image enhancement and edge detection techniques. These features are not very robust as evident from Fig. 7.2 and Fig. 7.3. ART feature based pattern matching technique used in the proposed algorithm gives significantly improved results. The approximated landmark positions in the query image obtained with the help of training set images are very close to the true positions. The mean error is below 10 mm in both x and y direction, which is not so with LAFC features, which gives MD as high as 22 mm in x direction and around 13 mm in y direction for some landmarks.



* LAFC: Length and angular features using automatically computed reference landmarks

Figure 7.2: ME (pixels) in x direction for the two methods for 18 landmarks.

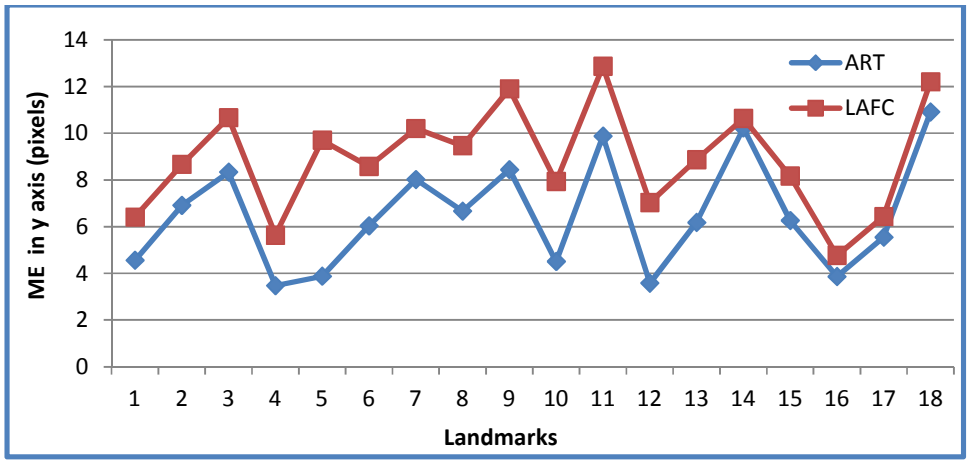


Figure 7.3: ME (pixels) in y direction for the two methods for 18 landmarks.

The quantitative results for the 18 landmarks in terms of percentage of success rate and the comparison with other cephalometric landmark detection algorithms are given in Table 7.1. 92% of the localization of 18 selected landmarks is within a window of $\leq \pm 2$ mm. The results for some typical landmarks like $Sella$, Or and Pr located within the confines of the skull with greater likelihood of being noisy are detected with high accuracy by the proposed algorithm. Results for soft tissue landmarks lying on low contrast regions (like Prn, Ls, Li) are quite promising.

Table 7.1: Accuracy of landmarks in percentage.

Method	Cardillo [5]	Eli-Feghi [22]	Grau [24]	Yue [17]	Proposed Algorithm (ZM)	Proposed Algorithm (ART)
Landmarks	Percentage (%) of landmarks detected successful					
Pr	-	-	-	-	95	97
Sella	53	77	65	76	90	93
Na	83	91	85	86	100	98
Or	40	74	65	-	85	85
Pns	86	100	80	83	88	95
Ans	68	92	90	79	86	95
Go	61	87	85	86	100	94
Uie	79	81		90	95	95
Me	78	84	100	98	84	90
Sor	-	-	-	83	80	90
Cd	89	91	-	69	72	85
Prn	94	100	-	94	85	95
Rgn	-	-	-	86	85	85
Pog	97	100	95	-	77	90
A	77	94	95	-	85	90
B	71	85	-	-	100	100
Ls	-	-	-	-	100	100
Li	-	-	-	-	95	95

The result in terms of accuracy corresponding to six landmarks that are common to all the studies are given in Fig. 7.4. It can be observed that except the proposed algorithm, the other techniques give high accuracy for certain landmarks, but the accuracy falls sharply for other landmarks.

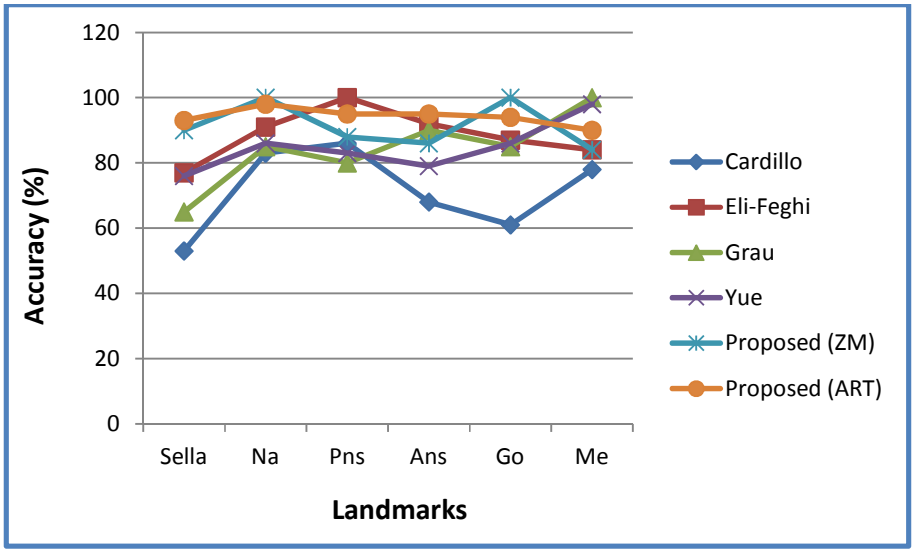


Figure 7.4: Assessment of accuracy of six landmarks common in all the studies.

The ME and SD of mean location error for each landmark in x and y direction is given in Table 7.2. Average of x and y direction ME and its SD for each landmark as obtained by the proposed algorithm and previous research methods when reported are given in Table 7.3.

Table 7.2: ME and SD for each landmark in x and y direction in mm.

Landmarks	ME (x axis)	SD (x axis)	ME (y axis)	SD (y axis)
Pr	1.68	0.97	0.98	0.97
Sella	2.00	1.28	0.92	0.62
Na	1.78	1.73	1.78	1.81
Or	1.32	1.11	2.14	1.68
Pns	1.92	1.57	1.92	1.49
Ans	1.99	1.37	1.36	2.00
Go	1.80	2.00	1.60	1.24
Uie	1.29	0.78	1.66	1.90
Me	1.15	0.87	1.42	0.84
Sor	1.94	1.94	2.37	2.54
Cd	2.55	1.32	1.31	1.39
Prn	1.03	0.90	1.19	0.75
Rgn	2.05	1.12	1.80	1.12
Pog	1.77	0.89	0.66	0.38
A	1.28	0.87	1.49	1.04
B	1.66	1.95	1.33	1.06
Ls	1.19	1.09	1.56	0.90
Li	1.77	0.92	0.85	0.87

Table 7.3: Comparison of recognition results in terms of ME and SD in mm when reported.

Method	Parthasa-rathy [9]	Tong [10]	Rudolph [14]	Hutton [16]	Grau [24]	Saad [20]	Proposed Method
Landmark	ME	ME	ME± SD	ME± SD	ME	ME± SD	ME± SD
Pr		5.13	5.67±3.37	7.3±6.5		3.48±2.46	1.33±0.97
Sella	1.41		5.06±3.37	5.5±6.8	1.92	3.24±2.85	1.46±0.95
Na	1.83		2.57±2.18	5.6±2.4	1.40	2.97±1.85	1.78±1.77
Or		1.28	2.46±3.77	5.5±3.4	1.92	3.42±2.43	1.75±1.34
Pns	2.16			5.0±4.1	1.32	3.04±1.38	1.78±1.53
Ans	2.36		2.64±3.06	3.8±2.2	0.75	2.70±1.05	1.49±1.50
Go	1.00	2.72		5.8±6.0	1.11	3.64±1.76	1.70±1.08
Uie	2.09	1.15	2.02±1.99	±	0.90	3.65±1.60	1.47±1.34
Me			3.09±3.46	2.7±3.6	0.48	4.40±2.03	1.28±0.85
Sor			1.85±2.26	2.7±3.4		3.66±1.74	2.15±2.04
Cd						2.55±0.97	2.05±2.09
Prn						2.23±1.24	1.11±0.82
Rgn							1.00±1.12
Pog	2.91						0.71±0.63
A		0.78	2.33±2.63	3.3±2.4	0.90		1.38±0.97
B	3.29		1.85±2.09	2.6±2.7	0.95		1.65±1.68
Ls		0.54					1.37±0.99
Li		0.34					1.31±0.89

7.4. Conclusions

The comparative results show that the ME, and SD for each landmark achieved using the proposed algorithm is either similar or show improvement over previous work. The bar graph of Fig. 7.5 compares the variability in landmark location using the proposed algorithm with the observer marked landmarks. As referred in [1] Miethke (1989) observed that some landmarks were localized more exactly by the observers than the other landmarks. A variability of ±2 mm was observed in most of the 33 landmarks they considered in their study. Out of the 33 landmarks used in this study 25% showed a variation of more than ±2 mm. Thus, a landmark detected within a range of ±2 mm is considered accurately detected. The high variability observed in some landmarks is due to high complexity of the region around them, ambiguity in their definition like landmarks identified as a point of change between convexity and concavity are generally unreliable, and observers lack of training and experience, working conditions, etc. The complex landmarks like Pr and Or and geometrically located landmarks like A and B are detected more reliable by the proposed algorithm than the human observer. However, landmarks like $Sella$ and Uie which can easily be detected by the humans are detected with a little higher variability by the proposed algorithm. As for finding the exact centre for $Sella$ needs further processing besides template matching. The landmark Uie has a very high variability as teeth move and grow in different direction they maybe braces on the teeth or blurring caused by patient movement (Fig. 7.6).

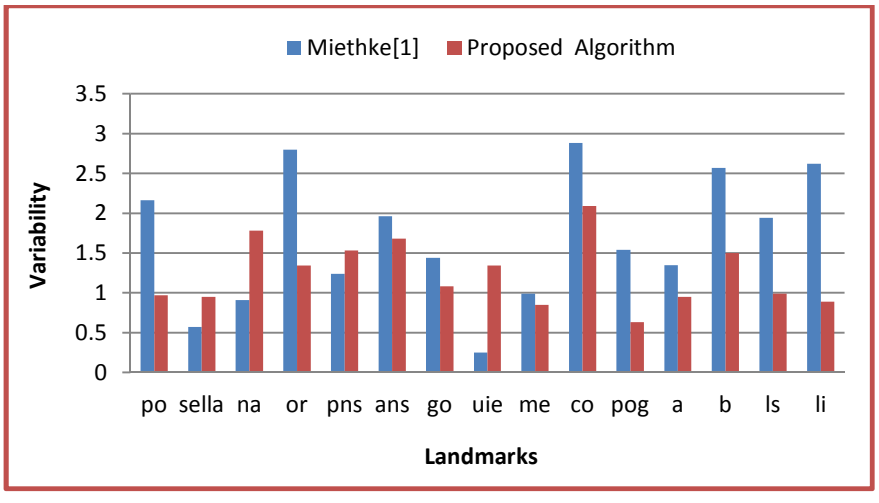


Figure 7.5: Comparison of variability (mm) of proposed method with variability of manually marked landmarks.

Figure 7.6: Cropped images of landmark Uie.

ART descriptor outperformed the ZM descriptor as evident from the results in Fig. 7.1 and Fig. 7.2. The reason for the inferior results of the ZM descriptor maybe its variance under translation and scaling as reported by [120]. Thus additional preprocessing will be needed for the ZM descriptor to cope with the translated and scaled images. The contributions of this work are listed below:

In this work to speed up algorithm

i. ART region shape descriptor used instead of Zernike moments.
ii. Coarse to fine template matching used.
iii. NCC and SSD, computed using FFT.

Hierarchical template matching approach applies templates of part of objects to the image is used. This strategy is less affected by distortions than a template of the entire object. Thus reasonable best matches are obtained. Landmark location accuracy improved by preprocessing by DoG, which helps suppress noise and uneven illumination effect.