University of Alberta

Protection of Arbitrarily Shaped Objects in Generalized Multiple Description Coding

by

Zhibin An 

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of

the requirements for the degree of Master of Science

Department of Computing Science

Edmonton, Alberta
Fall 2004

Your file Votre référence
ISBN: 0-612-95699-7
Our file Notre référence
ISBN: 0-612-95699-7

To my beloved wife Ting Shu and our families.

ACKNOWLEDGMENTS

I would like to thank Dr. Anup Basu for his supervision and help through this research. Without his timely guidance, inspiration and encouragement, this thesis could not have been done. I would also like to thank my examining committee, Dr. Mrinal Mandal and Dr. Janelle Harms for their careful reading and valuable comments on my thesis. Thanks also go to Dr. Liyan Yuan who chaired my defense.

I would like to thank Mr. Yufei Yuan for the OBDWT source code and Mrs. Agnieszka C. Miguel for providing the images "Little girl" and "Brain". Thanks also go to Ms. Anne J.R. Nield for her proofreading of my thesis.

My final thanks go to my wife Ting Shu and our families. They provided me endless support and they were always the source of my power and courage.

Table of Contents

List of Figures

GLOSSARY

bpp bits per pixel.

DCT Discrete Cosine Transform.

DWT Discrete Wavelet Transform.

EZW Embedded Zerotree Wavelet Algorithm.

GMDC Generalized Multiple Description Coding.

HVS Human Visual System.

JPEG Joint Photographic Experts Group.

MDC Multiple Description Coding.

MDCT Multiple Description Correlating Transform.

MDSQ Multiple Description Scalar Quantizer.

MPEG Moving Pictures Expert Group.

OBDWT Object-Based Discrete Wavelet Transform.

PDMD Polyphase Downsampling Multiple Description Coding.

PSNR Peak Signal to Noise Ratio

ROI Region of Interest.

SPIHT Set Partitioning In Hierarchical Trees.

Chapter 1

INTRODUCTION

1.1 Introduction to MDC

Multiple Description Coding (MDC) is a source coding technique that is robust against lossy transmission networks. MDC encodes a media source (video or image) into two or more sub-bitstreams (descriptions) that are of equal importance. These descriptions can be decoded independently to produce a basic quality of the original source. When more descriptions are received, the decoder can reconstruct the source using many descriptions and gradually increase the quality.

There are two types of MDC that have led to a flurry of proposed approaches. One is Multiple Description Scalar Quantizer (MDSQ) [1, 2, 3] that requires careful index assignments. The other is Multiple Description Transform Coding (MDTC)[4, 5] that necessitates an additional correlating transform to the conventional decorrelating transform. Since these approaches need to design specific quantizers or transforms, MDC is achieved at the expense of relatively complicated system design.

In [6], a new method is presented to separate description generation and redundancy addition that greatly reduces the implementation complexity specific to systems with more than two descriptions. In this method, description generation is accomplished using a polyphase transform, and each of the polyphase components is coded independently at a source coding rate. Redundancy is then explicitly added to each description by coding other descriptions at low redundancy coding rate, using selective quantization. Many methods [6, 7, 8, 9,10] are built on this framework.

1.2 Motivation

Although the algorithms in Section 1.1 achieve graceful degradation of image quality in the presence of increasing description loss, they improve the overall quality without taking into account the contents of an image. In practice, the foreground object often attracts more interest than the background does. For example, small important parts of medical images may be sufficient for doctors. The object-based coding technique that makes visual objects available in the compressed form has been an active research area in the past few years. It can not only provide great flexibility for operating arbitrarily shaped visual objects in multimedia applications but also potentially improve the quality of visual objects at low bit rates. The emerging image and video compression standards, such as JPEG-2000, MPEG-4 and MPEG-7 rely on a content-based representation of visual objects that code the region of interest (ROI) separately from the rest of an image.

In this thesis, my aim is to design and implement an algorithm to protect arbitrarily shaped regions of interest in Multiple Description Coding (MDC). The main idea is to treat an image as a composition of several layers, and set different layers with different priorities. Because many of the visual objects have arbitrary shapes and traditional Discrete Cosine Transform (DCT) and Discrete Wavelet Transform (DWT) algorithms can only be performed on a rectangular region, we need the Object-based DWT (OBDWT) to code arbitrarily shaped image objects. In MDC systems, if packet losses occur, it is possible to recover them by exploiting redundancy. There is a tradeoff between the coding efficiency and the image quality. How to add the redundancy in each descripition is the key point of the design of MDC systems. Here, we present two different methods capable of adapting to two kinds of network conditions.

The first method proposes a new scheme using the object-based discrete wavelet transform [11, 12] and Polyphase Downsampling Multiple Description Coding (PDMD) algorithm [13] in order to protect the region of interest. The main idea of [13] is that any given pixel can be reasonably predicted from the value of its neighbors. Therefore, we can take advantage of this natural correlation and split the image source into n descriptions by a polyphase downsampler along rows and columns. Each of the descriptions includes a region of interest and one component of the background. This method is suitable for error-prone networks.

In the second method, the MDC system is built on the framework proposed by [26]. The main idea of [26] is that wavelet coefficients corresponding to the ROI are scaled by a larger factor so that the information about ROI is sent in the earlier parts of the coding bitstream. Here, we use OBDWT shape mask coding to get wavelet coefficient trees belonging to arbitrarily shaped ROI. Using this method, the number of scaled wavelet coefficients can be reduced so that the coding efficiency of the Set Partitioning in Hierarchical Trees (SPIHT) algorithm will be improved, and the reconstructed quality of the ROI and background can also be improved. In the Experiments and Results section we show that this method is superior to the original one and it is suitable for relatively stable networks.

1.3 Thesis Organization

The thesis is organized as follows. Chapter 2 presents some background information on wavelet transform and zerotree coding to assist in understanding the content of this thesis.

Chapter 3 gives a description of related works. First, source coding techniques meeting the requirement of error-prone networks are introduced. Second, the definition of MDC is discussed. Finally, the different kinds of MDC frameworks are presented.

In Chapter 4, we propose the first method to protect arbitrarily shaped Region of Interests (ROI) in PDMD framework. Some experiments are designed and the results are compared with other papers in literature.

In Chapter 5, we present the second method, based on the MD-SPIHT framework. First, this method is compared with paper [26], demonstrating an improved result. Then I compare the results using arbitrarily shaped ROI with forevation techniques. Finally, I demonstrate the result of the multiple ROIs.

In Chapter 6, experimental results are analyzed and some conclusions are presented.

In Chapter 7, some interesting possible future works are discussed.

Chapter 2

BACKGROUND

This section presents some background information on wavelet transform and zerotree coding. This theory is important for understanding the material in other chapters in this thesis.

2.1 Wavelet Transform

A natural image can be modeled as piecewise polynomials, and a properly chosen polynomial function can lead to frequency domain analysis. Wavelets are functions that satisfy certain mathematical requirements and they are used in representing data or other functions. Wavelet algorithms process data at different *scales* or *resolutions*. The multiresolution representation of an image is the most important property of the wavelet transform.

2.1.1 Why use Wavelet Transform?

Compression techniques are used to reduce redundant information in image and video in order to facilitate data storage and network transmission. Coding of a still image under MPEG-4 and JPEG200 is based on the wavelet transform. However, in the original JPEG standard, the JPEG committee adopted discrete cosine transform (DCT) as the foundation of JPEG. Why should we use wavelet transform?

One reason for doing so is the shortcomings of DCT and other block-based transforms. Since the block-based transforms decompose the image into

non-overlapping blocks and process them separately, the blocking artifacts are visible at the block boundaries; these cause discontinuity in the reconstructed image.

The other factor is the development of methods to code the wavelet coefficients. The methods of coding the wavelet transform coefficients are less efficient than those of DCT, until the introduction of the Embedded Zerotree Wavelet (EZW) [15]. EZW algorithm takes advantage of the multiresolution features of the wavelet transform to achieve a high compression performance.

2.1.2 Wavelet Transform vs. Fourier Transform

The Fourier transform is based on frequency domain analysis of the image. In the Fourier transform, a signal is broken down into constituent sinusoids of different frequencies. These sins and cosines are the basis functions and elements of Fourier synthesis. Similarly, the wavelet transform can be viewed as transforming the signal from the time domain to the wavelet domain. This new domain contains more flexible basis functions called wavelets, mother wavelets or analyzing wavelets.

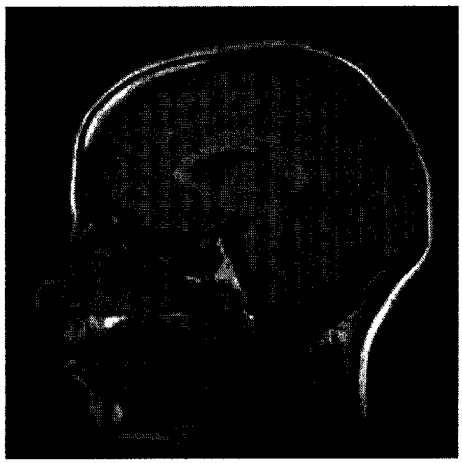
The most important difference between these two kinds of transform is that individual wavelet functions are localized in space. The basis functions of the Fourier transform are very exact in frequency, but are spatially not precise. This means that the signal energy is not concentrated at one frequency, but is spread over space. When pixels in images have low correlation, such as an edge or object boundaries, it will lead to some problems. After the Fourier transform, each coefficient corresponds to a fixed size spatial area and a fixed frequency bandwidth. Edge information tends to disperse and loses local representation with good fidelity. An advantage of wavelet transforms is that the windows can vary. Wavelet analysis allows us to use long time intervals when we want more precise low-frequency information, and use short time intervals when we want high-frequency information. This localization feature, along with the wavelets'

localization of frequency, makes many functions and operators using wavelets in data compression, detecting features in images and de-noising signals [16].

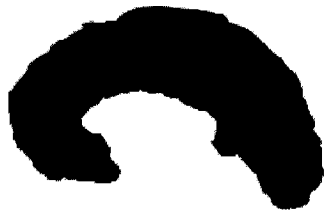
2.2 Object-based Discrete Wavelet Transform (OBDWT)

Object-based coding of images and video allows separate decoding and reconstruction of arbitrarily shaped video objects. The object-based coding technique that makes visual objects available in the compressed form has been an active research area in the past few years. It can provide great flexibility for operating visual objects in multimedia applications, and can potentially improve the quality of visual objects for low bit rate coding. The Object-Based Discrete Wavelet Transform (OBDWT) scheme can be directly applied to the arbitrarily shaped region and transforms the samples in this region into the same number of coefficients in the subband domain.

There are different kinds of methods to achieve object-based discrete wavelet transform. Generally, the arbitrarily shaped object in the image consists of a shape mask and a texture image. The shape mask is usually binary indicating whether a pixel belongs to an object. Coding of an arbitrarily shaped object involves coding the binary shape mask and the texture image. The shape mask can be coded with a chain-code, while coding of the texture image involves arbitrarily shaped object transform, quantization, and entropy coding. Figure 1 is an example of OBDWT.



a) The Brain Image



b) The binary shape mask.

c) the arbitrarily shaped object

Figure 1: An example of OBDWT.

2.3 Embedded Zerotree Wavelet (EZW) algorithm [15]

Shapiro's embedded image coding using zerotree of wavelet coefficients (EZW) [15], made a significant breakthrough in coding of wavelet coefficients. The EZW coder is designed for use with wavelet transforms; it exploits multiresolution features of the wavelet transform and provides a way to compress the image efficiently.

2.3.1 Zerotree

After wavelet transform, wavelet coefficients in the same orientation subbands have strong tree-like father-children dependency. A coefficient in a low subband can have four children in the next higher subband. Each father has its own four children, except for the root node, which has three. The zerotree structure is shown in Figure 2.

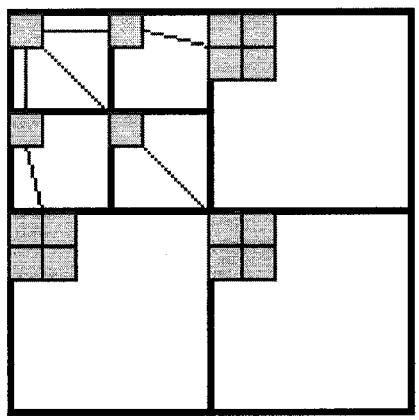


Figure 2: EZW's zerotree.

2.3.2 Coding Principle

The concept of zerotrees takes advantage of the self-similarity among wavelet coefficients magnitudes in different scales. The EZW encoder is based on two important observations:

1. After wavelet transforms, the energy in the subbands decreases as the scale decreases. The wavelet coefficients will be smaller in the higher subbands than in the lower ones.

2. All of the coefficients of the same orientation in the same spatial location are smaller than the coefficient of their parents.

From these observations, zerotree root is encoded with a special symbol indicating that the whole tree is insignificant. It can combine most of trees into zerotrees after the quantization step. As a result, zero coefficients are grouped together and the spatial position of nonzero coefficients are recorded by fewer side information.

2.3.3 Algorithm

EZW coder is a kind of embedded coder. It can terminate the encoding at any point and allow a target rate or target distortion to met exactly [15]. The bit stream generated by the embedded coder is sorted in order of importance. The coding method is based on first approximating an image with a few most important bits of data, and then improving the quality of approximation as more refined information is received.

The algorithm includes following steps:

1. Sorting pass (Significance Map Encoding). The significance map is a binary map indicating the positions of the significant coefficients. The coefficients are divides into two types: significance that is larger than the threshold and insignificance that is smaller that the threshold. The EZW coder takes advantage of the self-similarity zerotree structure to reduce the cost of encoding the significance map.
2. Refinement pass. This step is to refine the coefficients found significant in the sorting passes. It increases the precision of the coarse quantization by sending an additional bit.

2.4 Set Partitioning In Hierarchical Trees (SPIHT) [17]

2.4.1 An extension of the EZW Algorithm

The Set Partitioning in Hierarchical Trees (SPIHT) algorithm [17] is an extension of the EZW scheme proposed by A. Said and W. Pearlman in [15]. Both algorithms are zerotree based coding techniques that take advantage of self-similarity across different scales of an image wavelet transform. However, there are some differences between them. One of main differences is that the SPIHT algorithm uses a slightly different tree structure and combines parallel zerotrees in order to generate much more zerotree symbols. Another one is that the SPIHT algorithm uses a more efficient way to code the significance map. It uses a set partitioning sorting rule to partially sort the coefficients. Because of these alternative representations of the principles of EZW [15], the SPIHT algorithm can provide even better coding performance than the original EZW algorithm. The spatial orientation tree structure is shown in Figure 3. In this tree structure, the pixels in the highest level of the pyramid are the tree roots and are grouped in 2×2 adjacent pixels. One pixel (indicated by the star) in each group has no descendants. Compared with Figure 2, we can see that the tree structure in Figure 3 is a little different from the previous zerotree structure.

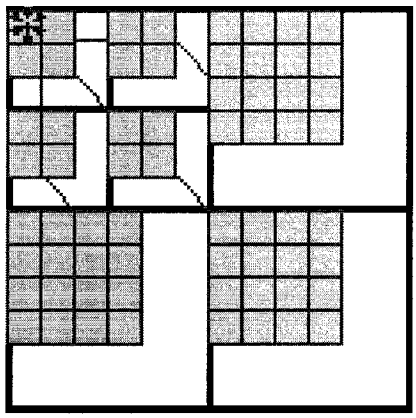


Figure 3: SPIHT's zerotree.

2.4.2 Data structure and Algorithm

SPIHT uses three lists:

LIP: List of Insignificant Pixels (individual insignificant coefficients);

LIS: List of Insignificant Sets (insignificant coefficient trees and sets);

LSP: List of Significant Pixels (significant coefficients).

Trees are of two types:

Type D - check all descendants for significance;

Type L - check all of the descendants with the exception of the immediate children.

The main parts of the algorithm [17]:

Set partitioning rules: [17] defines O(i, j) as a set of coordinates of direct descendants of node (i, j) and D(i, j) as a set of all descendants of the node (i, j) and L(i, j) = D(i, j) – O(i, j).

1. If D(i, j) is significant then it is partitioned into L(i, j) plus the four single element set with (k, l) $\in$ O(i, j).

2. If L(i, j) is significant then it is partitioned into the four sets D(k, l), with (k, l) $\in$ O(i, j).

Sorting pass:

1. Checks the significance of the LIP and LIS elements, and moves significant coefficients to the LSP.
2. When a significant coefficient is a member of a set, the set partitioning rule is used to split the set. Because the root node is much more likely to be significant then the rest of the tree, it is checked for significance independently.

Refinement pass: the encoder increases the precision of coefficients from the LSP by sending the next bit from the binary representation of their values.

Chapter 3

RELATED RESEARCH

3.1 Source Coding Techniques

With the increasing demand for network and wireless communication, congestion always occurs when packets move from a higher capacity link to a lower one. Because of congestion, networks are forced to discard some packets. In order to combat this lossy transmission, Forward Error Correction (FEC) and Automatic Repeat Request (ARQ) have been proposed as source coding techniques that are robust against inevitable transmission errors. In ARQ, the receiver can ask the transmitter to resend corrupted data in order to correct errors. However, if the environment does not provide a retransmission mechanism or if transmission is delay-constrained, it is difficult to achieve some delay sensitive real-time applications, such as interactive multimedia (Video on Demand) and video-conferencing. In FEC, the algorithm adds extra information along with the data and the receiver can use this extra information to check and correct the data. However, FEC needs a significant amount of redundancy to recover the original data from bit errors; it is much more suitable for the applications in which some parts of the multimedia data stream need considerably more protection.

Recently, Layered Coding (LC) and Multiple Description Coding (MDC) have received considerably more attention. LC and MDC encode a media source into two or more sub-bitstreams. For LC, layers split into a base layer and enhancement layers. The base-layer bitstream can be decoded to provide a basic quality of original source, while the enhancement layer depends on all the data in the lower layer and is used to refine the previous reconstruction. This type of multiple representations of original source is

very suitable for a heterogeneous network. The disadvantage of LC, however, is that when packet loss happens in lower layers, packets in higher layers become useless. For MDC, each multiple representation is of equal importance and can be decoded independently to reconstruct the basic quality of the original source. Each multiple representation is called a description and, when more descriptions are received, the decoder can gradually increase the reconstruction quality. Compared with MDC, LC more often uses hierarchical decomposition.

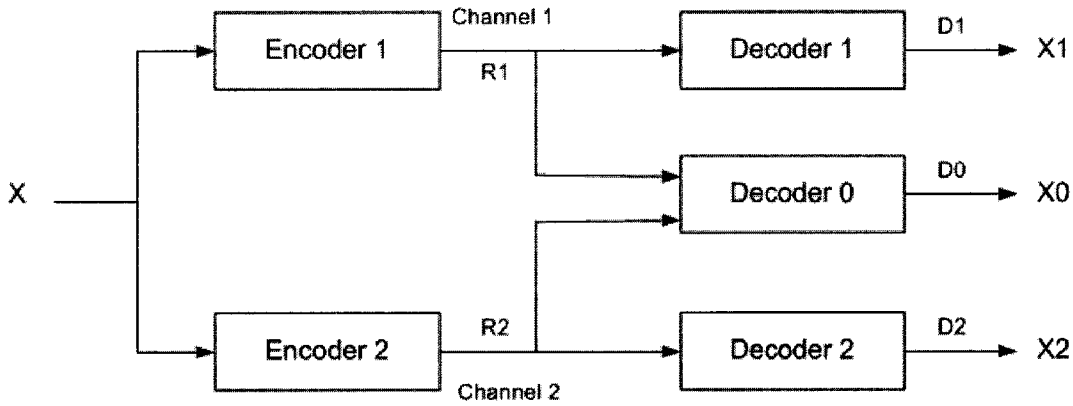
In [18], the authors compared LC and MDC over a wide range of loss rate. According to the simulation results, they drew following conclusions:

1. If no error protection is applied to both LC and MDC, MDC always has better performance than LC;
2. If ARQ-based error protection is applied to both LC and MDC, LC is preferred if the base-layer data is guaranteed to be received intact using a strong error protection method.
3. If FEC-based error protection is applied to both LC and MDC, both methods have almost equivalent performance up to 10% loss rates. MDC is preferred at higher loss rates.

3.2 Overview of Multiple Description Coding (MDC)

MDC is a source coding technique that is robust against inevitable transmission errors. MDC encodes a media source into two or more sub-bitstreams (descriptions) that are of equal importance. These descriptions can be decoded independently to produce a signal of basic quality. When more descriptions are received, the decoder can gradually increase the quality. The cost of this operation is the insertion of a certain amount of redundancy in the descriptions among the stream. The redundancy in different descriptions is used to estimate the loss description when packet loss happens. Figure 4 is a basic example of multiple description coding. R1 and R2 are the coding rate of the

two channels respectively. D0 is the average distortion of two-channel reconstruction, and D1 and D2 are the average distortion of one-channel reconstruction.



R1=R2 and D0 < D1=D2

Figure 4: The architecture of a basic MDC.

The basic principles can be summarized as below:

1. X is encoded into equally important streams

2. Decoding quality using any subset is acceptable

3. Better quality is obtained with more descriptions

There are several techniques for multiple description coding, but basically they can be grouped into two kinds of approaches. One is Multiple Description Scalar Quantizer (MDSQ) [1, 2, 3], which requires careful index assignments. The other is Multiple Description Transform Coding (MDTC) [4, 5], which necessitates a correlating transform in addition to the conventional decorrelating transform. In the following section, the basic ideas of the MDSQ and MDTC are introduced.

3.3 Multiple Description Scalar Quantizer (MDSQ)

The first paper on MDSQ was proposed by Vaishampayan [1]. Based on this method, a class of multiple description scalar quantizer is proposed [1, 2, 3]. Consider the case when a transmission uses two distinct channels. This paper proposed the design of two coarse side quantizers that can generate acceptable reconstruction when just one channel can be used. If both channels are used, two coarse quantizers are combined to generate a finer central quantizer that can lead to a better reconstruction. The figure below depicts the scheme.

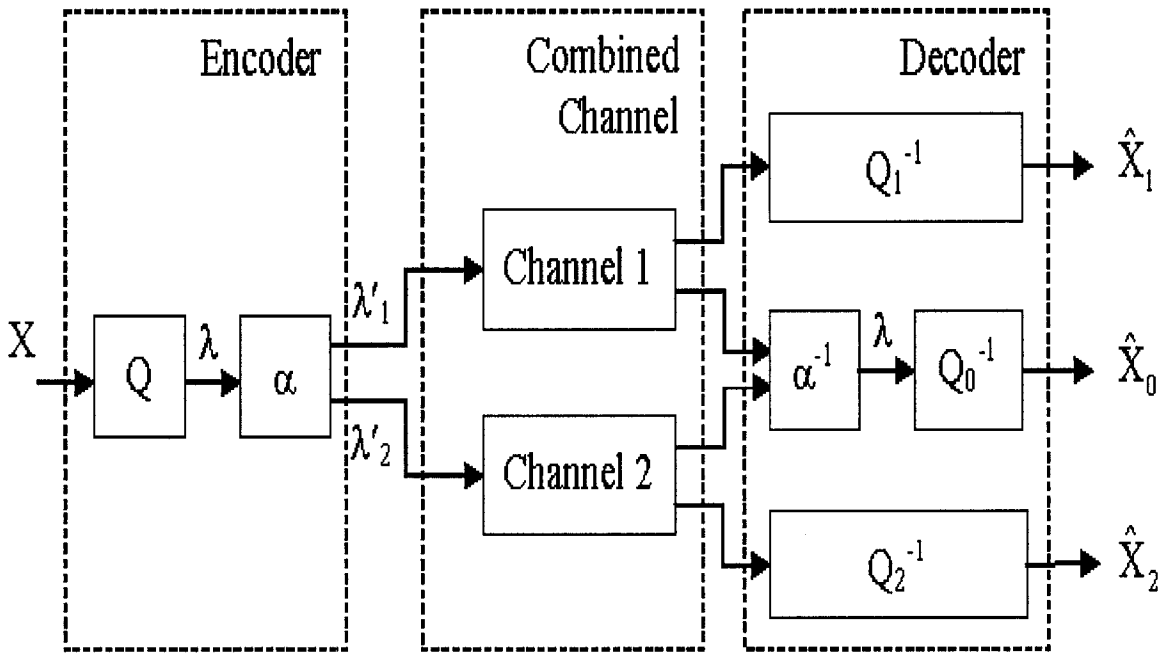


Figure 5: The architecture of MDSQ.

The architecture of MDSQ is shown in Figure 5. On the transmitter side, two steps are required to achieve the multiple descriptions: a quantization step Q and an index assignment step α. In the first step, the encoder quantizes the source sample using the central quantizer, and generates a central quantizer index, then a labeling (index assignment) function α produces two indices for each quantization output. Each index

should individually give a reasonable decoder output and combined indices should identify the output of the regular quantizer.

There are an infinite number of ways to design the index assignment. The objective of MDSQ is to find one kind of index assignment that minimizes the average distortion of the central decoder and side decoders under constraints. More details about the index assignment problem can be found in [1].

3.4 Multiple Description Correlating Transform (MDCT)

The Multiple Description Correlating Transform (MDCT) was introduced by Wang [4, 5]. The idea proposed is to use Pairwise Correlating Transform (PCT) to add statistical redundancy into the descriptions. The basic scheme proposed by [4] is shown in Figure 6. There are two transform procedures to achieve multiple descriptions. The first is a conventional decorrelating transform which decorrelates the source into multiple descriptions. The second is a Pairwise Correlating Transform (PCT), which reintroduces correlation in a controlled manner. Figure 6 below depicts the scheme.

$$\begin{pmatrix} C \\ D \end{pmatrix} = T \begin{pmatrix} A \\ B \end{pmatrix} \text{, T=} \begin{pmatrix} ac \\ bd \end{pmatrix} \qquad (1)$$

Pairwise MDC transform T in Equation 1 takes two independent input variables, A and B, and outputs two transformed variable, C and D. The transform T controls the correlation between C and D, which controls the redundancy of the MDC coder. If the decoder receives both bitstreams, it uses the inverse PCT to reconstruct the original source. If only one description is received, the decoder estimates the missing description and then performs inverse PCT reconstruction.

The objective of this method is to achieve a good reconstruction of lost data by using less additional redundancy. In order to achieve this goal, two main problems have to

be solved: 1. How should the input be split and which elements must be paired? 2. What does matrix T look like, and which one is optimal?

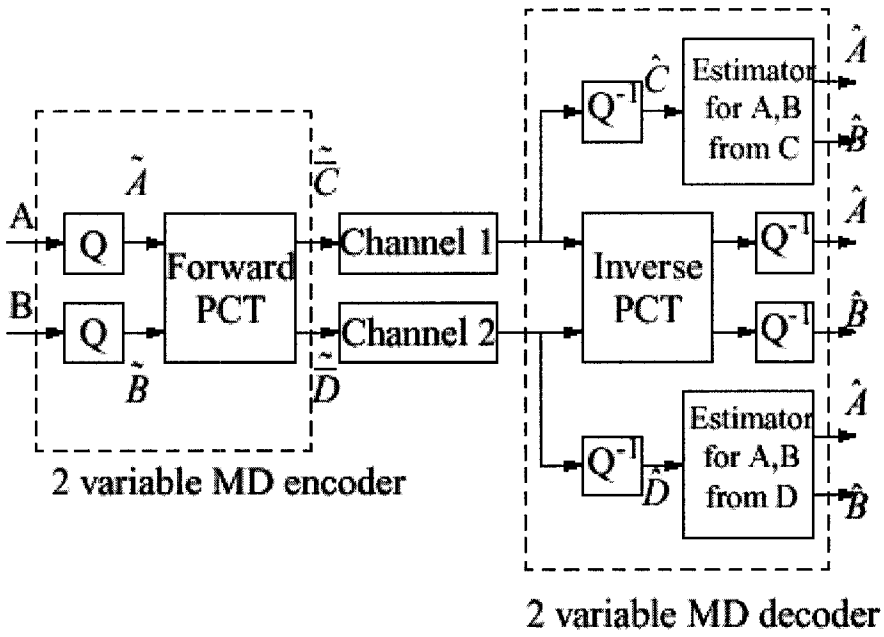


Figure 6: Coding and decoding process for a single pair: the basic scheme [4]

3.5 Generalized Multiple Description Coding (GMDC)

MDSQ and MDCT are specific to the two-channel MD problem and there are only three cases in which all, half or no descriptions are received by the decoder. However, when network communication has more than two channels or more than two packets, such as a multicast scenario, we cannot take advantage of the techniques of MDSQ and MDCT. This gives a motivation for using the Generalized Multiple Description Coding (GMDC) method [6, 7, 8, 9, 10]. In GMDC, N descriptions of a source are transmitted to a receiver, and less than N descriptions are received because of packet

losses. The goal of GMDC is to maximize the quality of the reconstruction with received descriptions.

3.5.1 GMDC via Polyphase Transform and Selective Quantization

In [6], a new method of GMDC is presented to separate description generation and redundancy addition, which greatly reduces the implementation complexity specific to systems with more than two descriptions. In this method, description generation is accomplished using a polyphase transform, and each of the polyphase components is coded independently at a source coding rate. Redundancy is then explicitly added to each description by coding other descriptions at a low redundancy coding rate, using selective quantization.

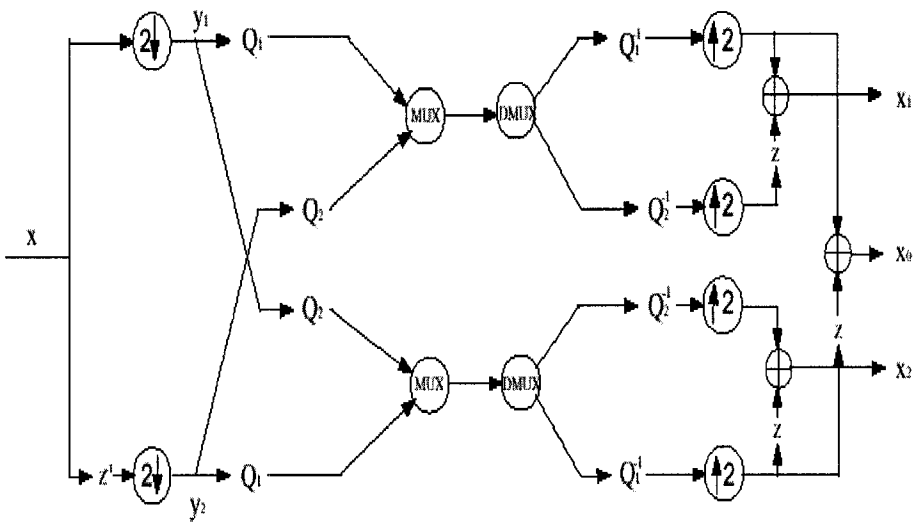


Figure 7: The MDC system proposed in [6].

Figure 7 shows the proposed MDC system. In the first step, the input X is decomposed into two components y1 and y2, via a polyphase transform. Each polyphase component is quantized independently by Q1, and forms the primary part

of a description. In order to estimate the data from the other channel in case of loss, another coarse quantizer, Q2, is then introduced to add redundancy into a description. Each of these two polyphase components is thereby quantized independently by a fine quantizer Q1 and a coarse quantizer Q2. The primary bitstream and the redundant bitstream are multiplexed together to form a description for transmission.

3.5.2 GMDC via SPIHT

The generalized MDC framework has received considerable attention and many new algorithms have been proposed. The method MD-SPIHT was proposed in [8]. This method is based on the principle provided by [6]. The main idea is to extend SPIHT to the generalized MDC framework. Since the SPIHT algorithm is a progressive source coder, it sends the globally most important information to the earlier part of the coded bitstream. This means that data coded earlier are much more important to image quality. Each description contains two parts. One part is the primary information of one input tree and the other is the redundancy consisting of repeated zerotrees.

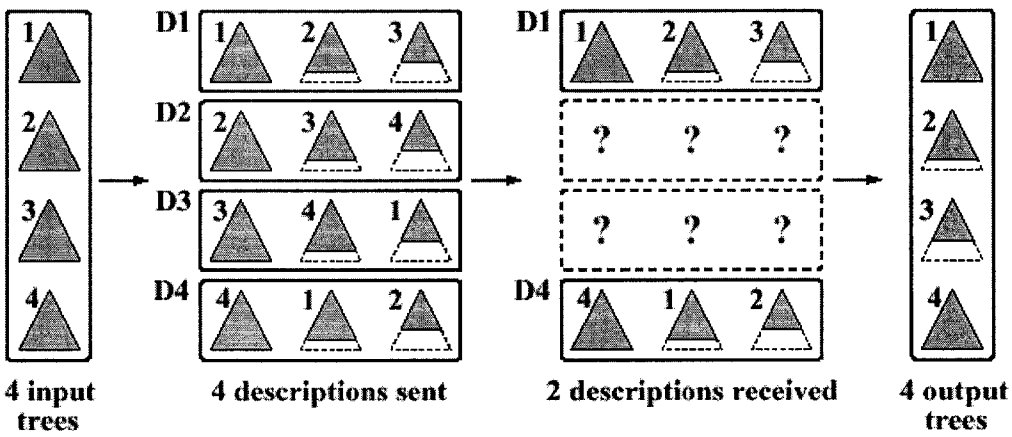


Figure 8: An example of MD-SPIHT [8].

Figure 8 shows an example of four wavelet coefficient trees (1-4). The primary information of one input tree and the redundant copies of other input trees are sent in four descriptions (D1-D4) with two sets of copies of the redundant trees. If two descriptions are lost (D2 and D3), the second and third trees are partially recovered from their copies in the received descriptions.

3.5.3 GMDC via Unequal Loss Protection

In [7], the proposed method uses explicit channel coding in the form of Unequal Loss Protection to achieve multiple descriptions. Reed-Solomon (RS) codes are used to generate FEC redundancy. These block codes are very effective in recovering erased symbols when the locations of the erased symbols are known [19]. Each block has an independent (N, k) RS code, where N is the block length and k is the number of source symbols. Unequal Loss Protection (ULP) [20] is the core of this method. In SPIHT code, data coded earlier are much more important to image quality. ULP is a system that assigns unequal amounts of forward error correction to progressive data, in order to provide a graceful degradation when packet losses increase.

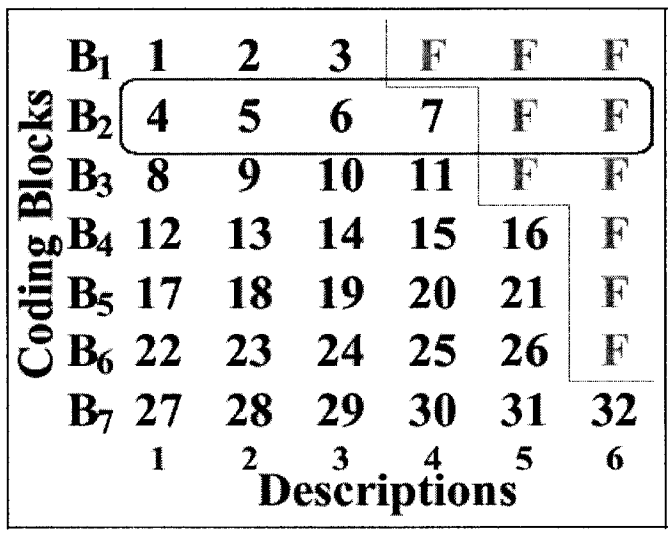


Figure 9: An example provided by [7].

Figure 9 shows an example provided by [7]. Each of N columns is one description, and each of the rows is an independent Reed-Solomon code block. This method is composed of two steps. First, it uses the progressive SPIHT algorithm to compress the source image, and 32 bytes bitstream is formed. Secondly, an ULP assignment algorithm of [20] is used to determine how many source symbols are in one block. We can see from block 1 in the example that more important information provides far more FEC redundancy and gets much better protection.

3.5.4 Polyphase Downsampling Multiple Description Coding

In [13], a very interesting method is proposed. The MD system is based on a Polyphase Downsampling algorithm (PDMD) to split an image source into balanced descriptions. It is a special kind of GMDC, since it does not need explicit redundancy. Coding, psychovisual, and interpixel are three basic data redundancies in digital natural images. The method proposed by [13] takes advantage of interpixel redundancy and lets the compression algorithm take care of the other two. Since the value of any given pixel can be reasonably predicted from the value of its neighbors, the information carried by an individual pixel is relatively small. It is possible to exploit this feature to create a multiple description of the image. At the encoder side, the source image is split into n descriptions by a polyphase downsampler along rows and columns. Each single description is then compressed by the progressive SPIHT algorithm with arithmetic coding and sent to the receivers. At the decoder side, we can use a simple interpolating filter or some other more advanced reconstruction filter to reconstruct the images.

3.6 Protection of Regions of Interest in GMDC

3.6.1 Overview of ROI in data compression

Regions of interest (ROI) are spatial regions in images that are the most important to the end user. The ROI research on data compression may be divided into three types.

Some ROI research concentrates on designing methods for coding ROI separately from the rest of the image. The emerging image and video standards, such as JPEG-2000, MEEG-4 and MPEG-7 require object-oriented coding and motion estimation.

Some ROI research concentrates on the foveated technique. Since the human visual system (HVS) is highly space-variant, the spatial resolution of the HVS is the highest around the point of fixation and decreases rapidly with increasing eccentricity. The concept of prioritizing data based on the distance from a ROI was first introduced in [21]. This particular prioritization method employs a foveated image and allows a gradual increase in peripheral quality loss during network congestion. In [22] the authors use the foveation technique in order to rearrange the packets in the code stream to place the regions of interest before the background coefficients. This is fully compatible with the JPEG2000 standard, and allows transmission of different regions of interest with different priorities.

Others code the ROI at a higher bit rate [9, 14, 23] than the rest of the image. With the recent popularity of zerotree based coding, some of these methods have been extended to Embedded Zerotree Wavelet (EZW) [15] and the Set Partitioning in Hierarchical Trees (SPIHT) [17] compression algorithms. These methods assign the ROI a higher bit rate than the rest of the image, by multiplying the wavelet coefficients

corresponding to ROI by some large factor. This way, the locally important information is sent in the earlier parts of the compressed bit stream.

3.6.2 Protection of Regions of Interest in GMDC

Recently, Generalized Multiple Description Coding (GMDC) has received considerable attention, and many new algorithms have been proposed. However, only a few current ROI coding methods have been applied in the GMDC framework.

The method suggested in [26] is the first to protect ROI in multiple description coding. This method is built on the MD-SPIHT framework. The basic idea is that when SPIHT is used with ROI coding, we assign the ROI a higher bit rate than the rest of the image by multiplying the wavelet coefficients corresponding to the ROI by some large factor. Therefore, not only is the globally important information sent in the earlier parts of the compressed bitstream, but also is the locally important information (ROI). This method requires both the encoder and the decoder to know the ROI parameters and the multiplication factor, but does not involve any modifications to the original algorithm.

Another method proposed in this paper is built on the MD-ULP framework. Again, the wavelet coefficients corresponding to the ROI are scaled by a large factor, and the ROI information is sent to the earlier part of the stream. Since much more FEC redundancy is added to the earlier part of the compressed bitstream than to the later part, the ROI information can receive better protection.

Chapter 4

PROTECTION OF ROI IN PDMD FRAMEWORK

4.1 System Overview

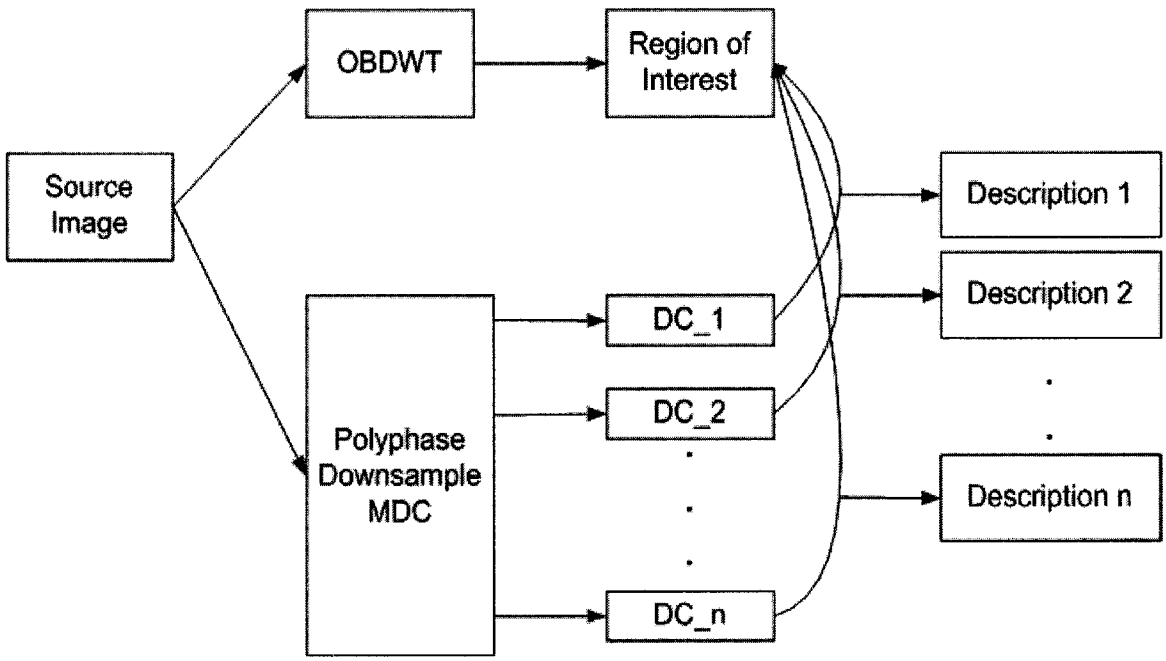


Figure 10: N descriptions are generated by obtaining n polyphase downsamples from an original image. Each description includes one component of region of interest and a compressed polyphase downsample component of the background.

Figure 10 is the block diagram of our proposed method. In this method, we treat the source image as two layers: one is the region of interest and the other is the background. The OBDWT algorithm [12] is used to code the region of interest and the Polyphase downsampling algorithm [13] is used to split the background into several

components. Each of my descriptions includes the region of interest and one component of the background.

The main contribution of this method is to protect any arbitrarily shaped objects in the image. In the experiments and results section, we can see that the proposed algorithm not only effectively protects arbitrarily shaped ROI in the image, but also gets satisfactory results over error-prone networks.

4.2 Proposed Algorithm

4.2.1 ROI Coding

Object-based coding techniques that make visual objects available in the compressed form have been an active research area in the past few years. Since it can provide great flexibility for operating visual objects in multimedia applications, and potentially improve the quality of visual objects at low bit coding, this technique is used in the multiple description coding method presented here.

We will treat an image as a composition of several layers. There are many visual objects with an arbitrary shape in the image and there are two steps in coding an arbitrarily shaped object. The first is to code the shape of the visual object and the second is to code the texture of the visual object (pixels inside the object region).

Here, the OBDWT proposed by [12] is used to transform arbitrarily shaped visual objects. OBDWT utilizes the signal extension method proposed in [24, 25], and makes conventional DWT easy to apply to arbitrarily shaped regions. Any arbitrarily shaped region can be decomposed using OBDWT, and perfectly reconstructed using inverse OBDWT, without processing more pixels than what is contained in the original visual object. This means that the number of coefficients after OBDWT is identical to the number of pixels in the original arbitrarily shaped visual object.

Outline of proposed approach:

Step1. Creating the mask of the arbitrarily shaped object to change the source image into two layers: Region of Interest and background.

In my implementation, the tools in Photoshop are used to generate a shape mask according to the ROI. The shape mask is usually binary indicating whether a pixel belongs to an object. Figure 11 is an example to create the masks and the black area is the region of interest.

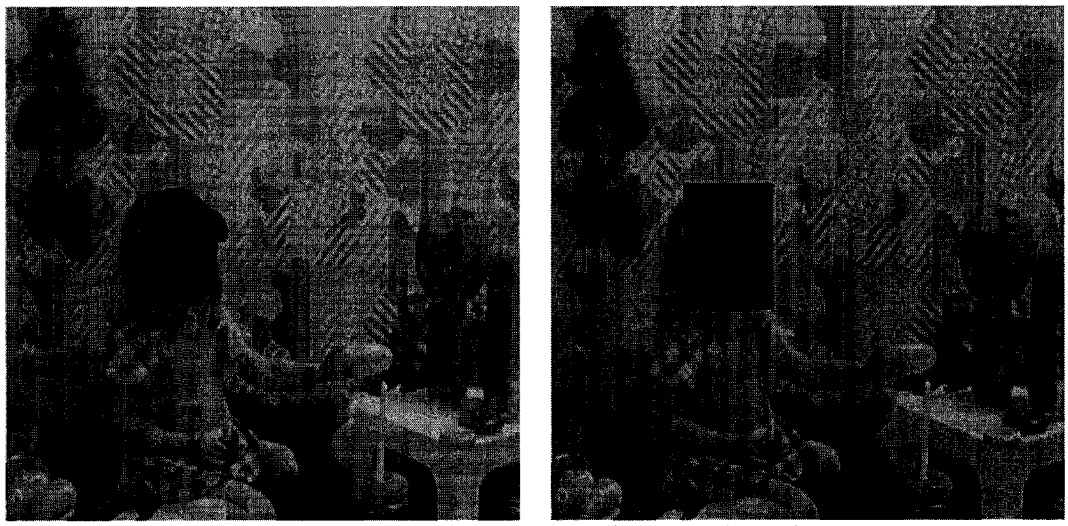


Figure 11: There are two kinds of masks of the little girl image: one is an arbitrarily shaped mask and the other is a rectangular mask.

Step 2. Using the OBDWT to code the region of interest.

Shape mask is a key factor in the OBDWT, and the decomposition and reconstruction procedures are supervised by the mask information. There are two kinds of information that need to be compressed: one is the binary shape mask that can be coded with a chain-code; the other is the ROI object. In the first, OBDWT decomposes the shape mask to determine whether a coefficient needs to be coded. The decomposition of the shape mask after OBDWT is shown in Figure 12. The coefficients belonging to the decomposition of the shape mask are coded by the

P-SPIHT algorithm [12], which is built on SPIHT. The only difference is that P-SPIHT codes the coefficients belonging to the decomposed shape mask. The result of the OBDWT coding is shown in Figure 12.

Figure 12: Decomposition of the shape mask and the result of OBDWT coding.

4.4.2 Background Coding

1. Generating components through a polyphase downsampling algorithm (PDMD framework) to split the background into *n* components.

In the polyphase downsampling algorithm, the source image is split into *n* components along the rows and columns. Figure 13 is an example illustrating the algorithm. It is assumed that the source image has a size of 4×4 and is split into four components. We can consider the source image to be composed of four blocks. The pixels in each sub-image are grouped according to the same spatial location in each block.

2. Coding each component of background.

Each component holds the main feature of the source image because of the interpixel redundancy in images. Figure 14 is an example of one component. We can see that the

component contains the information of the ROI and background. In order to improve the coding efficiency the binary shape mask is inverted and used to generate a new shape mask for each component. OBDWT and P-SPIHT can be used again to code the pixels in the background in every component. The coding result of one component of the background is represented in Figure 14.

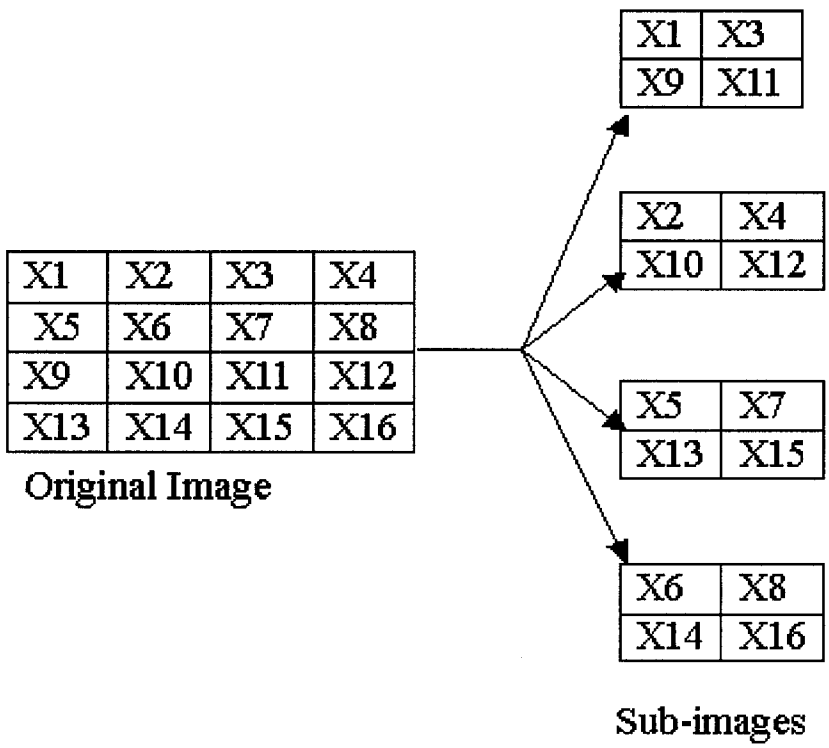


Figure 13: The method of polyphase downsampling.

Figure 14: The downsampling result and the result of one component of background.

4.2.3 Decoder

At the encoder part, each description consists of the ROI and one component of the background, and has the same importance for the reconstruction of the original image. Since any pixel in an image can be estimated from its neighbors the original image can be reconstructed from just one description, if more descriptions reach the decoder, a better reconstruction is possible.

The reconstruction procedure allows us to oversample the description and interpolate the missing pixels. A simple linear interpolation filter is used to reconstruct the images. The results can be seen in the experimental results section.

4.3 Experimental results & comparison with other work

The experiments are mainly conducted with "Lena" and "the little girl" grey-level images of size 512×512. The MDC system block diagram for this simulation is shown in Figure 10. Each description contains the Region of Interest (ROI) and one

polyphase component, corresponding to different components. The experiments can be divided into three parts.

4.3.1 Comparison of our method with the method proposed by [8].

The method provided by [26] is the best method to protect the ROI in literature. The algorithm in [26] follows the framework provided by [8]. In [8], the MDC is simulated using the standard 512×512 gray-scale Lena image without protection of the Region of Interest. We use the same image, and compare the whole image quality of our method with that of [8]. The total bit rate is fixed at 1.25 bpp.

(a) image 26.6059dB, ROI: 44.6072 (b) image 27.0393dB, ROI: 44.6072

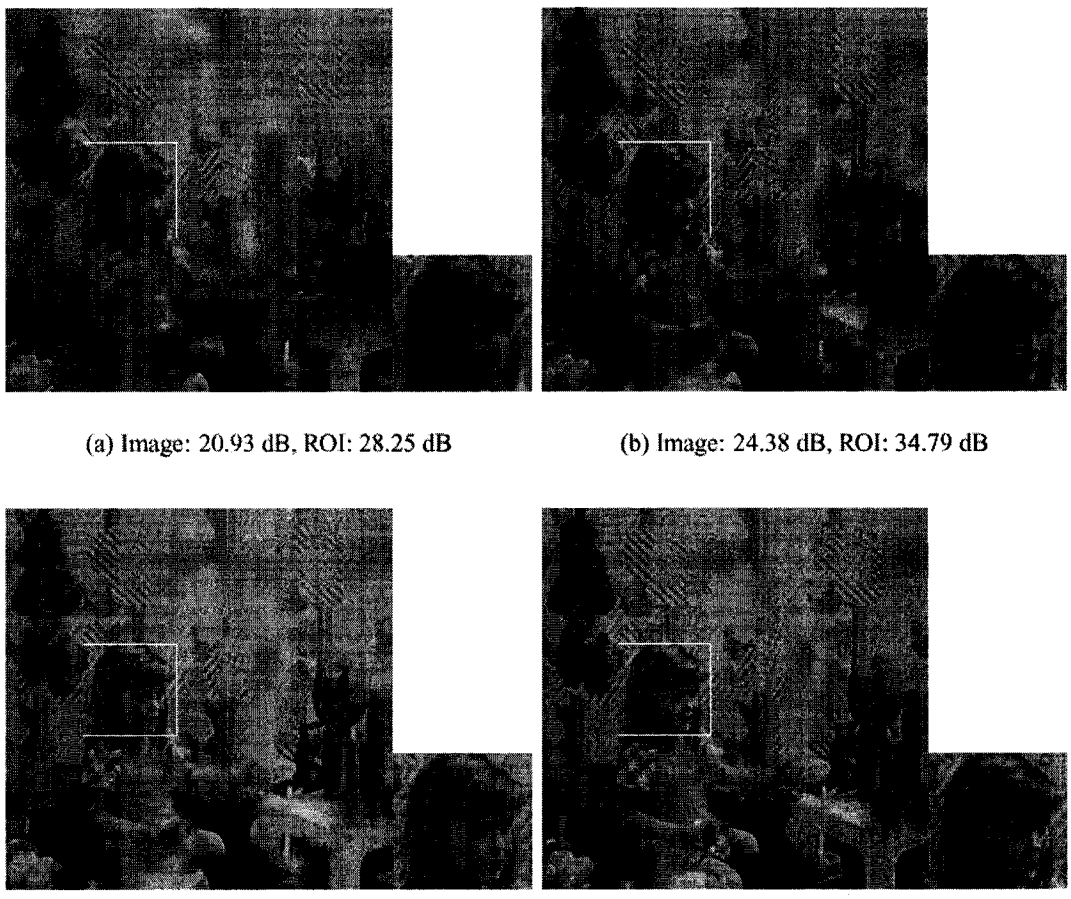
(c) image 27.7306dB, ROI: 44.6072 (d) image 29.6182, ROI: 44.6072

Figure 15: The results of our method.

From Figure 15, images a, b, c and d show 1, 2, 3, 4 descriptions received out of 8. In [8], we can read the PSNR from Figure 5 in [8]. When 1, 2, and 3 descriptions are received the PSNR equals 19.15dB, 24.43dB and 26.5dB respectively. In our method, the whole image quality is better when 1, 2 and 3 descriptions out of 8 are received, and part of the ROI achieves much higher quality than that of the whole image.

4.3.2 Comparison of our method with the method proposed in [26].

Our results are demonstrated on a 512×512 gray scale image of a little girl, where the girl's face is an 80×80 region of interest. The same image is used with the same region of interest. The number of descriptions is 8 and the total bit rate is 1.0 bpp.



(a) Image: 20.93 dB, ROI: 28.25 dB

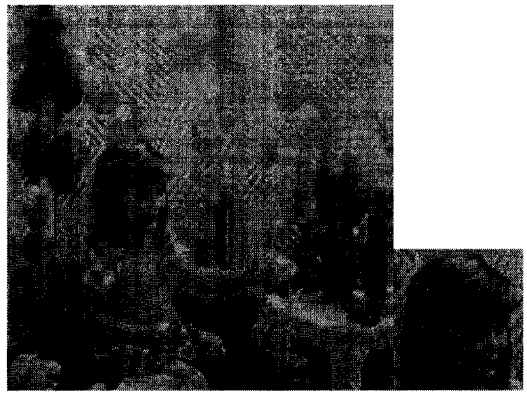
(b) Image: 24.38 dB, ROI: 34.79 dB

(c) Image: 26.36 dB, ROI: 38.54 dB

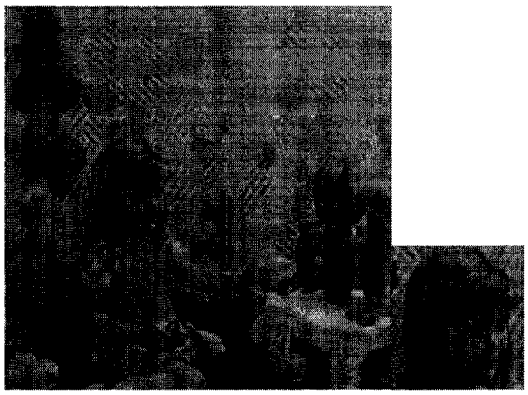
(d) Image: 27.57 dB, ROI: 40.47 dB

Figure 16: Results from [26].

(a) Image: 26.4490, ROI: 43.3705



(b) Image: 26.8211, ROI: 43.3705



(c) Image: 27.1208, ROI: 43.3705

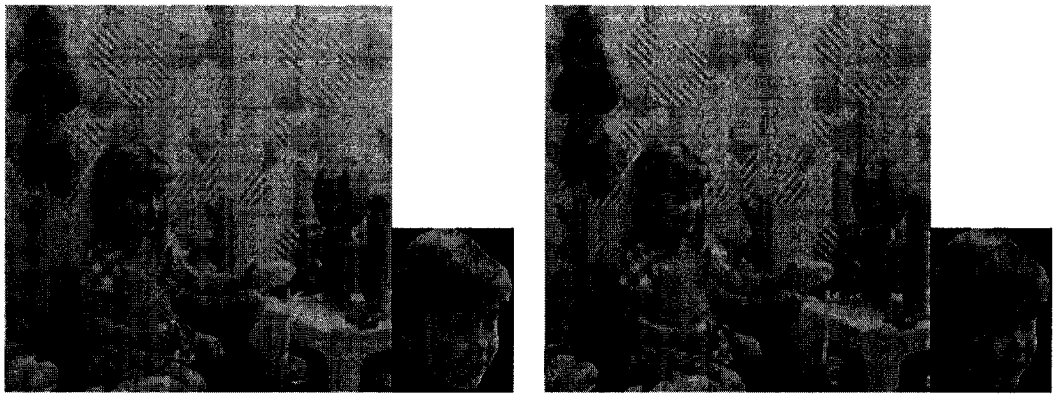
(d) Image: 27.7693, ROI: 43.3705

Figure 17: The results of our method, with rectangular region of interest.

I then change the rectangular region of interest into the arbitrarily shaped region of interest:

(a) Image: 27.5607, ROI: 46.9951

(b) Image: 28.3007, ROI: 46.9951



(c) Image: 30.2452, ROI: 46.9951 (d) Image: 31.1221, ROI: 46.995

Figure 18: The results of our method, with the arbitrarily shaped region of interest.

Figures 17 and 18 show the quality of the little girl images as 1, 2, 3, and 4 out of 8 descriptions are received. We see that the results of our method with the rectangular region of interest and the arbitrarily shaped region of interest are better than that of [26]. Furthermore, the results of the arbitrarily shaped ROI are better than that of the rectangular ROI. These results can be explained with the aid of mathematical analysis.

Let us assume that the coding rates of the ROI and background are R1 and R2, the size of ROI is X pixels, and the total bit rate of the image is 1.0 bpp. In this case, the total number of pixels in the image is 512×512 and 8 descriptions will be generated. Since each description includes 32768 bits, we get:

$$R2 = \frac{32768 - R1 \bullet X}{32768 - \frac{1}{8} \bullet X}$$

In this experiment, R1 is equal to 0.4, $(R1 \bullet X) < 32768$ and X<262144.

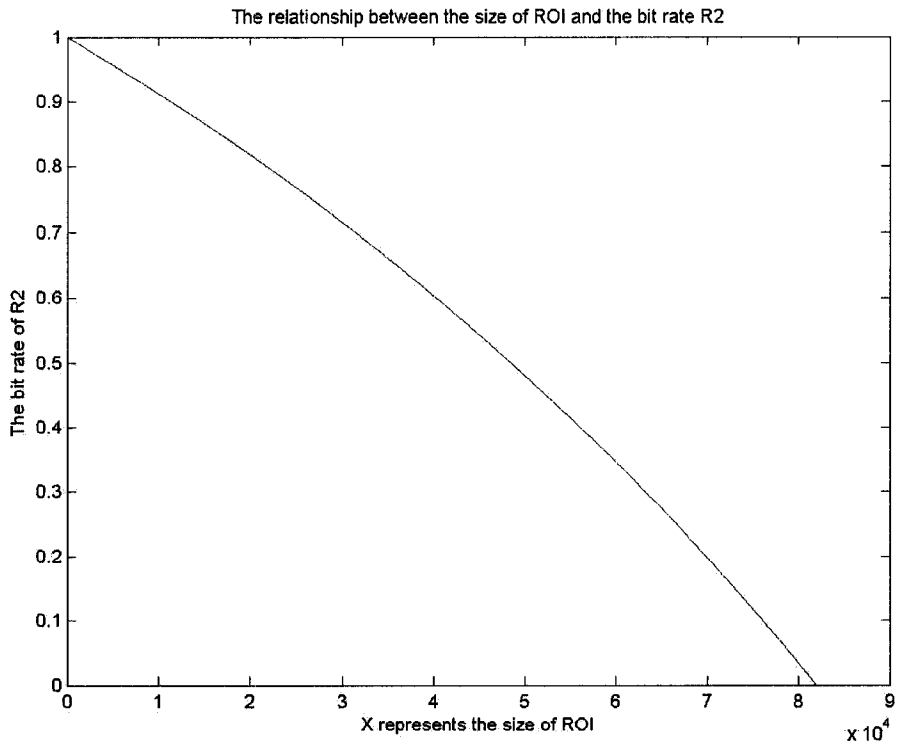


Figure 19: The horizontal axis represents X, the size of the ROI, and the vertical axis represents R2, the coding rate of the background.

Figure 19 shows us that when the size of ROI increases the coding rate of background R2 will decrease. This special case can be extended into a general model.

Let us assume the size of the image is I, the total bit rate of the image is B, and n descriptions are generated. The bit rate of R2 is represented in Equation (2).

$$R2 = \frac{I \bullet B - n \bullet R1 \bullet X}{I - X}. \qquad (2)$$

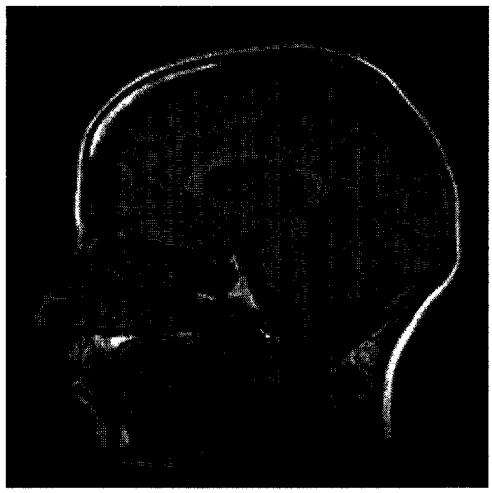
Then we differentiate this function with respect to X.

$$\frac{\partial R2}{\partial X} = -\frac{(I - X) \bullet n \bullet R1 + (I \bullet B - n \bullet R1 \bullet X)}{(I - X)^2} < 0 \qquad (3)$$

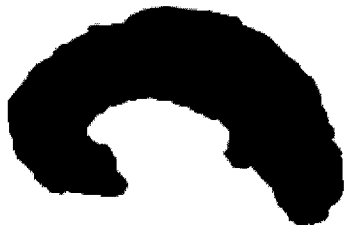
Since the function is decreasing, the differential of the function (3) is always negative. Therefore, when the size of ROI increases, the coding rate of the background, R2 will

decrease and the quality of the background will be reduced. This explains why some results of the arbitrarily shaped ROI are better than those of the rectangle ROI.

We show another image example to demonstrate this result.



(a) The original image brain



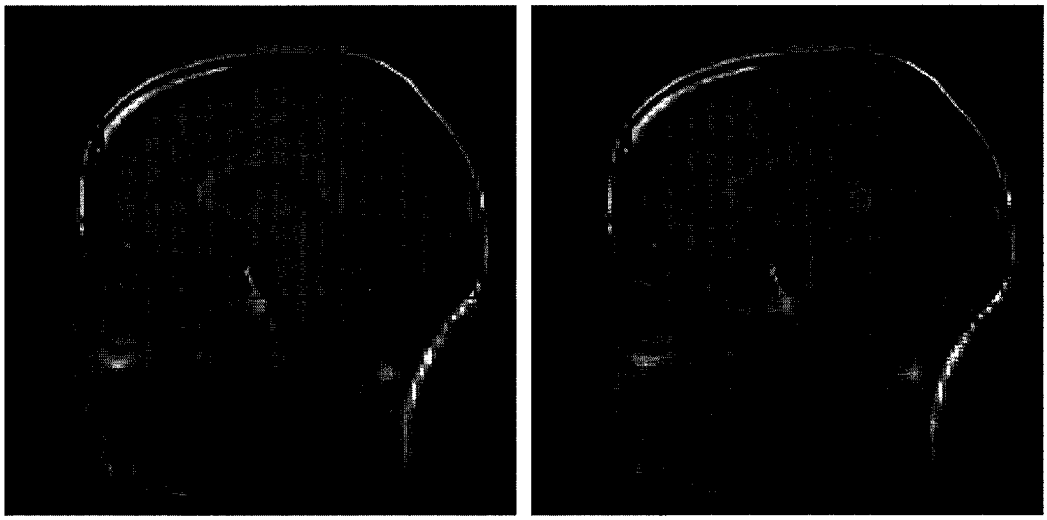
(b) rectangular ROI (c) arbitrarily shaped ROI

Figure 20: Two types of shape masks of the brain image.

Our results are demonstrated on a 256×256 Magnetic Resonance Image (MRI) of a brain, shown in Figure 20(a) using my method. Figure 20(b) is a rectangular ROI mask and comprises 22755 pixels — about 34% of the total image. Figure 20(c) is an arbitrarily shaped ROI mask and it comprises 11138 pixels — about 16% of the total image. The number of descriptions is set to four, and the total bit rate is 1.0 bpp. The

results in Figures 21 and 22 show the quality of the brain image as 1, 2, 3, and 4, of 4 descriptions that are received.

The results of the rectangular ROI

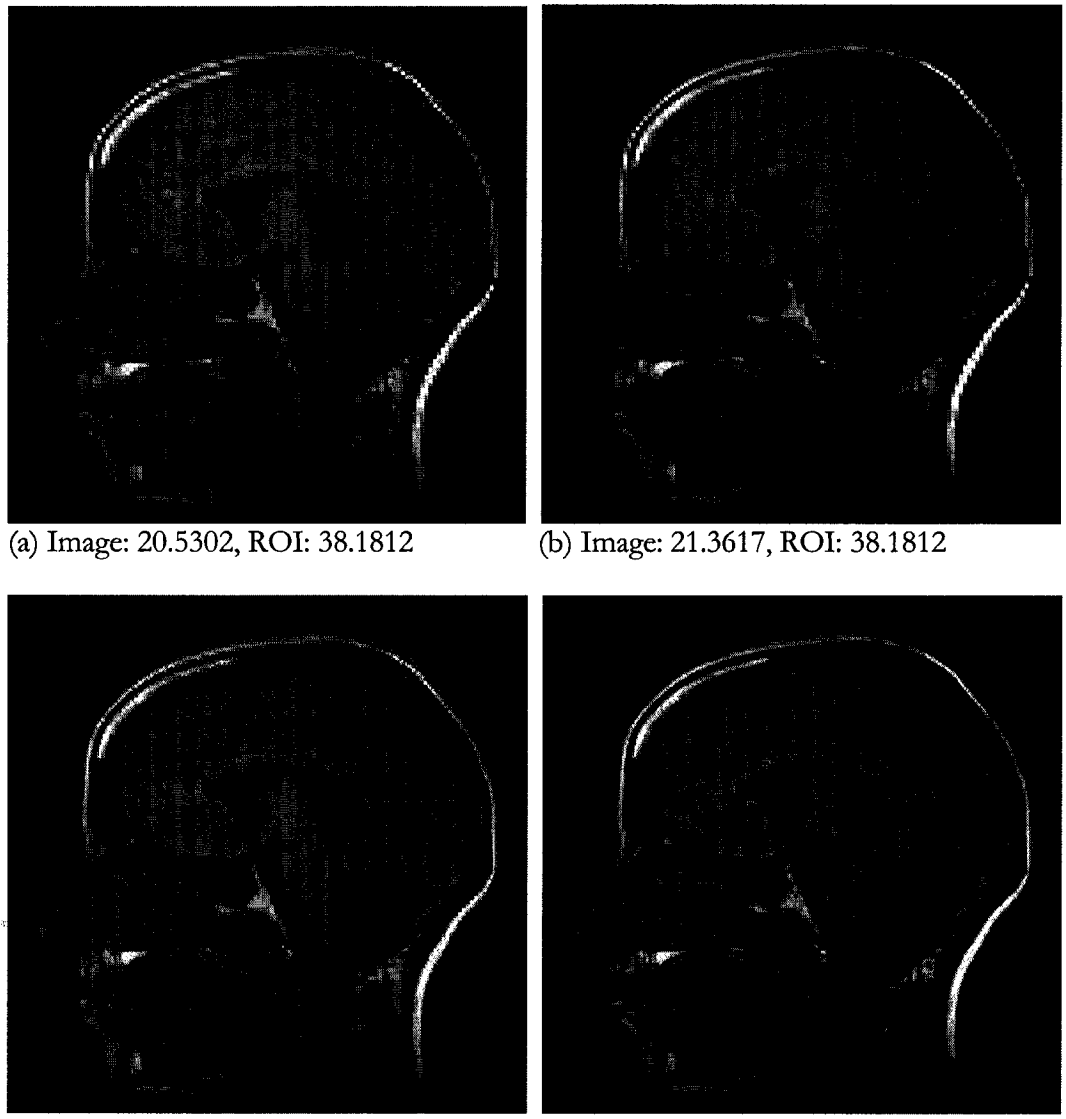


(a) Image: 20.3370, ROI: 35.3683 (b) Image: 20.6035, ROI: 35.3683

(c) Image: 20.7908, ROI: 35.3683 (d) Image: 21.0274, ROI: 35.3683

Figure 21: The results of the rectangular ROI.

The results of the arbitrarily shaped ROI



(a) Image: 20.5302, ROI: 38.1812 (b) Image: 21.3617, ROI: 38.1812

(c) Image: 21.8069, ROI: 38.1812 (d) Image: 22.2684, ROI: 38.1812

Figure 22: The results of an arbitrarily shaped ROI.

Comparing the results of the rectangular ROI in Figure 21 and the arbitrarily shaped ROI in Figure 22, we see that using the arbitrarily shaped coding technique can save half of the bits in the region of interest coding, and the saved bits can increase the quality of background, thus increasing the quality of the whole image.

4.3.3 Decomposition of additional descriptions

Result of decomposition of 16 descriptions

(a) Image: 25.6759, ROI: 46.9951

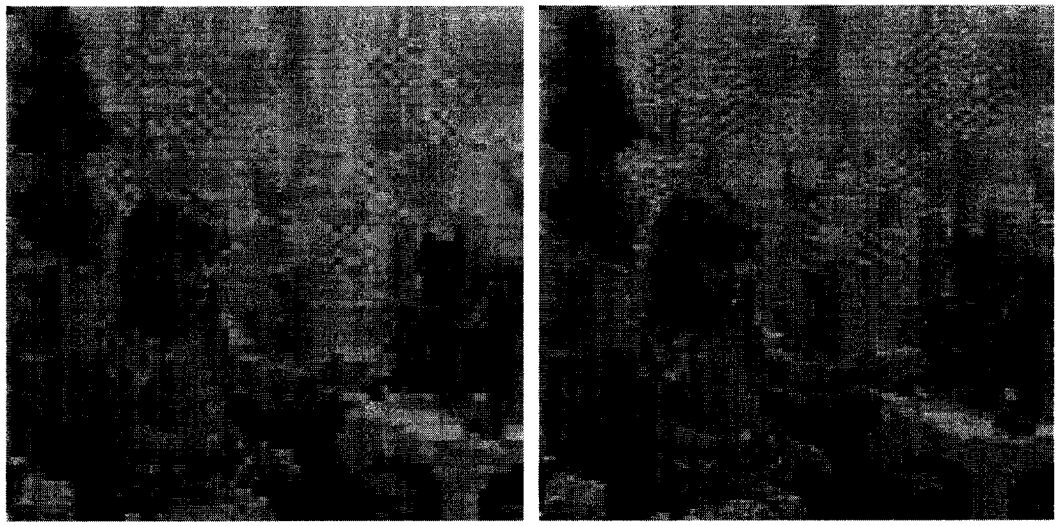
(b) Image: 26.9556, ROI: 46.9951

(c) Image: 27.1271, ROI: 46.9951

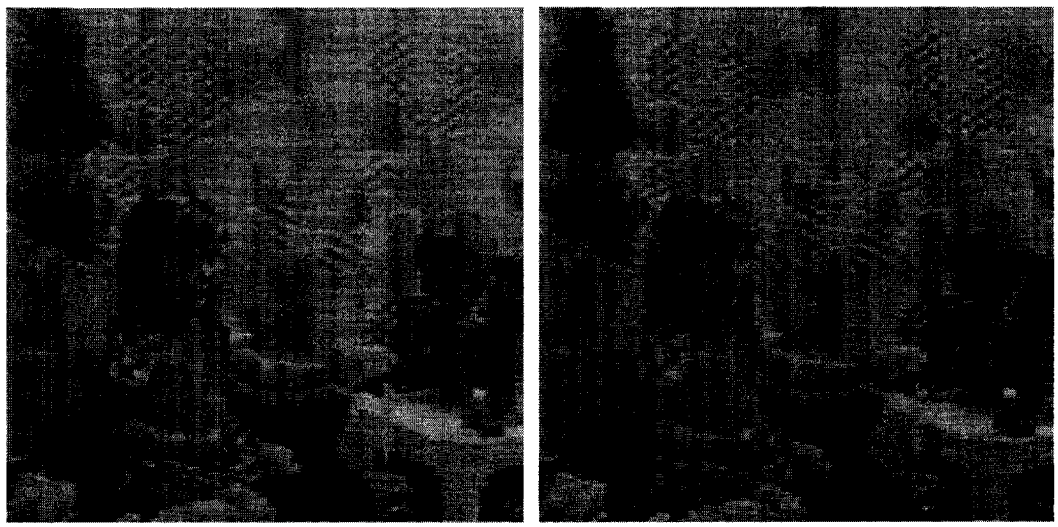
(d) Image: 27.7747, ROI: 46.9951

Figure 23: a, b, c and d represent the quality of the little girl image as 1, 2, 3 and 4 out of 16 descriptions are received.

Results of decomposition of 32 descriptions



(a) Image: 22.3545, ROI: 46.9951 (b) Image: 24.1528, ROI: 46.9951



(c) Image: 24.6228, ROI: 46.9951 (d) Image: 25.0006, ROI: 46.9951

Figure 24: a, b, c and d represent the quality of the little girl image as 1, 2, 3 and 4 out of 32 descriptions are received.

From Figure 23 and 24, we find that when the original image is decomposed into a greater number of descriptions with the same total bit rate, the results are still acceptable. We take advantage of the interpixel feature, and use a simple linear interpolating filter to reconstruct the background of the image. Although the quality of

the background is still not satisfactory, when applications pay more attentions to the ROI, the quality of the ROI and the whole image are still acceptable.

Chapter 5

PROTECTION OF ROI IN MD-SPIHT FRAMEWORK

5.1 System Overview

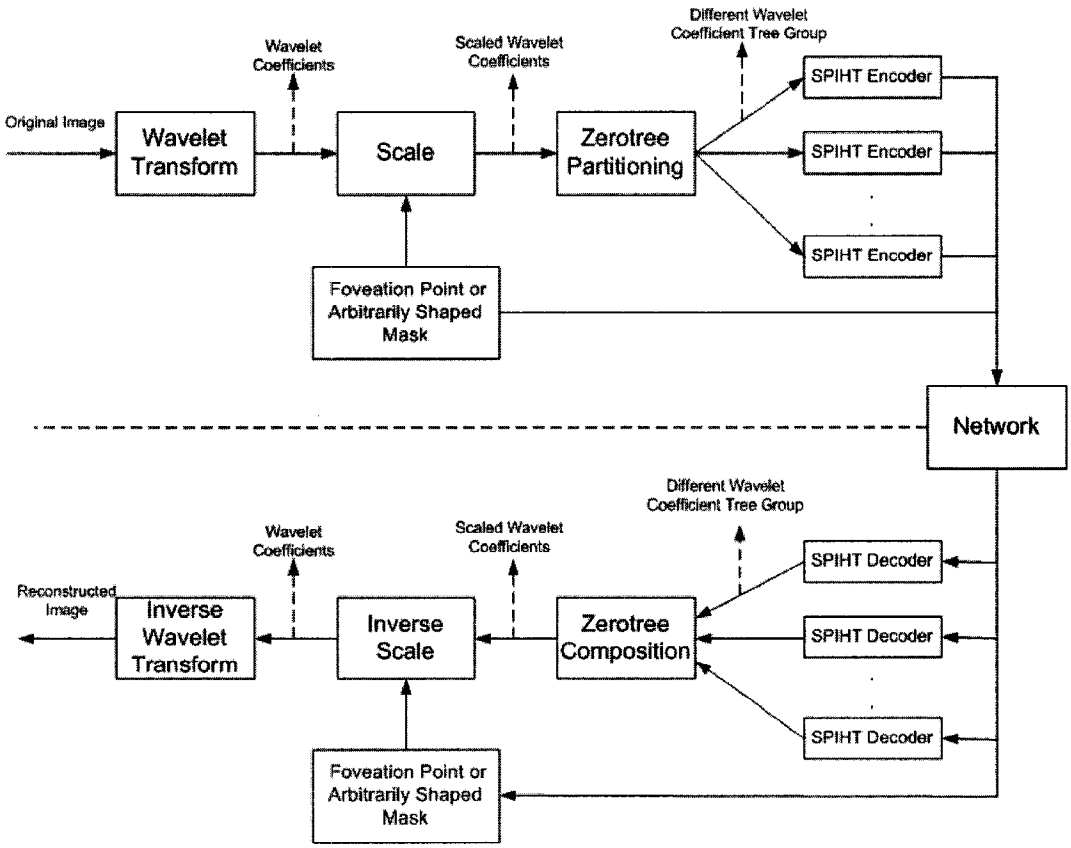


Figure 25: System architecture

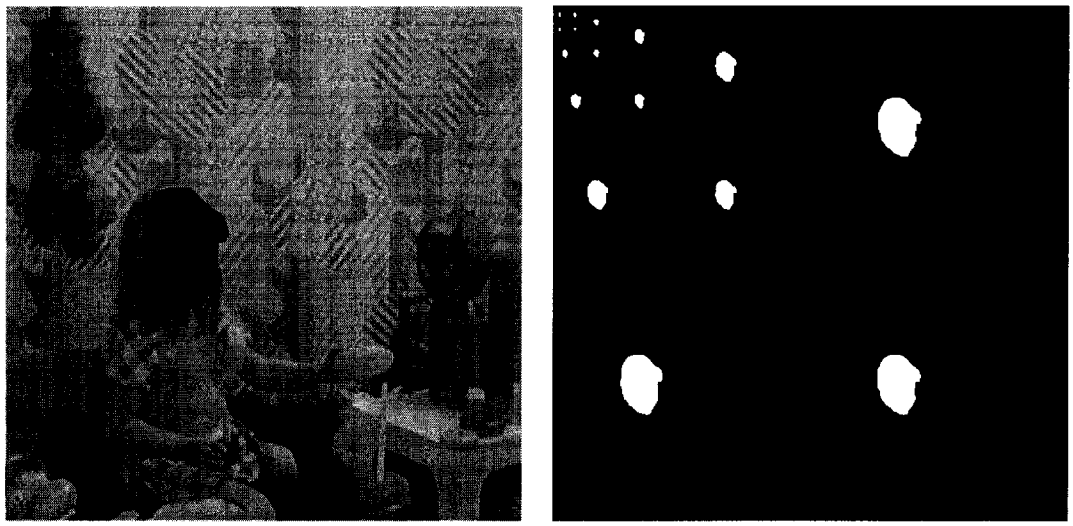
The human visual system (HVS) is highly space-variant. The spatial resolution of the HVS is highest around the point of fixation and decrease with increasing distance. The concept of prioritizing data based on the distance from a ROI was first introduced in [21]. This particular prioritization method employs a foveated image and allows a gradual increase in peripheral quality loss during network congestion. According to the human eye's perception, if the network cannot transmit all the data, we can discard some less important information in peripheral areas.

Figure 25 is the block diagram of the method we present here. I will propose a method combining the foveation technique with the zerotree ROI coding scheme to implement Multiple Description Coding in a Generalized Multiple Description Framework [8]. Two kinds of scaled strategies are used. In the first, only the coefficients in the face of the image are scaled. I choose 16 as the ROI scaling factor. In the second, the nose of the girl is chosen as the foveation point. In the LL sub-band, the roots of the zerotrees receive an importance level inversely proportional to their distance measured from the center of the ROI. According to the importance levels, the coefficients of different wavelet zerotrees will be scaled by different scale factors. Scale factors from 1 to 16 are used here.

5.2 Proposed Algorithm

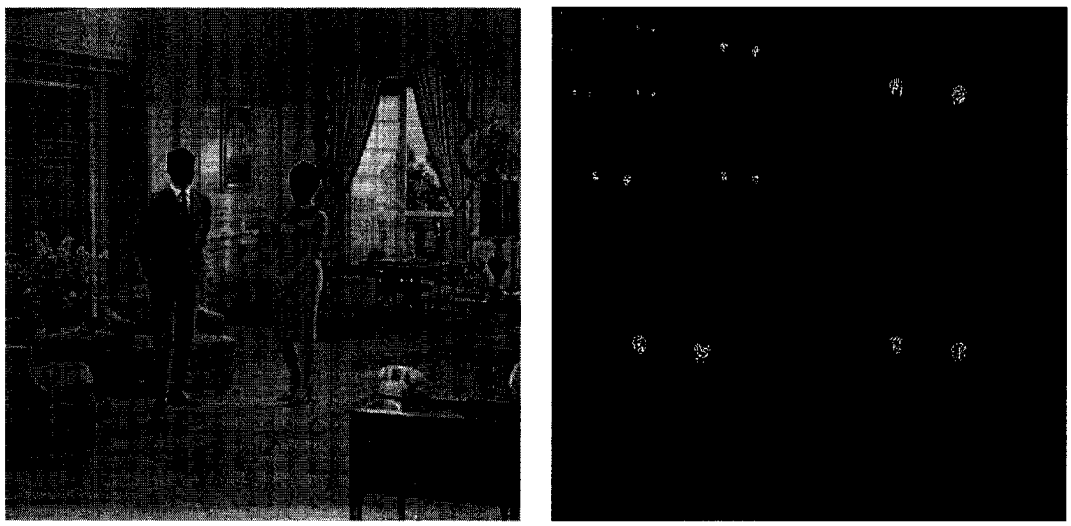
5.2.1 Object-based Discrete Wavelet Transform

We use the same method described in Section 4.2.1. Since the spatial correlation, locality properties of wavelet transforms, and self-similarity across the subbands are well preserved in the OBDWT, we can take advantage of these features to implement our method of MDC and protect the ROI in each description. In this case only OBDWT is performed to obtain the matrix of the decomposition of the shape mask. Figure 26 is an example of the OBDWT mask coding. And Figure 27 is an example of multiple ROIs mask coding.



(a) The original shape mask (b) The matrix after 5 level OBDWT

Figure 26: Examples of the OBDWT mask coding.



(a) The original shape masks (b) The matrix after 5 level OBDWT

Figure 27: The example of multiple ROIs mask coding.

5.2.2 Scale Factor Calculation

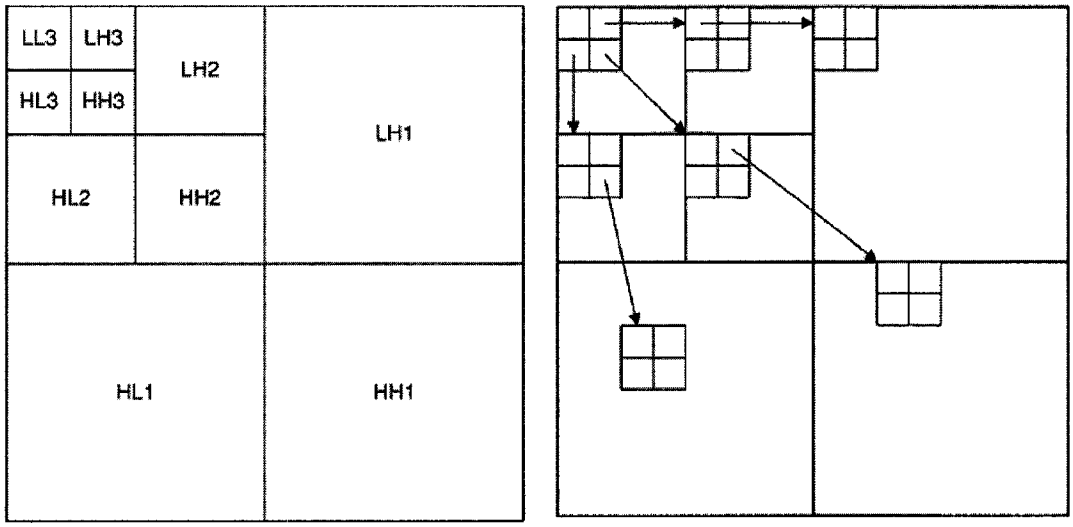


Figure 28: (a) DWT decomposition structure; (b) spatial orientation tree.

The typical DWT structure is shown in Figure 28(a). The wavelet coefficients have structural similarity across the wavelet sub-bands in the same spatial orientation. The zerotree structure in EZW and the spatial orientation tree structure in SPIHT capture this structural similarity very effectively. Figure 28(b) shows the spatial orientation tree used by SPIHT. In the SPIHT algorithm, the globally important information is allocated in the earlier parts of the compressed bit stream. We take advantage of the feature of SPIHT that allows for easy determination of the data importance for the overall quality of the image. Using the method proposed by Atsumi and Farvardin [23], the wavelet coefficients corresponding to the ROI are scaled by a larger factor. So the weighted coefficients will be considerably more important than the rest of the coefficients. As a result, these coefficients will be compressed in the earlier part of the bit stream.

In our implementation, after OBDWT, we obtain a matrix of a decomposed shape mask. According to the result of the shape mask coding, the coefficients related to the

ROI can easily be scaled with the scale factor, thus providing considerably more protection to the ROI.

5.2.3 Wavelet Coefficient Trees Partitioning

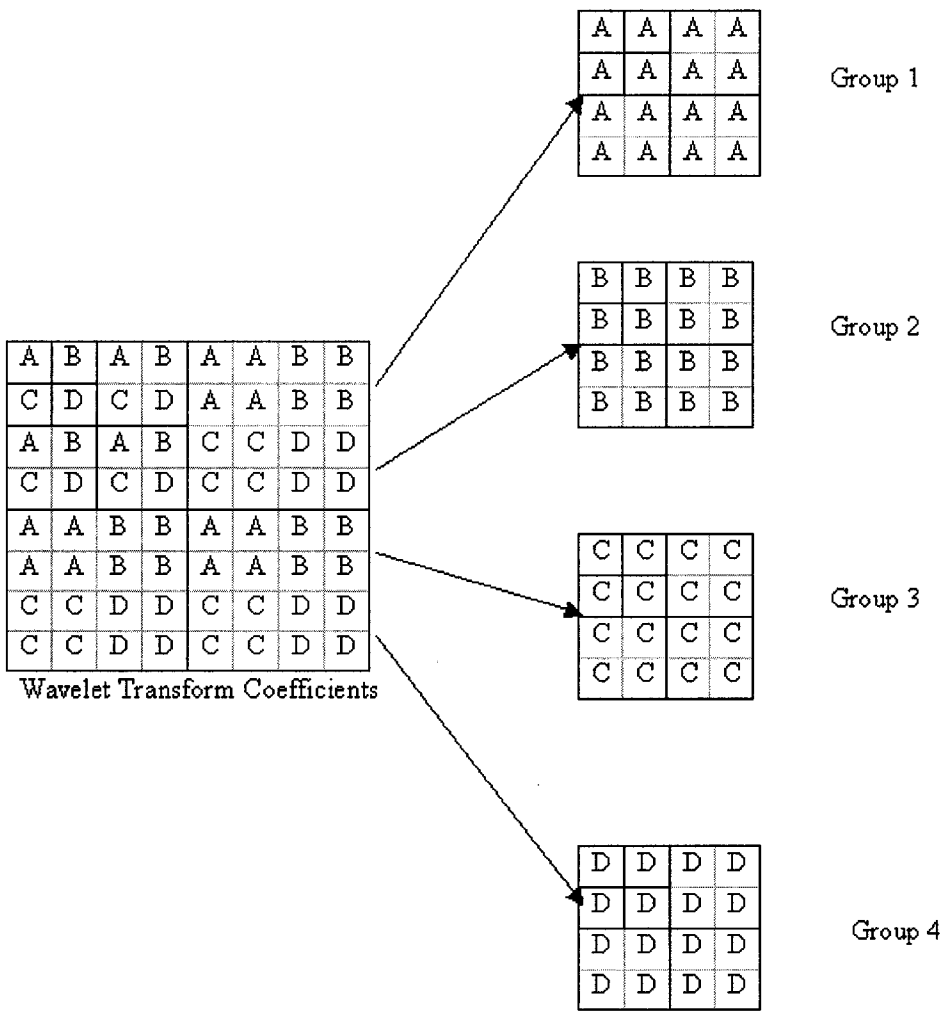


Figure 29: The procedure of wavelet coefficient trees partitioning.

After wavelet transform, the original image can generate M wavelet coefficient trees:

$M = \frac{X * Y}{2^{2L}}$, where X and Y denote rows and columns of the image, respectively, and L denotes the wavelet transform level. These wavelet coefficient trees are fed into the

same encoder, since the correlation of coefficients between different scales is exploited. We use polyphase transform to set the different wavelet coefficient trees into different groups, and then SPIHT is used to compress these groups to form multiple descriptions. In order to make every description with equal importance, N trees along the rows and the columns are sampled with a downsampling ratio. Downsampling procedure occurs in the lowest frequency subband (LL), since the roots of the trees are in this area. Each parent has four children nodes, except for the root node that has three. This means that the set of coordinates of all offspring of node (i, j) equals {(2i, 2j), (2i, 2j+1), (2i+1, 2j), (2i+1, 2j+1)}.

Figure 29 graphically illustrates the procedure of the wavelet coefficient partitioning. After 3-scale wavelet transform of an 8×8 image, four wavelet coefficient trees are generated. In the figure, every coefficients with the same letter are grouped together and processed by the same encoder. We note that spatially dispersed wavelet coefficient trees are grouped together and the zerotree structure in each group is preserved so we can obtain a good compression performance by using SPIHT algorithm.

5.2.4 Generated MDC framework

The method proposed by [8] is that each description has primary information of one group and a variable number (m) of partially coded redundant trees of other groups. In Figure 30, after wavelet coefficient tree partitioning, there are N zerotree groups. We use the generalized MDC framework provided by [4], and each description carries the information of the shape mask (SM), the primary copy of one group, *e.g.*, X1 in DC1, as well as redundant copies (X2 and Xm-1 in DC1) of some of other groups. R0, R1, and Rm are the bit rates of different copies. In this method, information about one zerotree is included in different descriptions and when some descriptions are erased the main information about these descriptions can be recovered by the copies located in other received descriptions.

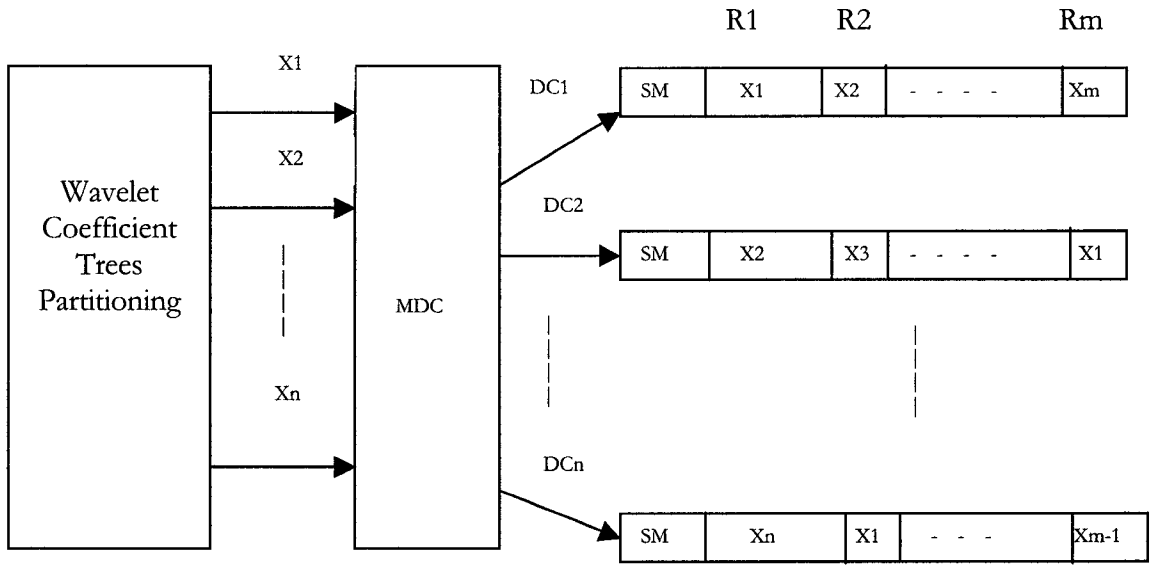


Figure 30: N groups are generated by wavelet coefficient trees partitioning. We use the method proposed by [8] to generate the N descriptions.

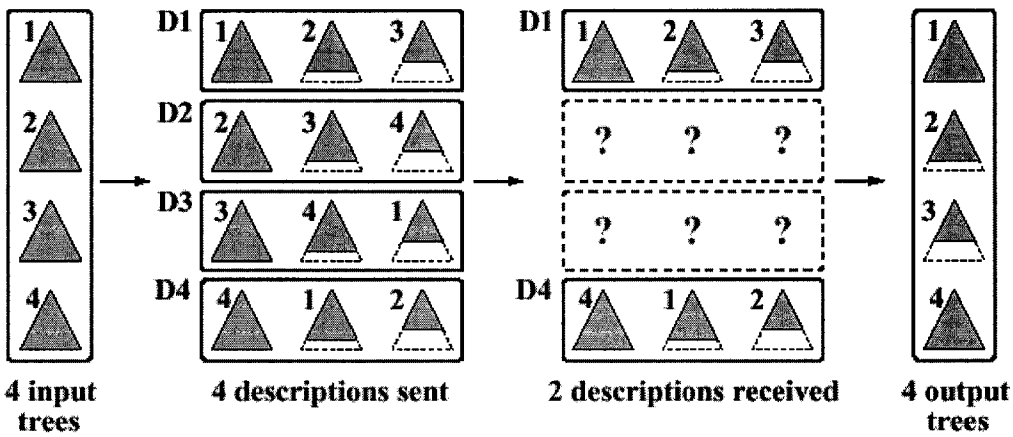


Figure 31: Coding example from [8].

In Figure 31 is an example of four groups (1-4) after the partitioning process. The primary copy of one group and the redundant copies of some of other groups are sent in four descriptions (D1-D4) , with two sets of redundant copies. If two descriptions are lost (D2 and D3), the second and third trees are partially recovered from their copies in the received descriptions.

5.2.5 Optimal Bit Allocation

If packet losses occur it is possible to recover them by exploiting the redundancy. As can be expected, if high redundancy is chosen, the performance will be good under severe error conditions. However the performance with high redundancy in an error free environment will be significantly worse than that of a non-redundant coder at the same rate [10]. It is therefore useful to adjust the amount of redundancy to adapt to different network conditions in our MDC system. We use the result of [10] to allocate the redundant bits in MDC.

In [10], an optimal bit allocation algorithm is introduced that allows us to select the amount of redundancy to best match a given target packet loss rate. The algorithm is based on rate-distortion tradeoff and uses statistical analysis tools to obtain the formula:

$$R_i^* = \frac{R}{M} + \tfrac{1}{2}(i - \frac{M-1}{2})\log_2(P)$$

In this formula, R is the total bit rate budget of the compression, M is the number of copies of each description, S is the number of descriptions, i is the tag of the copy, and P is the probability that a description is considered lost. The value obtained from the formula may be such that some of the R_i^* s are negative. This means that the copies from R_i^* to R_{M-1}^* are not needed, and we can choose M=i.

5.3 Experiments and Results

5.3.1 The size of the scaled area

From Figure 25, on the encoder side, wavelet coefficients in ROI are scaled and the scaled wavelet coefficients cause it to be coded in the earlier part of the bitstream. This way, the locally important information is protected. There is a trade off between the quality of ROI and the background for different ROI scaled factors. In [26], the authors analyzed this problem and obtained the results that "for every 5 dB increase in ROI quality, the overall image PSNR drops only by about 0.5 dB." In our case, we pay more attentions on the relation between the size of the scaled area and the reconstructed quality of the overall image. Before the simulation of MDC, we do an experiment to investigate this relation.

In this experiment we use a 512×512 gray image, "Little Girl." We use a 5-level DWT decomposition to obtain a total of 256 wavelet coefficient trees and the scale factor is set to 16. The result of the scaled number of trees and PSNR is presented in Figure 32.

We can see that, at the beginning, when the scaled number of wavelet coefficient trees increases, the PSNR of the image decreases; after the scaled number reaches half of all the trees the PSNR of the image begins to increase. The reason is that when the weighted coefficient trees increase, the weighted coefficients have a much larger dynamic range. When we use the SPIHT algorithm, it costs many more scan times and many more sign bits are increased. Since the coding efficiency is not good, the quality of the reconstructed image decreases. However, after the weighted trees reach half of the total trees, the range of the coefficients will become smaller. If all the coefficients are scaled by the same factor, the range is the same as the range of the original. The reconstructed image is nearly identical to the one reconstructed without scale. We can see that the PSNRs at 0 and 256 are nearly the same.

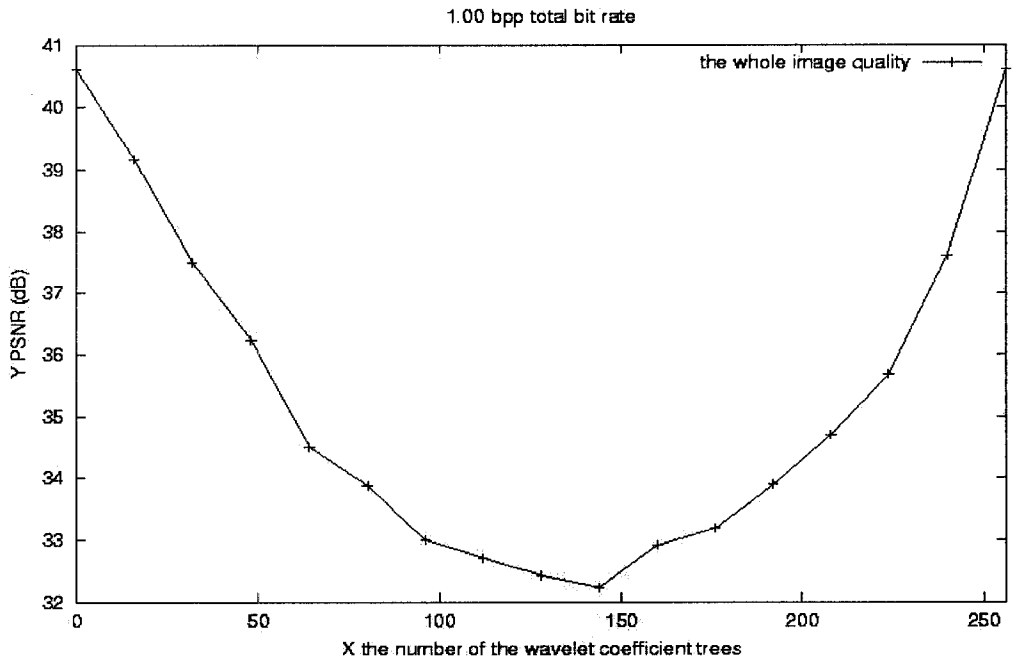


Figure 32: PSNR results for different number of scaled wavelet coefficient trees.

5.3.2 Protection of Arbitrarily Shaped ROI vs Rectangular ROI

First, we compare our method with the method proposed by [26]. In [26], the authors scale the coefficients corresponding to the ROI. They use the rectangular ROI that is about 5% of the whole image. In our method we use the arbitrarily shaped ROI that is 3% of the whole image.

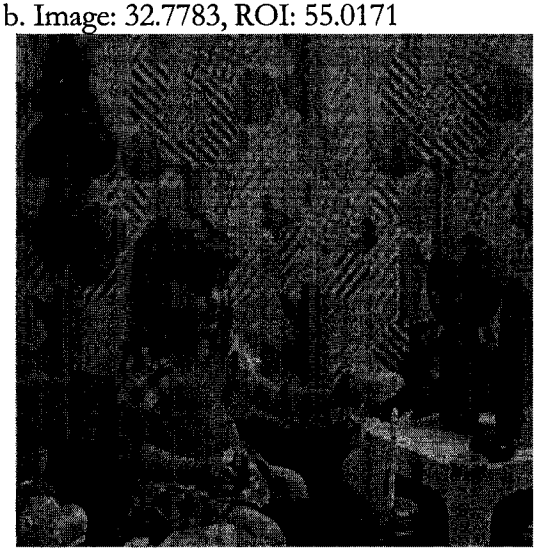
In our method only the trees corresponding to the arbitrarily shape ROI are scaled. Certainly, the ROI is less than half of the image. The total bit rate is 1.0 bpp and we add 35% redundancy. The results are shown in Figure 33. Images show 4, 5, 6, 7, 8 descriptions received out of 8. In [26], when 5 out 8 descriptions are received, the quality of image is 30.18 db and the quality of the ROI is 42.67 dB. We can see that

our reconstructed quality of the ROI and the whole image are much better than those of [26] (more details are given in Conclusion Chapter). The reason for improvement of the whole image has already been explained in Figure 32. The reason for improvement of the ROI is that since the number of the weighted trees is small, the trees belonging to the ROI can obtain better protection than many more weighted trees.

a. Image: 31.6205, ROI: 54.5193

b. Image: 32.7783, ROI: 55.0171

c. Image: 33.7790, ROI: 55.9216



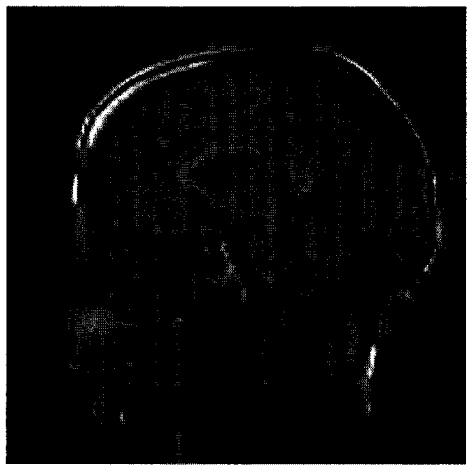
d. Image: 35.5769, ROI: 57.4708

e. Image: 36.6720, ROI: 57.6274

Figure 33: The results using the arbitrarily shaped ROI.

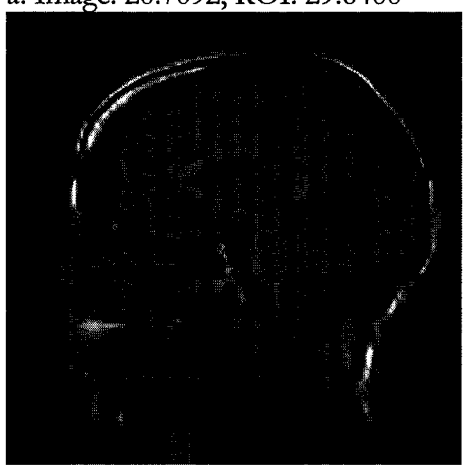
Secondly, we test our results again by using another 256×256 Magnetic Resonance Image (MRI). The MRI brain image is shown in Figure 20(a). Figure 20(b) is a rectangle ROI mask and comprises 22755 pixels — about 34% of the total image and Figure 20(c) is an arbitrarily shaped ROI mask and it comprises 11138 pixels — about 16% of the total image. The number of descriptions are set to four, and the total bit rate is 1.0bpp. The descriptions are coded with SPIHT with 35% redundancy. In this experiment we do not use the optimal bit allocation method provided by [10] because we only want to compare the influences at the different sizes of the scaled area. The results show the quality of the brain image as 1, 2, 3, and 4, of 4 descriptions are received.

Rectangular ROI **Arbitrarily Shaped ROI**

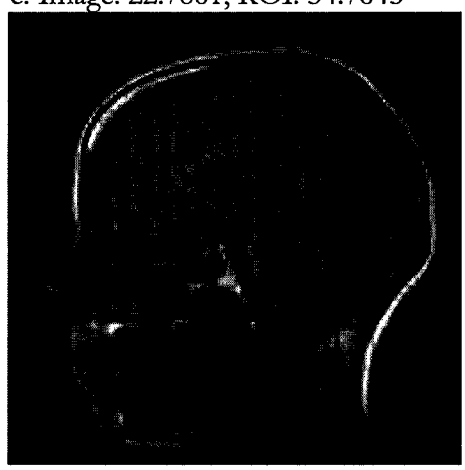


a. Image: 20.7092, ROI: 29.6400

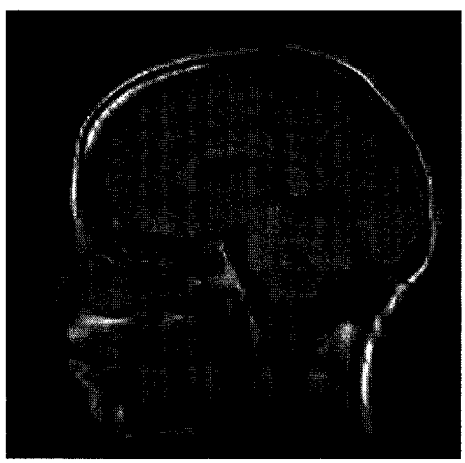
e. Image: 22.7861, ROI: 34.7645



b. Image: 21.9290, ROI: 31.1149

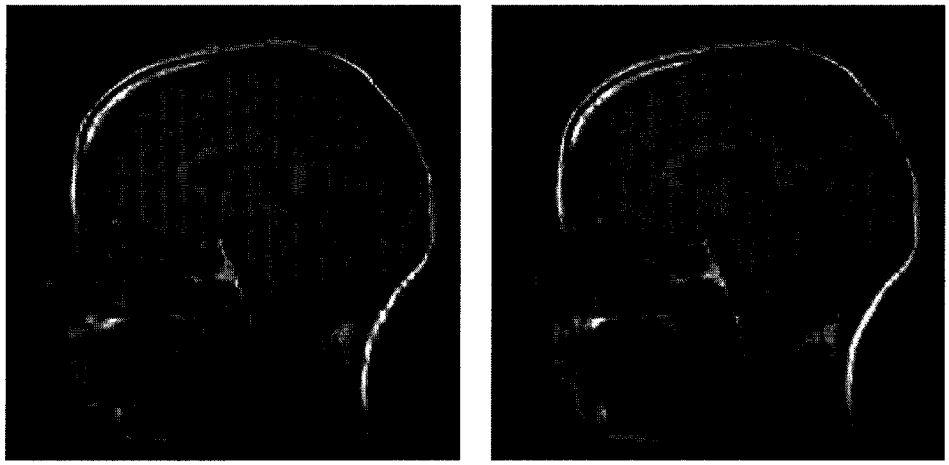


f. Image: 23.6876, ROI: 36.5357



c. Image: 23.8372, ROI: 32.9510

g. Image: 25.7311, ROI: 39.0251



d. Image: 26.4624, ROI: 37.0381 h. Image: 28.4556, ROI: 43.0381

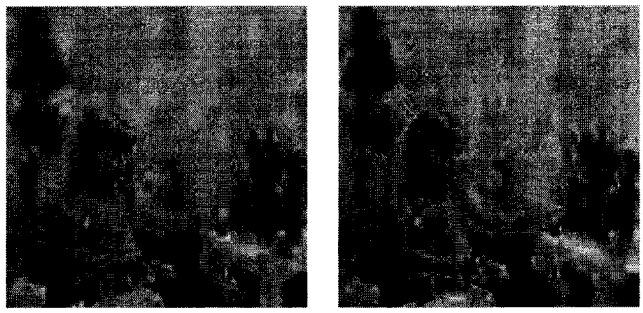
Figure 34: Comparison between the rectangular ROI and the Arbitrarily Shaped ROI.

In the left side of Figure 34, a, b, c, and d is represented as 1, 2, 3, and 4 out of 4 descriptions received in the MDC system that protects the rectangular ROI. The right side of the figure are results of the MDC system that protects the arbitrarily shaped ROI. From the comparison, we can obtain the same results as the previous one. Our method to protect ROI is demonstrated to be better than the method proposed in [26].

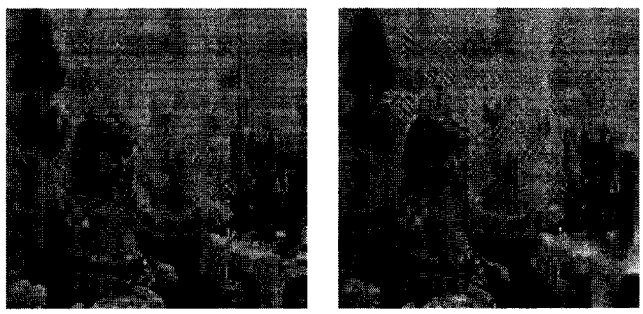
5.3.3 Foveation Technique

In foveated images, the spatial resolution is the highest around the foveation point and decreases with the increasing distance from it. Many methods are used to exploit the foveation feature to improve image and video coding algorithms. In [22], the authors use the foveation technique to rearrange the packets in the code stream to place the regions of interest before the background coefficients. It is fully compatible with JPEG2000 standard and allows transmission of different regions of interest with different priorities. Motivated by this method, we want to combine the foveation technique with the zerotree ROI coding scheme to implement Multiple Description Coding in a Generalized Multiple Description Framework [8].

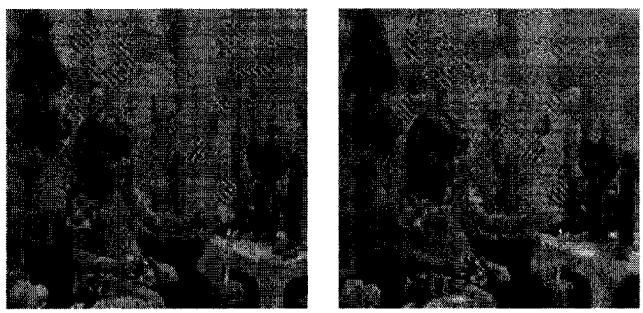
0.04bpp



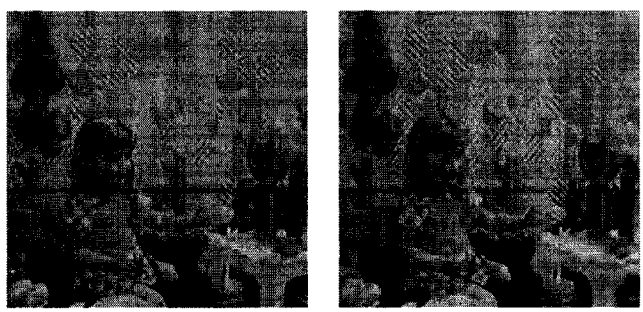
0.08bpp



0.12bpp



0.16bpp



(a). Method 1. (b). Method 2.

Figure 35: The comparison between the two methods.

In Figure 35, results are demonstrated on a 512×512 gray image, "Little Girl." I use a 5-level DWT decomposition to obtain a number of 256 wavelet coefficient trees. The Region Of Interest is the face of the little girl and the number of descriptions is set to 4. We compressed the image at 0.04bpp, 0.08bpp, 0.12bpp and 0.16bpp. We assume that all the descriptions are received. In Method 1, I only scale the coefficients in the face of the image. I choose 16 as the ROI scaling factor. In the Method 2, I choose the nose of the girl as the foveation point. In the LL sub-band, the roots of the zerotrees receive the important level inversely proportional to their distance measured from the center of the ROI. According to the importance levels the coefficients of different wavelet zerotrees will be scaled by different scale factor. I use the scale factors from 1 to 16.

The PSNR(dB) results are shown in the below table.

Coding Rate	Method 1		Method 2	
	Whole Image	ROI	Whole Image	ROI
0.04bpp	24.1988	46.4483	24.2528	43.1822
0.08bpp	26.2570	49.8950	25.8949	45.1948
0.12bpp	27.5071	51.4774	26.0891	46.0891
0.16bpp	28.8624	52.7456	27.5618	48.2879

From the comparison data in the table, we can see that the results of reconstructed quality of Method 1 are much better those of Method 2. Following are some reasons for this:

1. In Method 1, there are only two scale factors 1 and 16 so the ROI can be encoded with more bits. This means that the face of the little girl gets much more protections compared to the background of the image. However, in Method 2, there are 16 different kinds of scale factors. The area of the face will not get as much protection as it gets in Method 1 because the face area is not as relatively important as in Method 1.

2. Since there are 16 different kinds of scale factors in Method 2 the weighted coefficients have much larger dynamic range compared with that of Method 1. It will cost much more scan times and increase more sign bits. So the coding efficiency of Method 2 is less than the Method 1.

3. Most image quality measurement methods, such as PSNR, are designed for uniform resolution images. These methods are not always good indices for foveated images. We need to develop a new image quality metric considering the factors of the Human Visual System.

5.3.4 Multiple ROIs

This section shows the results of coding multiple ROIs using the proposed method. The wavelet coefficients belonging to the ROIs are scaled so the ROIs may be encoded in the earlier parts of the bitstreams. Figure 36(a) is the original couple image, and Figure 36(b) is the result of multiple ROI coding.

(a). Original image (b). Multiple ROIs (two faces)

Figure 36: The result of the multiple ROI Coding.

I demonstrate my results of multiple ROIs on a 512×512 gray image "Couple." I use a 5-level DWT decomposition to obtain a number of 256 wavelet coefficient trees. I set the number of descriptions to 4 and the total bit rate to 0.15 bpp. I assume that all the descriptions are received. I only scale the coefficients in the two faces of the image. I choose 16 as the ROI scaling factor. We can see that the two faces of the couple can get better protection compared with the background of the image.

Chapter 6

ANALYSIS AND CONCLUSIONS

6.1 Analysis of Results

The objective of this thesis is to design methods to protect arbitrarily shaped objects in generalized multiple description image coding. We propose two methods: one is based on the PDMD framework [13] and the other is based on the MD-SPIHT framework [8]. We call them ROI_PDMD and ROI_MD_SPIHT serperately. In our research, we investigated and improved the Multiple Description Coding technique to deal with network congestion. The main problem we faced in MDC is how to trade off between robustness and coding efficiency. If we add high redundancy in each description, we can exploit the redundancy to estimate the loss description, and achieve robust transmission. If we add low redundancy in each description, we may improve the coding efficiency but we cannot achieve system robustness.

In order to protect the ROI in Multiple Description Coding, there is a tradeoff between the quality of the ROI and the background. If more redundancy is added to protect the ROI, the quality of the ROI may be improved but it leads to the degradation of the whole image. In our first method, we code the ROI separately from the rest of the image. We take advantage of the features of the natural image and do not need to add explicit redundancy into each description. In our second method, we add more redundancy to the ROI and scale the wavelet coefficient trees belonging to the ROI. Both of these methods are designed to exploit the redundancy reasonably and effectively. Therefore, we not only can improve the quality of the ROI, but also improve the quality of the whole image.

The results are demonstrated on a 512×512 gray scale image of a little girl where the girl's face is a region of interest. The number of descriptions is 8, and the total bit rate is 1.0 bpp. Both of my methods are compared with Miguel'method proposed by [26]. The results are shown in Figure 37 and 38.

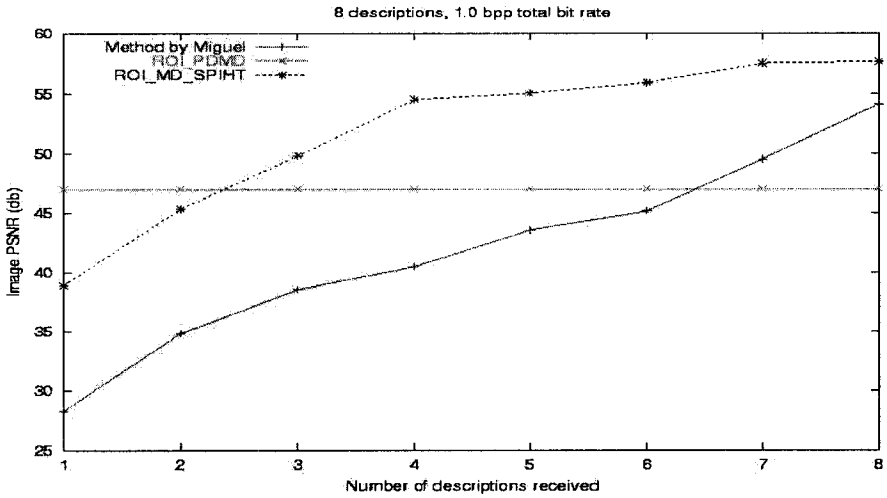


Figure 37: PSNR curve for ROI.

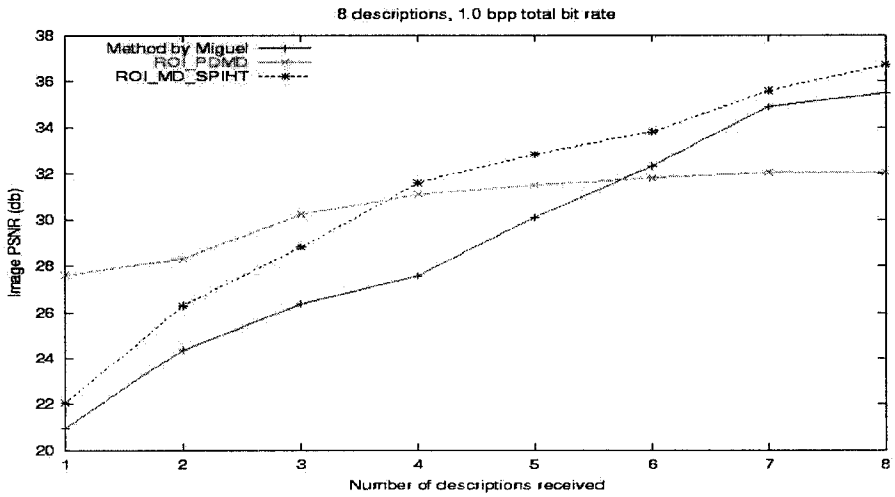


Figure 38: PSNR curve for the whole image quality.

We see that, although ROI_PDMD described in Charpter 4 yields comparable results at high losses it is somewhat worse at a loss of 4 packets or less. The downsampling method can take advantage of interpixel redundancy in the image to estimate values of neighbors, but it loses some compression efficiency with SPIHT. The reason is that the SPIHT algorithm also wants to exploit the interpixel feature of natural images to achieve good performance. Therefore, when we compress the downsampling image with SPIHT, if we increase the total bit rate of the image, the PSNR of the whole image increases relatively slowly, compared with compressing the original image with SPIHT.

We see that ROI_MD_SPIHT described in Chapter 5 is not as satisfactory as the first one for the quality of ROI and the whole image when packet losses are very high. In the second method, we achieve the MDC by adding explicit redundancy in the MD-SPIHT framework. When packet losses are very high, we need to add more redundancy to estimate the lost descriptions. So the coding efficiency is impaired and the quality of the ROI and the whole image will be degraded. When the packet losses are at a low level, we can achieve the MDC and protect the ROI with less redundancy, indicating that the second method can get better results at low packet losses.

6. 2 Conclusion

From the experiments and results discussed in Chapter 4 and Chapter 5, we can see that the methods presented here obtain better results than other systems described in the literature, under different network conditions. ROI_PDMD can achieve very good performance when the channel condition is poor. Under this scenario, our method obtains better results, since it has strong error-resilience capability hidden in the polyphase downsampling algorithm. ROI_MD_SPIHT is preferred compared with the former one when network transmission is at low-to-medium loss rates. In networks with higher loss rates, if we use the MD-SPIHT framework, we need to add much

more redundancy in case most descriptions are lost. This means we need to add more copies of other descriptions in one description. Under this scenario, the coding efficiency is degraded.

From Figures 37 and 38, we also find that the latter method can give considerably better results than the method proposed in [26] from ROI and the whole image aspects. Since we reduce the size of the ROI area, we can obtain better reconstructed quality in both the ROI and the whole image.

Chapter 7

FUTURE WORK

The work presented here should be seen as a first step towards investigating and developing Multiple Description Coding techniques for ROI processing. In our MDC approaches, some redundancy is preserved in the source coding. If packet losses occur, it is possible to recover them by exploiting the redundancy.

One interesting possible feature work is to do some research on the bit allocation in our MDC algorithm. The goal is to determine the optimal selection of the number of copies of a component, and the proportion of bits to assign to each copy (a copy is a compressed version of a polyphase component), given a known packet loss rate.

Another interesting possible future work is to use MDC in a Peer-to-Peer network. The basic idea is to design multiple trees to provide redundancy in network paths and also in the data. This means we can achieve robust network communication by introducing the redundancy on both the paths and the data. In [29], Microsoft provides a framework to combine MDC with the CoopNet system to achieve resilient Peer-to-Peer streaming. There is still considerable research to be done on the design of MDC in Peer-to-Peer systems.

BIBLIOGRAPHY

[1] V.A. Vaishampayan , "Design of multiple description scalar quantizers," *IEEE Trans. Information Theory*, Vol. 39, no. 3, pp. 821-834, 1993.

[2] V.A. Vaishampayan and J.-C. Batllo, "Asymptotic analysis of multiple description quantizers," *IEEE Trans. On Information Theory*, vol. 44, no. 1, pp. 278-284, Jan. 1998.

[3] S.D. Servetto, K. Ramchandran, V. Vaishampayan, and K. Nahrstedt, "Multiple description wavelet based image coding," in ICIP'98, 1998.

[4] Y. Wang, M. Orchard, and A.R. Reibman, "Multiple description image coding for noisy channels by pairing transform coefficients," *In Proc. IEEE 1997 First Workshop n Multimedia Signal Processing*, 1997.

[5] V.K. Goyal, J. Kovacevic, R. Arean, and M. Vetterli, "Multiple description transform coding of images," in ICIP'98, 1998.

[6] W. Jiang and A. Ortega. "Multiple description coding via polyphase transform and selective quantization", *In Proceedings of SPIE: Visual Communications and Image Proceeding*, January, Vol.3653, No.2, 1999, pp. 998-1008.

[7] A.E. Mohr, E.A. Riskin, and R. Ladner. " Generalized multiple description coding through unequal forward error protection", *In Proceedings of ICIP*, 1999, vol.1, pp. 411-415.

[8] Agnieszka C. Miguel, Alexander E. Mohr, Eve A. Riskin, "SPIHT for Generalized Multiple Description coding", *In Proceedings of ICIP*, 1999, vol.3, pp. 842-846.

[9] A. C. Miguel and E. A. Riskin, "Protection of Regions of Interest Against Data Loss in a Generalized Multiple Description Framework", *In Proceedings of the Data Compression Conference (DCC)*, 2000.

[10] P. Sagetong and A. Ortega, "Optimal Bit Allocation for Channel-Adaptive Multiple Description Coding," *In Image and Video Communications and Processing 2000 Conf.*, EI 2000, San Jose, CA, Jan 2000.

[11] Shipeng Li and Weiping Li, "Shape Adaptive Discrete Wavelet Transforms for Arbitrarily Shaped Visual Object Coding", IEEE Transaction on Circuits and System for Video Technology, August 2000.

[12] Yuan, Y., Choong Wah Chan, "Coding of arbitrarily shaped video objects based on SPIHT", Electronics Letters , Volume: 36 Issue: 13 , 22 June 2000, Page(s): 1105 –1106.

[13] Caramma Marcello, Fumingly Marco, Lancini Rosa, "Polyphase DownSampling Multiple Description Coding for IP transmission", *Proceedings of SPIE 2001*; Visual Communications and Image Processing, 21-26 January 2001, San Jose, California, USA.

[14] Franchi Nicola, Fumingly Marco, Lancini Rosa and Tubaro Stefano, "Multiple Description Video Coding for Scalable and Robust Transmission over IP", *Proceedings of 13th Packet Video Workshop*, 28-29 April 2003, Nanates (France).

[15] J.M. Shapiro, Embedded image coding using zerotrees of wavelet coefficients, IEEE Trans. *On Signal Processing*, vol.41, No.12, 1993, pp.3445-3462.

[16] A.Graps, "An introduction to Wavelets," *IEEE Computational Sciences and Engineering*, http://www.amara.com/IEEEwave/IEEEwavelet.html.

[17] A. Said and W.A. Pearlman, " A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Transactions on Circuits and Systems for video Technology*, vol. 6, pp. 243-250, June 1996.

[18] Yen-Chi Lee, Kim, J., Altunbasak, Y. and Mersereau, R.M. "Performance Comparisions of Layered and Multiple Description Coded Video Streaming over Error-Prone Network", 2003 IEEE International Conference on Communications.

[19] L. Rizzo, "Effective erasure codes for reliable computer communication protocol," *ACM Computer Communication Review*, vol. 27, pp.24-36, Apr. 1997.

[20] Alexander E. Mohr, Eve A. Riskin, and Richard E. Ladner, "Unequal Loss Protection: Graceful Degradation over Packet Erasure Channels through Forward Error Correction." *IEEE Journal on Selected Areas in Communication*, vol. 18, no. 7, pages 819-828. June 2000, 25 pages.

[21] A. Basu, A. Sullivan and K. Wiebe, "Variable resolution teleconferencing,", IEEE SMC Conference, pp. 170-175, 1993.

[22] V. Sanchez, A. Basu and M. K. Mandal , "Prioritized Region of Interest Transmission for JPEG2000," *IEEE Trans. on Circuits and Systems for Video Technology*, accepted for publication.

[23] E. Atsumi and N. Farvardin, "Lossy/lossless region of interest coding based on set partitioning in hierarchical trees," *in Proceedings of ICIP*, vol. 1, pp. 87-91, 1998.

[24] H. Freeman. "On the encoding of arbitrary geometric configurations." IRE Trans. Electronic Computers, EC-10(2):260-268.

[25] H.J. Barnard, J.H.Weber and J. Biemond. Efficient signal extension for subband/wavelet decomposition of arbitrary length signals. *In Proceedings SPIE VCIP'93*, volume 2094, pages 966-975, November 1993.

[26] Agnieszka C. Miguel and Eve A. Riskin, "Protection of Regions of Interest Against Data Loss in a Generalized Multiple Description Framework", UWEE Technique Report Number UWEETR-2002-0007, January 2002.

[27] J.K. Rogers and P.C. Cosman, "Robust wavelet zerotree image compression with fixed length packetization," *in Proceedings Data Compression Conference*, pp. 418-427, March-April 1998.

[28] G. Davis and A. Nosratinia, "Wavelet-based Image Coding: An Overview", Applied and Computational Control, Signals, and Circuits, vol. 1, no. 1, 1998.

[29] V.N. Padmanabhan, H. J. Wang, and P. A. Chou, "Resilient Peer-to-Peer Streaming," Microsoft Research Technical Report MSR-TR-2003-11, March 2003.