Scalable Multiple Description Coding for heterogeneous multimedia delivery networks

Enrico Baccaglini

Tutore
prof. Gabriella Olmo

Coordinatore del corso di dottorato
prof. Ivo Montrosset

Febbraio 2007
List of Acronyms

**AP** Access Point

**ARQ** Automatic Repeat reQuest

**BER** Bit Error Rate

**BPP** Bit per pixel

**CB** Code Block

**CIF** Common Intermediate Format

**DCT** Discrete Cousin Transform

**DPCM** Differential Pulse-Code Modulation

**DWT** Discrete Wavelet Transform

**EBCOT** Embedded Block Coding with Optimized Truncation

**FEC** Forward Error Correction

**FPS** Frames per second

**GSM** Global System for Mobile communications

**HCF** Hybrid Coordination Function

**IP** Internet Protocol
JPEG  Joint Photographic Experts Group
MAC  Media Access Control
MDC  Multiple Description Coding
MDSQ  Multiple Description Scalar Quantization
MDVQ  Multiple Description Vector Quantization
MJPEG  Motion JPEG
MPEG  Moving Picture Experts Group
MSE  Mean Squared Error
MT  Mobile terminal
MTU  Maximum Transfer Unit
OFDM  Orthogonal frequency-division multiplexing
PSNR  Peak Signal-to-Noise Ratio
QoS  Quality of Service
RS  Reed Solomon Codes
RD  Rate Distortion
SDC  Single Description Coding
SNR  Signal to Noise Ratio
TCP  Transmission Control Protocol
UEP  Unequal Error Protection
WLAN  Wireless Local Area Network
Table of contents

1 Introduction
   1.1 Network scenario ........................................... 4
   1.2 JPEG 2000 ..................................................... 5
   1.3 Applications of MD coding .................................... 7

2 The Multiple Description model .................................. 9
   2.1 Rate distortion regions ...................................... 11
   2.2 RD region for two descriptions ............................... 13
   2.3 RD region for many descriptions .............................. 14
   2.4 MD Techniques ................................................ 15

3 Network adaptive Multiple Description Coding for JPEG 2000 .... 18
   3.1 Two-rate approach for JPEG 2000 .............................. 19
   3.2 Theoretical analysis .......................................... 21
      3.2.1 Two-description case .................................... 21
      3.2.2 Arbitrary number of descriptions .......................... 24
   3.3 Experimental results .......................................... 25

4 A flexible MD multi-rate allocation scheme ....................... 30
   4.1 Multi-rate allocation scheme ................................ 30
   4.2 Redundancy optimization ...................................... 33
      4.2.1 Case I: Bounded worst-case end-user quality ........... 33
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.2.2</td>
<td>Case II: Maximization of average end-to-end quality</td>
<td>35</td>
</tr>
<tr>
<td>4.3</td>
<td>Multi-rate MD JPEG 2000 codec</td>
<td>39</td>
</tr>
<tr>
<td>4.3.1</td>
<td>Complexity</td>
<td>40</td>
</tr>
<tr>
<td>4.4</td>
<td>Experimental results</td>
<td>41</td>
</tr>
<tr>
<td>4.4.1</td>
<td>Gaussian source</td>
<td>41</td>
</tr>
<tr>
<td>4.4.2</td>
<td>JPEG 2000-based MD codec</td>
<td>43</td>
</tr>
<tr>
<td>4.5</td>
<td>Appendix A</td>
<td>46</td>
</tr>
<tr>
<td>5</td>
<td>MD Coding in sensor networks with finite buffers</td>
<td>57</td>
</tr>
<tr>
<td>5.1</td>
<td>MDC in a point-to-point link</td>
<td>58</td>
</tr>
<tr>
<td>5.1.1</td>
<td>MDC based on index assignment</td>
<td>60</td>
</tr>
<tr>
<td>5.2</td>
<td>MDC in a sensor network</td>
<td>61</td>
</tr>
<tr>
<td>6</td>
<td>MD Coding for MJPEG 2000 over congested 802.11e WLANs</td>
<td>68</td>
</tr>
<tr>
<td>6.1</td>
<td>System description</td>
<td>69</td>
</tr>
<tr>
<td>6.2</td>
<td>Proposed method</td>
<td>71</td>
</tr>
<tr>
<td>6.3</td>
<td>Experimental results</td>
<td>72</td>
</tr>
<tr>
<td>7</td>
<td>MD Coding for storage of Digital Cinema content</td>
<td>77</td>
</tr>
<tr>
<td>7.1</td>
<td>Introduction</td>
<td>77</td>
</tr>
<tr>
<td>7.2</td>
<td>Proposed approach</td>
<td>79</td>
</tr>
<tr>
<td>7.3</td>
<td>Experimental results</td>
<td>81</td>
</tr>
<tr>
<td>8</td>
<td>Conclusions</td>
<td>85</td>
</tr>
<tr>
<td>8.1</td>
<td>Redundancy tuning</td>
<td>85</td>
</tr>
<tr>
<td>8.2</td>
<td>Applications</td>
<td>87</td>
</tr>
</tbody>
</table>

Bibliography                                                                 | 89   |
Chapter 1

Introduction

The efficient transmission of video content over communication networks is a challenging goal, especially when considering heterogeneous networks where end-user terminals differ in terms of visual resolution, computational and storage capabilities. The reliable transmission of multimedia content requires new coding techniques able to face channel failures and, at the same time, to maintain a real-time communication between sender and receivers.

Current multimedia systems typically generate contents with a progressive coding. Quality improves with successive refinements as the number of consecutive received packets increases. Progressive transmission is effective when packets are received in order and without losses. When a loss occurs, the decoder requests a retransmission and the reconstruction stops until that particular packet is received. The delay in receiving a retransmitted packet may be much longer than the inter-arrival time between packets.

When a retransmission is required, the receiver tells the sender either which packet has arrived or which have not using a protocol of a family called Automatic Repeat reQuest (ARQ); then, the sender transmits the missing packet until the message is received correctly.

In real-time communication systems, retransmission is not always possible. It can happen that the transmission is strictly one-directional and so there is no way to get a
message back from the receiver to the sender to inform it about the received or missed packets. Moreover, even if the transmission is potentially bidirectional, the feedback could generate too much traffic. For example, in a broadcast communication (one sender, many receivers) acknowledgement messages could congest the network.

Furthermore, when packet losses are frequent, retransmission can create an even more congested environment and real-time services will particularly suffer from this situation. In fact, retransmissions imply an added delay of at least one round-trip transport time. This could be unacceptable in interactive communications and for audio or video streaming. In fact, the information contained in a retransmitted packet could be obsolete by the time it reaches the destination.

When retransmissions are not possible, the technique commonly used to protect data and to allow the receiver to deal with losses is called Forward Error Correction (FEC). Before sending bits on the transmission channel, the sender adds some redundancy to source bits so that, when a subset of them reaches the receiver, the original information can be recovered. Examples of FEC techniques (also called channel codes) are parity bits and Reed-Solomon Codes [27]. However, reliable use of these channel codes requires long block sizes and this creates difficulties associated with delay.

If losses cannot be avoided and retransmission is not possible, representations that make all of the received packets useful even though they are not consecutive, can be of great benefit. It could be useful to estimate the original information despite packet losses and obtain a reproduction quality proportional to the number of packets received.

Multiple Description (MD) coding applies to this situation. In conventional, “single description” (SD) source coding, a source encoder produces a single sequence of bits, which is received (without errors) by a source decoder. An MD source encoder produces bits that are partitioned into descriptions. We can associate these descriptions with packets sent over a network. The decoder is able to compute an estimate of the source from any subset of these descriptions. The quality of the
estimate will depend on how many descriptions have been received but, in contrast to the single-description case, the loss of packets does not lead to a failure.

MD coding requires some redundancy in encoded representations to gain robustness to combat packet losses. Some compression efficiency is thus sacrificed and so this technique should only be applied if the disadvantage in compression is offset by the advantage of mitigating transport failures. The amount of inserted redundancy is generally measured in terms of the extra rate required by the MDC scheme, compared to a single description coding scheme achieving the same average performance. As the redundancy impairs the system characteristics in case no loss occurs, it is clear that the number of descriptions and the introduced redundancy should be carefully tuned in order to match the actual network conditions.

In both SD and MD coding, encoder and decoder are oblivious to the transport mechanism and MD coding is usually described as a generalized source coding problem. With descriptions corresponding to packets, MD coding is plainly applicable to packet-based communications with the possibility of packet losses. Therefore, the growth of multimedia content on the Internet is spurring great interest in MD coding.

One of the first examples of a multiple description system has been proposed in [25]. As Fig. 1.1 shows, a transmitter samples a speech signal at 12 kHz and then splits into odd and even streams. These are separately DPCM encoded and each one is sent over an independent channels. If one of the two links fails, it is still possible
to reconstruct the original audio signal using a half-rate decoder made of a DPCM decoder and an interpolator. If both channels are available for decoding, a full-rate decoder reconstructs the original signal with higher fidelity.

1.1 Network scenario

In this work, we focus our attention on heterogeneous networks. With this term we denote a group of electronic devices with different capability in terms of memory and storage capability, visual resolution and computational capabilities. This is a typical scenario of a modern communication system. These devices can be desktop or laptop PCs, TV sets, handheld devices such as smart phones or personal digital assistants or classic mobile phones. All these devices, in fact, differ in terms of storage capability (few Megabytes for mobile phones and many Gigabytes for personal computers) but also visual resolution (few pixels for mobile phones and high-definition for TV sets and PCs). Similar considerations can be done for computational capabilities.

Most of the times, these devices are connected to the network in different ways. Some of them can be wired connected to the network using a Ethernet network card or a standard modem. Someone else could be connected with a wireless link. This poses significant open issues regarding the available bandwidth of each user.

In the rest of this work, we focus on heterogeneous networks where a subset of the devices are connected to the others with a wireless link. In fact, wireless communications have become very important in today’s life and they are finding their way into a wide variety of markets. This expansion is mainly due to the fact that wireless apparatus and networks can be installed nearly anywhere, and the associated services can be obtained anywhere and anytime in the covered area, resulting in high flexibility and mobility.

Transmission of multimedia information in mobile wireless networks poses many challenges, due from one hand to the huge amount of data to be transmitted, and on the other hand to the severe and non stationary channel conditions, resulting in
dynamically varying bit error rates (BER). Additional problems are related to the use of traditional transport-layer protocols, not suited to wireless applications, and to the heterogeneity of receivers with different capabilities in terms of visual quality requirements, processing capabilities, power and bandwidth limitation.

In fact, it is becoming important to efficiently exploit network capabilities to deliver to the devices multimedia contents on demand. From a service provider point of view, the desirable scenario would be a single file for each different content to be delivered, on demand, to these clients. A problem can arise when each of these users requires a dedicated stream that suits its capabilities. Then, the service provider has to encode the same multimedia streams in different formats (e.g.: one format for all the possible visual resolutions) and store them. This requires a long encoding time and significant capabilities in terms of storage for the service provider.

All these problems result in network scenarios characterized by large packet loss rates, associated with errors at the bit level. The bandwidth limitations of wireless applications limit the range and throughput that can be supported, meaning that high compression algorithms are to be addressed to cope with the massive multimedia information. Unfortunately, highly compressed multimedia data is even more sensitive to errors, and this requires the use of algorithms to protect the data from being corrupted. Efficient error control algorithms with low overhead are being explored; on the other hand, the size and battery power limitation of wireless mobile devices severely limit the complexity of error recovery algorithms. The limited resources of wireless applications and the time constraints of particular application make data retransmission not generally possible.

1.2 JPEG 2000

Our research has been focused on the reliable transmission of JPEG 2000 still images and Motion JPEG 2000 video sequences over unreliable heterogeneous networks.

JPEG 2000, as the original JPEG standard, applies a form of transform coding to compress images. However, JPEG 2000 uses the 9/7 wavelet transform for lossy coding, in contrast to JPEG, which uses an 8 × 8 block-size discrete cosine transform.

Initially, images are transformed from the RGB to the YCbCr color space leading to three components. The chrominance components can be down-scaled in resolution; since the wavelet transformation already separates images into scales, downsampling is more effectively handled by dropping the finest wavelet scale. This step is called *multiple component transformation.*

After color transformation, the image is split into *tiles,* rectangular regions of the image that are transformed and encoded separately. The purpose of tiles is to cope with memory limitations more easily. These tiles are then wavelet transformed. The result is a collection of subbands which represent several approximation scales. A subband is a set of coefficients—real numbers which represent aspects of the image associated with a certain frequency range as well as a spatial area of the image. These coefficients are scalar-quantized, yielding a set of integer numbers which have to be encoded bit-by-bit.

The quantized subbands are split further into *precincts,* rectangular regions in the wavelet domain. They are typically selected in a way that the coefficients within them across the subbands form approximately spatial blocks in the (reconstructed) image domain, though this is not a requirement.

Precincts are split further into *codeblocks* (CBs). CBs are located in a single subband and have equal sizes. The encoder has to encode the bits of all quantized coefficients of a CB, starting with the most significant bits and progressing to less significant bits by a process called the EBCOT scheme. In this encoding process, each bit-plane of the codeblock gets encoded in three coding passes, first encoding bits and signs of insignificant coefficients with significant neighbors, then refinement bits of significant coefficients and finally coefficients without significant neighbors. The bits selected by these coding passes then get encoded by a context-driven binary arithmetic coder, namely the binary MQ-coder. The context of a coefficient is formed
by the state of its nine neighbors in the codeblock.

The result is a bitstream that is split into packets where a packet groups selected passes of all codeblocks from a precinct into one indivisible unit. Packets from all subbands are then collected in so-called layers.

The goal of the encoding process is to find the optimal packet length for all CBs which minimizes the overall distortion in a way that the generated target bitrate equals the demanded bitrate. For each bit encoded by the EBCOT coder, the improvement in image quality, defined as mean squared error, is measured. Furthermore, the length of the resulting stream is measured. This forms for each codeblock a graph in the rate-distortion curve, giving image quality over bitstream length. The optimal selection for the truncation points, thus for the packet-build-up points is then given by defining critical slopes of these curves, and picking all those coding passes whose curve in the rate-distortion graph is steeper than the given critical slope. This method is an application of the method of Lagrange multiplier which turns out to be the critical slope, while the constraint is the demanded target bitrate, and the value to optimize is the overall distortion.

For further details about JPEG 2000, refer to [45].

1.3 Applications of MD coding

MD coding techniques can be naturally applied to packet networks. In such networks, packets are lost for many reasons. This can occur seemingly at random when an intermediate node along a packet path becomes congested, resulting in buffer overflows. This provides, for example in the Internet, a packet loss probability that varies with time of day and connection routing.

In addition, the Internet itself is becoming very heterogeneous as backbone capacities increase but more low-bandwidth, wireless devices are connected. Moving data from a higher to a lower-bandwidth link may require dropping packets to accommodate the lower capacity.
Using of a multiresolution or layered source coding system is however a good solution if the network is able to provide different treatment to some packets; in general, networks will not look inside packets and so they will be dropped at random.

In a network using Internet Protocol Version 6 (IPv6) [11], a node is required to handle at least 576-byte packets without fragmentation. Sending packets with less than 536-byte payloads is wasteful and so, with packets of this size, a typical image may be communicated in about ten packets; therefore, techniques that generate MD using many descriptions are important.

The distributed storage problem also matches the MD framework well. Typical users could view local images copies, but when the need for higher quality arises, one or more remote copies could be retrieved and combined with the local copy.

In many wireless systems, to provide robustness against bit errors, a transmitter can hop through a set of carrier frequencies in a manner known to the receiver. This is what happens, for example, in the GSM system. This gives protection against picking a bad carrier frequency. Some carriers will be good and others will be bad. Carrier frequencies can be considered separate channels for an MD source code. Channel codes can be applied separately for each carrier frequency. For some channels, all errors will be corrected, the other channels can be considered lost.

Finally, MD coding can be applied to ad-hoc networks. These networks are composed by a large number of unreliable devices. From any particular node, there are many possible paths to reach any other node. However, the probability of one of these paths to fail is not negligible. Multi path routing techniques have been found to be a good strategy to increase robustness under these conditions. These particular networks strongly call for specific coding techniques capable of exploiting the path diversity present in the network. MD coding fits very well in this scenario.
Chapter 2

The Multiple Description model

MD coding refers to the scenario depicted in Fig. 2.1 in which we consider the case of two descriptions. An encoder receives a sequence of source symbols \( \{X_k\}_{k=1}^{T} \) to communicate to three receivers over two noiseless (or error-corrected) channels. One decoder (the central decoder) receives information sent over both channels while the remaining two decoders (the side decoders) receive information only over their respective channel. The encoder sends each description along a different channel, therefore we can associate each channel with a particular description. The transmission rate over channel \( i \) is denoted by \( R_i \) (\( i = 1,2 \)) bits per source sample. The reconstruction sequence produced by decoder \( i \) is denoted by \( \{\hat{X}_k^{(i)}\}_{k=1}^{T} \).

If any description is received, the side decoder can recover the data with a given side distortion. The minimum distortion (central distortion) is obtained when all

![Figure 2.1. An MD system with 2 descriptions.](image-url)
descriptions are received. In the case of two descriptions, the performance of MD can be evaluated in terms of the central distortion $D_0$, and the side distortions $D_1, D_2$ when either description is received, as a functions of the rates $R_1$ and $R_2$ devoted to either description, corresponding to a total rate $R_0 = R_1 + R_2$. The main objective of the MD encoder is to find the optimal quintuple $(D_0, D_1, D_2, R_1, R_2)$. If an information source is described by separate descriptions, the MD problem consists in finding what are the concurrent limitations on qualities of these descriptions, taken separately and jointly.

The natural extension to more than two descriptions is to $N$ channels and $2^N - 1$ receivers, each corresponding to a particular subset of descriptions received. This generalization is of great practical importance for situations in which the network presents many possible paths (channels) to reach the destination.

An example with $N = 3$ descriptions is given in Fig. 2.2. The source encoder generates three different descriptions and send them along different channels. Decoder $Dec_i$ receives description $i$ (with $1 \leq i \leq 3$) and create a coarse version of the original source according to the description received. $Dec_{ij}$ receives descriptions $i$ and $j$ (with $1 \leq i \leq 3$, $i \neq j$) and it can recover a more accurate representation of the original information. Decoder $Dec_{123}$ receives all the descriptions.

To create an MD system, we have to find good representations to send on each channel so that the quality of the estimate increases with the number of channels.
(descriptions) that reach the destination.

## 2.1 Rate distortion regions

Assuming that a source produces a sequence of independent, identically distributed, real random variables $X_1, X_2, \ldots, X_T$, a *distortion measure* $d$ gives a nonnegative numerical rating $d(x, \hat{x})$ to how well a source letter $x \in \chi$ is approximated by a reproduction $\hat{x} \in \hat{\chi}$. The distortion between sequences $x^T = (x_1, x_2, \ldots, x_T)$ and $\hat{x}^T = (\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_T)$ is defined as:

$$d(x^T, \hat{x}^T) = \frac{1}{T} \sum_{i=1}^{T} d(x_i, \hat{x}_i) \tag{2.1}$$

The most common distortion measure is the squared error distortion defined as:

$$d(x, \hat{x}) = (x - \hat{x})^2 \tag{2.2}$$

This corresponds, for real sequences of length $T$, to the squared Euclidean norm between the two sequences, divided by $T$.

Given a code with dimension $T$, rate $R$ and expected distortion $D_T$, consisting of an encoding function

$$\alpha_T : \chi^T \rightarrow \{1, 2, \ldots, 2^{TR}\}$$

and a decoding function

$$\beta_T : \{1, 2, \ldots, 2^{TR}\} \rightarrow \hat{\chi}_T$$

that satisfy

$$D_T = E[d(X^T, \beta_T(\alpha_T(X^T)))]$$

then, a rate-distortion pair $(R, D)$ is called *achievable* if there exists a sequence of codes of rate $R$ such that

$$\lim_{T \to \infty} D_T \leq D$$

The rate distortion function $R(D)$ is the minimum rate such that $(R, D)$ is in the rate-distortion region. Conversely, the distortion rate function $D(R)$ is the minimum
distortion such that \((R,D)\) is in the rate-distortion region. These all depend on the source and distortion measure.

The boundary of a rate-distortion (RD) region is nearly impossible to determine by working directly from definitions. By construction, the RD function represents the limit of source coding and can be also determined by a constraint minimization as suggested in [40]. This problem has been solved for a few sources.

For a source that emits Gaussian random variables with variance \(\sigma^2\), the \(D(R)\) function subject to squared error distortion is [10]:

\[
D(R) = \sigma^2 2^{-2R} \tag{2.3}
\]

A graphical representation with \(\sigma^2 = 1\) is given in Fig. 2.3.
### 2.2 RD region for two descriptions

Referring to Fig. 2.1, given a distortion measure \( d \), we have three expected average distortions:

\[
D_i = E \left[ \frac{1}{T} \sum_{k=1}^{T} d_i(X_k, \hat{X}^{(i)}_k) \right] \text{ for } i = 0,1,2 \quad (2.4)
\]

The main problem is to determine the set of achievable values for the quintuple \((R_1, R_2, D_0, D_1, D_2)\). In general, the MD rate distortion region (for a particular source and distortion measure) is the closure of simultaneously achievable rates and distortions.

Side decoder \( i \) receives \( R_i \) bits per symbol and hence cannot have distortion less than \( D_i(R_i) \) where \( D_i \) is the distortion-rate function of the source for distortion measure \( d_i \). We obtain:

\[
\begin{align*}
D_0 & \geq D_0(R_1 + R_2) \quad (2.5) \\
D_1 & \geq D_1(R_1) \\
D_2 & \geq D_2(R_2)
\end{align*}
\]

Achieving equality simultaneously in bounds 2.5 would imply that an optimal rate \((R_1 + R_2)\) description can be partitioned into two optimal rates \( R_1 \) and \( R_2 \) descriptions. This is seldom possible because optimal individual descriptions at rates \( R_1 \) and \( R_2 \) are similar to each other and hence redundant when combined. Making descriptions individually good and yet not too similar is the fundamental trade-off of MD coding.

The set of achievable rates for a memoryless Gaussian source with variance \( \sigma^2 \) (with squared error distortion) has been obtained in [31]:

\[
\begin{align*}
D_i & \geq \sigma^2 2^{-2R_i} \text{ for } i = 1,2 \quad (2.6) \\
D_0 & \geq \sigma^2 2^{-2(R_1 + R_2)} \cdot \gamma_D(R_1, R_2, D_1, D_2) \quad (2.7)
\end{align*}
\]
where
\[ \gamma_D = \frac{1}{1 - (\sqrt{1 - D_1})(1 - D_2) - \sqrt{D_1 D_2 - 2^{-(R_1 + R_2)}}} \quad (2.8) \]

The term \( \gamma_D \) is the factor by which the central distortion must exceed the distortion-rate bound (2.6).

If the descriptions are individually very good, yielding \( D_1 = 2^{-2R_1} \) and \( D_2 = 2^{-2R_2} \), for \( R_1 \geq R_2 \) we have:
\[ \gamma_D = \frac{1}{1 - (1 - D_1)(1 - D_2)} \geq \frac{1}{2D_2} \quad (2.9) \]

that, using Eq. (2.7), yields:
\[ D_0 \geq D_1 D_2 \gamma_D \geq \frac{D_1}{2} \quad (2.10) \]

This means that, when the descriptions are individually good, the joint description is only slightly better than the best of the two.

We also note that, if the joint description is as good as possible, then \( \gamma_D = 1 \) and thus
\[ D_1 + D_2 = 1 + 2^{2(R_1 + R_2)} = 1 + D_0 \quad (2.11) \]

A distortion value of 1 is obtained with no information and so this expression implies a poor reconstruction for at least one of the side decoders. When one or both side distortions are large, \( \gamma_D = 1 \) and the central distortion can be low. Otherwise, there is a penalty in the central distortion.

Under the assumption \( R_1 = R_2 >> 1 \) and \( D_1 = D_2 = 2^{-2(1-\alpha)R_1} \) with \( 0 \leq \alpha \leq 1 \) (the balanced case), \( \gamma_D \) can be estimated as \( (4D_1)^{-1} \) [20] and, thus, \( D_0 \geq 2^{-4R_1(4D_1)^{-1}} \). Therefore, the product of central and side distortions is approximately lower bounded by:
\[ D_0 \cdot D_1 \geq \frac{1}{4} \sigma^2 2^{-4R_1} \quad (2.12) \]

**2.3 RD region for many descriptions**

While for a Gaussian source the complete MD rate-distortion region for two descriptions is known, the general case of \( N \) descriptions is an open research problem. The
first achievable rate region for general $N$ was presented in [50] where the rate region relies on a conventional conditional successive refinement framework and is in that sense a generalization of [18].

A different approach to obtain bounds for many descriptions has been proposed in [33] which has adopted the coding with side information refinement framework. There, it is shown that it is possible to encode a unit variance i.i.d. Gaussian source into $N$ descriptions with each description containing $R$ bits/sample such that the reconstruction fidelity with the reception of any $(k + r)$ descriptions (for $0 \leq r \leq N - k$) is given by:

$$D_{k+r} = \frac{k}{2^{2kR(k + r) - r}}$$  \hspace{1cm} (2.13)

When $r = 0$, $D_k = 2^{-2kR}$, which exactly attains the information theoretic optimal RD performance of the corresponding Gaussian source.

### 2.4 MD Techniques

The theoretical problem of determining the achievable rate/redundancy/distortion regions for MDC with a given source statistical model has been addressed by many authors [53, 18, 5, 54].

As for practical realizations, many techniques have been proposed to generate multiple descriptions. Most of these deal only with the 2-description case. In [49], for example, MD scalar quantization (MDSQ) was introduced to represent a single scalar random variable with two descriptions. A 2-description system can be seen as a central quantizer and an index assignment in 2 dimensions. The encoder produces, for each source scalar input, 2 quantization indices, each one sent over a channel. These indices can be seen as indices of a matrix. The goal is to fill the matrix so that the difference in all the rows/columns is as low as possible. This guarantees that if one description is lost, the receiver can reconstruct the best possible coarse version of the original value. Fig. 2.4 shows an example with 10 possible different values.
sent using two indexes from 1 to 4. If the transmitter wants to send the value 4, it
transmits on both channels the value 2. If only one of the two values is received,
the encoder is required to choose between three different possible values. Both the
encoder and the decoder have the index assignment matrix stored in memory.

The general scheme employing MD quantization followed by entropy coding has
been further developed in several papers, among which we can mention [39] and
references therein. In [24], a generalization is provided for MD with trellis codes.

MDSQ can be formally extended to vector quantization (MDVQ) where the MD
encoder receives many samples at once but the actual design and implementation
is significantly more complicated than MDSQ. This problem has been addressed
in [28] where a framework for optimization is proposed for an arbitrary number of
channels.

In [8], a multistage index assignment strategy is devised; however, besides being
suboptimal, this allocation strategy allows for generating only a dyadic number of
descriptions.

Recently, techniques based on transform coding [21] and coding with frames [23]
have emerged together with methods which employ pairwise correlating transform
operating on the coefficients in order to introduce redundancy among the descrip-
tions [52]; however, for more than two descriptions, grouping coefficients and allo-
cating redundancy is not trivial [22].

In [36], the authors propose an MDC scheme using redundant expansions by
partitioning a generic redundant dictionary. The most significant terms in the signal
representation are drawn from the partitions that better approximate the signal,
and distributed into the different descriptions, while the less important ones are
alternatively split between the descriptions. However, the allocation is empirical and the redundancy cannot be finely tuned.

In the perspective of generating an arbitrary number of balanced descriptions\(^1\) and controlling the trade off between central and side qualities, it is worth mentioning methods based on the polyphase decomposition followed by selective quantization [26] or the optimized rate-distortion MDC approach [48]. This latter algorithm represents a good choice for applications where the probability of description loss widely changes, such as in the wireless scenario.

A powerful technique to generate multiple descriptions and gain robustness to the loss of descriptions is called Unequal Error Protection (UEP). This technique consists in inserting in each description an optimal level of redundancy using channel codes to combat the unreliability of the network. If any descriptions are lost, a coarse version of the original information can be decoded with available descriptions using the features of channel codes. This technique can be generalized to deal with more than two descriptions as described in [29, 34]. The problem here is that, in order to yield at least the minimum level of quality, at least \(k_{\text{min}}\) descriptions out of \(N\) should be received, with \((k_{\text{min}},N)\) corresponding to the most powerful RS code employed.

Most of the mentioned methods are conceived as stand-alone co-decoding algorithms. Therefore, when applied to deliver multimedia contents, they are not compatible with standard codecs. In other words, it is not generally possible to configure classical MDC as pre- and post-processing stages to be combined with any co-decoding tool; this can represent a limitation, due to the widespread diffusion of standard tools, such as JPEG or JPEG 2000 [42].

\(^1\)Equal rate descriptions are said to be balanced if they yield the same average distortion when individually decoded [49].
Chapter 3

Network adaptive Multiple Description Coding for JPEG 2000

Given a sample vector $X$ of a zero mean uncorrelated process, a straight way to create $N$ descriptions starts generating two RD optimized streams $X_1$ and $X_2$, at rates $R_1$ and $R_2 < R_1$ and distortion $D(R_1)$ and $D(R_2)$ respectively, $D(r)$ being the RD function of the source data $X$. Then, samples are grouped into $N$ disjoint subsets $S_l$, $l = 1,..,N$, so that $X = \bigcup_{l=1}^{N} S_l$, and descriptions are finally built by combining the subsets from the two streams. For each subset, a description will contain its representation from $X_1$ whereas all the other descriptions will have it represented from $X_2$ (i.e. encoded at lower rate). In other words, if we define $S_{j,l}$ as the $l$-th subset belonging to stream $X_j$, $j = 1,2$, the first description is obtained by combining subset $S_{1,1}$, with $S_{2,l}$, $\forall l \neq 1$. In general, the $i$-th description is built by taking the coefficients of subset $S_{1,i}$ and those of the remaining sets from $X_2$: $S_{1,i} \bigcup \{\bigcup_{l=1,l\neq i}^{N} S_{2,l}\}$.

Let us define $\mathcal{R}(\cdot)$ and $\mathcal{D}(\cdot)$ as functions that determine the rate (in bits) and distortion$^1$ associated to data subsets, assuming (as common practice) an additive distortion model, the rate and distortion of the $i$-th description can be written as

\[ \mathcal{R}(S_{j,i}) = |S_i - S_{j,i}|^2 \]

\[ \mathcal{D}(S_{j,i}) = |S_i - S_{j,i}|^2 \]

\[ \mathcal{D}(r) = \text{squ} \]

---

$^1$Unless otherwise stated, the distortion is measured as the squared Euclidean distance:

$D(S_{j,i}) = |S_i - S_{j,i}|^2$
\[ R_i = R(S_{1,i}) + \sum_{l=1, l \neq i}^{N} R(S_{2,l}) \quad \text{and} \quad D_i = D(S_{1,i}) + \sum_{l=1, l \neq i}^{N} D(S_{2,l}) \]

respectively. Therefore, a sufficient condition to create \( N \) balanced descriptions is to identify subsets so that, if two subsets have the same rate, they yield the same distortion when used to recover the data:

\[ \forall l, m \in [1, N] : R(S_{j,l}) = R(S_{j,m}) \Rightarrow D(S_{j,l}) = D(S_{j,m}) \quad (3.1) \]

In this case, \( R(S_{j,l}) = R_j/N \) and \( D(S_{j,l}) = D(R_j)/N \) with \( l \in [1, N] \) and \( j = 1, 2 \). As a consequence, each description is encoded at approximately \( \frac{R_1 + (N-1)R_2}{N} \) bits, and distortion \( \frac{D(R_1) + (N-1)D(R_2)}{N} \), with a total output rate of

\[ R_t = R_1 + (N - 1)R_2 \quad (3.2) \]

Practical guidelines about how to match condition Eq. (3.1) will be discussed in Sec. 4.3 when dealing with JPEG 2000 encoded data.

With the outlined procedure, which avoids encoding any subset at high rate in any two descriptions, when a subset of \( k \) descriptions out of \( N \) is received, \( k \) subsets belong to the stream \( X_1 \) and the remaining \( (N - k) \) belong to \( X_2 \).

### 3.1 Two-rate approach for JPEG 2000

In order to build \( N \) balanced descriptions for a given image, we generate two different JPEG 2000 bitstreams of rates \( R_1 \) bits per pixel (bpp) and \( R_2 < R_1 \) bpp.

The codeblocks (CBs) generated after the wavelet decomposition of the encoding procedure [35] are grouped into \( N \) sets. These sets are built so that they have similar characteristics in terms of size and distortion contribution. In principle, to obtain \( N \) balanced descriptions, one should identify subsets of CBs with similar RD characteristics and allocate them to the descriptions. Nevertheless, we have verified that, if the number of CBs per subband is at least equal to \( N \), it is sufficient that each description contains the same number of CBs encoded at rate \( R_1 \) (and consequently at rate \( R_2 \)) for each subband.
The first description is obtained by combining the CBs of set 1 taken from the first stream, with the CBs of all the other sets taken from the second one. Analogously, the \(i\)-th description is built by taking the CBs of set \(i\) from the first stream and those of the remaining sets from the second stream.

This procedure generates an arbitrary number \(N\) of balanced descriptions fully backward compatible with Part 1 of the JPEG 2000 standard [48]. We assume that the total output rate is fixed by the application requirements and kept constant during the data transmission.

The decoder pre-processes all the received descriptions and selects, for each CB, the finest representation. Given that the procedure yields balanced descriptions, when a subset of \(k\) descriptions out of \(N\) is received, a side distortion \(D_{k,N}\) is obtained, representing a weighted average of the quality yielded by two bitstreams at rates \(R_1\) and \(R_2\). As the distortion of JPEG 2000 CBs can be assumed additive [35], we can write that:

\[
D_{k,N}(R_2) = \left(\frac{N - k}{N}\right)D(R_2) + kD(R - (N - 1)R_2)
\]

where \(1 \leq k \leq N\) and \(D(r)\) is the RD curve of the JPEG 2000 image encoded at rate \(r\). When all descriptions are received, the quality is a function of \(R_1\) only, because all CBs are taken from the bitstream coded at that rate. Therefore, the central distortion can be evaluated as \(D(R_1) = D(R - (N - 1)R_2)\). The extra rate, which amounts to \((N - 1)R_2\), is the redundancy of the MDC scheme. It is worth noticing that \(D_{k,N}(R_1,R_2)\) depends on the number of received descriptions and not on the specific loss pattern. Due to the bound on the overall rate \(R\), when all descriptions are received the redundancy impairs the RD performance of the system, but it becomes beneficial when a subset of descriptions is received. As a consequence, the redundancy must be carefully tuned according to the network conditions and this can be accomplished by modifying the fraction of total rate to be assigned to the stream at lower rate \(R_2\).
Usually, the redundancy is normalized with $R_{tot}$ leading to the expression:

$$\rho = \frac{R_2(M - 1)}{R_1 + R_2(M - 1)}$$

The parameter $\rho$ can vary between 0 and 0.5.

### 3.2 Theoretical analysis

Our goal is to maximize the quality at each receiver in terms of PSNR, according to the probability $p$ of a description being dropped or lost. The parameter $p$ depends on the overload in relay nodes that leads to packet dropping or delaying and to bit errors in the received packets as we are dealing with an unreliable physical link.

#### 3.2.1 Two-description case

In the two description case, the previously described procedure yields descriptions encoded at rate $(R_1 + R_2)/2$ bpp. According to Eq. (3.3), the side distortion is:

$$D_{1,2}(R_2) = \frac{D(R_2) + D(R - R_2)}{2} \quad (3.4)$$

Assuming that each description fits a single packet, packets are randomly dropped, and no Automatic Repeat reQuest (ARQ) is used, we can work out the expected end-to-end distortion for two descriptions:

$$\overline{D}_2(R_2) = (1 - p)^2 D(R - R_2) + 2p(1 - p)D_{1,2}(R_2) + p^2 \sigma^2 \quad (3.5)$$

where $p$ is the probability of packet loss$^2$ and $\sigma^2$ is the variance of the image. Using Eq. (3.4), we can write that:

$$\overline{D}_2(R_2) = (1 - p) [D(R - R_2) + pD(R_2)] + p^2 \sigma^2 \quad (3.6)$$

$^2$If a description does not fit a single packet, the probability $p$ should be replaced by the probability of description loss. The relationship between probability of packet loss and probability of description loss depends on the actual packetization scheme and description size.
The MDC scheme should be optimized selecting the value of redundancy $R_2$ that minimizes Eq. (3.6), given $p$ and the total rate $R$. This optimization task will be addressed in Sec. 3.2.2 in the general case of $N$ descriptions.

In order to make sensible evaluations of the effectiveness of the proposed two description scheme, we compare it with a single description coding (SDC) scheme where the source is encoded at rate $R = R_1 + R_2$ bpp and fit into two packets.

When the source image is single description encoded, the expected end-to-end distortion can be evaluated as:

$$
\overline{D}_1 = (1 - p)^2D(R) + p(1 - p)D_{pkt_1} + p(1 - p)D_{pkt_2} + p^2\sigma^2
$$

(3.7)

where $D_{pkt_i}$ is the distortion when only packet $i$ ($i = 1, 2$) is received and $\sigma^2 \geq D_{pkt_i} \geq D(R)$, $i = 1, 2$. If only one of the two packets (e.g. pkt$_i$) can be decoded independently of the other (e.g. pkt$_j$), then $D_{pkt_i} \geq D(R)$ and $D_{pkt_j} = \sigma^2$. This is a typical situation of layered coding, when the information contained in packet $j$ is usable if and only if packet $i$ has been properly received. In this situation, Eq. (3.7) is lower-bounded by $\overline{D}_{LB}$:

$$
\overline{D}_1 \geq (1 - p)D(R) + p\sigma^2 = \overline{D}_{LB}
$$

(3.8)

$\overline{D}_{LB}$ also represents the expected end-to-end distortion of the SDC scheme when all data fit a single packet, and assuming that $p$ is almost independent of the packet length (as inferred in [43]).

If both packets need to be received in order to decode the source, then $D_{pkt_i} = D_{pkt_j} = \sigma^2$ and the lower bound Eq. (3.8) still holds. If both packets can be independently decoded, then the system turns out to be an MDC one.

Assuming that the header of each description or packet requires a rate $R_h$, we can write the total available rate $R_{tot}$ as:

$$
R_{tot} = R + R_h = R_1 + R_2 + 2R_h
$$

(3.9)

In order to determine if MDC outperforms SDC, according to Eqs. (3.6) and (3.8)
we work out the condition under which $\mathcal{D}_{LB} - \mathcal{D}_2 > 0$ obtaining:

$$\frac{D(R - R_2) - D(R)}{\sigma^2 - D(R_2)} = f(R_2) < p$$

(3.10)

The optimal solution is then obtained selecting $R_2$ so that $f(R_2)$ is minimized. If it happens that $p_2 = \min_{R_2} f(R_2) \geq p$, then the SDC scheme always outperforms MDC.

The previous inequality means that, in order to decide whether it is better to transmit data as one or two descriptions, it is sufficient to estimate the probability of packet loss $p$ on the network and the RD function of the data to be transmitted; if $p$ is larger than the probability bound $p_2$, it is worth using MDC. The probability $p$ can be estimated or obtained using a feedback channel; as for the RD function, one can rely on the evaluated function in some of the state of the art encoder (such as JPEG 2000) or using the estimated curve at some operational points. Assuming that the RD function is approximated by a monotonic decreasing exponential function $D(r) = \sigma^2 e^{-br}$, where $b$ is a constant$^3$, we can write that:

$$\frac{\mathcal{D}_1 - \mathcal{D}_2}{(1 - p)\sigma^2} = e^{-bR} - e^{-b(R - R_2 - R_h)} + p(1 - e^{-bR_2})$$

(3.11)

Therefore, using two descriptions performs better than SDC when $\mathcal{D}_1 - \mathcal{D}_2 \geq 0$, which leads to the condition:

$$p \geq \frac{e^{-bR}(e^{bR_2 + R_h} - 1)}{1 - e^{-bR_2}}$$

(3.12)

When $R_2 \gg R_h$, e.g. headers are negligible in comparison with the introduced redundancy $R_2$, we can write that:

$$p \geq \frac{e^{-bR}(e^{bR_2} - 1)}{1 - e^{-bR_2}}$$

(3.13)

When $e^{-bR} \geq p$ the value $R_2$ that satisfies the previous inequality is $R_2 = 0$, which means that $\mathcal{D}_1 = \mathcal{D}_2$. For the values of $p \in (e^{-bR}, 1]$ the two description

$^3$For iid Gaussian source $b = 2\log 2$.
configuration is better than SDC. When \( R_2 \gg R_h \) we can summarize the obtained results in:

\[
\forall p \in [0,e^{-bR}] \text{ and } R_2 = 0 \Rightarrow \overline{D}_1 = \overline{D}_2
\]

\[
\forall p \in (e^{-bR},1] \Rightarrow \exists R_2 : \overline{D}_1 > \overline{D}_2
\]

which means that, for negligible headers in comparison with the inserted redundancy, the two descriptions configuration is better than SDC when \( p > e^{-bR} \).

3.2.2 Arbitrary number of descriptions

In this section we study the general case of an arbitrary number \( N \) of descriptions generated using the procedure of Sec. 3.2.1. We identify the optimal number of descriptions and the related redundancy according to the network conditions and the available resources. As stated in Sec. 3.2.1, assuming that each description fits a packet and that no ARQ is employed, we can work out the following expression of the expected end-to-end distortion:

\[
\overline{D}_N(R_2) = \sum_{k=1}^{N} \binom{N}{k} (1-p)^k p^{N-k} D_{k,N}(R_2) + p^N \sigma^2
\]

(3.15)

where \( D_{k,N}(R_2) \) is reported in Eq. (3.3). Taking into account that

\[
\sum_{k=1}^{N} \binom{N}{k} \frac{N-k}{N} \frac{N-k}{N} (1-p)^k p^{N-k} = p - p^N
\]

\[
\sum_{k=1}^{N} \binom{N}{k} \frac{k}{N} (1-p)^k p^{N-k} = 1 - p
\]

we can rewrite Eq. (3.15) as

\[
\overline{D}_N(R_2) = (p - p^N) D(R_2) + (1-p) D(R_1) + p^N \sigma^2
\]

(3.16)

where \( R_1 = R - (N-1)R_2 \). We want to find out how to optimally split the total available rate \( R \) between \( R_1 \) and \( R_2 \). To this end, we work out the minimum of
Eq. (3.16), given $N$, $p$ and $R$. Evaluating the derivative of Eq. (3.16), with respect to $R_2$, we obtain that:

$$\frac{\partial \tilde{D}_N(R_2)}{\partial R_2} = (p - p^N) \frac{\partial D(R_2)}{\partial R_2} + (1 - p) \frac{\partial D(R_1)}{\partial R_1} \frac{\partial R_1}{\partial R_2} \tag{3.17}$$

The minimum expected end-to-end distortion $(\tilde{D}_N)_{\text{min}}$ is then obtained from Eq. (3.16) selecting $R_2$ so that the following relationship is satisfied:

$$\frac{\partial D(R_2)}{\partial R_2} = (1 - p)(N - 1) \frac{\partial D(R_1)}{\partial R_1} = p - p^N \tag{3.18}$$

In order to find the optimal number of descriptions $N_{\text{opt}}$ from a given set of candidate values, an exhaustive search may be feasible if the set is not too large. In fact, using Eq. (3.16) and Eq. (3.18) the optimal rates $R_1$, $R_2$ and the minimum possible distortion for that $N$, can be analytically worked out without the need of actual co-decoding. In case exhaustive search cannot be afforded, the greedy approach sketched below has shown to yield satisfactory performance:

1. set $N = 2$.
2. initialize $M$ to a candidate value larger than $N$
3. $(\tilde{D}_N)_{\text{min}} \leq (\tilde{D}_M)_{\text{min}}$?
   3.a NO: set $N = M$; assign a larger candidate value to $M$ and goto 3
   3.b YES: $N_{\text{opt}} = N$; end

### 3.3 Experimental results

In [48], the 2-description case of the algorithm has been compared to state-of-the-art procedures for JPEG 2000. We now focus on the possibility of tuning redundancy and number of descriptions. The algorithm has been tested using the JPEG 2000 Verification Model 8.6 co-decoder and the test images Lenna, Goldhill and Elaine of
Table 3.1. Simulation and analytical bounds for images Lenna, Goldhill and Elaine at rates 0.4 and 1.2 bpp.

<table>
<thead>
<tr>
<th>Image</th>
<th>$R$</th>
<th>$p_2$ Simulation</th>
<th>$p_2$ Analytic</th>
<th>$p_1$ Simulation</th>
<th>$p_1$ Analytic</th>
<th>$p_8$ Simulation</th>
<th>$p_8$ Analytic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lenna</td>
<td>0.4</td>
<td>$4 \times 10^{-4}$</td>
<td>$5.5 \times 10^{-4}$</td>
<td>0.063</td>
<td>0.059</td>
<td>0.36</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>1.2</td>
<td>$1 \times 10^{-4}$</td>
<td>$0.7 \times 10^{-4}$</td>
<td>0.032</td>
<td>0.029</td>
<td>0.27</td>
<td>0.27</td>
</tr>
<tr>
<td>Goldhill</td>
<td>0.4</td>
<td>$0.8 \times 10^{-3}$</td>
<td>$1.6 \times 10^{-4}$</td>
<td>0.051</td>
<td>0.057</td>
<td>0.33</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td>1.2</td>
<td>$2.5 \times 10^{-4}$</td>
<td>$2.5 \times 10^{-4}$</td>
<td>0.032</td>
<td>0.036</td>
<td>0.27</td>
<td>0.29</td>
</tr>
<tr>
<td>Elaine</td>
<td>0.4</td>
<td>$3.5 \times 10^{-4}$</td>
<td>$4.5 \times 10^{-4}$</td>
<td>0.049</td>
<td>0.043</td>
<td>0.36</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td>1.2</td>
<td>$3 \times 10^{-4}$</td>
<td>$2 \times 10^{-4}$</td>
<td>0.032</td>
<td>0.032</td>
<td>0.26</td>
<td>0.24</td>
</tr>
</tbody>
</table>

dimension $512 \times 512$ pixels. The wavelet transform has been applied with 4 levels of decomposition and, for each image, the total output rate has been set to $R = 0.4$ and $R = 1.2$ bpp. SDC and MDC with $N = 2, 4, 8$ descriptions have been compared. Each description fits a packet as discussed in Sec. 3.2.1. The SDC performance bound is evaluated according to Eq. (3.8).

In Fig. 3.1, the end-to-end PSNR is reported versus $p$ for image Lenna. The curves have been worked out by simulation, given the total rate $R$, and selecting the best $R_2$ for each value of $p$. It can be noticed that, as $p$ increases, a higher number $N$ of descriptions leads to higher end-to-end PSNR. For example, at 0.4 bpp and $0.06 \leq p \leq 0.36$, the encoder should create 4 descriptions to ensure a gain up to 4 dB over the two-description configuration and up to 9 dB if compared to SDC. Then, for $p > 0.36$, the 8-description scheme outperforms those with a smaller number of descriptions. In fact, many descriptions allow the transmission to rely on many representations of the source, and the reception of a subset of them can be sufficient to reconstruct the original information with decent quality.

It can also be noticed that the scheme gets inefficient in the high redundancy region. This is due to the fact that most information is duplicated in this case. A possible solution to improve the algorithm performance could be the use of interleaved quantizers (MDSQ-like [49]) but, as a consequence, the descriptions would not be backward compatible with Part 1 of the JPEG 2000 any more. Fig. 3.1
also reports the bounds $p_j$ that delimit the regions where $j$ outperform $i$ descriptions, with $i < j$. Such boundary values (with possible exception of $p_2$, which can be analytically worked out using Eq. (3.10)) are obtained selecting the intersections between curves $D_i(R_2)$ and $D_j(R_2)$ of Eq. (3.16) with $R_2$ optimized using Eq. (3.18).

Fig. 3.2 focuses on low values of $p$, comparing SDC and MDC with $N = 2$. The probability $p_2$ derived from Eq. (3.10) is also reported, which represents the analytical lower bound of region where MDC is advantageous with respect to SDC. It can be noticed that the values of $p$ for which SDC outperforms MDC are extremely low. This is a consequence of the fact that, in our scheme, the redundancy can be finely tuned; this allows for an almost continuous transition between the single and multiple description cases. The SDC is nearly included in MDC scheme in the limit $R_2 \rightarrow 0$.

Table 3.1 reports, for the three test images, region bounds and intersections obtained from simulation. A very good match between such boundaries can be appreciated; the difference is mainly due to the approximations of the level of redundancy in the simulation chain. This implies that, once the probability of packet loss has been estimated, it is possible to recursively run the algorithm of Sec. 3.2.2 varying the parameter $N$ to get its optimal value in order to dynamically adapt the encoder to the network conditions.

Finally, we tested the performance sensitivity to estimated packet loss probability; in fact, this parameter may be difficult to precisely estimate, or it may change dynamically. Therefore, it is important to verify the effects of a possible mismatch between the actual $p_{\text{act}}$ and the estimated value $p$. Results are reported in Table 3.2. We analyzed the transmission of images Lenna and Goldhill at 1.2 bpp using two descriptions and evaluated the average quality (in terms of PSNR) when receiving one or two descriptions. As expected, if the actual loss rate $p_{\text{act}}$ matches $p$, the system yields the best results compared to other redundancy allocations. If $p$ is overestimated with respect to $p_{\text{act}}$, the introduced redundancy is too high and leads
to performance loss, increasing with the gap between the two. Similar considerations hold in case $p$ is underestimated. However, we can appreciate the fact that the performance impairment is limited.

Finally, we check the system when only one description out of $N$ is received, in order to evaluate performance when the receiver has no MDC-capabilities, and one of the generated descriptions, being fully backward compatible with Part 1 of the JPEG 2000 standard [48], is simply decoded. As Fig. 3.3 shows, for low values of $p$, if only one description is considered at the decoder, SDC yields better performance.
Figure 3.2. Expected end-to-end PSNR versus low values of probability of packet loss \( p \); image Lenna; (a) \( R = 0.4 \) bpp, (b) \( R = 1.2 \) bpp.

Figure 3.3. Simulated end-to-end PSNR for JPEG 2000 decoders with and without MDC-capabilities for 2,4,8 descriptions.

When \( p \) increases, the optimal PSNR is obtained with higher level of redundancy and then even decoding only one description leads to higher quality than SDC.
Chapter 4

A flexible MD multi-rate allocation scheme

The main drawback of the MDC scheme proposed in Chapter 3 is that, when a new description is received, the quality of only one set of coefficients is improved, as the decoder already has a representation at rate larger than or equal to $R_2$ of all other sets. Therefore, this scheme implies a high amount of unused redundancy. To overcome this limitation, we propose a multi-rate allocation scheme making use of $N$ different rates to encode the source data.

4.1 Multi-rate allocation scheme

Let us assume that the encoder generates $N$ bitstreams, each of which is encoded at a different rate according to a rate vector $\mathbf{R} = \{R_j; j = 1,...,N\}$, $R_1 \geq R_2 \geq R_3... \geq R_N$ and $R_t = \sum_{j=1}^{N} R_j$. Then, let $N$ different subsets be selected as in Chapter 3. $N$ descriptions may be generated by combining the subsets from the $N$ streams following a rate allocation pattern based on a circular-shift of the vector $\mathbf{R}$.

In Tab. 4.1, an allocation example is reported for $N = 4$ descriptions. The rate allocation patterns generated by the left-circular-shift of $\mathbf{R} = \{R_1,R_2,R_3,R_4\}$ are reported in terms of rates devoted to the $l$-th subset $S_l$, whereas $F_k(R_j)$ represents
Table 4.1. A first example of multi-rate allocation pattern for $N = 4$ descriptions.

<table>
<thead>
<tr>
<th>$k$</th>
<th>received des.</th>
<th>$S_1$</th>
<th>$S_2$</th>
<th>$S_3$</th>
<th>$S_4$</th>
<th>$F_k(R_1)$</th>
<th>$F_k(R_2)$</th>
<th>$F_k(R_3)$</th>
<th>$F_k(R_4)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1/4</td>
<td>1/4</td>
<td>1/4</td>
<td>1/4</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>$R_1$</td>
<td>$R_2$</td>
<td>$R_3$</td>
<td>$R_4$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>$R_2$</td>
<td>$R_3$</td>
<td>$R_4$</td>
<td>$R_1$</td>
<td>1/4</td>
<td>1/4</td>
<td>1/4</td>
<td>1/4</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>$R_3$</td>
<td>$R_4$</td>
<td>$R_1$</td>
<td>$R_2$</td>
<td>1/4</td>
<td>1/4</td>
<td>1/4</td>
<td>1/4</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>$R_4$</td>
<td>$R_1$</td>
<td>$R_2$</td>
<td>$R_3$</td>
<td>1/4</td>
<td>1/4</td>
<td>1/4</td>
<td>1/4</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1/2</td>
<td>1/4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1,2</td>
<td>$R_1$</td>
<td>$R_2$</td>
<td>$R_3$</td>
<td>$R_1$</td>
<td>1/2</td>
<td>1/4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1,3</td>
<td>$R_1$</td>
<td>$R_2$</td>
<td>$R_1$</td>
<td>$R_2$</td>
<td>1/2</td>
<td>1/2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1,4</td>
<td>$R_1$</td>
<td>$R_1$</td>
<td>$R_2$</td>
<td>$R_3$</td>
<td>1/2</td>
<td>1/4</td>
<td>1/4</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2,3</td>
<td>$R_2$</td>
<td>$R_3$</td>
<td>$R_1$</td>
<td>$R_1$</td>
<td>1/2</td>
<td>1/4</td>
<td>1/4</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2,4</td>
<td>$R_2$</td>
<td>$R_1$</td>
<td>$R_2$</td>
<td>$R_1$</td>
<td>1/2</td>
<td>1/2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>3,4</td>
<td>$R_3$</td>
<td>$R_1$</td>
<td>$R_1$</td>
<td>$R_2$</td>
<td>1/2</td>
<td>1/4</td>
<td>1/4</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3/4</td>
<td>1/4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1,2,3</td>
<td>$R_1$</td>
<td>$R_2$</td>
<td>$R_1$</td>
<td>$R_1$</td>
<td>3/4</td>
<td>1/4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1,2,4</td>
<td>$R_1$</td>
<td>$R_1$</td>
<td>$R_2$</td>
<td>$R_1$</td>
<td>3/4</td>
<td>1/4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1,3,4</td>
<td>$R_1$</td>
<td>$R_1$</td>
<td>$R_1$</td>
<td>$R_2$</td>
<td>3/4</td>
<td>1/4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2,3,4</td>
<td>$R_2$</td>
<td>$R_1$</td>
<td>$R_1$</td>
<td>$R_1$</td>
<td>3/4</td>
<td>1/4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1,2,3,4</td>
<td>$R_1$</td>
<td>$R_1$</td>
<td>$R_1$</td>
<td>$R_1$</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

the statistical frequency of subsets encoded at rate $R_j$ when $k$ descriptions are received. The case $k = 1$ represents both the allocation scheme for each description and the rate of each subset when decoding only that description.

At the decoder side, all the received descriptions are preprocessed so as to select the finest representation of each subset. When all packets are received ($k = 4$), the quality is only a function of $R_1$, because the decoder selects all subsets encoded at rate $R_1$. When $k$ descriptions out of $N$ are received, a side distortion $D_{k,N}(R)$ is obtained, that is a weighted average of the distortions yielded by the $N$ bitstreams encoded at rates $R$:

$$D_{k,N}(R) = \sum_{j=1}^{N} F_k(R_j) D(R_j)$$ (4.1)

Since no subset is encoded at the same quality in any two different descriptions, any received description can improve the quality of at least one subset.

With this allocation procedure, the quality yielded by any subset of $k$ received...
descriptions may be unbalanced. In fact, \( F_k(R_j) \) depends on the specific received pattern and not only on \( k \). As an example, referring to Tab. 4.1, if descriptions 1 and 4 are received and merged, then the first 2 subsets will be available at high rate while the remaining ones will be represented at lower rates (\( R_2, R_3 \)). On the other hand, when descriptions 1 and 3 are received, subsets are coded at rates \( R_1 \) and \( R_2 \), and no subset is coded at rates \( R_3 \) and \( R_4 \) thus yielding a higher quality. In order to guarantee fairness among the different users we do not want the quality at the receiver to depend on the particular set of received descriptions.

Therefore, the procedure may be modified, so that all the possible permutations of the \( N \) rates in vector \( \mathbf{R} \) are used in order to generate balanced descriptions. To this end, \( N! \) different subsets are identified as discussed in Chapter 3, and all the \( N! \) permutations of \( \mathbf{R} \) are applied to the \( N! \) subsets to create the descriptions. Subset \( S_1 \) is encoded along descriptions using a permutation \( \pi_1 \) of rates \( \mathbf{R} \), and so on for subset \( S_2 \) (\( \pi_2 \)), \ldots, \( S_N! \) (\( \pi(N!) \)). As an example, in Tab. 4.2 the rate allocation pattern for 4 descriptions is reported, along with the decoding patterns for some of the possible reception cases. As shown, subset \( S_1 \) is encoded at rate \( R_1 \) in the first description, and at rates \( R_2, R_3 \) and \( R_4 \) in the second, third and forth description respectively. The permutation of vector \( \mathbf{R} \) allows to obtain balanced descriptions in terms of both rate and distortion. In fact, the contribution of each description to the
end quality is the same, regardless of the particular received description. We refer to
this scheme as the multi-rate allocation strategy. In Appendix 4.5, we demonstrate
that, for this rate allocation procedure, $F_k(R_j)$ can be evaluated as

$$ F_k(R_j) = \begin{cases} 
\frac{k(N-j)!(N-k)!}{(N-k-j+1)!N!} & \text{for } j \leq N - k + 1 \\
0 & \text{otherwise}
\end{cases} \quad (4.2) $$

Using Eq. (4.1), the side distortion $D_{k,N}(R)$ can be worked out as:

$$ D_{k,N}(R) = \frac{k(N-k)!}{N!} \sum_{j=1}^{N-k+1} \frac{(N-j)!}{(N-k-j+1)!} D(R_j) \quad (4.3) $$

This expression allows one to identify the optimal rate vector $R$ for a given transmis-
sion scenario. In fact, Eq. (4.3) can be minimized according to particular constraints,
such as the maximum allowed end-to-end distortion or the mean quality required
by users; this will be discussed in detail in Sec. 4.2.

It is worth noticing that the two-rate scheme of Chapter 3 is a special case of
the $N$-rate scheme where $R_j = R_2, \forall j \in \{2,N\}$.

4.2 Redundancy optimization

The redundancy of an MD coding scheme can be tuned in different ways according to
the network scenario at hand. Here, we focus on two particular situations. In Case
I, the goal is to guarantee a minimum end-user quality. In Case II, the objective is
to minimize the end-to-end mean squared error (MSE) among the different users.

4.2.1 Case I: Bounded worst-case end-user quality

In this scenario, one should guarantee a minimum side quality, corresponding to
the reception of a single description. This could be the case of many users access-
ing the same information, but equipped with different terminals and/or available
bandwidth. In this case, each user should be guaranteed to obtain at least a coarse
resolution version of the source; then, in case it has more bandwidth resources, quality may be increased. In this context, the effectiveness of the MD scheme can be evaluated in terms of the central quality achieved while satisfying the constraint on the minimum side quality. In the following, the dependency of the distortion terms on the rate vector $\mathbf{R}$ is omitted for the sake of brevity.

The objective of allocating $\mathbf{R}$ for this problem can be formalized as the minimization of the Lagrangian cost function:

$$J = D_{N,N} + \lambda_1 (D_{1,N} - \overline{D}_{1,N}) + \lambda_2 \left( \sum_{j=1}^{N} R_j - R_t \right)$$

where $\lambda_1, \lambda_2$ represent Lagrange multipliers, $D_{N,N} = D(R_1)$ is the central distortion and $R_t$ is the total rate available for transmission (assumed to be known and imposed by the application). The side distortion $D_{1,N}$ of Eq. (4.1) can be written as:

$$D_{1,N} = \sum_{j=1}^{N} F_1(R_j) D(R_j) = \frac{1}{N} \sum_{j=1}^{N} D(R_j)$$

which can be reworked out as:

$$D_{1,N} = \frac{D_{N,N}}{N} + \frac{1}{N} \sum_{j=2}^{N} D(R_j)$$

The necessary condition of optimality [41] requires that $\nabla J = 0$ (i.e. $\frac{\partial J}{\partial R_j} = 0$), meaning that:

$$\begin{cases}
\frac{\partial D(R_1)}{\partial R_1} + \lambda_1 \frac{1}{N} \frac{\partial D(R_1)}{\partial R_1} + \lambda_2 &= 0 \\
\lambda_1 \frac{1}{N} \frac{\partial D(R_j)}{\partial R_j} + \lambda_2 &= 0, \text{ for } j \in [2,N]
\end{cases} \quad (4.4)$$

As a consequence:

$$\frac{\partial D(R_j)}{\partial R_j} = \begin{cases}
-\frac{N\lambda_2}{N+\lambda_1} & \text{for } j = 1 \\
-\frac{N\lambda_2}{\lambda_1} & \text{for } j \in [2,N]
\end{cases} \quad (4.5)$$

Since the rate vector $\mathbf{R}$ is applied on $N!$ subsets of equal RD characteristics, the equal-slope condition for $j \in [2,N]$ yields to $R_j = R_2$ for $j \in [2,N]$. We can conclude
that for this transmission scenario:

\[
\begin{align*}
D_{1,N} &= \frac{D(R_1)}{N} + \frac{N-1}{N} D \left( \frac{R_t - R_1}{N-1} \right) \quad \text{for } R_t/N \leq R_1 \leq R_t \\
D_{N,N} &= D(R_1)
\end{align*}
\]

What we have obtained is equivalent to the two-rate scheme of Chapter 3. This implies that the two-rate allocation is well suited to transmission scenarios with bounded worst-case end-user quality.

### 4.2.2 Case II: Maximization of average end-to-end quality

In this scenario, given the total available transmission rate \( R_t \) and the probability of packet loss, one needs to evaluate the rate allocation that minimizes the end-to-end MSE. Furthermore, testing different number of descriptions, one can obtain the best configuration settings in terms of number of descriptions for a given transmission scenario. This could be, for instance, the goal of a service provider, which wants to guarantee fairness among users, in not critical applications when a non zero outage probability can be tolerated.

Let \( p \) be the probability of description loss. It can be easily related to the size of the Maximum Transfer Unit (MTU) used at the physical layer, the probability of MTU (packet) loss \( p_m \) and description size. This relationship is important, as it allows to enter the physical network parameters (MTU size and loss probability) into the MDC optimization procedure, which is related to the probability of description loss as explained in the following. In fact, assuming that the correct reception of a description requires that all the MTUs composing that description are received, then

\[
p = 1 - (1 - p_m)^{\alpha/N}
\]

with \( \alpha \geq N \) being the minimum number of MTUs required to transmit the whole data. When \( \alpha < N \) (i.e. the description size is less than one MTU), we can assume that \( p = p_m \).
Rate allocation

Assuming that descriptions are randomly dropped, and no Automatic Repeat re-Quest (ARQ) is used, the expected end-to-end distortion for the $N$-description case can be worked out as:

$$
\overline{D}_N = \sum_{k=1}^{N} \binom{N}{k} (1 - p)^k p^{N-k} D_{k,N} + p^N \varepsilon
$$

(4.7)

where $\varepsilon$ is the distortion when no description is received. By using (4.3) and with some algebraic manipulation, the following expression can be obtained:

$$
\overline{D}_N = (1 - p) \sum_{j=1}^{N} p^{j-1} D(R_j) + p^N \varepsilon
$$

(4.8)

We want to find out how to optimally distribute the total available rate $R_t$ over $\mathbf{R}$ given $N$ and $p$, so as to minimize the expected distortion. To this end, we define the Lagrangian cost function:

$$
J = \overline{D}_N + \lambda \left[ \sum_{j=1}^{N} R_j - R_t \right]
$$

where $\lambda$ represents the Lagrange multiplier. The necessary condition of optimality [41] requires that $\nabla J = 0$, meaning that:

$$
\frac{\partial J}{\partial R_j} = (1 - p)p^{j-1} \frac{\partial D(R_j)}{\partial R_j} + \lambda = 0
$$

(4.9)

This leads to

$$
\frac{\partial D(R_j)}{\partial R_j} = \frac{-\lambda}{(1 - p)p^{j-1}}
$$

(4.10)

Considering two rates $R_j$ and $R_m$ from vector $\mathbf{R}$, the minimum expected end-to-end distortion $(\overline{D}_N)_{\text{min}}$ is obtained selecting such rates so that:

$$
p^{j-m} = \frac{\frac{\partial D(R_m)}{\partial R_m}}{\frac{\partial D(R_j)}{\partial R_j}}
$$

(4.11)
Equation (4.11) implies that the ratio between the RD slopes of a subset coded at rate $R_j$ and a subset coded at coarser rate $R_{j+1}$, should be equal to the probability of losing the description. This result is particularly significant, since it establishes a relationship between the description loss probability (straightforwardly related to the probability of packet loss $p_m$) and the optimal RD slopes. Moreover, it indicates that, for $N$ descriptions, the optimal performance is obtained when the rates in vector $\mathbf{R}$ are in general different. In order to get more insight in this formula, in the following we analyze some particular limit situations.

- **Low failure probability:** When $p \to 0$ and $R_t$ is less than the lossless coding rate, Eq. (4.11) leads to $\frac{\partial D(R_j)}{\partial R_j} \to -\infty$ and consequently $R_j = 0$ for $j \in \{2,N\}$. This means that all the available rate is allocated to the finest coded set ($R_1 = R_t$). Actually, when $p \to 0$, the probability of receiving any description is extremely high; therefore, the redundant part should be minimal.

- **High failure probability:** We analyze the case $p \to 1$, which leads to $\frac{\partial D(R_m)}{\partial R_m} \approx \frac{\partial D(R_j)}{\partial R_j}$. This means that the descriptions should be “similar”, so as to deliver a maximum amount of information in case some of them get lost, which is very likely in this situation.

Given the total rate $R_t$ and the probability of packet loss $p_m$, the MD scheme should be optimized selecting the rate vector $\mathbf{R}$ which satisfies Eq. (4.11). This objective can be fulfilled by the following procedure:
**Procedure 1:** Proposed redundancy allocation procedure

**Given** \( \epsilon > 0 \) (numerical tolerance)

**Given** \( \Delta R > 0 \) (rate accuracy)

**Given** \( p_m, R_t \) and \( N \)

work out \( p \) from \( p_m \) using Eq. (4.6)

\[ R_s := 0 \]

repeat

\[ R_N := R_s \]

for \( j = N - 1 \) downto 1 do

evaluate \( R_j \) using Eq. (4.11)

end for

\[ R_s = R_s + \Delta R \]

until \( \left| R_t - \sum_{j=1}^{N} R_j \right| < \epsilon \)

The parameters \( \epsilon \) and \( \Delta R \) determine the convergence speed of the algorithm.

**Optimal number of descriptions**

In this section we propose a procedure to identify the optimal number of descriptions \( (N_{opt}) \) and the related rate vector \( (R_{opt}) \) in order to minimize the expected distortion according to the network conditions and the available resources. To achieve this objective, we devise this procedure:
### Procedure 2: Redundancy and $N$ allocation procedure

| Given $R_t$, $N_{\text{max}}$ and $p_m$  
| estimate the data RD curve $D(R)$  
| $N := 1$  
| work out $p$ from $p_m$ using Eq. (4.6)  
| $R(1) = \{R_t\}$  
| evaluate $D_1$ using Eq. (4.8)  
| while ($N \leq N_{\text{max}}$)  
| $N := N + 1$  
| work out $p$ from $p_m$ using Eq. (4.6)  
| evaluate $R(N)$ using Procedure 1  
| evaluate $D_N$ using Eq. (4.8)  
| repeat  
| $N_{\text{opt}} := \arg \min_N D_N$  
| $R_{\text{opt}} := R(N_{\text{opt}})$  

### 4.3 Multi-rate MD JPEG 2000 codec

In the following, the ideas discussed in the previous sections are applied to a practical system using JPEG 2000 as the main compression engine. In JPEG 2000, the image is first DWT transformed. Then, the generated coefficients are quantized by a high rate quantizer, and the coefficients of each subband are divided into non-overlapping rectangular areas called codeblocks (CBs), these latter are bit-plane encoded in the Tier-1 module. The bitstream is organized by the rate allocator (Tier-2 module) into a sequence of layers, each layer containing contributions from each CB. The block truncation points associated with each layer are optimized in the RD sense [2]. The proposed rate allocation MD scheme is well suited to such rate allocation procedure, and is fully backward compatible with Part 1 of the JPEG 2000 standard [48].

In order to implement the MDC scheme, the Tier-2 module of the JPEG 2000 encoder is used to generate $N$ JPEG 2000 bitstreams, each of which is coded at a
different rate according to the array $\mathbf{R}$ of rates $\{R_j; j = 1,\ldots,N\}$, $R_1 \geq R_2 \geq R_3 \ldots \geq R_N$ (from now on, $R_j$ is expressed in bits per pixel -bpp). Rates are chosen so as to satisfy the constraints relevant to the application at hand (e.g. the MSE ...).

Let us discuss how to identify the subsets used to blend the descriptions. As shown in Fig. 4.1, for each level of wavelet decomposition we assume that each detail subband is made of CBs having roughly similar RD characteristics. So, we apply the procedure sketched in Sec. 4.1 to these groups. CBs are sequentially put into descriptions according to the allocation table reported in Tab. 4.2. Such a procedure is heuristic, but yields good results especially in the practical case when the CBs are numerous with respect to $N$. However, it is worth noticing that, for those subbands where the number of CBs is less than $N!$, the algorithm does not guarantee perfectly balanced descriptions. This issue will be further discussed when simulation results are presented.

At the decoder side, if all the descriptions are received, they are pre-processed and merged into a single bitstream, where, for each CB, the best representation is selected\(^1\); the resulting stream is then JPEG 2000 decoded. On the other hand, if some descriptions are lost, the received ones are merged and JPEG 2000 decoded, yielding inferior quality. Similarly, a standard JPEG 2000 decoder is still able to decode any single description.

### 4.3.1 Complexity

The proposed encoding procedure can be obtained by modifying the rate allocation and Tier-2 modules of a standard JPEG 2000 encoder. In fact, after the wavelet decomposition, $N$ RD-optimized streams of the transformed image are generated, and then subsets are combined from such streams. The DWT and Tier-1 stages are evaluated only once for each image (as in a standard JPEG 2000 encoder), so that no modification to these stages is required.

\(^1\)The best CB representation can be identified by simply determining the CB length.
Table 4.3. Encoding and decoding times of Lenna and Women images (ε = negligible).

<table>
<thead>
<tr>
<th></th>
<th>Enc.</th>
<th>Dec.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DWT Tier-1 Rate allocation Tier-2</td>
<td>DWT Tier-1 Rate allocation Tier-2</td>
</tr>
<tr>
<td>Image</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lenna (512 × 512)</td>
<td>12.50 87.50 ε</td>
<td>2.65 97.35 ε</td>
</tr>
<tr>
<td>Women (2048 × 2560)</td>
<td>48.40 51.50 0.10 ε</td>
<td>12.60 87.40 ε</td>
</tr>
</tbody>
</table>

At the decoder side, the MD decoder requires $N$ Tier-2 decoding modules, followed by a merging stage that selects, for each CB, the best representation available. Then, as in a standard JPEG 2000 decoder, a Tier-1 decoding module performs entropy decoding and inverse quantization, followed by the inverse DWT.

In Tab. 4.3, the encoding and decoding times of images Lenna (512 × 512) and Women (2048 × 2560) are reported, focusing on the DWT, Tier-1, rate allocation, and Tier-2 percentage times, on a 1.8 GHz Pentium 4 machine with 1 GB of RAM. As shown, the most expensive procedure is Tier-1, followed by DWT. Both are executed only once (as in a standard JPEG 2000 encoder) regardless of the number of descriptions. The rate-allocation and Tier-2 modules are performed $N$ times, but, as shown, their contribution to the overall system complexity is negligible.

### 4.4 Experimental results

#### 4.4.1 Gaussian source

In this section, we study the general case of an arbitrary number $N$ of descriptions generated using the rate allocation procedure sketched in Sec. 4.1. We identify the optimal number of descriptions and the related redundancy according to the packet loss probability $p_m$ and the total rate $R_t$. Let us assume that the RD function is

\footnote{Another possible implementation (sequential configuration) employs a single Tier-2 decoding module with sequential decoding of the received descriptions.}
$D(R) = \sigma^2 2^{-2R}$, where $\sigma^2$ is the source variance. Actually this is the RD bound of an iid Gaussian source. In this case, Eq. (4.11) leads to:

$$R_j = \frac{j - m}{2} \log_2 p + R_m, \quad m, j \in [1,N]$$

(4.12)

where the probability of description loss $p$ is related to $p_m$ by means of Eq. (4.6). Taking into account that $\sum_{j=1}^N R_j = R_t$ and $R_j \geq R_{j+1} \geq 0; \forall j$ and combining Eq. (4.8) and Eq. (4.12) we obtain:

$$\left(\frac{D_N}{\sigma^2}\right)_{\min} = L(1-p)p^{\frac{L-1}{2}}2^{-2R_t/L} + p^L$$

(4.13)

where $L = \min \left( N, \left[ \frac{1+\sqrt{1-16R_t/\log_2 p}}{2} \right] \right)$. It is worth pointing out that $R_j = 0; \forall j \in [L+1,N]$. In order to analyze the expected distortion as a function of the number of descriptions and the probability of packet loss as in Sec. 4.2.2, Fig. 4.2 represents SNR = $10 \log \left( \frac{\sigma^2}{(D_N)_{\min}} \right)$ as a function of $p_m$ and $N$.

In order to properly evaluate the effect of MTU size, different sizes were considered, leading to (a) $\alpha = 3$ and (b) $\alpha = 5$. In both cases, results reveal that the number of descriptions to use should be larger than or equal to $\alpha$. In fact, with this configuration (i.e. $N = \alpha$), each description fits an MTU. As a consequence, losing one MTU causes the loss of the corresponding description only, and the remaining ones will still be merged together. On the other hand, if $N < \alpha$, many MTUs are required to transmit one description. As a consequence, if one of these MTUs is lost, the whole description becomes useless and should be discarded. For instance, when $\alpha = 3$ and $p_m = 0.1$, the two-description scheme yields SNR=12.06 dB, whereas, with three or more descriptions, the performance saturates at SNR=13.41 dB. The saturation of the SNR value is due to the fact that no header information has been considered both for MTUs and descriptions. In real systems, one should take into account an increasing amount of overhead data in the same budget of total output rate.

In Fig. 4.3 the system RD performance is compared to the $N$-description bound for a Gaussian source [51] and $N = 3$. The best central distortion performance is
represented as a function of the rate devoted to each description \( R_d = R_i/N \), and several values of side distortion \( D_{1,N} \) given by the \( N \)-description bound. For the sake of comparison, we also report the best achievable performance of the proposed method, assuming that a quantizer can reach the RD bound for a Gaussian source. By analyzing the Gaussian case, we can conclude that in the low redundancy region (that is, with high values for \( D_{1,N} \)), the proposed rate allocation algorithm approaches the \( N \)-description bound, whereas in the high redundancy region (i.e. low \( D_{1,N} \) values) the gap tends to increase. For example, for \( R_i = 2.1 \) bps, the gap is nearly 3 dB for \( D_{1,N} = -4 \) dB, and 0.7 dB for \( D_{1,N} = -2 \) dB. This can be explained noticing that the redundant part encoded at lower rate is neglected when \( k \) descriptions are received; therefore, for small redundancy values, the algorithm approaches the bound. The neglected part could be exploited to correct those descriptions which are received affected by bit errors, so achieving a larger overall efficiency as shown is [46], this subject is left to future research.

### 4.4.2 JPEG 2000-based MD codec

The goal of the experiments reported in this section is to validate the proposed MDC scheme applied to the JPEG 2000 codec, and to provide comparisons with similar schemes in [44]. For this set of simulations, we used the JPEG 2000 codec engine from the OpenJPEG libraries [30], generalized to support the encoding/decoding procedure of an arbitrary number of descriptions. We considered Lenna and Goldhill images of dimension 512 × 512 pixels as test images, the first one being a smooth image whereas the second one being very detailed. The DWT has been applied with 4 levels of decomposition and the total output rate has been set to \( R = 0.4 \) and \( R = 1.2 \) bpp. First of all, we validate the allocation procedure in order to verify that the generated descriptions are balanced in terms of PSNR and rate. To this end, we use the multi-rate allocation scheme proposed in Sec. 4.1 in order to generate four descriptions optimized in terms of average expected PSNR. The maximum and minimum PSNR achieved when \( k \) descriptions are received are reported versus \( p \).
in Fig. 4.4.a. As \( p \) is related to \( p_m \) and the MTU size, from these curves one is able to obtain the actual performance of the scheme for any \( p_m \) and MTU size, by simply working out the corresponding value of \( p \) using Eq. (4.6). The total coding rate \( R_t \) is 0.4 bpp and the codeblock size is \( 64 \times 64 \) pixels. We can notice that the gap between the maximum and minimum PSNR decreases with \( p \). This is due to the fact that the redundancy increases with \( p \), and consequently the four intermediate bitstreams are more similar, yielding more balanced descriptions. When the redundancy decreases, the four bitstreams become unbalanced and the merging of codeblocks into a description becomes unfair from a RD point of view. With this configuration, the size of the DWT LL band is \( 32 \times 32 \), so it happens that only one description contains the LL band coded at high quality. In the case of single-description reception (curves marked with circles), if the one that contains the finely represented LL band is received, then the maximum quality is obtained; otherwise, a coarse representation is given, and this explains the significant performance gap at low redundancy levels.

This effect can be mitigated by reducing the size of the CBs. For example, if descriptions are coded using \( 16 \times 16 \) CBs, then each description contains a high quality version of a segment of the LL band. The results can be appreciated in Fig. 4.4.b. As a side effect of generating smaller CBs, each description carries more header information. This slightly decreases performance, as can be seen at \( p = 0.25 \) and 4-description reception. When the CB size is \( 64 \times 64 \), the PSNR is approximately 0.5 dB higher than in the \( 16 \times 16 \) case.

It is also important to notice that, as \( p \) increases, the central quality gets impaired (curve marked with small dots) since redundancy increases with \( p \). Similarly, the side quality gets improved (in particular for one and two descriptions) because CBs are better encoded.

The fact that the CB size affects the possibility of creating well-balanced descriptions in terms of rate can also be appreciated in Fig. 4.5. Fig. 4.5.a refers to \( R_t = 0.4 \) bpp, and reports the minimum and maximum relative rate deviation of the
descriptions $\eta = (\overline{R}_d - R_d)/R_d$, where $R_d = R_t/N$ is the theoretic description rate and $\overline{R}_d$ is the actual description rate. As $p$ and redundancy decrease, descriptions tend to be unbalanced. For example, at $p = 0.05$ and CB size $64 \times 64$, the largest description is 30% larger than expected. The gap is reduced to 13% when using $16 \times 16$ CBs. This behaviour is even more visible when increasing the total output rate to $R_t = 1.2$ bpp (thus allocating 0.3 bpp to each description), as shown in Fig. 4.5.b. At $p = 0.05$ the size of a description can be 45% larger than expected for big CB size, while it is approximately 10% larger for small CB size. In both plots, the gap between the largest and the smallest description decreases with $p$, because redundancy increases and then the four intermediate bitstreams become similar.

Further experiments aim at comparing the multi-rate allocation of Sec. 4.1 with the two-rate allocation described in Chapter 3. Comparisons are made in terms of the normalized rate allocation of the intermediate bitstreams $\left(\frac{R_j}{R_t}\right)$. Fig. 4.6.a and Fig. 4.6.b, referring to $R_t = 0.4$ bpp and $R_t = 1.2$ bpp respectively for Lenna image, show that, when $p$ increases, the encoder requires more redundancy to face these losses. This affects the allocation scheme, resulting in a reduction of the rate devoted to the highest quality, $R_1$, and at the same time, increasing other rate values. The behaviour is similar for $R_t = 0.4$ bpp and $R_t = 1.2$ bpp. The multi-rate scheme, however, relies on more granularity to allocate rates to the intermediate bitstreams. Similar considerations can be made from Goldhill curves reported in Fig. 4.6.c and Fig. 4.6.d for $R_t = 0.4$ bpp and $R_t = 1.2$ bpp respectively.

We now focus on the quality achieved when only a subset of descriptions is available at the receiver, both for multi-rate and two-rate allocations. In Fig. 4.7.a ($R_t = 0.4$ bpp) and 4.7.b ($R_t = 1.2$ bpp), the average PSNR obtained at the receiver side is plotted for each probability of description loss up to $p = 0.5$ for Lenna image. All possible configurations of received descriptions for decoding are considered. We can notice that, for a subset of $2, 3, 4$ descriptions, the multi-rate scheme yields better performance in terms of PSNR. The gain is particularly significant when only 2 or 3 descriptions are decoded and for several values of $p$. It is also worth noticing that,
when the decoder relies on only one description, the two-rate scheme outperforms the multi-rate algorithm. This is mainly due to the fact that, as previously discussed, the two-rate scheme inserts a higher amount of redundancy (in terms of encoding rate of the low-quality bitstream) and this redundancy is then successfully exploited at receiver-side. Similar considerations can be made from Goldhill curves.

Finally, in Fig. 4.8.a ($R_t = 0.4$ bpp) and 4.8.b ($R_t = 1.2$ bpp), the multi-rate allocation scheme is compared to the two-rate procedure that minimizes the end-to-end distortion, in terms of the expected PSNR versus the probability $p$ of description loss. We can notice that the multi-rate scheme outperforms the two-rate scheme by approximately 0.6 dB, for both total output rates. A curve where the allocation rates have been chosen randomly is also reported for comparison. We can notice that no allocation outperforms the multi-rate one, whereas it can sometimes happen that the random allocation performs better than the two-rate scheme. Similar considerations hold true for Goldhill curves reported in Fig. 4.8.c and Fig. 4.8.d for $R_t = 0.4$ bpp and $R_t = 1.2$ bpp respectively.

4.5 Appendix A

We want to evaluate the statistical frequency of subsets encoded at rate $R_j$ ($1 \leq j \leq N$) when receiving $k$ descriptions out of $N$, according to the multi-rate allocation procedure of Sec. 4.1.

**Theorem 1** When using the multi-rate allocation procedure, the statistical frequency of subsets encoded at rate $R_j$ when receiving $k$ descriptions out of $N$ is:

$$F_k(R_j) = \begin{cases} \frac{k(N-j)!}{(N-k-j+1)!} \frac{N!}{N-N-k+1} & \text{for } j \leq N - k + 1 \\ 0 & \text{otherwise} \end{cases}$$

(4.14)

In order to evaluate $F_k(R_j)$, let us begin generating the rate allocation patterns for an information source fragmented into $N!$ subsets. In the first description, all the possible combinations of rates $R = \{R_j; j = 1,\ldots,N\}$ are applied, with $R_1$ as the first
element. As shown in Tab. 4.4, \((N - 1)!\) different groups are generated, each one representing a different permutation of \(R\). The rate allocation pattern of the other descriptions are obtained from the previous one by a left-circular-shift within each group.

When \(k\) descriptions are received, given that each description carries one subset coded at rate \(R_1\) per group, \(k(N - 1)!\) subsets are available at rate \(R_1\) and, as a consequence, \(F_k(R_1) = \frac{k(N - 1)!}{N} = \frac{k}{N}\).

Subsets coded at rate \(R_1\) override the coarse-coded representations of the same subsets \((R_j, \forall j \in [2, N])\). In fact, for each received description, \((k - 1)(N - 2)!\) \(R_1\)-coded subsets override \(R_2\)-coded subsets in \(k - 1\) descriptions. Then, \(k(k-1)(N-2)!\) subsets are discarded over \(k(N - 1)!\) subsets coded at \(R_2\). Consequently, the number of \(R_2\)-coded subsets obtained by merging \(k\) descriptions is

\[ k(N-1)! - k(k-1)(N-2)! = k(N-k)(N-2)! \] (4.15)

Equation (4.15) also represents the number of \(R_3\)-coded subsets not overridden by \(R_1\)-coded subsets when receiving \(k\) descriptions. However, some of these subsets are also overridden by \(R_2\)-coded subsets. The number of subsets coded at \(R_2\) and \(R_3\) but not at \(R_1\) in the \(k\) received descriptions is \(k(k-1)(N-k)(N-3)!\) out of \(N!\) subsets. As a consequence, the number of \(R_3\)-coded subsets obtained by merging \(k\) descriptions is

\[ k(N-1)! - k(k-1)(N-2)! - (N-k)k(k-1)(N-3)! = k(N-3)!(N-k)(N-k-1) \]
Analogously, it is possible to evaluate the number of $R_4$-coded subsets when $k$ descriptions are received as:

$$k(N - 4)! (N - k)(N - k - 1)(N - k - 2)$$

By generalizing this process the number of $R_j$-coded subsets when receiving $k$ descriptions out of $N$ is:

$$k(N - j)! (N - k)(N - k - 1)(N - k - 2)\ldots(N - k - j + 2) = \frac{k(N - j)! (N - k)!}{(N - k - j + 1)!}$$

It is worth noticing that, when $k$ descriptions are merged, no subset is coded at $R_j$ with $j > N - k + 1$, then $F_k(R_j) = 0, j > N - k + 1$. As a consequence, the statistical frequency of subsets encoded at rate $R_j$ when receiving $k$ descriptions out of $N$ is:

$$F_k(R_j) = \begin{cases} \frac{k(N-j)! (N-k)!}{(N-k-j+1)! N!} & \text{for } j \leq N - k + 1 \\ 0 & \text{otherwise} \end{cases}$$
Figure 4.1. Proposed procedure to identify the subsets used to compose the various descriptions for a JPEG 2000 encoder with a multi-rate allocation scheme. For each level of wavelet decomposition, the detail subbands are assumed to be composed of subsets of CBs having nearly similar RD characteristics. Subsets are coded at different rates according to the permutations of the rate vector $\mathbf{R} = \{R_j; j = 1,\ldots,N\}$. Here, $N = 4$ descriptions.
Figure 4.2. SNR [dB] as a function of the probability of MTU loss $p_m$ and number of descriptions $N$. The total output rate is $R_t = 4$ bps; (a) $\alpha = 3$; (b) $\alpha = 5$. 
Figure 4.3. Best achievable central RD performance with several values of side distortion for the $N$-description bound [51]; $N = 3$. 
Figure 4.4. The maximum and minimum PSNR when $k$ descriptions out of $N$ are received, as a function of the probability of description loss $p$ for Lenna image with $N = 4$. (a) CB $64 \times 64$, $R_t = 0.4$bpp; (b) CB $16 \times 16$, $R_t = 0.4$bpp; (c) CB $64 \times 64$, $R_t = 1.2$bpp; (d) CB $16 \times 16$, $R_t = 1.2$bpp.
Figure 4.5. Relative rate deviation ($\eta$) of the descriptions as a function of the probability $p$ of description loss for Lenna image, various CB sizes with $N = 4$. (a) $R_t = 0.4$bpp; (b) $R_t = 1.2$bpp.
Figure 4.6. Normalized rate allocation \( \left( \frac{R_j}{R_t} \right) \) of the two-rate and multi-rate algorithms versus the probability \( p \) of description loss, CB size 64 × 64 and \( N = 4 \). (a) Lenna image, \( R_t = 0.4 \text{bpp} \); (b) Lenna image, \( R_t = 1.2 \text{bpp} \); (c) Goldhill image, \( R_t = 0.4 \text{bpp} \); (d) Goldhill image, \( R_t = 1.2 \text{bpp} \).
Figure 4.7. Quality achieved when a subset of descriptions is available at the receiver, versus the probability $p$ of description loss. Multi-rate and two-rate allocations, CB size $64 \times 64$ and $N = 4$. (a) Lenna image, $R_t = 0.4$bpp; (b) Lenna image, $R_t = 1.2$bpp; (c) Goldhill image, $R_t = 0.4$bpp; (d) Goldhill image, $R_t = 1.2$bpp.
Figure 4.8. Comparison of the multi-rate and two-rate procedures in terms of expected PSNR versus the probability $p$ of description loss; CB size $64 \times 64$ and $N = 4$. (a) Lenna image, $R_t = 0.4$bpp; (b) Lenna image, $R_t = 1.2$bpp; (c) Goldhill image, $R_t = 0.4$bpp; (d) Goldhill image, $R_t = 1.2$bpp.
Chapter 5

MD Coding in sensor networks with finite buffers

A subnetwork of a heterogeneous networks can be made of autonomous wireless devices which cooperatively monitor physical or environmental conditions, such as temperature, sound, vibration, pressure or motion at different locations. These subnetworks, often called sensor networks, are usually composed of a large number of individually-unreliable nodes with many connecting paths between them. Transmissions over these networks are subject to failures such as buffer overflow and nodes failures. This results in packet losses that motivates the use of specific coding techniques capable of exploiting network diversity in order to face node unreliability. We focus our attention to losses caused by the fact that, in practice, sensors are equipped with finite buffers (queues), which causes dropping of packets when the rate of packets injected in the sensor network is sufficiently high [37]. Retransmission could not be possible due to time constraints or expensive feedback and so representations that make all of the received packets useful, such as multiple description, can be of great benefit.

As a first theoretical step, for a point-to-point link and a given fixed probability of
failure, we analyze the dependency between the number of descriptions and the end-to-end distortion using both UEP and MDSQ based techniques. Then, for a practical sensor network performing a data-gathering task (simultaneous transmissions case) having finite buffers, we analyze the optimal transmission strategy as a function of practical parameters such as packet sizes and transmission rates in addition to the number of descriptions. Although from a theoretical point of view, the use of more descriptions usually provides a better performance, this is not in general the case when we introduce practical constraints (finite buffers and packets containing both header information and payload). Since in practice the presence of the header information (sensor location, sequence numbers, etc...) in each packet containing a description increases the traffic generated for the network (thus increasing the packet losses due to overflow), it is shown experimentally that the optimal number of descriptions decreases when the fraction of header information increases.

5.1 MDC in a point-to-point link

For the analytical study of an MD coding system in a sensor network, we first consider an information source that transmits, every time slot, real Gaussian variables with zero mean and variance $\sigma^2$. The point-to-point network model is characterized by the probability $p$ of losing a packet as it moves from the source to the destination following a particular path and it is caused by buffer overload. We assume that if the packet reaches the destination, it does not contain bit errors, and, thus, it can be correctly decoded and we also assume that $p$ does not depend on the length of the packet. The probability $p$ can also be seen as the probability for a packet to reach the destination before a given time useful for representation in a real-time environment.

The information generated by the source is quantized by the encoder that generates a progressive bitstream of length $L = N \cdot R$ bits and then marks it at $N$ different positions, each one corresponding to the attainment of a distortion level
\( D_{Sk} \) (\( 1 \leq k \leq N \)). We denote with \( N \) the number of descriptions and with \( R \) the rate of each of them.

Denoting with \( k \) the number of descriptions out of \( N \) that are correctly received, the mean end-to-end distortion is given by:

\[
D = \sum_{k=1}^{N} \binom{N}{k} \cdot D_{Sk} \cdot (1 - p)^k \cdot p^{N-k} + \sigma^2 \cdot p^N
\]  \hspace{1cm} (5.1)

Our goal is to choose the \( N \) values of \( D_{Sk} \) that minimizes, for a given \( p \), Eq. (5.1).

The main technique we used is based on [34] and it optimally tunes the redundancy in each description with linear complexity using a Lagrange multiplier. For our practical implementation of this algorithm its output is rounded to the nearest integer and the theoretical rate allocation is guaranteed over a long sequence of samples.

In order to get the optimal number of description to transmit for a given \( p \), the output rate \( L \) of the MD coder is fixed while the number of descriptions changes. Being the RD function exponential with the rate for a Gaussian source, relatively high values of \( L \) are tested to allow systems that rely on many descriptions to work with large values of \( R \) on the RD curve. This implies an upper bound on the number of descriptions to be transmitted for a given \( L \) and a certain lower bound on the rates \( R \) that one wants to use.

An example for a source bitstream of 48 bits is given in Fig. 5.1, where the end-to-end distortion is plotted as function of \( p \). In this range of values of \( p \), a 8-description scheme always outperforms systems that rely on a smaller number of packets. But when \( p \) is low, low values of \( N \) should be chosen to guarantee the best end-to-end distortion. This has been proven analytically substituting different values of \( N \) in Eq. (5.1) together with the corresponding values of optimal redundancy (and consequently \( D_{sk} \)) found with the Lagrangian method. Using Eq. (5.1) is thus possible to obtain the ranges of \( p \) where it is worth using a particular \( N \). The same has been verified with simulations as shown in Fig. (5.2). As \( p \) decreases, a smaller number of descriptions should be used.
Figure 5.1. Experimental results for an MD-UEP system with a source bitstream of $L = 48$ bits and no header information.

Figure 5.2. Experimental results for an MD-UEP system with a source bitstream of $L = 12$ bits.

5.1.1 MDC based on index assignment

Together with UEP, another well-known MD coding technique is Multiple Description Scalar Quantization (MDSQ) and it consists in creating multiple descriptions using $N$ different quantizers. We deal with the generalization of an MDSQ system for more than two descriptions according to Vaishampayan [49] where only the
two-description case is addressed. A \( N \)-description system can be seen as a central quantizer and an index assignment in \( N \) dimensions. The encoder produces, for each source scalar input, \( N \) quantization indices, each one sent over a channel. These indices can be seen as indices of a \( N \)-dimensional cube denoting an hyper-plane of the index assignment hyper-cube. Each dimension has size \( S = 2^R \) cells. In the hyper-cube one can insert up to \( S^N = 2^{NR} \) quantization indices. The goal is to fill this hyper-cube with no more than \( 2^{NR} \) numbers so that the difference in all the hyper-planes is as low as possible. This guarantees that if any description is lost, the receiver can reconstruct the best coarse version of the original value and quality increases as the number of received descriptions increase.

With the same information source used for UEP, we apply the method proposed in [7] to completely fill the hyper-cube. The algorithm has proved to be nearly optimal but arrangements of less than \( 2^{NR} \) numbers are not addressed. In fact, as explained in [7], the problem of filling out completely an hyper-cube to minimize the difference in each hyper-plane is NP-complete. Moreover, in practice, since we need to fill out only partially the hyper-cube in order to achieve good trade-offs between central and side distortions, and there does not exist an efficient optimization algorithm, the only possibility is to try all possible ways to fill the hyper-cube with less than \( 2^{NR} \) numbers, which has a very high computational complexity.

The complexity of dynamically adapting the coder to different operating points discourages a real-time implementation of the index assignment method of [49] using an arbitrary number of descriptions. Anyway, it is worth noticing that, from an analytical point of view, MDSQ, at least for two descriptions, guarantees better performance in terms of mean end-to-end distortion, as proved in [19].

### 5.2 MDC in a sensor network

The network model that we assumed in the previous section was characterized by a constant probability \( p \) of losing a packet, where this event is uncorrelated among
channels and independent of the number of descriptions. Under these assumptions, as it is shown experimentally in the previous section for UEP, the more descriptions we generate, the lower distortion we achieve for a wide range of values of $p$. However, in many real applications, this model assumption is not valid and, as we will see now, a higher number of descriptions does not lead necessarily to a lower distortion.

In this section, we analyze the performance of MD coding schemes in a real network scenario, namely, a data gathering lattice sensor network.

Suppose we want to measure a Gaussian random field $X$ over a square area. We assume the process to be uncorrelated in space and time. We uniformly place $W^2$ sensor devices that sample the field and send all the information to a single device, denoted as base-station, which gathers all the information generated by the network (Fig. 5.3(a)).

We assume that these devices generate samples of $B$ bits following a Bernoulli distribution with parameter $R$. Note that $R$ also represents the average information rate generated per device. To transmit the data, devices wait to have a number $S$ of samples to put all in one single data packet, and send it to the base-station. We
assume that $S$ is determined by the application requirements independently of the number of descriptions generated. In practice, notice that in addition to the samples (payload of the packet), data packets need also a header that includes information such as the device location, sampling times or sequence numbers. We denote by $H$ the size of the header in bits.

Devices code the samples and produce $N$ descriptions per sample. To make the comparison between SD coding and MD coding fair, we assume that each description contains $B/N$ bits. As $S$ descriptions are included in the same packet, the total packet size $K$ is given by $K = H + SB/N$.

In terms of communication capability, we assume that the devices can only communicate to their four adjacent neighbors (see Fig. 5.3(a)). Therefore, to send data to the base-station, devices act also as relays for other communications. Packets are routed to the base-station using a random routing algorithm that uses all paths between each source device and the base-station with equal probability [38]. Note that this routing algorithm allows to exploit the network diversity present in the network, making the use of multiple description coding very convenient (see Fig. 5.3(b)).

The devices found in real sensor networks generally have a small buffer (queue) for the temporary storage of the packets that need to be transmitted, that is, packets generated by the device itself or packets travelling through it. From a queueing point of view, we can model the system as a queueing network with a service time proportional to the packet size [6] where packet losses occur due to buffer overflow at the devices. Therefore, the probability $p$ of losing a packet depends on the network parameters such as the transmission rate $\mathcal{R}$, the packet size $K$, and the number of descriptions $N$.

With all the information generated by the network, we reconstruct the random field $X$ at the base-station. We measure the distortion $\tilde{D}$ between $X$ and the reconstructed field $\hat{X}$ as the average mean-square-error (MSE) per device and unit
Figure 5.4. \( \hat{D} \) as a function of the transmission rate \( R \) for \( B = 8 \) bits and \( N = [1, 2, 4] \).

We want to investigate the optimal strategy that achieves the lowest \( \hat{D} \), that is, 
\[
\min_{\{R, N, K\}} \hat{D}.
\]
First, note that for fixed values of \( N \) and \( K \), there exists an optimal transmission rate \( R^{\text{opt}}(N) \) that minimizes \( \hat{D} \). As we increase \( R \), the devices generate more information packets and \( \hat{D} \) decreases. However, if we continue to increase \( R \), the network becomes congested and devices start losing packets. Fig. 5.4 shows \( \hat{D} \) as a function of \( R \) in a \( 25 \times 25 \) network with \( B = 8 \) bits per sample, \( H = 36 \) bits, and \( S = 20 \).

We examine now how \( \hat{D} \) is affected by the number of descriptions. First, note that because of the practical constraints in a real application, every information packet needs to include a header, and therefore, the traffic transmitted through the network increases with the number of descriptions. Moreover, the characteristics of the input traffic also changes. In MD coding, the devices generate several packets at the same time and the traffic becomes more bursty. These two factors contribute to increase the overflow probability and consequently, to increase \( \hat{D} \).
On the other hand, when we generate several descriptions, the size $K$ of each packet is reduced. This packet size reduction decreases network congestion and, equivalently, decreases $\tilde{D}$. This can be illustrated with an example. Consider just a single $M/M/1/Q$ queue with a buffer size of $Q$ packets, average arrival rate $\lambda$, and average service time $\mu$. Applying $M/M/1/K$ formulas, the utilization factor is given by $\rho = \lambda / \mu$, and the probability of packet overflow $p$ by:

$$p = \frac{1 - \rho}{1 - \rho^{Q+1}} \rho^Q.$$ 

Suppose now that we generate $M$ times more packets of $1/M$ the size. This is equivalent to consider $\lambda' = M\lambda$ and $\mu' = M\mu$ and $Q' = MQ$. Obviously, the utilization factor does not change, that is, $\rho' = \rho$. However, the new probability of packet overflow $p'$ is clearly reduced:

$$p' = \rho^{(M-1)Q} \frac{1 - \rho^{Q+1}}{1 - \rho^{MQ+1}} p < p.$$ 

Note also that as the average waiting time $T$ in the queue is proportional to $1/\mu$, $T$ is also reduced when we consider smaller packets.

This suggests that there exists an optimal number of descriptions that depends on the packet size $K$, more particularly on the ratio $H/K$.

We compare now the $\tilde{D}$ achieved by transmitting at $R_{\text{opt}}(N)$ for different values of $N$ as a function of the ratio $H/K$. Fig. 5.5 shows the gain in $\tilde{D}$ achieved by using $N$ descriptions with respect to SD coding as a function of $H/K$ in a $25 \times 25$ network with $S = 20$ and $B = 8$ bits per sample for $N = 1, 2, 4$. Similarly, Fig. 5.6 shows the same gain for $B = 12$ bits and $N = 1, 2, 3, 4$ and 6.

In the case of $B = 8$ bits, the optimal strategy for $H/K < 0.5$ consists in generating 2 descriptions using MDSQ. However, the benefit of using 2 descriptions decreases as the rate $H/K$ increases due to the extra traffic generated, and when $H/K > 0.5$, the lowest $\tilde{D}$ is achieved by SD coding. Note that due to its exponential complexity, we considered MDSQ only for 2 descriptions.

For $B = 12$ bits, we considered only UEP coding to generate several descriptions. As we discussed in the previous section, the complexity of MDSQ is exponential not
only with $N$, but also with $B$. Even for the simplest case of $N = 2$, each device would have to generate a $2^{B/2} \times 2^{B/2}$ matrix where $2^{B/2+1} - 1$ diagonals can be filled. For small $H/K$ values, the lowest $\tilde{D}$ is achieved by generating three descriptions, for which a gain of almost 2.5 dB is achieved with respect to SD coding. To generate more than three descriptions does not decrease the distortion: first, we have the
practical constraints of creating descriptions with very few bits, and second, there is the penalty due to the extra traffic generated. As the header size increases, the performance of MD degrades due to this extra traffic. Note that the more descriptions we generate, the more rapidly it degrades. Therefore, the optimal number of descriptions decreases when the ratio $H/K$ increases. For instance, when $B = 12$, we have that for values of $H/K$ between 0 and 0.15, the optimal strategy consists in generating 3 descriptions. When $H/K$ is between 0.15 and 0.35, the optimal is $N = 2$. Finally, when $H/K > 0.35$ we use SD coding.
Chapter 6

MD Coding for MJPEG 2000 over congested 802.11e WLANs

The efficient transmission of video content over wireless communication networks is a challenging goal, especially when considering multiple mobile users equipped with energy-constrained handheld devices and sharing the same channel resources. Here, considering the scenario of the real-time delivery of multiple Motion JPEG 2000 (MJPEG 2000) video streams over a QoS-enabled 802.11e wireless LAN, we investigate the performance of multiple description coding and compare it with different video transmission techniques. We analyze if and when data redundancy of the information source can improve the end-to-end perceived quality in this realistic multi-user scenario.

The considered setup is shown in Fig. 6.1 and consists of multiple independent Mobile Terminals (MTs) which upstream real-time MJPEG 2000 video traffic to the Access Point (AP) of an 802.11e wireless local area network (WLAN). These terminals transmit their data over a shared wireless channel, assumed to be slowly fading (typical of indoor propagation conditions). The 802.11e-based scheduler introduced in [32] is considered so as to offer QoS support for the energy-efficient delivery of the different video streams.
Figure 6.1. The considered setup: many Mobile Terminals who upstream real-time video traffic to an Access Point.

This scenario strongly calls for advanced video coding techniques suited to error-prone transmission conditions. The purpose is to investigate the performance of multiple description coding in the aforementioned realistic network setup and to compare it with other state-of-the-art video techniques, such as layered coding and conventional single description coding.

Here, we apply to video coding the methods previously introduced in order to build an arbitrary number of descriptions and we analyze MD performance considering the realistic implementation of a wireless network mentioned above.

6.1 System description

A WiFi-based WLAN is considered: the IEEE 802.11a standard is taken for the physical layer (OFDM-based transmission in the 5GHz band), and the QoS-enabled IEEE 802.11e standard is considered for the MAC functionalities. The considered setup consists of multiple independent users who upstream real-time MJPEG 2000 video traffic to the AP of the WLAN. At the AP, a scheduler that relies on the Hybrid Coordination Function (HCF) of the IEEE 802.11e MAC protocol [4] steers
the whole WLAN network. Capitalizing on appropriate design-time performance-energy modelling of the whole system, this controller decides at run-time on a regular basis how to allocate channel resources to the different mobile terminals and how to configure their radio parameters. The scheduler is designed so as to reach a given transmission packet error rate under a given delay constraint, assuming no congestion is present in the network. Further details can be found in [32].

In case congestion takes place, packets have to be dropped and the packet error rate then exceeds the targeted error rate. In our design, according to the video coder installed at the MT side, each transmitter has as many 802.11e priority queues as the number of layers (for layered coding scheme) or descriptions (in the MD case). The AP keeps track of the state (in terms of size) of each of these queues and, as explained in [4], regularly polls queues using specific control frames (QoS CF-Poll frame). If a polled MT has no queued traffic to send or if the packets available to send are too long to be transmitted within the specified timeslot (referred to as TXOP limit in the standard), it informs the AP using another control frame (QoS Null frame). If the packet fits into the specified TXOP, the transmission can then happen. This means that there exists in the network a background traffic associated with these control aspects (we will refer to it as overhead traffic), which depends on the number of transmitters and the number of layers/descriptions sent by each MT. This overhead traffic, as shown later, can influence performance.

A dropping policy is implemented in case congestion happens (i.e., when the amount of data to be transmitted cannot be scheduled because it exceeds the network bandwidth). The optimal strategy from a network fairness criterion is to drop the biggest data demand first. The scheduler then sends a message to the appropriate MT forcing it to free its biggest queue (per queue dropping policy). If the network resource is still not enough, other queues are purged according to the same principle. For MDC, this is optimal because each packet is treated with an equal priority.

We also remind that in the wireless network data loss not only depends on the level of congestion (that causes packet dropping) but also on the channel quality that,
as previously stated, is assumed to be slowly fading and then, with its fluctuations, can cause errors into the video stream.

### 6.2 Proposed method

Each transmitter of the wireless LAN encodes and sends to the Access Point a MJPEG 2000 video sequence. The MJPEG 2000 standard [3] is an intra-frame scalable video coder based on wavelet transform. Relying on intra-frame encoding, MJPEG 2000 has less coding efficiency than an inter-frame encoder but, being a low-complexity encoder, it suits well in devices where complexity is a key point (i.e., battery-constrained mobile terminals). Moreover, because frames are independently coded, the spread of transmission errors is effectively prevented across consecutive frames. Scalable video codecs can offer significant advantages in error-prone wireless network applications. A comparison of the performance of MJPEG 2000 and the well-known MPEG-4 standard in the framework of video transmission over low bit-rate error-prone wireless channels has already been presented in [14]. In our design a full protocol stack, combined with state-of-the-art radio transceivers and an advanced scheduling algorithm were set up, as opposed to raw transmission of [14]. Moreover, both packet error patterns caused by link failure and congestion are considered.

We generalized the method proposed in Chapter 3 to deal with MJPEG 2000 video sequences and an arbitrary number of descriptions. For each video frame, a description is a composition of codeblocks (CBs) taken from either bitstream. For each CB, one description contains its representation at rate $R_1$ whereas all the others have the same CB represented from the stream at rate $R_2$. In order to generate $N$ balanced descriptions, for each layer of either bitstream, CBs are grouped into $N$ sets having similar RD characteristics as explained in Chapter 3. The procedure yields balanced descriptions, each of which is encoded at approximately $\frac{R_1 + (N-1)R_2}{N}$ bit per pixel (bpp). The total available output rate, denoted with $R_{tot} = R_1 + (N - 1)R_2$,
is a system constraint in our setup and the MD redundancy $\rho$ is normalized with $R_{tot}$ leading to the expression:

$$\rho = \frac{R_2(N - 1)}{R_1 + R_2(N - 1)}$$

For each video frame, the decoder pre-processes all the received descriptions and selects, for each CB, the finest representation. The resulting stream is then JPEG 2000 decoded. When all packets are received, the quality is only a function of $R_1$ because all CBs are taken from the bitstream coded at that rate. When only a subset of $k \geq 1$ descriptions out of $N$ is properly received, a side distortion is obtained which represents the quality as a weighted average of the quality yielded by two bitstreams at rates $R_1$ and $R_2$.

We compare MD to layered coding and, in order to have a fair comparison, we allocate bits to layers according to [45]. Each packet sent over the network contains a layer so that when a packet is dropped due to congestion or bad channel state, then the corresponding layer is lost. The size of each layer is approximately constant among layers so the scheduler does not priorioritize any particular layer. This has been done to fairly compare layered coding with multiple descriptions. A key point of layered coding is the possibility for the decoder to reconstruct the video frame with a quality proportional to the last consecutive received layer. If the first layer is lost, no information is obtained at the receiver, even if other layers are received.

### 6.3 Experimental results

For our tests with MD, we used the JPEG 2000 codec engine from the OpenJPEG libraries [30]. We generalized it to support the encoding procedure of an arbitrary number of descriptions per frame (up to 8 descriptions). We first fix $\rho$ to the lowest feasible level (thus leading to good central performance and poor side performance) and then to $\rho \approx 0.5$, suitable when many packets are lost. Regarding the JPEG 2000 encoder, we set 5 DWT decomposition levels for the wavelet transform and a code-block size of $64 \times 64$. 

72
In the 802.11e wireless network we set up, each UDP datagram at application level is encapsulated into a single IP packet. Then each IP packet is fragmented into MAC units of 1024 bytes.

We compare MD with layered and traditional (Single Description) coding. In our scheme, when a MT encodes data with SDC, it transmits each JPEG 2000 video frame in a single UDP packet using a single priority queue.

For our tests, we used Bus sequence (CIF, 150 frames, 30fps) at different bitrates and we ran it for approximately 40s to have significant statistical information. We also tested other sequences to verify the behaviour we obtained. The number of MTs varies between 2 and 8.

First of all, we focus on the relation between the number of transmitters and the probability of packet loss. Here, the number of video transmitters influences the probability of packet dropping as we can see from Fig. 6.2 and Fig. 6.3. The transmission rate, in fact, is comparable with the total available bandwidth. At 0.7 bpp, the bandwidth required for each transmitter is approximately 2 Mbps when the actual bandwidth is approximately 5 Mbps. When 3 users are transmitting using $N = 8$, then the scheduler already drops some queues to face congestion.

In the 0.7 bpp case, the highest packet dropping is with $N = 8$ descriptions
whereas in the 1.2 bpp case, SDC performs worse. This behaviour is mainly caused by the overhead traffic as explained in Sec. 6.1 and by MAC fragmentation. Concerning the overhead traffic at 0.7 bpp, we see that, for 3 users, the percentage of packet loss with \( N = 8 \) is higher than \( N = 2 \). In the 8-description case, in fact, the overhead traffic is four times higher than the 2-description case and this heavily increases network traffic.

As previously stated, also the MAC fragmentation affects packet dropping. In fact, at 0.7 bpp, the size of a CIF frame is approximately 8870 bytes. When using 8 descriptions, the size of each of them is approximately 1109 bytes so each description, on average, is sent using 2 MAC units. In the 4-description case, 3 MAC units are generated for each description. This implies that the 4-description case is favored compared to 8 descriptions because it is relatively transmitted using less MAC units than it should to be fairly compared. A fair comparison would be \( N \) descriptions encapsulated in \( L \) MAC units each, and \( N/2 \) descriptions encapsulated in \( 2L \) MAC units each. It is thus important to have a strong cross-layer link between application and MAC layer so that it would be possible to tune one’s parameters according to the others.

When using a coding rate of 1.2 bpp (Fig. 6.3) the comparison among different
number of descriptions and SDC is fair. For a given number of transmitting users, the more the descriptions, the lower the probability of packet dropping. When congestion happens, the scheduler drops packets from the first queue on and it stops when the total required rate is below the available bandwidth. Using a high number of queues implies that the scheduler can rely on more granularity to face congestion so that it can purge only the required amount of information as it drops on per-queue basis. In the SDC case with congestion the whole queue is removed. On the other side, using a high number of queues implies a significant amount of overhead traffic. For 1.2 bpp, its rate is smaller if compared to the data rate so the main contribution on the packet loss probability is given by the video traffic.

In Fig. 6.4 we plot the best MDC configuration versus the best layered coding for Bus sequence. When congestion happens, MD can effectively outperform layered coding. In fact, it relies on many independent layers of resolution and, even if a subset is lost during transmission, the inserted redundancy is exploited to reconstruct with good quality the original video stream. When no congestion happens, layered coding performs better because no redundancy is inserted. The layered coder inserts information instead of redundancy in the same budget of available rate. At 1.2 bpp, congestion is present since 2 transmitting users so traditional coding performs badly because the scheduler drops its whole single queue.

Fig. 6.5 shows curves of SDC and MDC at different rates. It is worth noticing that the encoding rate plays a crucial role in the probability of packet loss. When 2 users are transmitting, the available bandwidth is higher than the amount of transmitted data. In general, then, no packets are dropped due to congestion and packet loss happens due to channel fluctuation. In this case, redundancy inserted in MD and a rate of 1.2 bpp can face well network unreliability and lead to good quality. But when the number of users increases, packets are dropped also to face congestion. Then, first we should decrease the transmitting rate of each users and then, with MD, adding a significant amount of redundancy. Anyway, the gap between MD and SDC is limited and this is caused by the high traffic in the network.
Figure 6.4. Comparison between SDC, MDC and layered coding for Bus sequence at 1.2 bpp.

Figure 6.5. Comparison of MDC and SDC for Bus sequence at different rates and redundancy levels.
Chapter 7

MD Coding for storage of Digital Cinema content

We propose a multiple-description-based distributed system in order to store Digital Cinema contents in distribution centers and in theaters. Different representations of the high quality version of a movie are stored in a decentralized system in order to allow both data redundancy in case of server breakdowns and efficient and fast distribution of video content. Independent and mutually refinable representations of the movie, compliant with Part 3 of JPEG 2000 standard, are generated. These can be used autonomously in case of damage of a portion of the content or mixed together for traditional high-quality screening.

7.1 Introduction

In the last ten years, Digital Cinema (DC) has attracted much attention from diverse disciplines, for its high quality, security, and the saving of resources in long term. The two bodies most actively involved in producing standards for Digital Cinema are the Society of Motion Picture and Television Engineers (SMPTE) and the DCI (Digital Cinema Initiative). SMPTE is the official standards-setting organization for cinema and television; the DCI is an ad-hoc organization created by several
major Hollywood studios, which in July 2005 released the first version of Digital Cinema System Specification [9]. This represents a guide for companies interested in developing products, which can be compatible with others, pointing out many key points and research topics. Among others, we mention video formats and compression, packaging and composition of audio/video tracks, transport to theaters, theater system and projection requirements, security.

As discussed in [13, 15], there are still open issues that prevent Digital Cinema from a wide-scale distribution. As an example, resolution of digital projectors is constrained by the modulators responsible for creating pixels of the screen image.

An other issue that should be carefully taken into account in a digital cinema system is the method used for distribution of video content to the theaters and, once delivered to their locations, to different screens in a multiplex facility. DCI specification suggests many ways, such as physical media (such as hard-disks), Virtual Private Networks or satellites. Distributors will select the method that is both economical and technically robust to ship their content to the theaters, but it is worth noticing that if movies are transported using a physical media, this does not take advantage of the money-saving potentials of digital content distribution.

Independent of the transport method, the output interface of the transport system is required to be ingested into the Digital Cinema Storage of the theater. The interface is required to be Gigabit or 1000Base-T Ethernet interface using a TCP/IP protocol. The system is also required to provide for intra-theater movement of content for multiple screens. Emergency moves due to equipment failure between different auditoriums are required to allow playback to start within 15 minutes after the start of the movement.

Other issues regard video compression and storage. In particular, for a 2048 × 1080 video with a frame rate of 24fps and 12 bit per pixel per color component, a 3-hour uncompressed movie requires approximately more than 2 Terabytes of disk space for storage. It is worth noticing that, besides the 2048 × 1080 resolution (called 2K resolution), the Technical Specification also proposes a 4K resolution.
(4096 × 2160 pixels) that dramatically increases disk space required for storage. This strongly calls for compression techniques able to reduce the total output bitrate while maintaining the high-quality required by digital cinema. JPEG 2000 has been chosen as standard for video frame compression. It allows for scalability both in resolution and quality (PSNR), achieving higher compression rate than JPEG. Part 3 of JPEG 2000 standard addresses video sequences (Motion JPEG 2000) [3]. Based on JPEG 2000, it represents an intra-frame wavelet-based scalable video coder. Frames are independently coded and the spread of transmission errors is effectively prevented across consecutive frames. Moreover, intra-coding allows for random access to frames (for editing) and reduced complexity. Relying on intra-frame encoding, Motion JPEG 2000 has less coding efficiency than a motion-compensated encoder but, at the high bit rates required for Digital Cinema applications, the compression efficiency loss with respect to state-of-the-art inter-frame coders is not relevant. More details about Motion JPEG 2000 can be found in [16]. In [12], authors address practical hardware implementation of a JPEG 2000 decoder for Digital Cinema, further focusing on issues regarding complexity and projectors.

We propose multiple description coding for digital cinema data backup. Being the amount of data huge, it can be hazardous to store it all in a single location. Then, the intrinsic scalability of JPEG 2000 coder, together with the capability of multiple descriptions to fragment data into independent portions, allows content producers and theater operators to safely backup data with existing storage systems.

### 7.2 Proposed approach

As pointed out in [47], MDC also allows one to store data in a distributed fashion, distributing network load over different servers, thus reducing overload of links and nodes. Concerning digital cinema and management of movies in a multiplex theater, MD can be of great benefit. First, if a video is fragmented along different servers, in case of a system breakdown, only a portion of the movie is lost. Theater owner only...
has to ask the distributor for the missing part (thus reducing the time of delivery if fiber-optic network delivery is supported). Second, people can still watch a lower-resolution version of the movie (possibly paying a cheaper ticket) while the theater management tries to recover the missing part. This requires that the lower-resolution version is still at a good quality.

In order to build two balanced descriptions of a movie, we adopt the rate-distortion MD approach suggested in Chapter 3 and we generalize it to support digital video sequences. For each video frame, two different JPEG 2000 bitstreams of rates $R_1$ bits per pixel (bpp) and $R_2 < R_1$ bpp respectively are generated.

![Multiple Description allocation procedure for 2 descriptions.](image)

For each JPEG 2000 bitstream, codeblocks (CBs) generated after the wavelet decomposition [35] are grouped into two disjoint sets. These sets are built so that they have similar characteristics in terms of size and distortion contribution. In principle, to obtain two balanced descriptions, one should identify subsets of CBs with similar RD characteristics and allocate them to the descriptions. Nevertheless, we have verified that it is sufficient that each description contains the same number of
CBs encoded at rate $R_1$ (and consequently at rate $R_2$) for each subband. This is due to the fact that these video frames contain a high number of CBs. As a consequence, we can approximately generate the two subsets using a sort of checkerboard pattern as shown in Fig. 7.1.

The first description is thus obtained by combining the CBs of set 1 taken from the first stream, with the CBs of the second set taken from the second bitstream. Analogously, the second description is built by taking the CBs of set 2 from the first stream and those of the first set from the second stream.

This procedure generates two balanced descriptions, each of which is fully backward compatible with Part 3 of the JPEG 2000 standard [48]. Each description is encoded at $(R_1 + R_2)/2$ bpp, yielding a total output rate of $R = R_1 + R_2$ bpp (per frame).

The decoder, for each video frame, pre-processes all the received descriptions and selects, for each CB, the finest representation. Given that the procedure yields balanced descriptions, when only one description is received, a side distortion $D_1$ is obtained, representing a weighted average of the quality yielded by two bitstreams at rates $R_1$ and $R_2$. As the distortion of JPEG 2000 CBs can be assumed additive [35], we can write that:

$$D_1(R_1,R_2) = \frac{D(R_2) + D(R - R_2)}{2}$$

(7.1)

where $D(r)$ is the RD curve of the JPEG 2000 image encoded at rate $r$. When both descriptions are received, the quality is a function of $R_1$ only, because all CBs are taken from the bitstream coded at that rate as explained in Chapter 3.

### 7.3 Experimental results

We analyzed the performance of a 2K distribution system with a resolution of $2048 \times 1080$ pixels at 24fps. For our tests, we used the JPEG 2000 codec engine from the OpenJPEG libraries [30]. According to the DCI Digital Cinema Specification [9],
Table 7.1. Mean quality (PSNR) [dB] of test sequences coded at 1.8bpp for traditional (SDC) coding and MD coding with \( \rho = [0.05, 0.1, 0.3] \).

<table>
<thead>
<tr>
<th>Sequence</th>
<th>SDC</th>
<th>MDC ( \rho = 0.05 )</th>
<th>MDC ( \rho = 0.1 )</th>
<th>MDC ( \rho = 0.3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blue Sky</td>
<td>44.96</td>
<td>44.41</td>
<td>33</td>
<td>43.97</td>
</tr>
<tr>
<td>Ped. Area</td>
<td>46.76</td>
<td>46.13</td>
<td>38.69</td>
<td>45.88</td>
</tr>
<tr>
<td>Rush hour</td>
<td>46.45</td>
<td>46.35</td>
<td>40.82</td>
<td>46.06</td>
</tr>
</tbody>
</table>

and in conformity with [1], each frame is encoded as a single tile and 5 wavelet transform levels are generated in the 9/7 DWT process. Test sequences have been downloaded from [17]. For the three test sequences, a high-quality stream is obtained when coding each frame at 1.8bpp, generating a video stream with a rate of 90Mbps (thus compliant with [9] which sets to 250 Mbps the maximum rate).

In Table 7.1, we report the mean quality of three test sequences when using a traditional JPEG 2000 encoder, together with central and side quality of an MD-JPEG 2000 encoder. Results with \( \rho = 0.05 \), \( \rho = 0.1 \) and \( \rho = 0.3 \) are reported. Let us consider Blue Sky test image. With \( \rho = 0.05 \), when the two representations are mixed together, the quality obtained is approximately the same of a standard JPEG 2000 encoder. When \( \rho \) increases, the central quality is reduced in favor of the side quality. Thus, it is possible to obtain a side quality of 35 dB and a central quality of 38 dB without increasing the total amount of stored information. This is paid by a reduction of only 2 dB in the full (central) quality. Similar considerations can be done for the other sequences. The higher the redundancy, the higher the side quality but, in digital cinema, it makes little sense to insert high redundancy because it reduces the central quality (i.e. the quality experienced by the audience).

Fig. 7.2 shows the quality (expressed as PSNR) of the first ten frames of Pedestrian Area sequence. The best performance is achieved with a standard JPEG 2000 encoder. For MD coding, in fact, the redundancy is set to \( \rho = 0.1 \) so that, in the same amount of total output bitrate (1.8bpp), a portion of the information is redundant. The main advantage of MD coding, anyway, is that it also provides a
low-quality version of the movie (shown as \textit{MDC - side} in the figure). If a slight degradation of the central quality is permitted (approximately 0.5 dB for each RGB component with these settings), then a representation at approximately 40 dB is available for screening. In the figure, we plotted the mean quality of the two descriptions. It is worth pointing out that the quality gap between the two descriptions is less than 0.2 dB.

In Fig. 7.3 we plot the quality of the first 25 frames of \textit{Rush hour} sequence coded at 1.8bpp. The best performance is achieved by the JPEG 2000 encoder but, if a quality degradation of less than 0.2 dB is tolerated, then MD with a redundancy of $\rho = 0.05$ encodes two descriptions, each one with a quality nearly 40.8 dB (useful for backup and preview). If they are merged together the average central quality is 46.35 dB according to Table 7.1. If more degradation in the central quality is acceptable (1.6 dB with these settings), then two high-quality MD streams with $\rho = 0.3$ can be encoded. If merged together, the quality is approximately 44.8 dB; if only one description is available, then the quality is almost 43.5 dB.

Fig. 7.4 represents the tradeoff between central and side quality for \textit{Pedestrian}
Figure 7.3. A comparison between traditional coding and MDC with different values of redundancy ($\rho = 0.05$, $\rho = 0.30$) for Rush hour sequence coded at 1.8bpp.

Figure 7.4. Tradeoff between MD central and side quality for Pedestrian sequence coded at 1.8bpp.

*Area* sequence at 1.8bpp. If it is required a central quality of no less than 45 dB for a high-quality experience of video content, then it is possible to tune the amount of redundancy information in order to obtain up to 42.5 dB for the side quality.
Chapter 8

Conclusions

As discussed in the previous chapters, we have shown that Multiple Description coding is a powerful technique to protect multimedia information over heterogeneous networks. In fact, it allows to protect data with tunable redundancy according to the importance of each particular data packet.

The main challenges posed by heterogeneous networks are the implementation of new transport layer protocols to support real-time traffic, the possibility of enabling Quality-of-Service features for different types of data and an efficient exploitation of the different paths between nodes to face packet loss. Furthermore, a good management system to efficiently handle multiple data streams could be very useful for the video content provider.

Our research has been focused on the application of Multiple Description for JPEG 2000 and Motion JPEG 2000. Furthermore, we analyzed different scenarios in which this coding technique can be successfully applied.

8.1 Redundancy tuning

Concerning JPEG 2000 and Motion JPEG 2000, we pointed out that the MD redundancy tuning task is particularly significant in the encoding process. The value of
this parameter should be finely tuned according to network conditions and dynamically changed according to them. As a matter of fact, when too much redundancy is inserted in the descriptions, the performance of the scheme is considerably reduced because data information carried by each description is similar along them. On the contrary, when few redundancy is inserted and many packets are lost, the received information can be inadequate to recover the data source. We have shown that knowing the rate-distortion function of the source in some points can effectively help to tune this parameter.

Particular effort in our research has been made to better exploit of redundancy using backward compatible co/decoding schemes. As for JPEG 2000, in fact, the proposed schemes are backward compatible with the standard. This permits a not-MD enhanced receiver to decode the first received description and to obtain an acceptable representation of the original image.

Furthermore, we proposed schemes to encode an arbitrary number of descriptions. In general, the more the descriptions, the higher the probability that a subset of them can be successfully decoded at the receiver. A bound on their number can be given by the available bandwidth and on encoding capabilities on the transmitter side. We have shown that a range of network conditions exists in which a particular number of descriptions should be used to guarantee the best performance. We have also analyzed the effect of the header size of each description and we have verified that also the header-payload ratio should be carefully taken into account when setting up a MD coding system.

Using a multi-rate allocation scheme, in particular, we investigated the possibility to guarantee a minimum end quality to each user of the network properly configuring the allocation scheme of the combined MD-JPEG 2000 encoder.

We also focused our research on the significant importance of generating balanced descriptions. In fact, if descriptions are balanced (from a RD point of view) it is then possible to transmit data so that only the number of received packet is relevant for quality measurement purposes and not the particular received set.
8.2 Applications

Concerning the applications of MD coding we focused on sensor networks, wireless LAN for video transmission and distributed storage for digital cinema. Multiple Description, in general, can be successfully applied in packet networks where retransmission is not possible or data is highly sensitive to delays such as multimedia streaming applications over Internet.

We discussed about heterogeneous 802.11e wireless networks and we compared multiple description, layered coding and traditional coding as a mean to transmit real-time video traffic over these networks.

As a first consideration for MD coding we highlighted that it is important to well balance each description in terms of size so that they all approximately have the same probability of being dropped by the scheduler that polls each transmitter. If descriptions are not well balanced, as the biggest packets contain the highest amount of information, performance significantly decreases.

As previously stated, the number of descriptions in MD coding can affect performance. In a WLAN with priority queues, a significant amount of data is used to manage queues at transmitters. The more the queues, the more the overhead traffic influences performance. This is especially evident at low rates where the data rate is comparable to the channel bandwidth and performance are reduced due to this background traffic. Having a high number of queues per transmitter not only causes a degradation in performance. In fact, the scheduler can rely on many queues and then it has a higher granularity when dropping data traffic. With many descriptions, only the required amount of information is dropped to face congestion.

Another point that should be taken into account when developing an MD scheme for wireless heterogeneous networks is the encoding rate that significantly modify network behaviour. The probability of packet loss is strongly influenced by the encoding rate because the required bandwidth for MJPEG 2000 is significant compared to the total available bandwidth.
It is worth noting that a good cross-layer integration between application and MAC layers should be considered to guarantee fair transmission of video streams using multiple descriptions not only in WLANs. The encoding rate should be chosen according to MAC unit size.

The last application for MD coding we considered is for the distributed storage of digital cinema content both for distributors and theaters. Here, we do not deal with real-time transmission of data but with its storage for backup. We pointed out that multiple description coding allows to store movies on different servers in order to increase robustness to system breakdowns. Moreover, another particularly significative advantage of this coding technique is that it allows to obtain two (or more) independent and mutually refinable representations of the video, both compliant with Part 3 of JPEG 2000 standard. These can be also used independently to obtain a high-quality movie even in case of damage of a portion of the content, for low-cost screenings and preview.
Bibliography


[19] V.K. Goyal. Single and multiple description transform coding with bases and frames. to be published by SIAM.


[22] V.K. Goyal, J. Kovacevic, R. Arean, and M. Vetterli. Multiple description


[33] S.S. Pradhan, R. Puri, and K. Ramchandran. (n,k) source channel erasure


