CHAPTER 2

IMAGE CODING TECHNIQUES

A digital image could be viewed as a matrix of dimension $N \times M$, which has $N \times M$ -byte representation. Typical value for N is 256 or it may have higher spatial resolution up to 1024 for medical images. When color images are considered, the required storage is tripled. Well-known statement, “A picture is worth a thousand words” by Descartes, appears in many books on image processing and computer vision to express the observation that human being percept the images in a highly intelligent way, unfortunately images require much more storage space than it sounds. Image coding has focused on this problem. Though many image-coding methods for image compression have been proposed since late fifties, the main objective of all is to provide the best possible quality image for the minimum data rate. This chapter is devoted to review these efforts in image compression. Particularly, model-based image coding techniques will be investigated.

2.1 Introduction

Image coding techniques can be classified into *three* groups according to what they actually code. In waveform image-coding techniques an image itself or some simple variation of that such as difference in intensity at consecutive pixels is coded. In transform-based image-coding techniques, an image is transformed into another domain such as frequency domain and transform coefficients are coded. In model-based image-coding techniques, an image is modeled and model parameters are coded. Waveform image coding techniques has used information theory end its results.

2.2 Waveform Image Coding

Waveform image-coding systems consist of three elements. The first and most important element is the transformation of the image to the most suitable domain for quantization and code-word assignment. In waveform coding, we code the image intensity itself or some simple variation of image brightness such as the difference between consecutive pixel brightness. One major advantage of waveform coding is its simplicity. Since the waveform itself is coded, the coders are very simple both conceptually and computationally. Waveform coders do not generally perform as well as transform coders.

2.2.1 Pulse-Code Modulation (PCM)

The simplest waveform coding method is the basic pulse code modulation system, in which the image intensity is quantized by a uniform or a non-uniform quantizer. PCM systems introduce a noise to the image which can be modeled as additive random noise. A way of improving the performance of a PCM system is to remove the signal dependence of the quantization noise, which appears as false contours at low bit rates. Roberts`s psedonoise technique, also known as dithering, is a method that removes the signal dependence of the quantization noise. In this method, a known random noise is added to the original image before the quantization at the transmitter and then the same random noise is subtracted at the receiver. Since random noise is known at both the receiver and the transmitter prior to image transmission, it does not have to be transmitted. The compression ratio is about 3:1.

2.2.2 Delta Modulation (DM)

In the PCM system, the image intensity is coded bit scalar quantization, and the correlation among pixel intensities is not exploited. One way of exploiting some of the correlation is delta modulation. In the DM system, the difference between two consecutive pixel intensities is coded by a one-bit quantizer. Although the dynamic range of the difference signal is doubled as a result of differentiation, the variance of the difference signal is significantly reduced due to the strong correlation typically present in the intensities of two pixels that are spatially close.

An important design parameter in DM is the step size. In the region where the signal varies slowly, the reconstructed signal varies rapidly around the original signal. This is called *granular noise*. A large step size results in a correspondingly large amount of granular noise. When the signal increases or decreases rapidly, it may take many pixels to catch up original signal using small step size. The reconstructed signal will appear blurred in such regions. This is called *slope overload distortion*. The compression ratio is about 4:1.

2.2.3 Differential Pulse Code Modulation (DPCM)

Differential pulse code modulation (DPCM) can be viewed as a generalization of DM. In DM, the difference signal is quantized by a one-bit quantizer. In DPCM, more than one bit can be used in coding the error. Since a PCM is a component of a DPCM system, it is possible to use Robert`s pseudonoise technique in a DPCM system. Since we reduce the number of bits available to encode each pixel, the quantization noise will be less if we use DPCM rather than PCM at the same rate. The compression ratio is between 2:1 and 3.5:1 depending on the image statistics.

This approach can easily be extended to the two dimensions. Two dimensional DPCM performs better than PCM and one dimensional DPCM by about 3.75:1.

2.2.4 Predictive Coding Techniques

There exists a statistical dependence between gray values at consecutive pixels. Previous transmitted signals convey some sort of information about oncoming signals. Prediction techniques are used to exploit the dependency and prediction sequence is defined as the error between estimated and actual signals. Now, prediction error signal is quantized instead of transmitted signal. Image compression ability depends on the prediction technique used and correlation exists among neighboring pixels. If all individual signals are mutually independent, then there is no advantage over PCM or DPCM. Prediction image coding assumes some amount of dependency and small variance of prediction error sequence.

2.3 Transform Coding

In transform image coding, an image is transformed to a domain significantly different from the image intensity domain, and the transform coefficients are then coded. In low bit rate applications, transform coding techniques with scalar quantization typically perform significantly better than waveform coding techniques with scalar quantization. However they are more expensive computationally.

Transform coding is significantly different from DPCM and achieves compression in transform domain. Transform coding techniques attempt to reduce the correlation that exists among image pixel intensities more fully than do waveform coding techniques. Transform coding techniques also exploits the observation that for typical images a large amount of energy is concentrated in a small fraction of the transform coefficients. This is called the *energy compaction* property. Because of this property, it is possible to code only a fraction of the transform coefficients without seriously affecting the image quality.

The statistically optimal linear block transform, in the sense that it minimizes the mean squared reconstruction error, for coding images is well known to be the Karhunen-Loeve transformation (KLT) [1]. The KLT is related to principal component analysis, since the basis vectors are also the principal components of data. Because the KLT is an orthonormal transformation, its inverse is simply its transpose.

$$\bar{y} = W\bar{x}$$
$$\sum = E[\bar{x}\bar{x}^T]$$

A number of practical difficulties exist when trying to implement the KLT. The calculation of the covariance estimate, Σ, and its eigendecomposition is not practical even with today's computing resources. The algorithms used to do these computations are complex and therefore not suitable for hardware implementation. The calculation of the covariance estimate requires $O(N^2)$ calculations. Due to these difficulties, fixed basis transforms such as discrete cosine transform (DCT) [2], which can be computed in order $O(N \log N)$, are typically used.

2.3.1 Hybrid Coding

The term refers to techniques that combine transform coding which performs very well in low bit rate applications and waveform coding which is very simple to implement. In hybrid coding, a two-dimensional image or its row or column is transformed to obtain statistically independent sequence of transform coefficients accumulating in a narrow energy band known as energy compaction property and then this sequence is coded by waveform coders such as DPCM. Hybrid coding can achieve the compression ratio up to 8.0:1.

Hybrid coding of a single image is not practical since it does not decorrelate the image as much as 2D transform coding and implementation complexity of transform coding is not much more than hybrid coders. Hybrid coding is useful in interframe image coding. Exploiting the temporal correlation as well as spatial correlation to code a sequence is called *interframe coding*. In intraframe hybrid coding, every frame is computed by 2D transform and then waveform coding is applied to 2D transform coefficients along temporal direction.

2.3.2 Adaptive Coding and Vector Quantization (VQ)

Transform coding can be made adaptive to the local characteristics within a block of particular size. For example, transform block size can be chosen small in regions containing edges. Adaptive coding significantly improves the performance of transform coding while adding little to its complexity.

Transform coefficients are generally coded with scalar quantizers. It is also possible to use vector quantizers in which designing goal is to obtain a quantizer consisting of N reproduction vectors, such that it minimizes the expected distortion rate.

Vector quantization is the joint quantization of the components of a vector. Unlike scalar quantization, it is often used to requantize signals that are already digital, for the purpose of compression. When vector quantization is used for image compression, the image is partitioned into blocks of $N = n \times n$ pixels, which form an N-dimensional vector. This vector is encoded by searching a codebook of representative quantization vectors. Let $x = \left[x_1,...,x_N\right]^T$ denote a vector that is formed of a block of size N of the image. In VQ, x is mapped into another N-dimensional vector $q = \left[q_1,...,q_N\right]^T$. The vector q is chosen from L possible quantization codewords in q way that minimizes an Euclidean distance measure. The quantization codewords are chosen so as to be optimal for a given distribution. The techniques for choosing the quantization symbols and the different search algorithms for the encoding stage have been reviewed and detailed in [4]. One of the most popular VQ algorithms makes use of the K-means algorithm and is known as the LBG algorithm [5]. Given the number of clusters to be formed, it iteratively refines the cluster centers and boundaries. The iteration is stopped when the mean square error falls below a threshold or remains constant between iterations.

2.3.3 Two-Channel Coders

In a two-channel coder [6], an image, $f(n_1,n_2)$, is divided into two components: low and high frequency components. The lows component $f_L(n_1,n_2)$ consists of low-frequency components and represents the local luminance mean of $f(n_1,n_2)$. The high component $f_H(n_1,n_2)$ consists of high-frequency components and represents the local contrast of $f(n_1,n_2)$. Since low component is a low-pass filtered version of $f(n_1,n_2)$, it can be down sampled and interpolated by low-order polynomials or splines. High components can be coarsely quantized by a PCM system which may use Robert's pseudorandom noise technique.

Two-channel coding system is shown in Figure 2.1. The original image $f(n_1,n_2)$ is low-pass filtered by a FIR filter. Since lows component is smooth enough to be represented by a polynomial at coarser levels, the lows component $f_L(n_1,n_2)$ is sub sampled and estimated by a polynomial. The high component is obtained by subtracting the $f_L(n_1,n_2)$ from $f(n_1,n_2)$ and quantized by a PCM.

A two-channel coding system can be viewed as a special case of a subband image coder. Coding methods in which a signal is divided into many channels and each channel is then coded with own coder is called *subband image coding* [7].

2.3.4 Fractal Image Compression

A fractal is considered as a function, *F*, having the following properties:

1. *F* has detail at every scale,
2. *F* is self-similar,
3. The fractal dimension of *F* is greater than its topological dimension.
4. There is a simple algorithmic definition of *F*.

Fractal image coding [8] is based on the assumption that images are exactly or at least statistically self-similar. By using self-similarity, some regions called *domains* are transformed into such regions called *range* that means square error between ranges and transformed domain is minimum. Then the transformation coefficients are coded. Reconstruction is done by iteratively applying the transform onto itself. This is called *iterated function system*. Due to the property (1), iterations result in an image called *attractor* and final image, attractor, is not affected by initial image. Different transformations lead to different attractors. The only limitation on the selection of transformation is that the transformation must be contractive. In practice, the transformation is of the form

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} a_i & b_i \\ c_i & d_i \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} e_i \\ f_i \end{bmatrix}$$

called *affine* transformation. Each affine transformation is defined by six numbers $a_i, b_i, c_i, d_i, e_i, f_i$.

Performance of fractal image compression is restricted with the self similarity of the image. Real images are not exactly self-similar which makes the compression ratio lowered.

2.3.5 Pyramid Coding

A pyramid is a pyramid-like grid structure where bottom of the pyramid corresponds to finest scale and top of the pyramid corresponds to coarsest grid. An image is described in the scale space on the pyramid from finer scales to coarser by convolving a scale-space filter like gaussian function or solving diffusion equation in time. The difference image between two consecutive levels is called *residual*. Since residuals are smooth signals, they can be represented finely at one coarser level by subsampling and interpolating. This process is iterated at every level of the pyramid.

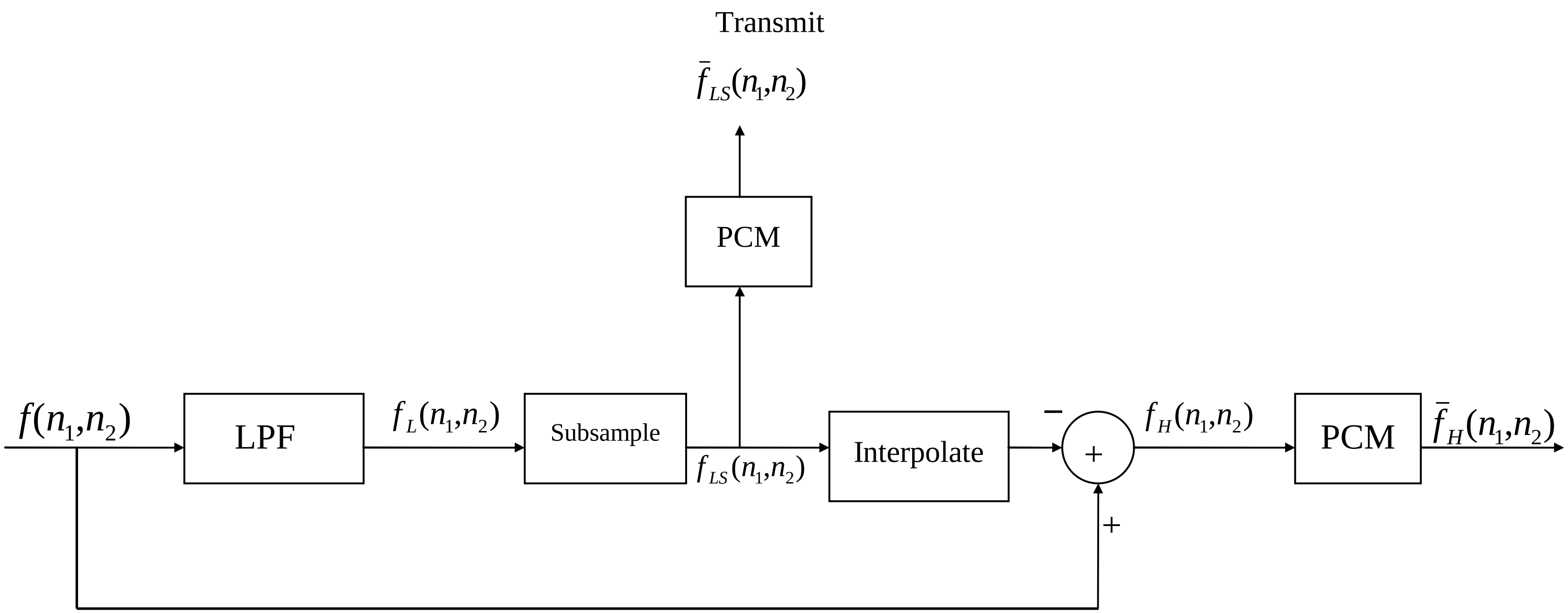


Figure 2.1 Two-channel image coding system

The image at the coarsest level is the blurry version the original. Each residual is coded by DPCM and transmitted. At the receiver, residuals from coarsest level to finest level are interpolated and added to the previous received residual. As the process is repeated for each level on the receiver, the reconstructed image will become higher spatial resolution image. One major advantage of the pyramid is its ability to progressive image transmission.

2.3.6 Wavelet Image Coding

A wavelet transform is the decomposition of a signal into a set of basis functions consisting of contractions, expansions and translations of a mother function $\psi(t)$ called the wavelet [11,12]. It is equivalent to the analysis of a signal into several frequency bands having the same bandwidth on a logarithmic scale, each one representing a tradeoff between time and frequency resolution:

$$\text{Time - Bandwidth Product} = \Delta t \times \Delta f \geq \frac{1}{4\pi}$$

This is referred to as the uncertainty principle, or Heisenberg inequality. It means that one can only trade time resolution, or vice versa. The wavelet transform represents a different interpretation of subband coding. In this interpretation, other features of the filter banks than frequency selectivity, such as regularity and number of vanishing moments, are considered [13,14].

2.4 Model-based Image Coding Techniques

Common feature of all coding methods is to exploit the redundancies exist in images by decorrelating the image. Since techniques in first and second class explained above make use of information theory and its results, decorrelation is based on the entropy and entropy coding. Information theory suggests that upper level of coding gain is bounded by the entropy for a given signal and its distribution. Fortunately, the entropy of an image is unknown and depends mostly on the model used. Model-based image coding techniques [15,16] , contrast to previous class methods, make use of some features of images such as edges, segments, textures. This is based on the fact that an image can be described by its discontinuities where sharp changes occur in brightness and by regions where brightness is equally distributed. In fact, most of the computer vision high level processes are partially based on the edges, regions and texture information. Human brain also processes the visual information in the form of edges and regions.

The first study which inspired such methods classified as model-based techniques is Graham`s paper [17] on edge-based compression and actually goes back to late sixties. There are two reasons why edge-based techniques did not take so much interest of researchers as it deserves. First, from sixties to late seventies information theory was on the way to develop and, information theory tools were considered as the optimal way of coding. Second, there were no study on coding of arbitrary curves effectively. In fact, in middle of seventies Freeman [18] put forward a method called chain code for curve representation. Later, Kaneko and Okudaira [19] developed chain code representation by utilizing edge link concept. Schreiber [6] considered an image as being composed of low and high frequency parts, which are encoded separately. The high frequency part corresponds to contour and low frequency part corresponds to regions. Kunt et al. [15] use region growing to segment the image into regions and gray level intensity is modeled and encoded by polynomials. Region-based methods suffer from the artifacts that the transitions at the region borders are step edges. Kunt offered a solution that tries to model the edge profile by wavelets. Carlson [20] presented a sketch-based coding scheme for gray level images. Carlson proposed to code the intensity on either side of each edge, and interpolate image intensity between contours by diffusing intensity estimates from these edges. The results were not good in perceptual quality. There are many reasons for the distortions on the reconstructed image. Distortion is mostly caused by the incomplete information on data. Recently, Elder and Zucker [21] have proposed an algorithm for reliable edge detection and blur estimation. Scale is estimated only on edges and approximated elsewhere by diffusing them. The reconstructed image is obtained by solving anisotropic heat equation with space-varying scale parameter. The results are perceptually very close to the original images. But they did not mention how to code the scale estimates at edges, intensities and contrast information.

Acar and Gençata [22,23] used a model called weak membrane. They formulated the edge detection and surface reconstruction problems as a regularization problem with a non-convex energy functional. They proposed to code the intensity on either side of each edge in the image modeled by weak membrane. Since the image is modeled by weak membrane, the resultant image is far from being perfect.

Gökmen and Ersoy [24] has recently developed a model-based coding method for fingerprint images. They use the regular structure of the fingerprint images and hybrid model. They proposed to detect ridges and valleys by adaptive thresholding and to code the intensities on ridges and valleys. Reconstruction was done by minimizing the hybrid energy functional. One advantage of the model is its ability of preserving the fingerprint structure in high compression ratios. Because, what they code is fingerprint structure itself. Another advantage is that one can process even on the compressed data without reconstruction for the recognition or classification purpose. The method suffers from two problems. First, coding and decoding processes require heavy computation load. Second, the reconstructed image is mostly blurry version of the original fingerprint.

J. Rabinson [26] presented a different model which uses ridges and valleys as perceptual image primitives instead of edges and regions. The primitive curves (ridges and valleys) are called “threads”. Transmitted data are the thread locations, shapes and profiles. Reconstruction is done by interpolating the transmitted data. The image is first filtered with a valley/ridge detection operator such as a Gaussian-smoothed second derivative. He reported that LOG with Gaussian variance of 1.4 pixels had been used. The threads were obtained by a valley/ridge follower. Location of valley/ridge was coded by standard chain code. Intensities and profiles on valleys/ridges were compressed by fractal coding. The unknown pixels are interpolated by C^0 Natural Neighbor Interpolation [27]. Since the method is based on ridge/valley primitives, the approach preserves texture rather than edges. The model has a drawback that since the valleys/ridges are mostly discontinues, the interpolation causes the reconstructed image to blur at the discontinues.

U. Y. Desai et al. [28] has recently proposed an edge and mean-based compression method for color images. They used Sobel operator to detect edges. Contour intensity was coded by line-fitting in one dimension for both vertical and horizontal directions. In order to enhance the reconstructed image quality while keeping the bit rate down, mean of the image in a block of size 10x10 was also coded. In decoding stage, the mean in a block is used to modify the mean of the interpolated image. They reported acceptable quality images at 0.1 to 0.3 bpp for (256x256) color images.

Dijk and Martens [29] presented a method combining the transform coding and model-based coding. They introduced steered Hermite Transform described as a weighted sum of orthogonal polynomials. The term “steerable” is used to describe a class of filters in which a filter of arbitrary orientation is synthesized as a linear combination of a set of basis filters. Hermite filters of order n form a steerable basis for every individual filter of order n. It had been shown that Hermite transform has the advantage that high energy compaction can be achieved by adaptively steering the transform. They used Gaussian edge model and expressed explicitly the local edge parameters as a function of the Hermite Transform coefficients. Unfortunately, they did not deal with the problem of compressing these parameters.

Salembier et al. [30] has presented morphologic tools to compress images. They defined four morphologic operators rather than complete coding scheme. These operators are as the following:

- Connected Operators [31]: The operator solves the problem of image simplification while preserving the contour information.
- Region-growing version of the watershed [32] : The watershed transformation is the classical morphological tool for segmentation
- Geodesic skeleton [33]: In region-based image coding, there is a problem of assigning a contour to a region while the contour belongs to at least two different regions.
- Morphological interpolation: Morphological interpolation can be used very efficiently to interpolate on irregular grids.

There are many studies on region-based image coding [34]. In [35], an image is partitioned into distinct regions by using a segmentation algorithm, the contents of the regions are then coded using polynomials. The effect of polynomial order was studied by comparing the segmentation and rate-distortion performance produced by different order approximations. In contrast to the previous studies on region-based coding, texture on each region was modeled.

2.5 Centipede Model

We have seen the following problems with the previous edge-based coding techniques

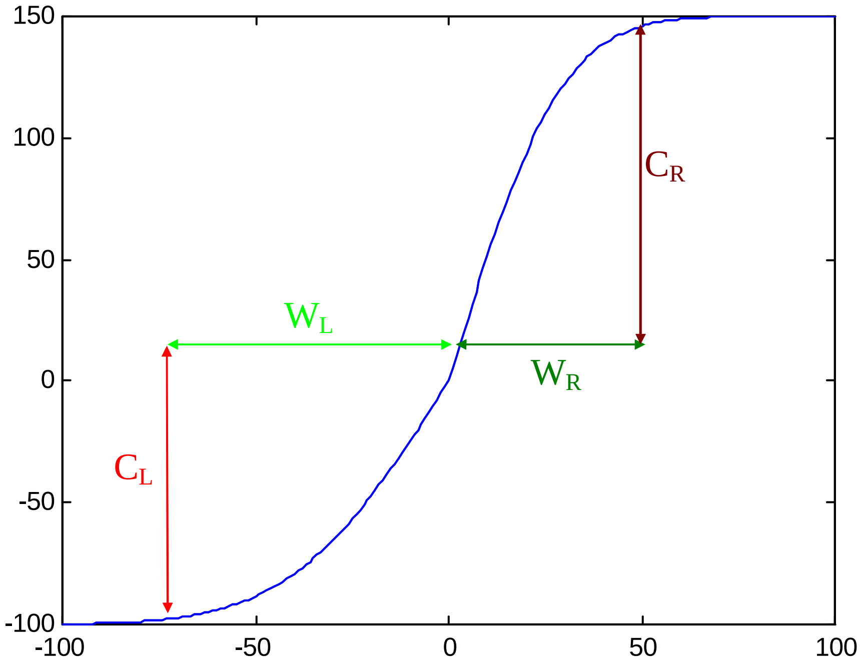
1) Edges are generally modeled by symmetric Gaussian (or similar) functions, causes that reconstructed image looking like artificial.

2) Since they do not attempt to control the selection of contours or the selection is done randomly, some contextually important features may disappear.

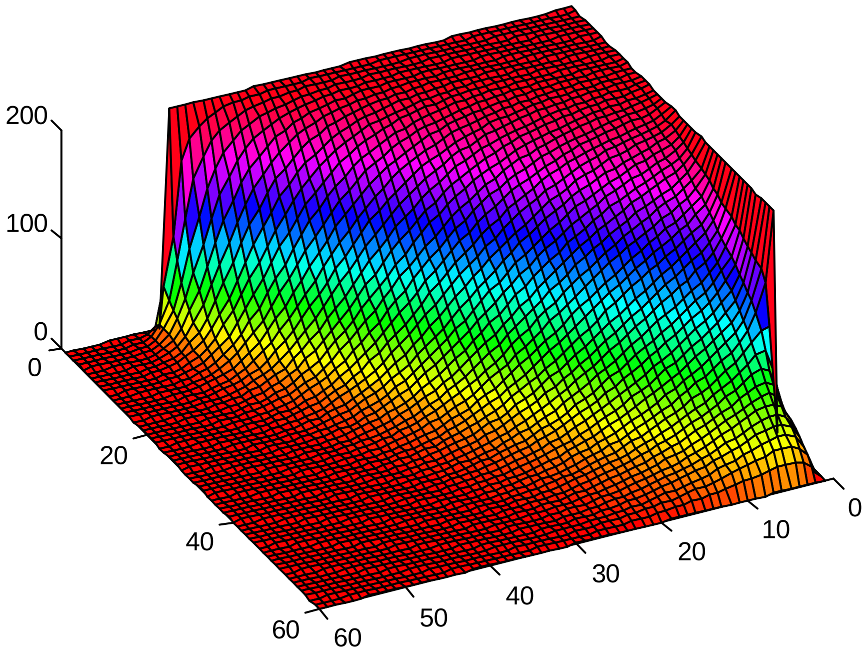
3) None of the techniques develops a unified approach for edge detection, contour selection and coding, edge representation, and reconstruction.

We have developed a model which we call "Centipede Model". Centipede model [25] consists of five parameters : Intensity on edge (I_L), Left and Right Contrast (C_L, C_R) and Left and Right Width (W_L, W_R). The model parameters are drawn on an edge profile given in Figure 2.2 (a). The model parameters are extracted on each edge element on the contour shown in Figure 2.2 (c). Two-dimensional edge profile is given in Figure 2.2 (b). Width is defined as the distance at which difference in consecutive pixel is lower than a given threshold. Threshold is determined from the SNR (dB)ratio for an image. Width is a rough estimate of edge scale. In chapter 4, it has been shown that the tuple (I_L,C_L, C_R, W_L, W_R) with the edge map constructs a powerful representation of the image (solve the problem (1)). How the parameters are extracted and coded are explained in chapter 4. We present an approximation scheme for the model parameters by polynomials with varying order and constant block length. We also investigate how the selection of polynomial order and block length affect the quality of the reconstructed image.

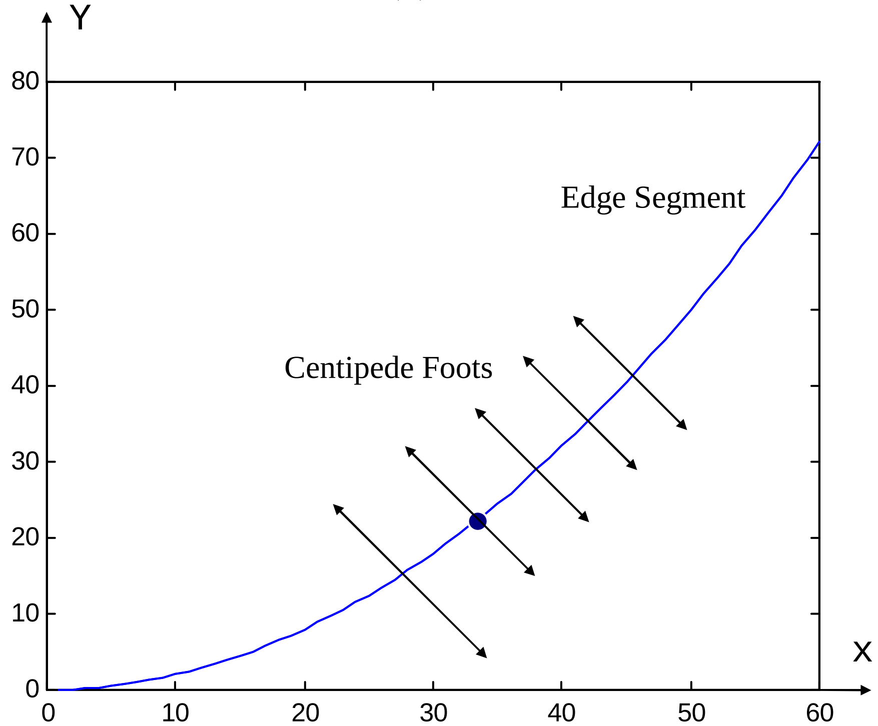
We propose a contour selection method based on the ordering of edge segments regarding to normalized feature set containing length, contrast and curvature in chapter 3. The method is experimentally tested and verified that for high compression ratios all perceptually most important features are still kept (solve the problem (2)).



(a)



(b)



(c)

Figure 2.2

(a) Centipede Model Parameters for the edge on the contour shown in (c),
(b) 2D Edge Profile,
(c) Contour.