CHAPTER 2

COMPREHENSIVE LITERATURE REVIEW OF TARGET SYSTEMS FOR PREPROCESSING AND LANDMARK DETECTION

Computer technology is having an increasing impact on the practice of orthodontics. Orthodontist needs software that may enhance the cephalograms thereby increasing their diagnostic value. Suppression of noise and enhancement of contrast improve the visibility of details in the image. Efficient enhancement algorithms will help reduce the radiation exposure and avoid the need of multiple exposures of the patients to the harmful X-rays. The development of automatic cephalometric analysis algorithms will help save the precious time of the doctors and reduce the subjectivity involved when the cephalograms are traced manually. In this Chapter, we present a detailed review of image enhancement techniques, and a systematic introduction and review of all the methods used in landmark detection.

2.1. Cephalometric Image Enhancement

A good quality cephalogram is needed to clearly show dental and skeletal structures. Generally, the X-rays suffer from low contrast, blurred edges and noise. In cephalometric X-rays, the contrast is determined by the level of kilovolts (KV) set in the X-ray machine. High values of KV tend to reduce the difference in radiation absorbed by different tissues and hence reduce the contrast between them. If KV is reduced to improve the contrast, it adversely affects the signal to noise ratio (SNR) of the resulting X-ray. Other characteristics that affect the quality of X-rays are specific to the technique of capturing the X-ray. In case of film based technique, the contrast is low in high and low density areas. Noise is added during scanning of films to convert to digital form. In case of image intensifiers, noise is added due to conversion from light to photons and photons to light and suffers from distortions toward the border of the image. In direct digital methods, SNR is lower.

Retake of X-rays may be needed if the quality of obtained X-ray is not high due to incorrect setting of the parameters. Repeated X-ray exposures may lead to cancer. But high quality X-rays should not compromise the need for minimal radiation exposure. Image enhancement methods can help avoid these retakes and multiple exposures of

the patients. Thus to improve the quality of captured cephalograms, efficient contrast enhancement methods are needed. This will help improve the contrast of meaningful details and allow the orthodontist to easily perceive them. Sharpness is the subjective perception of the distinctness of boundaries enhanced when the contrast is improved. This section discusses various techniques used for noise suppression and contrast enhancement [1-3].

2.1.1. Image noise suppression

X-ray images generally suffer from low SNR and are blurry in nature. All X-rays, either film based or digital, suffer from Poisson's noise. In digital X-rays, the electric detector systems add Gaussian noise, and their SNR is low owing to reduced dose. Thus, it is desirable to increase the SNR and reduce blur in X-ray images to improve visibility of details and make the image suitable for further processing.

2.1.1.1. Variance stabilization transform

Most existing noise suppression algorithms are designed for Gaussian noise. X-rays images suffer from Poisson's noise. We need to convert Poisson's noise to Gaussian additive noise before applying the noise suppression algorithm. Considering each pixel as independent random Poisson's variable, and g_i as the intensity values of pixel corresponding to grayscale image with Poisson's noise, the discrete Poisson probability of each pixel is given as

$$P(f_i; e_\lambda) = \frac{{e_\lambda}^{f_i} exp\,(-e_\lambda)}{g_i!} \tag{2.1}$$

where e_λ is the expected value of g_i that is equal to the variance [124].

$$\lambda = E(g_i) = var(g_i) \tag{2.2}$$

From this Poisson's noise maybe deduced as

$$\eta_i = g_i - e_\lambda \tag{2.3}$$

To suppress Poisson's noise that is a signal dependent noise, the common approach is to modify this noise using variance stabilizing transformation (VST) that makes the noise variance constant throughout the image. Anscombe VST [38] is the commonly used stabilization technique. Using this transformation the Poisson's noise is converted to signal independent Gaussian noise. For images with both Gaussian and Poisson's noise, generalized Anscombe transformation [39] is used to stabilize the noise variance. After this, any Gaussian noise suppression method may be applied to

reduce the noise. Finally, inverse transformation is applied to estimate the enhanced output image.

VST forward transform:

Image corrupted with Poisson's noise (I_N) is converted to image with Gaussian noise (I_G) as given in the following equation

$$I_G = T(I_N) = 2\sqrt{I_N + \frac{3}{8}} \tag{2.4}$$

The denoising of I_G using any of the Gaussian noise suppression method produces denoised image I_D.

Inverse VST transform:

Inverse transform is applied to I_D to obtain the desired estimate image I_E. The direct algebraic inverse of Eq. (2.4) is given below

$$I_E = T^{-1}(I_D) = \left(\frac{I_D}{2}\right)^2 - \frac{3}{8} \tag{2.5}$$

This estimate I_E is biased as

$$E[T(I_N)] \neq T[E(I_N)] \tag{2.6}$$

where E is the expected value.

Adjusted inverse is an alternative specified in [39] to provide unbiased results. However, this method gives improved results only for high-intensity images.

$$I_E = T^{-1}(I_D) = \left(\frac{I_D}{2}\right)^2 - \frac{1}{8} \tag{2.7}$$

The choice of proper inverse transformation is very important to obtain good numerical results. Makitalo et al. [39] have proposed an optimized inverse transformation and its close form approximation in [40]. The close form approximation is given as

$$I_E = T^{-1}(I_D) = \frac{1}{4}{I_D}^2 + \frac{1}{4}\sqrt{\frac{3}{2}}{I_D}^{-1} - \frac{11}{8}{I_D}^{-2} + \frac{5}{8}\sqrt{\frac{3}{2}}{I_D}^{-3} - \frac{1}{8} \tag{2.8}$$

2.1.1.2. Edge preserving noise suppression

Gaussian noise suppression methods are well-established and are widely used. Linear noise suppression techniques like averaging and Gaussian filters smooth Gaussian noise but increase the image blur. Thus, the focus shifts to nonlinear techniques that can remove noise without sacrificing the edge detail. Median filtering is the commonly used nonlinear smoothing technique that eliminates noise without blurring

the edges of the image, but 2D median filtering destroys thin lines and clips the corners. Its performance is poor when the number of noise pixels is higher than or equal to half the pixels within the window [35-37]. In 1990, Perona and Malik [41] introduced a nonlinear diffusion based powerful edge sensitive noise removal technique. It results in smoothing within the region while preserving the edges. The proposed algorithm is derived by modifying the linear heat equation to nonlinear diffusion equation. The method is briefly explained below. Partial differential equation based on nonlinear diffusion is given as

$$\frac{\partial I}{\partial t} = div[c(|\nabla I|)\nabla I] \tag{2.9}$$

Where ∇ is the gradient operator, div is the divergence operator, and the function $c(\cdot)$ is used to compute the diffusion coefficients. The initial condition is given by the original image $I(t=0) = I_0$. The function $c(\cdot)$ is monotonically decreasing and gives diffusion value that decreases as gradient strength increases. Several diffusion functions are proposed in literature. Two most commonly used functions are given below

$$c(|\nabla I|) = exp\left(-\frac{1}{\sqrt{1+|\nabla I|^2/k^2}}\right) \tag{2.10}$$

$$c(|\nabla I|) = exp\left(-\frac{|\nabla \mathrm{I}|}{k^2}\right) \tag{2.11}$$

where k is the conductance parameter.

For the discrete formulation, using the Euler formula the left side of Eq. (2.9) becomes

$$\frac{\partial I}{\partial t} = \frac{I(x,y,t+\Delta t) - I(x,y,t)}{\Delta t} \tag{2.12}$$

Substitute this value in Eq. (2.9) we get

$$\frac{I(x,y,t+\Delta t) - I(x,y,t)}{\Delta t} = div[c(\nabla I)\nabla I] \tag{2.13}$$

Thus discrete time implementation derived by modifying Eq. (2.13) is given as

$$I(x,y,t+\Delta t) = div[c(\nabla I)\nabla I] \times \Delta t + I(x,y,t) \tag{2.14}$$

The introduction of a discrete time affects the stability of the of the PM model. The time step Δt needs to be small for convergence of the model to a steady state. Weickert [43] suggested that for stable evolution of PM equation, time step should not

be higher than $\Delta t = 0.25$. In this method, the noise at edges is not eliminated nor are the edges enhanced. For achieving this, Weickert proposed two representations of anisotropic diffusion methods that prefer diffusion along edges to diffusion perpendicular to them. Their methods consider gradient ∇I_σ at scale σ and its direction. For this diffusion, tensors are developed using structural tensors that reflects the edge structure. The first method named edge enhancing anisotropic diffusion helps in diffusion along edges and reduces diffusion perpendicular to the edges. This method thus preserves and enhances the edges. The second method called coherence enhancing diffusion helps preserve finer details of the image.

The anisotropic diffusion equation is written as

$$\frac{\partial I(x,y,t)}{\partial t} = \boldsymbol{div}(D(x,y,t)\nabla I(x,y,t)) \tag{2.15}$$

$D(x,y,t)$ is a symmetric positive definite diffusion tensor derived using structure tensor with eigenvalues $\mu_1$and μ_2 and corresponding eigenvectors $\boldsymbol{\nu_1}$ and $\boldsymbol{\nu_2}$.

$$D = [\boldsymbol{\nu_1}\, \boldsymbol{\nu_2}] \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\nu_1}^T \\ \boldsymbol{\nu_2}^T \end{bmatrix} \tag{2.16}$$

D uses eigenvectors ν_1 and ν_2 from structure tensor to reflect the image structure and eigenvalue λ_1 and λ_2 derived based on the desired representation edge enhancing diffusion (EED) or coherence enhancing diffusion (CED).

Values of λ_1 and λ_2 for EED Method:

To smooth along edges λ_2 is kept as

$$\lambda_2 = 1 \tag{2.17}$$

λ_1 is chosen using the diffusion function to reduce the diffusibility perpendicular to the edges, which is given by

$$\lambda_1 = c(|\nabla I_\sigma|) \tag{2.18}$$

$$c(|\nabla I_\sigma|) = \begin{cases} 1, & if\ \ |\nabla I_\sigma| = 0 \\ 1 - exp\left(\frac{-c_m}{(|\nabla I_\sigma|^2/k^2)^m}\right), & otherwise \end{cases} \tag{2.19}$$

where C_m is the threshold constant and k is the contrast parameter. For $m = 4$ C_m is given as $C_4 = 3.31488$ [43].

Values of λ_1 and λ_2 for CED Method:

Eigenvalues μ_1 and μ_2 of structure tensor measure the image contrast in the direction of eigenvectors ν_1 and ν_2, respectively. Vector ν_1 corresponds to the orientation with the highest gray value fluctuations and ν_2 gives the local orientation. Difference in

eigenvalues μ_1 and μ_2 gives the measure of localized coherence direction. To enhance smaller structures and textures, smoothing has to be done in direction v_2 with diffusibility λ_2. In this case

$$\lambda_1 = \alpha \tag{2.20}$$

$$\lambda_2 = \begin{cases} \alpha, & if\ \mu_1 = \mu_2 \\ \alpha + (1-\alpha)exp\left(\frac{c_m}{(\mu_1 - \mu_2)^{2m}}\right), & otherwise \end{cases} \tag{2.21}$$

where $\alpha\, \epsilon\, (0,1)$ is a small positive parameter to keeps diffusion tensor uniformly positive definite [43].

Medical images are very complex and carry structures of varying sizes. If EED is used, it will preserve larger structure but blur the finer details while suppressing the noise. The opposite is true for CED. Thus, Frangakis et al. [44] proposed a 3D noise suppression method by combining EED and CED using a discrete switch to preserve both larger structures and finer textures in the electronic tomographic images. Kroon and Slump [45] tested various numerical schemes of the tensor diffusion of Weickert using CED for denoising cone beam CT scans. Mendrik et al. [46] proposed an improvement over Frangakis [44] approach. In their paper, a hybrid diffusion filter with continuous switch (HDCS) is introduced that exploits the benefits of three dimensional EED and CED. The results were tested using dental CT images acquired at reduced radiation doses.

Value of hybrid eigenvalue λ_{hi} is given as

$$\lambda_{hi} = (1-\varepsilon)\lambda_{ci} + \varepsilon\lambda_{ei}, \qquad i = 1,2 \tag{2.22}$$

where λ_{ei} is the EED contrast parameter and λ_{ci} is the CED contrast parameter and ε is computed as follows

$$\varepsilon = \exp\left(\frac{\mu_2(\lambda_{hi}^2(\xi - |\xi|) - 2\mu_3)}{2\lambda_{hi}^4}\right) \tag{2.23}$$

$$\xi = \frac{\mu_1}{\alpha + \mu_2} - \frac{\mu_2}{\alpha + \mu_3} \tag{2.24}$$

Considering the 2D case the modified value of ε is computed as suggested in Image Edge Enhancing Coherence Filter Toolbox developed by Dirk-Jan Kroon [121].

$$\varepsilon = \exp\left(\frac{\mu_2\left(\lambda_{hi}^2(\xi - |\xi|)\right)}{2\lambda_{hi}^4}\right) \tag{2.25}$$

$$\xi = \mu_1 - \mu_2 \tag{2.26}$$

2.1.1.3. Three dimensional transform domain collaborative filtering for noise suppression

Dabov et al. [47] proposed a novel technique to suppress noise using 3D transform domain collaborative filtering (TDCF). Their algorithm is divided into two parts basic estimate and final estimate. Both the parts follow the same three-step procedure except some minor difference in the method of thresholding and use of different transforms.

In the first step 3D arrays are formed by grouping 2D fragments that are similar to each other. The 3D groups are formed by block matching in transform domain using normalized 2D transform. The transform domain coefficients are thresholded before the application of distance measure to reduce the effect of noise. Blocks those distances from the reference block are smaller than a fixed threshold are considered being the part of the group. In the next step collaborative filtering is applied to these groups by first applying 3D separable transforms (normal profile or fast profile) to the groups. The transform used in the two profiles are shown in the Table 2.1 below:

Table 2.1: Transforms used in two profiles.

	Fast Profile	**Normal Profile**	
		$\sigma \leq 40$ Low levels of noise	$\sigma > 40$ High levels of noise
Basic Estimate	2D Bior1.5, 1D Haar transform	2D Bior1.5, 1D Haar transform	2D DCT, 1D Haar Transform
Final Estimate	2D DCT, 1D Haar transform	2D DCT, 1D Haar transform	2D DCT, 1D Haar Transform

The derived coefficients are thresholded (hard thresholded for the initial estimate or Weiner filtered for the final estimate). Finally, inverse 3D transformation is applied, and the blocks are returned to their original positions. Weighted averaging is applied on overlapping blocks to estimate the image. For a final estimate, improved grouping is performed using the basic estimates. Further, with the help of these locations, grouping is performed on noisy image. The algorithm gives one of the best results in terms of PSNR and subjective visual quality.

2.1.2. Contrast enhancement

Image that has a broadly distributed histogram tends to be more visually distinctive. Owing to the nonlinear nature of the image, the linear contrast enhancement

techniques are not always satisfactory and may require rescaling, which may result in information loss or distortion. Linear techniques not only enhance the image features but also enhance the noise. Histogram equalization (HE) is an effective nonlinear technique that modifies the image through pixel mapping such that the histogram of the processed image is uniform than that of the original image. The technique is fully automatic and straightforward that may either be applied globally or locally to enhance the contrast of an image. Thus most existing techniques for automatic landmark detection use, HE to improve the contrast of cephalograms.

The traditional histogram equalization (THE) is a global technique, which sometimes leads to excessive contrast enhancement. That may cause extreme modifications on the histogram, resulting in visual artifacts and giving the image an unnatural look [48]. Thus, some kind of regularization is needed to avoid excessive enhancement. Global THE produces images with mean intensity approximately in the middle of the dynamic range [48]. Thus, it does not preserve the brightness of the image that leads to washed-out appearance of some images.

Several advance HE methods has been proposed to avoid these limitations. Bi-histogram equalization was proposed by Kim et al. [49] to preserve the brightness by separating the histogram of the input image into two parts based on its mean gray level. The algorithm preserves the mean brightness of a given image very well. However, this method cannot enhance some images. Chen and Ramli proposed minimum mean brightness error bi-histogram equalization (MMBEBHE) [50] with improved performance. Their method uses the minimum absolute mean brightness error (AMBE) to compute the threshold to partition the input histogram.

Ooi et al. [51] proposed bi-histogram equalization plateau limit (BHEPL) a hybrid of the BBHE and CHE. Abdullah-Al-Wadud et al. [52] proposed dynamic histogram equalization (DHE) technique that partitions the image histogram into sub-histograms based on the local minima of the smoothed histogram and assigns a specified gray level range to each partition before equalizing them separately. However, it does not preserve the image brightness. Ibrahim and Kong [53] proposed brightness preserving DHE (BPDHE). This method partitions the image histogram based on the local maxima of the smoothed histogram. It then assigns a new dynamic range to each partition. Finally, the output intensity is normalized to make the mean intensity of the resulting image equal to the input one.

A clipped HE (CHE) is proposed in [57] to overcome the unwanted over-enhancement and noise amplifying. In this method, a clipping level is determined, and the histogram is clipped based on this clipping level. The clipped portion is then redistributed to the entire dynamic range. Although the CHE can control the over-enhancement and noise amplifying artifacts, it requires the user to determine the clipping level. Furthermore, the redistributed process may require more time for processing, which could increase the complexity in the real time system. Yan-feng et al. [54] presented a double-plateau HE using two thresholds, upper threshold to constrain background noise and a lower threshold to protect and enhance the details. A critical issue of double-plateau HE is how to properly choose the upper and lower threshold values. Wang and Ward [55] proposed an enhancement process, called weighted thresholded histogram equalization (WTHE). The results using the WTHE method show well enhanced contrast and little artifacts. Chen et al. [48] proposed an unconventional technique gray level grouping (GLG) that is based on grouping and ungrouping histogram bins to avoid histogram spikes.

The properties of pixels in an image vary spatially, thus using a single global histogram mapping cannot enhance the local contrast optimally. A solution to this has been proposed in [56]. In this histogram of a local window about a pixel is used to determine the mapping for that pixel. Pizer et al. [57] has proposed an improvement on [56]. They developed a procedure called Contrast-Limited Adaptive Histogram Equalization (CLAHE). In their technique, the localized contrast gain is limited by restricting the height of local histograms. It gives improved noise performance.

In this method, the image is first divided into blocks, and then mapping functions for each block are computed. To handle histogram spikes the derived histogram is clipped and renormalized before obtaining the mapping function. This clipping and renormalization add additional computation burden, especially in the image with very high spikes. To enhance a specific pixel, the mapping function is interpolated using four mapping functions associated with its four neighboring blocks. Clipping and renormalization of the histogram reduce noise amplification and ringing artifacts. Interpolation helps improve speed. The size of the pixel contextual region and the clip level of the histogram are the parameters of CLAHE.

Traditional histogram is computed using the frequency of gray levels in the image. But the frequency of the gray levels is not consistent with their contribution to the representation of the image [59]. In the traditional histogram, spikes occur because of

the presence of many similar pixels in the homogeneous regions of the image. This histogram spikes results in enhancement of noise and addition of artifacts in the image [48]. Common approaches to handle histogram spikes are histogram clipping and redistribution [57], histogram weighting [55], and low pass filtering [58] and. These techniques are computationally expensive. To overcome the drawback of the traditional histogram [59-60] use only contrasted pixels for the computation of the histogram. Arici et al. [59] presented a low complexity histogram modification framework (HMF) that adjusts the level of enhancement adaptively. The image histogram is obtained by considering pixels based on the horizontal variation measure. The number of pixels included is counted for proper normalization. Finally, the modified histogram is computed using the weighted average of input histogram and uniform histogram. Zeng et al. [60] propose a new form of the histogram called the gray level information histogram (GLIH). In their technique, they divide the input image into several equal sized regions according to the intensities of gradients. The processed histogram for the whole image is obtained by the summation of all the weighted values of regions. The amplitudes of this new form of the histogram reflect the contribution of the gray levels to the representation of image information. They suggest that the performance of many histogram based enhancement techniques might be improved dramatically using the proposed histogram. GLIH gives improved results when used to enhance the cephalogram images in combination with CLAHE.

Steps for computing gray level information histogram

i. Noise suppression using image smoothing by Gaussian filter

ii. Calculation of the image gradient G_{xy} using Sobel's operator

iii. In this step, the image is divided into five regions based on the ascending order of G_{xy}. Each pixel location in the input image has specific statistical weighting coefficient depending on its G_{xy}. The coefficients are as given below:

$$W(x,y) = \begin{cases} w_1 & if\ G_{xy} \leq T_1 \\ w_2 & if\ T_1 < G_{xy} \leq T_2 \\ w_3 & if\ T_2 < G_{xy} \leq T_3 \\ w_4 & if\ T_3 < G_{xy} \leq T_4 \\ w_5 & otherwise \end{cases} \tag{2.27}$$

where T_i denotes the limits of the gradient intervals and w_i is the statistical weighting coefficients for $i = 1, 2, 3, 4, 5$.

iv. The cumulative summation of the weighted statistical value of gray levels in different regions is computed to get the modified gray level information histogram H_m.

$$H_m(r) = \sum_{i=1}^{5} w_i n_i(r) \tag{2.28}$$

where n_i is the number of pixels at the gray level r in a certain region. The weights w_i is computed using Canny edge detection.

$$w_i = \frac{N_i}{N_5} \tag{2.29}$$

In Eq. (2.29) N_i is the number of edge points in a specific region, and N_5 denotes the number of edge points in region 5. They assume that region five has the maximum number of edge points. This is not always true, maximum edge points may be present in any of the regions. Especially in cephalometric X-rays owing to low contrast between soft tissue and background and between soft tissue and bony structure. Instead of incurring the over head of the Canny edge detector the weights can be directly computed from the image gradients. In X-rays, edges are already blurry. Gaussian smoothing in step one help reduce the image noise but further blurs the edges. Thus, it is unsuitable for such images. Sobel's operator is unable to differentiate between desired edges and the image noise and gives high gradients for both. Accordingly if the image is noisy, the noise is further enhanced using the described modified histogram.

2.2 Edge Detection

Human visual system uses edges as key features to comprehend the contents of an image. Similarly, many computer vision applications use edge features to achieve higher level goals. In manual identification of landmark, edges of the cephalograms are traced then landmarks are marked using these traced edges. The initial automatic landmark detection algorithms [8, 9] that tried to imitate the manual process used edge detection and a-priori knowledge to locate the landmarks. These algorithms were only able to locate landmarks on clear edges. Later, algorithms [16, 23, 25, 26, and 32] used a combination of techniques to detect landmarks both on the edges and in the interiors of the structures. Edge enhancement and detection are the first steps in these algorithms after image preprocessing.

Parthasarathy et al. [9] in their work enhanced edges using three different operators namely Kirsch operator, Prewitt's operator and Mero-Vassey operator. Tong et al. [10] applied Prewitt's operator to detect the edges. Grau et al. [24] applied Laplacian of Gaussian for edge detection. El. Feghi et al. [26] used Sobel's operator to enhance the edges. Yue et al. [17] extracted edges using the Canny edge detector. Kafieh et al. [27] applied edge detection using Susan edge detection and morphological opening. Favaedi et al. [32] applied Canny edge detection to isolate the edges. Most papers use first derivative operators to enhance and detect the edges.

The local gradient operators like Sobel and Prewitt only detected edges having certain orientations and performed poorly when the edges were blurred and noisy. Since then, more sophisticated operators like Canny [61] and Susan [63] were developed to provide better edge detection results. The majority of these are linear operators that are derivatives of some sort of smoothing filter. The Gaussian filter is by far the most widely used smoothing filter in these techniques, but it tends to blur the image, which leads to shift in the edge locations. Canny is the most widely used operator in this category [35, 62]. The presence of noise can significantly degrade the performance of this operator, and it may require many more runs using different combinations of parameters (the upper and lower threshold values and sigma value) although a very good tradeoff can be found after enough of runs. Furthermore, at times this edge detector tends to connect lines into closed contours and may give rise to unnecessary edges thus cluttering the main shape details.

Smith et al. [63] proposed a non derivative edge detector named Susan. The proposed technique is noise resistant and fast. The detector uses a pixel's gray value similarity to its neighbors under a circular mask to find the area called univalue segment assimilating nucleus (USAN). Edges are found by using the size, centroid and second moment of the derived area. This detector gives results similar to Canny edge detector but is less sensitive to noise and is suited to handle both step and ramp edges unlike Canny that can handle just step edges. Canny uses three adaptable parameters whereas Susan uses one regulating parameter (brightness threshold).

Image processing algorithms are based on crisp logic, thus information uncertainty and subjectivity are hard to model in the algorithms. Fuzzy logic can deal with situations in, which making a sharp distinction between the boundaries is difficult. Cephalometric images are blurry in nature. Thus, we explore fuzzy edge detection for these images as fuzzy logic allows to model uncertainty in a better form than the crisp

models. Several fuzzy edge detection techniques have been proposed. Initial fuzzy techniques characterized edge detection as a fuzzy reasoning problem but these techniques were very expensive in computing [64, 65]. The work of Russo [65] proposes a noise protected operator for noisy images that combines effective rules for noise cancellation and edge detection in the same structure. However, the price paid is the nonisotropic structure of the operator used, which leads to slight disuniformatity in the responses to the borders having different orientations. The work of Tizhoosh [66] introduces uncomplicated and fast fuzzy edge detectors using heuristic membership functions, simple fuzzy rules and fuzzy complement based edge maps. Liang and Looney [67] uses extended epanechnikov function as a fuzzy set membership function for each class where the class assigned to each pixel is the one with the greatest fuzzy truth of membership. After this classification, a competition is run as a second step to thin the edges. The performance of this algorithm is given to be the same as Canny but is shown to be faster. In this case, the tradeoff used is more edges with more noise versus fewer noisy edges. Wu et al. [68] proposes a fast multi-level fuzzy edge detection algorithm. This algorithm first enhances image contrast then extracts the edges by two stage edge detector. This algorithm extracts thin edges and tends to remove false edges from the image. Another fuzzy detector, proposed by Hanmandlu et al. [69] is based on USAN area. The proposed detector drops some edges and enhances the important edges that contribute to the principal shape. Nevertheless, in certain cases these dropped edges may be required.

2.2.1. Edge detection using fuzzy complement approach

The first step in any fuzzy based algorithm is to transform the image to the fuzzy domain. In this domain, all the values are in range of 0 and 1. The mapping function can be defined using a fuzzy membership function. For fuzzy edge detection, we need to find the degree of edginess for each pixel in the image. The intersection of a fuzzy image with its complement gives high membership value to edge pixels than nonedge pixels [66]. To find the fuzzy image consider $m \times m$ spatial window S_W about the pixel to be fuzzified. The following fuzzy membership function is used to find the fuzzy value of a pixel $g_i(x, y)$ in image $f(M, N)$.

$$\mu_F(x, y) = \frac{g_i(x, y) \times (\max(S_W) - \min(S_W))}{\max(S_W) \times \max(f(M, N))} \tag{2.30}$$

An edgy fuzzy membership value at each pixel is given by taking the intersection between the fuzzy value for each pixel in the spatial window and its complement and taking the average value. This makes the edge detector robust against noise.

$$\widehat{\mu_F}(x,y) = min\left(1, \frac{2}{m^2}\sum_{i=1}^{m}\sum_{\mathrm{j}=1}^{\mathrm{m}} \min\left(\mu_F(i,j), 1-\mu_F(i,j)\right)\right) \tag{2.31}$$

After defining the degree of edginess as fuzzy membership function, fuzzy set operations, linguistic variable or fuzzy rules are applied to detect the edges.

2.2.2. Biorthogonal wavelet transform

Wavelets are functions generated from a single function called the mother wavelet by scaling and shifting along time/frequency.

$$\psi_{d,t}(x) = \frac{1}{\sqrt{|d|}}\psi\left(\frac{x-t}{d}\right) \tag{2.32}$$

where d represents the parameter for dilation and t for translation. For digital image processing, we need to define these parameters in the discrete form as given below:

$$d = d_0^m \tag{2.33}$$

$$t = nt_0 d_0^m \tag{2.34}$$

where m and n are integers.

Considering the commonly used values of $d_0 = 2$ and $t_0 = 1$. Discrete wavelet can be represented by

$$\psi_{d,t}(x) = 2^{\frac{-m}{2}}\psi(2^{-m}x - n) \tag{2.35}$$

These functions allow both time and frequency analysis of images simultaneously. In this a scaling function defined similarly to the wavelet function is used to create a series of approximations of an image. Wavelets are then used to encode the difference in information between adjacent approximations.

$$\varphi_{d,t}(x) = 2^{\frac{-m}{2}}\varphi(2^{-m}x - n) \tag{2.36}$$

In the recent year's discrete wavelet transform (DWT) have become an attractive technique in image compression, enhancement and analysis. DWT is used to decompose digital image into a low frequency approximation and high-frequency detailed components. The forward DWT coefficients for any sequence $f(x)$ are computed as follows

$$a_\varphi(d_0,t) = \frac{1}{\sqrt{M}}\sum_n f(x)\varphi_{(d_0,t)}(x) \tag{2.37}$$

$$d_{\psi}(d,t) = \frac{1}{\sqrt{M}} \sum_{n} f(x)\psi_{(d,t)}(x) \quad (2.38)$$

Where a_{φ} corresponds to approximation coefficients, d_{ψ} corresponds to detailed coefficients and $x = 0, 1, 2, \ldots, M-1$, $d = 0, 1, 2, \ldots, D-1$, t= $0, 1, 2, \ldots, 2^{d}-1$.

Inverse DWT is applied to obtain a perfectly reconstructed input signal $f(x)$

$$f(x) = \frac{1}{\sqrt{M}} \sum_{n} a_{\varphi}(d_0,t)\varphi_{(d_0,t)}(x) + \frac{1}{\sqrt{M}} \sum_{n} w_{\psi}(d,t)\psi_{(d,t)}(x) \quad (2.39)$$

In 2D signals like images, the signal components are represented by 2D wavelets and 2D approximation functions. The signal is divided into one approximation function and three wavelet functions. The functions are given by 1D separable function in terms of rows and columns.

Scaling function

$$\varphi_{i,j}(x,y) = \varphi(x-i)\varphi(y-j) \quad (2.40)$$

The three wavelet functions are

$$\psi_{i,j}^{[H]}(x,y) = \varphi(x-i)\psi(y-j) \quad (2.41)$$

$$\psi_{i,j}^{[V]}(x,y) = \psi(x-i)\varphi(y-j) \quad (2.42)$$

$$\psi_{i,j}^{[D]}(x,y) = \psi(x-i)\psi(y-j) \quad (2.43)$$

The three wavelets represent details in horizontal, vertical and diagonal directions, respectively. By choosing a proper decomposition level it is possible to select high-frequency coefficients that are rich in edge detail and low in noise [35].

Decimation of the wavelet coefficients in DWT removes every other of the coefficients of the current level. This decimation causes shift variance of the wavelet transform. Thus, DWT suffers from aliasing effects owing to down-sampling and may lead to artifacts in the reconstructed image. This is a disadvantage in applications like edge detection and pattern recognition. Stationary wavelet transform (SWT) is justified in such applications. SWT is identical to DWT in terms of the decomposition structure except that no down-sampling is involved. The key features of this transform are that it is redundant and shift invariant and does not suffer from aliasing effects. These properties help in the efficient edge localization [70].

Images mostly have homogeneous regions except some edges. Thus, wavelets for image processing should be reasonably smooth, which requires the associated filters to be long enough to obtain good smoothness and energy compaction capability. However, longer filters increase the computational cost of the transformation [71].

Wavelets are of two types orthogonal and biorthogonal. DWT was originally derived for orthogonal wavelets. Transforms are constructed using orthogonal filters that can be easily inverted and coded. Orthogonal wavelets are often irregular and implicitly defined. Thus, the low and high pass filters are implemented using infinite impulse response (IIR) filter banks. If the orthogonality constraints are removed wavelets with more desirable properties in analysis and decomposition part may be developed [72-73]. In biorthogonal wavelets, two scaling and two wavelet functions are used. Cohen et al. [74] introduced biorthogonal wavelet by removing the orthogonality requirement. They obtain wavelet pairs that are symmetric, regular and compactly supported [75]. Compact support helps improve the computational complexity of the transform. The symmetry property of the filter coefficients is desirable since it results in linear phase transfer function. It is a general belief in image processing that linear phase filters produce fewer visual artifacts. Finite impulse response (FIR) filters are linear phase filters with compact support. Longer filters make the sharp edges to be detected as wide events that affect edge localization. In image processing applications, it is desirable to have FIR filters with a few taps [72-74]. Spline wavelets have an explicit representation as piecewise polynomials that make the computation of the function values easy [76].

2.3. Subpixel Edge Detection

A pixel is the smallest unit of an image. The image edges that are continuous in nature do not always fall on the exact pixel boundaries as shown in Fig. 2.1.

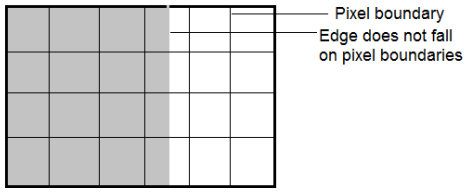


Figure 2.1: An image showing subpixel location of an edge.

Subpixel location accuracy is higher than the image resolution. Thus, subpixel edge algorithms utilize floating point operations to realize higher precision edges. The common subpixel edge detection methods fall into three categories: reconstruction techniques, interpolation techniques and moment based techniques. One of the earliest reported techniques in subpixel edge detection is due to Hueckel [77]. However, this

technique was designed for implementation over large window, and its complexity is high. Macvicar and Binford [78] proposed subpixel edge detection based on linear interpolation. However, the precision of the method is low owing to interpolation. Jensen and Anastassiou [79] utilized a nonlinear interpolation scheme based on a source model emphasizing the visual integrity of the detected edges and incorporated a novel edge fitting operator for subpixel edge location. Among the moment based operators introduced to locate subpixel edges, there are methods based on gray level moments, geometric moments, and Zernike moments. Tabatabai and Mitchell [80] determined edge locations by fitting first three gray level moments to the edge data. Lyvers et al. [81] proposed an efficient method based on geometric moments for subpixel level edge detection. The method uses six spatial moments and a 3-D edge model. The computational load of this method is high. Moments of orthogonal basis functions such as Zernike polynomials were introduced by Teague [82]. They can be used to represent the image features by a set of mutually independent descriptors with a minimal amount of information redundancy. Zernike moments are defined in polar coordinates, and their magnitude information remains unchanged under rotation only the phase changes which cannot be achieved by geometric moments. Ghosal and Mehrotra [83] introduced Zernike moments for subpixel edge detection and later presented a new parametric model-based approach to high-precision composite edge detection using orthogonal Zernike moment based operators [84]. As it is evident from the current literature [85-88] on moment based edge detectors most methods are still largely based on the technique proposed by Ghosal. Bin et al. [89] proposed a method based on Fourier-Mellin moments for subpixel edge detection. Their results show that the proposed algorithm can be used to achieve a location accuracy of 0.16 pixels for straight line with noise and 0.23 pixels for curves with noise. They have suggested in their paper that their method is more capable than Zernike moment based methods to describe small objects in the images.

2.3.1. Zernike moments

Zernike moments (ZMs) introduced by Frits Zernike is a set of complex integral functions with a simple rotational invariance property and forms a complete orthogonal basis over the class of square integral functions defined over a unit circle. Zernike polynomials were introduced by Teague [82]. They can be used to represent the image features by a set of mutually independent descriptors with the minimal amount of information redundancy. ZMs are defined in polar coordinates, and their

magnitude information remains unchanged under rotation, only the phase changes. This cannot be achieved by geometric moments. ZMs extract spatial information from an image. The features are in complex form and are represented by their phase and magnitude [123]. The ZM of order p and repetition q of an image function $f(x,y)$ in two dimensions over a unit disc are defined by

$$Z_{pq} = \frac{p+1}{\pi} \iint_{x^2+y^2 \le 1} f(x,y) V_{pq}^*(x,y) dxdy \tag{2.44}$$

Where the image function $f(x,y)$ is defined over a discrete square domain $N \times N$ and $V_{pq}^*(x,y)$ is the complex conjugate of the complex Zernike polynomials $V_{pq}(x,y)$ given by

$$V_{pq}^*(x,y) = R_{pq}(r) \exp(-jq\theta) \tag{2.45}$$

where

$r = \sqrt{x^2+y^2}$, $j = \sqrt{-1}$, $p \ge 0$, $|q| \le p$, $p - |q| = even$, and $\theta = tan^{-1}(y/x)$, $\theta \in [0, 2\pi]$

The function R_{pq}(r) is a radial polynomial defined as

$$R_{pq}(r) = \sum_{s=0}^{(p-|q|)/2} \frac{(-1)^s (p-s)!\, r^{p-2s}}{s!\left(\frac{p-|q|}{2}-s\right)!\left(\frac{p+|q|}{2}-s\right)!} \tag{2.46}$$

Clearly $R_{pq}(r) = R_{p,-q}(r)$

In discrete domain Z_{pq} are approximated as

$$Z_{pq} = \frac{p+1}{\pi} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} f(x_i, y_j) V_{pq}^*(x_i, y_j) \Delta x_i \Delta y_j, \;\; x_i^2 + y_i^2 \le 1 \tag{2.47}$$

with

$$x_i = \frac{2i+1-N}{N\sqrt{2}}, y_j = \frac{2j+1-N}{N\sqrt{2}}, i,j = 0,1,2,\dots,N-1, and\ \Delta x_i = \Delta y_j = \frac{2}{N/\sqrt{2}}$$

An important property of ZMs is that their values for an image, and its rotated version have a simple relationship. If an image is rotated by an angle ϕ, the ZMs of the rotated image, Z'_{pq}, are as given below

$$Z'_{pq} = Z_{pq} \exp(-iq\phi) \tag{2.48}$$

For more detailed information on fast computation of ZMs refer to [90]. To estimate these moments the corresponding masks of any desired size can be obtained by evaluating the associated integral over each pixel assuming $f(x,y)$ to be constant

over the pixel. Furthermore, we need to project the square discrete image onto the unit disk of Zernike polynomial. Two mapping techniques are possible. The conventional technique is based on inner circle mapping as shown in Fig. 2.2 (a). The conventional technique causes geometrical error as the pixels whose centres fall outside of unit disk are not included in the ZM computation. This error can be avoided by using the outer circle mapping as shown in Fig. 2.2 (b).

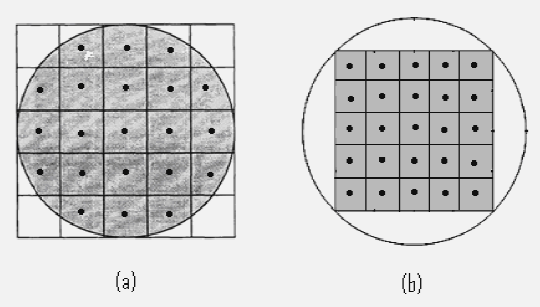


Figure 2.2: (a) Inner circle mapping and (b) outer circle mapping.

2.3.2 Pseudo-Zernike moments

Pseudo-Zernike moments (PZMs) are orthogonal moments derived in a similar fashion to the ZMs after releasing the condition $p - |q| = even$ [123]. PZMs have been proven to be superior to other moment functions such as ZMs in terms of their feature representation capabilities. They are less sensitive to image noise than the conventional ZMs [91, 92]. The 2-D PZMs of order p with repetition q of an image $f(x, y)$ are defined as

$$PZ_{pq} = \frac{p+1}{\pi} \iint_{x^2+y^2 \le 1} f(x,y) V_{pq}^*(x,y) dxdy \tag{2.49}$$

The features are in complex form and are represented by their phase and magnitude. $V_{pq}(x, y)$ is the pseudo-Zernike polynomial in polar coordinates and is defined as

$$V_{pq}^*(x,y) = R_{pq}(r) \exp(-jq\theta) \tag{2.50}$$

where $r = \sqrt{x^2 + y^2}$, $j = \sqrt{-1}$, $p \ge 0$, $|q| \le p$, , and $\theta = tan^{-1}(y/x)$, $\theta \in [0, 2\pi]$

The function $R_{pq}(\mathrm{r})$ is a radial polynomial defined as

$$R_{pq}(r) = \sum_{s=0}^{(p-|q|)} \frac{(-1)^s (2p+1-s)!\, r^{p-s}}{s!\,(p-|q|-s)!\,(p+|q|+1-s)!} \tag{2.51}$$

In discrete domain PZ_{pq} are approximated as

$$PZ_{pq} = \frac{p+1}{\pi} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} f(x_i, y_j) V_{pq}^*(x_i, y_j) \Delta x_i \Delta y_j, \quad x_i^2 + y_j^2 \leq 1 \tag{2.52}$$

$$x_i = \frac{2i+1-N}{N\sqrt{2}}, y_j = \frac{2j+1-N}{N\sqrt{2}}, i, j = 0, 1, 2, \ldots, N-1, and\ \Delta x_i = \Delta y_j = \frac{2}{N/\sqrt{2}}$$

PZMs can provide more features than ZMs. PZMs have $(p+1)^2$ linearly independent moments of order $\leq p$, whereas ZMs contain only $(p+1)(p+2)/2$ linearly independent moments. Moreover, like ZMs, an important property of PZMs is that their values within an image and its rotated version have a simple relationship. If an image is rotated by an angle ϕ, the PZMs of the rotated image, PZ'_{pq}, are as given in Eq. (2.53).

$$PZ'_{pq} = PZ_{pq} \exp(-iq\phi) \tag{2.53}$$

The masks for PZM are obtained in a similar fashion to ZM.

2.3.3 Subpixel edge detection using Zernike moments

A step edge pattern is shown in Fig. 2.3 where h is the height of the edge, b is the background grayscale, ℓ is the perpendicular distance between the centre of unit circle and the edge and β is the edge direction. These four parameters (h, b, ℓ, β) can be calculated using Zernike polynomials of different orders. For finding edge parameters h, ℓ, and β, three masks Z_{00}, Z_{11}, and Z_{20}, are calculated using orthogonal complex polynomials.

$$V_{00} = 1 \tag{2.54}$$

$$V_{11} = x - jy \tag{2.55}$$

$$V_{20} = (2x^2 + 2y^2 - 1) \tag{2.56}$$

If we rotate the edge by an angle $-\beta$, it will be aligned parallel to the y axis. So we have

$$\iint_{x^2+y^2 \leq 1} f'(x, y) y dy dx = 0 \tag{2.57}$$

$f'(x, y)$ is the edge function after it is rotated. The corresponding Zernike moments Z_{pq} of the original image and Z'_{pq} of the rotated image are related by $Z'_{00} = Z_{00}$ $Z'_{20} = Z_{20}$ and $Z'_{11} = Z_{11} \exp(j\beta)$.

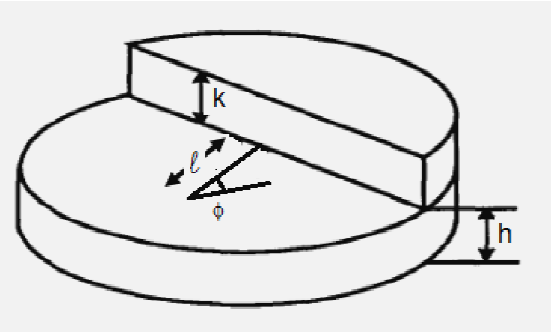


Figure 2.3: Ideal step edge.

After the edge is rotated through an angle β, the imaginary part of Z_{11} is zero while the edge is parallel with y axis, so the rotation angle of the edge is

$$\beta = tan^{-1}\left(\frac{Im[Z_{11}]}{Re[Z_{11}]}\right) \tag{2.58}$$

Further, the following equations can be deduced based upon the theory of ZM

$$Z'_{00} = b\pi + \frac{h\pi}{2} - hsin^{-1}\ell - h\ell\sqrt{1-\ell^2} \tag{2.59}$$

$$Z'_{11} = \iint f'(x,y)(x-jy)dydx = \frac{2h(1-\ell^2)^{3/2}}{3} \tag{2.60}$$

$$Z'_{20} = \iint f'(x,y)(2x^2+2y^2-1)dydx = \frac{2h\ell(1-\ell^2)^{3/2}}{3} \tag{2.61}$$

Solving the above two equations the edge parameter ℓ can be deduced as follows

$$\ell = \frac{Z_{20}}{Z'_{11}} \tag{2.62}$$

Criterion k is obtained by solving Eq. (2.60)

$$h = \frac{1.5Z'_{11}}{(1-\ell^2)^{3/2}} \tag{2.63}$$

Finally b is found using Eq. (2.59)

$$b = \frac{Z_{00} - h\pi/2 + hsin^{-1}(\ell) + h\ell\sqrt{(1-\ell^2)}}{\pi} \tag{2.64}$$

Using edge parameter h decision is made if a pixel is an edge or a nonedge pixel.

Subpixel edge location are found using ℓ and β.

$For \quad \ell < \delta$ (where 2δ is less than the size of pixel)

$$x_{new} = x + \ell\cos(\beta) \tag{2.65}$$

$$y_{new} = y + \ell\sin(\beta) \tag{2.66}$$

x_{new} and y_{new} are the new edge locations with subpixel accuracy [83].

2.4. Region Shape Descriptors

The representations of images require a large amount of numeric data. Many alternative representations of images that provide nearly the same accuracy, while requiring fewer data are proposed in literature. Pattern matching using these representations can be performed with higher speed as less data must be worked on. This provides significant storage saving and improves their efficiency and effectiveness [93]. A key factor in the effectiveness of the 2-D shape recognition system is the selected shape representation. If the representation is not robust to noise, is ambiguous or does not adapt to geometric transformations, then the accuracy of the algorithm will be naturally poor. Among the statistical approaches for feature extraction, orthogonal moments are very important in pattern matching and classifications due to their invariant properties. The term invariant denotes that if an image goes through one or a combination of the changes, such as size, position and rotation, its features remain unchanged. Numerical features reduce the dimensions of the image space by removing redundant data.

Region based image descriptors capture the global properties of the pixel distribution in the entire image, and hence they are extensively used in many pattern recognition and computer vision applications. In the present case, we are interested in finding the shape information from the cephalograms, which help in the localization of landmarks. Region based descriptor describes the shape of an object in an image very efficiently. ZM based descriptors are good shape descriptors. This shape descriptor has proven its superiority with respect to description capability and robustness to noise or deformations over other moment functions. Thus, it is the most commonly used rotation invariant pattern recognition technique used in image shape feature extraction and description [94, 95]. However, their computational complexity is high. The other important shape descriptor belonging to this class is ART. The radial basis functions of ART [96, 97] are the sinusoidal functions unlike the Zernike moments, whose radial basis functions are radial polynomials. The transform has the same characteristics as the moments: minimum information redundancy, and robustness to image noise and invariant to rotation. Two of the most important characteristics of ART that distinguishes it from the moments are that it is computationally very fast, and the high order transforms do not suffer from numerical instability like Zernike and other orthogonal moments.

Recently Yap et al. [98] proposed three orthogonal rotation invariant transforms: polar complex exponential transforms (PCETs), polar cosine transforms (PCTs) and polar sine transforms (PSTs). These transforms are collectively known as polar harmonic transforms (PHTs). The difference between orthogonal rotation invariant moments (ORIMs) and PHTs is that the radial parts of the kernel functions in ORIMs are polynomials and in PHTs these are sinusoidal functions. The PHTs are preferred to ORIMs because these are computationally very fast [99] and the high order transforms are numerically stable, whereas the ORIMs are less efficient and high order moments are numerically unstable. ZMs are already discussed in Section 2.4.2. Here we discuss ARTs and PHTs in the following sections.

2.4.1. Angular radial transforms

The ART is a transform based region shape descriptor of MPEG-7 [95]. The ART coefficients of order p and repetitions q for a continuous image function $f(x,y)$ in a unit disk are defined by

$$A_{pq} = \iint\limits_{x^2+y^2\leq 1} f(x,y)V_{pq}^{*}(x,y)dxdy \tag{2.67}$$

Where the function $V_{pq}^{*}(x,y)$ is the complex conjugate of the ART basis function $V_{pq}^{*}(x,y)$ defined by

$$V_{pq}(x,y) = \frac{1}{2\pi}R_p(r)\exp(jq\theta) \tag{2.68}$$

where $R_p(r) = \begin{cases} 1, & n = 0 \\ 2\cos(\pi pr), & n > 0 \end{cases}$

n is a nonnegative integer, m is an integer, $r = \sqrt{x^2+y^2}$, $j = \sqrt{-1}$ and $\theta = tan^{-1}(y/x)$. The ART coefficients for a discrete image function $f(x,y)$ of the size $N \times N$ pixels are approximated as

$$A_{pq} = \frac{1}{2\pi}\sum_{i=0}^{N-1}\sum_{k=0}^{N-1} f(x_i,y_k)R_p(r_{ik})\exp(-jq\theta_{ik})\Delta x_i \Delta y_k, \quad x_i^2 + y_k^2 \leq 1 \tag{2.69}$$

$$x_i = \frac{2i+1-N}{D}, y_k = \frac{2k+1-N}{D}, i,k = 0,1,2,\dots,N-1, and\ \Delta x_i = \Delta y_k = \frac{2}{D}$$

and $D = \begin{cases} N, \text{for inscribed disk} \\ N\sqrt{2}, \text{for outer disk} \end{cases}$

We take $D = N$ in our further discussions and experiments. The ART is a complex function and can be expressed in terms of its real and imaginary parts as given in Eq. (2.67) and Eq. (2.68). Let $a_{pq,ik}$ be the contribution to the $(p,q)^{th}$ coefficient of the ART transform by a pixel (i,k).The contribution of all pixels within the image to the $(p,q)^{th}$ coefficient can be expressed as

$$A_{pq} = \sum_{i=0}^{N-1} \sum_{k=0}^{N-1} a_{pq,ik} \ , \qquad x_i^2 + y_k^2 \leq 1 \tag{2.70}$$

$$A_{pq} = AR_{pq} + jAI_{pq} \tag{2.71}$$

$$A_{pq} = \frac{1}{2\pi} \sum_{i=0}^{N-1} \sum_{k=0}^{N-1} f(x_i, y_k) R_p(r_{ik}) \left(\cos(q\theta_{ik}) - j \sin(q\theta_{ik})\right) \Delta x_i \Delta y_k \tag{2.72}$$

where $x_i^2 + y_k^2 \leq 1$

2.4.2. Polar harmonic transform

The PHTs consist of the PCET, PCT and PST [98]. They have identical mathematical representation with a difference in radial part of the kernel function. A PHT is defined as

$$A_{pq} = \lambda \iint_{0\,0}^{2\pi 1} f(r,\theta) V_{pq}^{*}(r,\theta) r dr d\theta \tag{2.73}$$

where $|p| = |q| = 0,1,2,\ldots$ The kernel function $V_{pq}^{*}(r,\theta)$ is the complex conjugate of the function $V_{pq}(r,\theta)$ defined by

$$V_{pq}(r,\theta) = R_p(r) \exp(jq\theta) \tag{2.74}$$

where $j = \sqrt{-1}$

The radial kernel and the parameter λ are defined as

PCET:

$$R_p(r) = \exp(j2\pi pr^2) \tag{2.75}$$

$$\lambda = \frac{1}{\pi} \tag{2.76}$$

PCT and PST:

$$\mathrm{R_p(r)} = \begin{cases} cos(\pi pr^2), \text{for } PCT \\ sin(\pi pr^2), \text{for } PST \end{cases} \tag{2.77}$$

$$\lambda = \begin{cases} \frac{1}{\pi}, p = 0 \\ \frac{2}{\pi}, p \neq 0 \end{cases} \quad (2.78)$$

Since the image function $f(r, \theta)$ is discrete and defined in the Cartesian domain, the zero[th] order approximation of Eq. (2.73) for the image function $f(x, y)$ of size $N \times N$ pixel in Cartesian domain is expressed as follows

$$A_{pq} = \lambda \sum_{i=0}^{N-1} \sum_{k=0}^{N-1} f(x_i, y_k) V_{pq}^*(x_i, y_k) \Delta x_i \Delta y_k \quad (2.79)$$

with $|p| = 0,1, \dots, p_{max}$, $|q| = 0,1, \dots, q_{max}$, where p_{max} and q_{max} are the maximum order and repetition, respectively, and

$$x_i = \frac{2i + 1 - N}{N}, y_k = \frac{2k + 1 - N}{N}, \Delta x_i = \Delta y_k = \frac{2}{N}, i, k = 0,1, \dots, N - 1$$

The image is reconstructed using the following inverse transform

$$f(x_i, y_k) = \sum_{p=-p_{max}}^{p_{max}} \sum_{q=-q_{max}}^{q_{max}} A_{pq} V(x_{i,}, y_k) \quad i, k = 0,1, \dots, N - 1 \quad (2.80)$$

2.5. Shape Alignment

Shape is interpreted as what remains when the location, scale and rotational effects are filtered out (Kendall 1977). The object shape is described by locating a finite number of points on each image, called landmarks. To bring the shapes into a common frame before further analysis, it is essential to scale, rotate and translate the shapes described by its landmarks. Procrustes analysis is an effective and a simple way to shape correspondence. However, this analysis requires a one to one landmark correspondence.

The Procrustes mean shape of similar objects is simply the average of their Procrustes fitted shapes. It is based upon a set of mathematical least square solutions to directly estimate and perform simultaneous similarity transformations among the model point coordinates matrices up to their maximal agreement. PA has many variations and forms. The generalized orthogonal PA (GPA) is most useful in shape correspondence because of the orthogonal nature of the rotation matrix. GPA is used when correspondence among more than two shapes is required. The convergence of mean is not guaranteed in GPA. Therefore, convergence is signified when the change in the mean is insignificant. No prior information is required for the geometrical relationship

existing among the different model object components. By this approach, the transformation parameters are computed in a direct and efficient way based on a selected set of corresponding point coordinates [100-102].

Let y be the reference shape and w is the shape to be aligned to shape y. For alignment of the shapes, we need to find the parameters (scale s, rotation ϕ, transation (t_x, t_y)) of the transformation $T(\cdot)$ which when applied to w best aligns it to y.

$$y = T(w) \tag{2.81}$$

To carry out the superimposition of the shapes, we need to minimize sum of squared error as given in Eq. (2.82) with respect to the parameters. (s, ϕ, t_x, t_y)

$$e^2 = \|y - T(w)\|^2 \tag{2.82}$$

The minimizing values of the parameters are given below

$$t_x = 0, \qquad t_y = 0, \qquad s = \frac{|y \times w|}{w^2}, \qquad \phi = \arg(w \times y)$$

Procrustes analysis is used to align the training shapes in a common coordinate frame to obtain the mean shape to be used as an approximate location of landmarks on the test image. In the present study based on the anatomical knowledge each shape is represented in a biological meaningful way by using three landmarks. Thus from 18 landmarks considered in the study six shapes were derived for the craniofacial structures.

2.6. Template Matching

2.6.1. Rotation invariant template matching

The classical template matching process involves cross-correlating the template with the test image and computing a measure of similarity between them. Evaluation of the correlation is computationally expensive [95]. When objects within the image are rotated with respect to each other, the methods described above cannot be used and a set of templates at different orientations are to be used. This procedure is hardly practical for real time processing when the rotation angle is arbitrary or unconstrained. In such applications, rotation invariant template methods can be used. Rotation invariant feature extraction has been achieved using moment invariants and Fourier descriptors. Moment invariants are sensitive to noise, and Fourier descriptor

of an image is considered a closed curve. Sometimes it is not easy to obtain a closed curve directly [37, 103].

In 1991, Tang et al. [104] proposed a method of transforming 2-D patterns into 1-D patterns through ring projections (RPs). It reduces the dimensionality of two dimensional patterns by performing projection along circles with distinct radius and is rotation invariant. In 2000, Tao et al. [105] proposed a method called central projections (CPs), which reduces the dimensional of a two dimensional pattern by performing projection along lines with different polar angles.

In this work, we explore CPs and RPs to extract rotation invariant features with reduced dimensionality for optimal matching. These features are used to search the landmark's location within the expectation window due to their rotational invariance property, ease of implementation and low computational complexity. The object detection by this method gives a very close match to the actual location of landmarks.

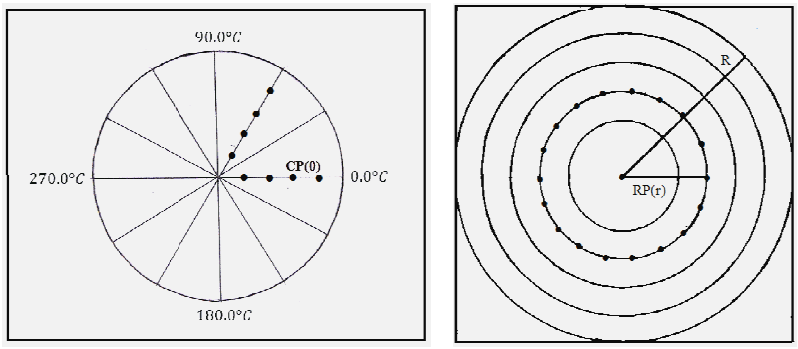
In template matching, CP and RP, transformations can be used for reduction of computational burden in the matching process and making the match rotation invariant [106]. They transform a 2D gray level image into a 1D projection space.

2.6.1.1. Central projection transforms

The CPs of image at angle ϕ_k denoted by $\boldsymbol{PC}(\boldsymbol{\phi_k})$ is defined as the mean of $f(r\ cos\phi_k, r\ sin\phi_k)$ at a specific angle ϕ_k refer Fig. 2.4 (a). The grayscale intensity values of pixels falling on discrete values of x and y coordinates are considered to compute the projections.

$$\boldsymbol{PC}(\boldsymbol{\phi_k}) = \frac{1}{n_r}\sum_{r=0}^{R} f(r\ cos\phi_k, r\ sin\phi_k) \tag{2.83}$$

Where $\phi_k \in [0,2\pi]$ and has N intervals.



(a) (b)

Figure 2.4: (a) CPs along lines with different angles and (b) RPs along different circles.

2.6.1.2. Ring projection transform

The RPs of image at radius r denoted by $\boldsymbol{PR}(\boldsymbol{r})$ is defined as the mean of $f(rcos\phi_k, rsin\phi_k)$ at a specific radius r Fig. 2.4 (b). The grayscale intensity values of pixels falling on discrete values of x and y coordinates are considered for computation of the projections.

$$\boldsymbol{PR}(\boldsymbol{r}) = \frac{1}{n_r}\sum_{\theta=0}^{2\pi} f(rcos\phi_k, rsin\phi_k) \tag{2.84}$$

where, $r = [1,2 \dots R]$ and n_r is the total number of pixels falling on the circle of radius r. The $\boldsymbol{PR}(\boldsymbol{r})$ values for all r in the window have equal importance in the computation of similarity measure. Since the projection is constructed along the different radius, the derived 1D CP pattern is invariant to rotation of its corresponding 2D image pattern. In the matching phase, the measure of similarity is given by the Euclidian distance.

2.6.2. Robust local template matching

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 correlation. Normalized cross correlation (NCC), sum of squared difference (SSD) are widely used in a pixel by pixel template matching. NCC captures the statistical difference effectively. However, it fails, when either of the windows contains constant image intensity. This results in division by zero. NCC is invariant to linear brightness changes, but is very sensitive to 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. SSD is sensitive to a few pixels presenting very large intensity differences between the two images that are compared [35, 95]. Thus, a single similarity measure may result in suboptimal solution. In case of cephalometric landmark detection, gray profile around each landmark greatly differs, accordingly using a combination of NCC and SSD as a similarity measure gives improved results then using these measures individually. In addition, the shape variability in

cephalometric radiographs 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. SSD can be viewed as the squared Euclidean distance and is calculated as follows

$$ssd[i,j] = \sum_{x=-N/2}^{x<N/2} \sum_{y=-N/2}^{y<N/2} (I_{(i+x,j+y)} - T_{(x+N/2,y+N/2)})^2 \tag{2.85}$$

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

NCC is given by

$$\text{ncc[i, j]} = \frac{\sum_{x=-N/2}^{x<N/2} \sum_{y=-N/2}^{y<N/2} \left(T_{(x+N/2,y+N/2)} - \bar{T}\right)(I_{(i+x,j+y)} - \bar{I})}{\sqrt{\sum_{x=-N/2}^{x<N/2} \sum_{y=-N/2}^{y<N/2} \left(T_{(x+N/2,y+N/2)} - \bar{T}\right)^2 (I_{(i+x,j+y)} - \bar{I})^2}} \tag{2.86}$$

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