CHAPTER 1

INTRODUCTION

1.1 OVERVIEW OF IMAGE COMPRESSION

Digital image and image sequences in uncompressed formats require excessive storage capacity and huge transmission bandwidth. Therefore compression of digital image and image sequence is necessary for efficient storage and transmission. A typical lossy digital image compression scheme consists of signal decomposition, quantization and coding. Image compression is possible due to redundancy and irrelevancy. Redundancy (Jain 1981) is a characteristic which is related to predictability, smoothness, etc., in the image data. All natural images exhibit some form of internal structures, which can be explained in terms of statistical dependencies between pixels. This dependency is utilized by signal decomposition. The irrelevancy relies on the fact that the human eye cannot perceive certain image degradations due to the visual perception mechanisms. This is utilized by the quantization scheme. Thus, the role of the signal decomposition is to decorrelate the signal and make the subsequent quantization process easier. Once the decomposed signal samples are quantized, an appropriate symbol coding strategy is used to represent the quantized symbols by bits. Image compression can be achieved either by means of spatial domain techniques or by means of frequency domain techniques.

The two major classes of image compression algorithms are lossy and lossless algorithms. Lossless algorithms preserve the image data in such a way that the original and the reconstructed images are exactly the same. In
lossy image compression, the original and the reconstructed images may or may not be identical in a strict mathematical sense, but to a human observer they may look the same; so the goal is to achieve compression that is visually lossless. Both lossy and lossless compression algorithms are used today in a broad range of applications right from transmitting satellite images to web browsing to image printing and scanning. For image, and particularly video sources, some loss of information can usually be tolerated. There are three reasons for this. Firstly, significant loss of information can often be tolerated by the human visual system without interfering with perception of the image or video sequence. Secondly, in many cases the digital input to the compression algorithm is itself an imperfect representation of the real-world scene. Thirdly, lossless image or video compression is usually incapable of satisfying the high compression requirements of most storage and distribution applications. In this work, the focus is on lossy image and video compression.

1.2 NEED FOR COMPRESSION

In the current digital era, family photo albums are stored not only on the bookshelf, but also on the personal computer. A medical doctor can make a diagnosis using a full three dimensional image on a computer screen. Satellite images, which convey information regarding earth resources, are continually transmitted over communication channels. The amount of data transmitted through the Internet almost doubles every year, and a large section of data comprises images. Uncompressed image and video data require considerable storage capacity and high bandwidth networks for their transmission. Reducing the bandwidth needs of any given device will result in significant cost reductions and will make use of the device more affordable. Magnetic hard discs, CD’s, DVD’s of larger capacity are released every year, in response to greater demand for storage of digital data. While the advancement of computer storage technology continues at a rapid pace, a means for reducing the storage requirements of an image is still needed in
most situations. Image compression offers ways to represent an image in a more compact way, so that one can store more images and transmit images faster. The advantages of image compression come at the expense of numerical computations, and therefore one can trade off computations for storage and bandwidth. Before storing or transmitting an image, it is processed in such a way that it will require fewer bits for its representation.

1.3 TRANSFORM-BASED COMPRESSION

Image compression methods can be broadly classified into two types which are:

- Transform based compression
- Predictive based compression

Predictive based compression removes the redundancy between successive pixels by encoding only the residual between the actual and the predicted values. Since the residue value usually has a much smaller dynamic range, this leads to fewer encoding bits and fewer quantization levels needed; thus, indirectly leads to compression. Some predictive based compression methods are Huffman coding, Differential Huffman coding, Differential Arithmetic codes, and Delta Pulse Code Modulation or Linear Predictive Encoding. The block diagram that depicts the basic transform based image compression/decompression model is shown in Figure 1.1.

![Block Diagram](image_url)

**Figure 1.1 The basic image compression/decompression model**
In the transform stage, the main goal is to decorrelate the original image data, so that the original image energy is redistributed among only a small set of transform coefficients. This decorrelation eliminates the inter-pixel redundancy, thereby, providing a representation that can be coded more efficiently. The zeroth order entropy of the transformed coefficients is much lower than that of the original image. The transforms used in coding are reversible, so that the original image can be recovered by the inverse transform, provided no quantization has been performed on the forward transform coefficients. Theoretically, the forward/inverse transform dual is a lossless process. The choice of transform used depends on a number of factors; in particular, computational complexity and coding gain. Computational complexity is measured in terms of number of multiplications and additions required for the implementation of the transform. Coding gain is a measure of how well the transformation compacts signal energy into a smaller number of coefficients. The Karhunen-Loeve Transform (KLT) is optimal in the sense that it minimizes the mean square distortion of the reconstructed image for a given number of bits (Jain 1989). However, KLT is rarely used in practice because of its substantial computational time and the basis vectors must be calculated for each image. The two-dimensional Fourier Transform is not used for compression because of the complex arithmetic involved. Nevertheless, it is useful if a frequency domain representation is needed for analysis. The Discrete Cosine Transform (DCT) is often used in practical applications as its basis vectors are two-dimensional cosine functions and it yields performance near that of the KLT. Also, the arithmetic is real and fast computational algorithms are also available. Xiong et al (1999) have done a good study on the benefits, drawbacks, and performance of DCT and wavelet transforms. It is observed by the authors that the DCT used in the JPEG standard is computationally less complex than the wavelet transforms for the same number of image samples. However, it is also recognized that wavelet transforms can achieve coding gain superior to that of
the DCT. Wavelet transforms also allow additional freedom in the selection of particular filter used; in contrast, there is only one DCT. The transforms considered in this work are Wavelet Transform, Wavelet Packet Transform, Multiwavelet Transform, and Contourlet Transform.

The second stage, quantization / dequantization is the process that leads to lossy compression. The dynamic range of the transform coefficients are narrowed by the quantization process; thus achieving high compression. Quantization can be either scalar quantization or vector quantization. In scalar quantization, a single coefficient ‘C’ is divided by a quantization factor ‘Q’ and rounded off to the nearest integer in order to obtain the quantized coefficient which is given by equation (1.1)

\[ T(C) = \text{Round}\left( \frac{C}{Q} \right) \]  

In vector quantization, the coefficient set is divided into 1-D or 2-D blocks, that are called code vectors, and a codebook is used to find a pattern for each block. The approximated pattern for each block is coded using a lookup value in the code book. In the quantization section, psycho visual redundancy in the image is reduced by throwing away the unwanted bits from the transform coefficients. This leads to high compression ratio and distortion in image fidelity. Image-Adaptive Multistage Vector Quantization scheme is used in this thesis.

The third stage, namely, entropy coding, determines the number of bits required to represent a particular image at a given quantization level. The combined process of entropy coding and entropy decoding is lossless. It maps the quantized transform coefficients into a bit stream using variable length codes, thus exploiting the coding redundancy.
1.4 LITERATURE SURVEY

Image compression algorithms have been the subject of research both in academia and industry for many years. The first widely adopted international image compression standard was Joint Photographic Expert Group (JPEG) (Pannebaker and Mitchell 1994) which was introduced in the late eighties. JPEG is based on DCT followed by entropy coding based on either Huffman coding (Huffman 1952) or binary arithmetic coding (Cover and Thomas 1991). It has been widely used right from the printing industry to Internet applications. JPEG is intended for continuous tone images of more than one bit depth. Algorithms for binary images work in a different way; JBIG-1 and JBIG-2 are the standards covering this area. JPEG and JBIG are part of other standards such as facsimile transmission standards, FlashPix file format (Hauf and Houchin 1996), the TIFF file format, and page description languages like PDF. In recent years, researchers have been using the Discrete Wavelet Transform (DWT) in image compression. Wavelet based coding does not require partitioning of the image. Hence, the typical blocking artifacts, like the ones occurring in JPEG, are avoided and the computation time is hardly increased Munteanu et al (1999). Burt and Adelson (1983) were the first to introduce multiresolution analysis in image compression. While their approach seemed counter intuitive at the first glance, given that it increased the number of samples to be coded, their results were promising. Mallat (1989) was the first to point out the connection between multiresolutional analysis and the wavelet transform. Daubechies (1988) studied the Discrete Wavelet Transform and popularized this tool among the scientific community. Some of the first papers on wavelet image compression (Antonini et al 1992) claimed excellent compression performance results and gave a lot of intuition behind the use of the wavelet transform in image compression. Numerous organizations have been using wavelet compression algorithms as their own, internal compression standards. An example is the
FBI (Hooper et al 1993) where there is a need for storing large databases of fingerprints and JPEG did not satisfy their requirements. However, wavelets are inefficient in representing texture-rich images. The work of Xiong et al (1998) shows that using wavelet packet decomposition to an image can improve PSNR by 0.8-1.3 dB for some texture-rich images, when compared with the standard wavelet transform. Rajpoot et al (1999) and Khalil et al (1999) applied zerotree technique to wavelet packet transforms to generate an embedded bitstream. Unfortunately, these techniques cannot avoid the difficulty of defining zerotree set for wavelet packet transform data of images. Motivated by this, this thesis applies multistage vector quantization scheme to wavelet packet coefficients, where it is not required to define the zerotree set. In particular, an attempt has been made to select the best basis of Wavelet Packet Transform through Singular Value Decomposition.

Recently, multiwavelets have been introduced as a more powerful multi-scale analysis tool. A scalar wavelet system is based on a single scaling function and mother wavelet. On the other hand, a multiwavelet uses several scaling functions and mother wavelets. This adds several degrees of freedom in multiwavelet design and generates useful properties such as symmetry, orthogonality, short support and higher number of vanishing moments, simultaneously. Strela et al (1999) proposed an approach for separable 2D multiwavelet decomposition which triggered the use of multiwavelet in the field of image processing. Geronimo, Hardin, and Massopust constructed one of the most well-known multiwavelets, called GHM (Geronimo et al 1994). GHM multiwavelet is used in this thesis. GHM scaling functions have short support and are symmetric about their centers. In addition to GHM multiwavelet, two other multiwavelets, namely, CL constructed by Chui and Lian (1996) and SA4 constructed by Shen et al (2000) have been used.
Multiscale and multidirectional are properties that are desirable for image decomposition. A multiresolution discrete transform represents image data in a hierarchical manner; each added level contributes to a finer representation of the image. Images have efficiently been modeled using the wavelet transform, which offers a multiscale and time-frequency localized image representation. The major drawback for wavelets in two-dimensions is their limited ability in capturing directional information. Research in human visual system shows that directional information plays an important role in the human visual perception process. Therefore, the transformation of image into directional representations should be the first step in the extraction and understanding of visual information. Researchers have proposed multiscale and directional representations that can capture the intrinsic geometrical structures such as smooth contours in natural images. Some examples include the steerable pyramid (Simoncelli et al 1992), bushlets (Meyer and Coifman 1997), complex wavelets (Kingsbury 2001), and curvelet transform (Candes and Donoho 2003). In particular, the curvelet transform, pioneered by Candes and Donoho, was shown to be optimal in a certain sense for functions in the continuous domain with curved singularities. Inspired by curvelets, Minh Do and Martin Vetterli, developed the contourlet transform based on efficient two-dimensional multiscale and directional filterbank that can deal effectively with images having smooth contours. Contourlets not only possess the main features of wavelets such as multiscale and time-frequency localization, but also offer a high degree of directionality and anisotropy. The main difference between contourlets and other multiscale directional systems is that the contourlet transform allows for different and flexible number of directions at each scale, while achieving nearly critical sampling. In addition, the contourlet transform uses iterated filterbanks, which makes it computationally efficient; specifically, it requires $O(N)$ operations for an $N$-pixel image. All these significant features of contourlet transform have helped this thesis to
apply image-adaptive multistage vector quantization technique to contourlet coefficients.

Vector Quantization (VQ) is a source coding technique that approximates block of input data by one of a finite number of pre-stored vectors in a codebook. The challenge is to find the set of vectors such that the total distortion between the actual source and the quantized source is as small as possible. Since distortion depends on the codebook design, design of vector quantizer is vital for the performance of the VQ-based system. The traditional VQ design approach is the Generalized Lloyd algorithm (GLA), also known as the LBG algorithm (Linde et al 1980). The GLA is an extension of the Lloyd’s algorithm to VQ design, where the original Lloyd algorithm was proposed for scalar quantizer design. The following two conditions are necessary for a quantizer to be locally optimal:

- The quantizer partition must be optimal for a given set of codevectors
- The set of codevectors must be optimal for the partition

For the mean squared error distortion criterion, the first condition implies nearest neighbor quantization rule and the second condition implies that the codevectors are located at the centroid of the corresponding partition. Shannon theory states that VQ can perform close to the theoretical optimal performance for a given rate, if the vectors have sufficiently large dimension. Unfortunately, code complexity grows exponentially with vector dimension (Cosmon et al 1993). A class of structural constraints that simultaneously reduces both computation and memory is the class of product code VQ structures (Barnes et al 1996). A product code VQ is a structured vector quantizer, where different components of the VQ system quantize different attributes or features of the source outputs. Examples include mean-residual
VQ (Baker and Gray 1983), gain-shape VQ (Sabin and Gray 1984), and mean-gain-shape VQ (Oehler and Gray 1993). Often implicit in the phrase “product code VQ” is the assumption that the components of the product code are sequentially searched in a predetermined order so as to find the best codevector match for each source vector attribute. A simple type of product code VQ is a VQ with a direct sum codebook structure. When the direct sum product code structure is combined with a sequential search procedure, the resulting quantizer has a sequence of encoder stages, where each stage encodes the residual or error vector of the previous stage. Such sequential search product code VQ’s are called Residual Vector Quantizers. Residual Vector Quantizers are also called multiple stage VQ (Juang and Gray 1982), multiple step VQ (Gray 1984), cascaded VQ (Makhoul et al 1985), and summation product code VQ.

Subband decomposition and VQ are the two powerful tools for image compression. Recent years have seen an explosion in the research efforts to put the two tools together in useful ways. Both the multiresolution nature of subband decompositions and the progressive nature of the quantization methods provide means of making reconstruction quality scalable or progressive. Scalability implies that there is a “successive approximation” property in the bit stream. As the decoder gets more bits from the encoder, it can decode a progressively better reconstruction of the image. With scalable coding, a single encoder can provide a variety of rates to customers with different channels or display capabilities. Since images can be reconstructed to increasing quality as additional bits arrive, it provides a natural means of adjusting to changing channel capacities and a more effective means of using a relatively slow channel. Vector quantization has been successively applied to the wavelet transformed coefficients, resulting in high quality image compression (Westernick et al 1998). In wavelet-based vector quantization, multiple codebooks are used such that each subband has
its own sub-codebook. This method reduces the computational requirements for the codebook design and codeword reconstruction. Application of MSVQ was first found in speech coding. Wang and Goldberg (1998) applied MSVQ to image coding. Nonadaptive image vector quantization employs a single codebook generated from a large set of training images. To maintain coding fidelity, the codebook size must be large, which, in turn, entails more bits for coding the labels and increased coder/decoder complexity (Goldberg et al 1986). Adaptive methods which track the local behavior of the image statistics, not only from image to image, but within an image, can improve the overall performance (Habibi 1977). VQ is asymptotically optimal for the coding of a data source whose statistics are stationary in time. If the data source is non-stationary (Fowler 1998), there is a gap between the performance predicted by theory and that actually obtained in real implementations. Indeed, the non-stationary nature of the sources common in practical applications has prompted a search for more general VQ algorithms that are capable of adapting to changing source statistics as coding progresses. Adaptive Vector Quantization (AVQ) algorithms are capable of adapting to changing source statistics as the coding process progresses (Guobin Shen et al 2003). There are generally two basic approaches to training set selection and codebook generation: universal and adaptive. In the universal approach (Gersho and Ramamurthi 1982), a codebook is generated from a set of training images and remains fixed at both the transmitter and the receiver. The problem with this universal approach is that there is no guarantee that images outside the training set are well represented. Furthermore, to ensure high fidelity, a large codebook may be needed; thus increasing both the encoding complexity and bit rate required for transmitting the labels. In contrast, in image-adaptive vector quantization (Wang et al 1992), the codebook is generated directly from the input image. Therefore, the input image is better represented and, for a given distortion, a smaller codebook is sufficient. This has lead to the development of image-adaptive multistage vector quantization
scheme, in this thesis, where the centroid of the codebook is decided by the dynamic range of the input image.

1.5 PROBLEM STATEMENT AND SCOPE OF THE THESIS

Numerous techniques have already been developed for still image and image sequence compression. However, it is very difficult for a compression technique to actually achieve simultaneously the optimality in compression performance and the computational simplicity due to the complexity of the still image/image sequence compression process itself. Wavelet transforms have attracted considerable attention, especially with applications to image coding, due to their ability to provide attractive space-frequency tradeoffs for natural images. Some of the drawbacks of wavelet based image compression are listed below:

- Wavelets with several vanishing moments are effective for coding locally smooth images. However, wavelets are not efficient in representing texture rich images such as fingerprint images and remote sensing images. The reason is that such images are mainly described by the smaller scale wavelet coefficients. Those smaller scale coefficients carry very little energy, and are often quantized to zero, even at high bit rates. A new representation which is capable of providing arbitrary frequency resolution so as to meet a signal’s spectral behavior is needed.

- The signal processing properties such as compact support, orthogonality, symmetry, and higher order approximation are to be satisfied by the filters used for compression scheme. These properties are not met by wavelet filters simultaneously. A new representation in which these properties can be met
simultaneously is necessary so as to compress the image efficiently at low bit rate.

- Natural images contain intrinsic geometrical structures (smooth edges) that carry a key feature in visual information. Many wavelet coefficients are required to reconstruct the edges in an image properly. Reducing the number of coefficients will introduce artifacts on the edges of the reconstructed image. Two dimensional wavelet is constructed by taking the tensor product of 1-D wavelets. In discrete domain, this is equal to using 2-D separable filter banks, which consist of the direct product of two independent one dimensional filter banks in the horizontal and vertical directions. This 2-D wavelet is effective at approximating point singularities, but not the line singularities like edges in an image. This is because, the tensor product wavelets do not adapt to the boundaries, due to isotropic scaling of basis functions. A new representation which can provide flexible multiresolution and directional image expansion is necessary.

The problem, in summary would be:

- To find out an efficient image compression scheme that will provide good reconstructed image quality with almost the same computational complexity as that of wavelet based scheme.

- To introduce and address new compression schemes other than wavelets for the above said criteria.
1.6 OBJECTIVES OF THE PRESENT WORK

The objective of this dissertation is to investigate the rate-distortion performance of the proposed image-adaptive multistage vector quantization scheme for different multiresolution and multi-directional image representations.

- To represent texture rich images such as fingerprints, wavelet packet transform is used in this work. Wavelet packet transform can be thought of as adaptive wavelet transform and is capable of providing arbitrary frequency resolution. The advantage of wavelet packet framework is its universality in adapting the transform to a signal without assuming any statistical property of the signal. The extra adaptivity of the wavelet packet framework is obtained at a price of added computation in searching for the best wavelet packet basis. The problem of choosing the best basis is handled in two different ways in this thesis work. First approach is by Nonlinear approximation. In this approach, all the possible wavelet packet bases are arranged in the descending order of their energy and a percentage of the wavelet packet coefficients are quantized by image-adaptive multistage vector quantization technique. In the second approach, the best basis is selected using Singular Value Decomposition. In this approach, compression ratio is used as a tool to limit the number of subbands. The best basis is then selected through Singular Value Decomposition.

- To simultaneously satisfy all the signal processing properties such as compact support, orthogonality, symmetry, and higher order approximation, Multiwavelet is used in this work. Multiwavelets are capable of providing reasonable
reconstructed image quality at low bit rate. The Multiwavelet coefficients are quantized using image-adaptive multistage vector quantization technique.

- The Multiwavelet based compression scheme necessitates a prefilter as a front end and a post filter at the back end. The role of the prefilter is to match the characteristics of the applied signal with that of the multifilter and the post filter performs the reverse operation of matching the multifilter to that of the combined output. However, the prefilter and the post filter pair exists only for a certain multifilter structures. Moreover, multiwavelets are not effective in capturing the geometrical structure of the image.

- To represent the geometrical structure efficiently, Contourlet transform is used in this work. Contourlet transform provides flexible multiresolution, local, and directional image expansion using contour segments. Different pyramidal and directional filters are considered in this work. Contourlet coefficients are then quantized using image-adaptive multistage vector quantization technique and the experimental results are compared with those of the wavelet based scheme.

- Efficient video coding can be achieved by exploiting both the spatial redundancy and the temporal redundancy. Multiwavelet transform and Contourlet transforms are considered for minimizing the spatial redundancy. The temporal redundancy has been minimized using Kite Cross Diamond Search algorithm.
In summary:

- Different multiresolution representations taken into consideration in this work are Wavelet, Wavelet Packet, Multiwavelet, and Contourlet transform. The coefficients of these transforms are quantized using image-adaptive multistage vector quantization technique.

- The transforms considered for the coding of video sequence is Multiwavelet and Contourlet transform.

1.7 THESIS ORGANIZATION

This thesis is organized into eight main chapters, including this first chapter which motivates the demand for efficient digital image and video compression methods so as to meet the growing number of heterogeneous multiparty and disparate network environments.

Chapter 2 begins with the preliminary ideas of vector quantization. The concept of optimal vector quantization is discussed in this chapter. Then, the motivation of multistage vector quantization scheme is given. This chapter concludes with the proposed image-adaptive vector quantization scheme.

Chapter 3 deals with the application of image-adaptive multistage vector quantization scheme to wavelet packet basis. The chapter begins with an introduction to wavelet transform. It is then followed by two-dimensional discrete wavelet transform. The limitation of the wavelet transform and the need for wavelet packet transform is discussed in this chapter. The impact of nonlinear approximation of wavelet and wavelet packet coefficients are also given in this chapter.
In Chapter 4, the selection of the best basis using Singular Value Decomposition is discussed. This chapter begins with the need for the selection of the best basis in the case of wavelet packet transform. It is then followed by the proposed algorithm in which the idea of using Singular Value Decomposition as a tool to select the best basis is discussed in detail.

Chapter 5 discusses the theory and application of multiwavelets. The chapter begins with the need for multiwavelets. It is then followed by implementation issues of multiwavelet. Then, the nonlinear approximation of multiwavelet coefficients is discussed. The application of image-adaptive multistage vector quantization scheme to the approximate multiwavelet coefficients is given in this chapter. Extensive image compression results are shown to verify the usefulness and efficiency of the proposed scheme. The results are compared with those of the wavelet based coder.

Chapter 6 deals with Contourlet transform which can efficiently capture and represent singularities along smooth object boundaries in natural images. In this chapter, image-adaptive multistage vector quantization scheme is applied to contourlet coefficients and the results are compared with wavelets.

Chapter 7 deals with the application of adaptive multistage vector quantization to multiwavelet and contourlet transformed video sequence and their results are compared with those of the wavelet based ones. This chapter begins with the motivation for video compression, which is followed by a review of block based motion estimation algorithms which is used to exploit temporal redundancy present in video sequences. In this dissertation, Kite Cross Diamond Search Algorithm (KCDS) is used for fast and accurate motion estimation. Then, the integration of KCDS strategy with multiwavelet and contourlet based adaptive vector quantization technique is discussed.
Elaborate experimental simulations are presented so as to verify the efficiency of the proposed video coder with that of the wavelet based video coder.

Chapter 8 concludes the thesis with a summary of the results and the directions for future work.