

# Stellingen

behorende bij het proefschrift

*Hierarchical Video Compression using SBC*

door F. Bosveld

23 september 1996

## I

De keuze tussen 'embedded' kwantisatoren en 'multistage' kwantisatoren in hiërarchische compressiesystemen is afhankelijk van de 'bit error rate' van het transmissiekanaal.

*(Dit proefschrift.)*

## II

Schaalbaarheid en compressie zijn moeilijk verenigbare doelen.

*(Dit proefschrift.)*

## III

Standaarden voor digitale video- en audiotoeepassingen zijn geen garantie voor winstgevende producten.

## IV

De introductie van digitale betaal-TV behelst meer dan de publiekelijk uitgevochten strijd rondom de decoder.

## V

Het begrip multimedia is ook voor veel traditionele audiovisuele apparatuur van toepassing.

## VI

Technische universiteiten zullen kwaliteitssystemen voor derde geldstroomprojecten moeten introduceren om op langere termijn de inkomsten uit deze geldstroom te garanderen.

## VII

Elementaire kennis over het ontwikkelen van producten mag niet ontbreken in de opleiding tot elektrotechnisch ingenieur.

## VIII

De profit sector is niet altijd winstgevend voor de werknemer.

## IX

'Open deur management' is alleen effectief indien de drempels laag zijn.

## X

Voor de consument is elektronisch thuisbankieren niet de goedkoopste vorm van bankieren.

## XI

De kracht van Internet ligt in de 1 promille zinvolle informatie.

## XII

Het verplicht stellen van vuilwatertanks voor de pleziervaart zal op termijn leiden tot een toename van de gemiddelde lengte van pleziervaartboten.

Hierarchical Video Compression  
using SBC

F. Bosveld

# Hierarchical Video Compression using SBC



## Proefschrift

ter verkrijging van de graad van doctor  
aan de Technische Universiteit Delft,  
op gezag van de Rector Magnificus Prof.ir. K.F. Wakker,  
in het openbaar te verdedigen ten overstaan van een commissie,  
door het College van Dekanen aangewezen,  
op maandag 23 september 1996 te 13:30 uur  
door

**Frank BOSVELD**

elektrotechnisch ingenieur,

geboren te Rotterdam.

Dit proefschrift is goedgekeurd door de promotor:  
Prof.dr.ir. J. Biemond.

Toegevoegd promotor:  
Dr.ir. R.L. Lagendijk.

Promotiecommissie:

Rector Magnificus, Technische Universiteit Delft, voorzitter  
Prof.dr.ir. J. Biemond, Technische Universiteit Delft, promotor  
Dr.ir. R.L. Lagendijk, Technische Universiteit Delft, toegevoegd promotor  
Prof.dr. J.W. Woods, Rensselaer Polytechnic Institute, USA  
Prof.dr.ir. H. Vinck, Universität GHS Essen, Duitsland  
Prof.dr. F.C. Schoute, Technische Universiteit Delft  
Prof.dr.ir. P.M. Dewilde, Technische Universiteit Delft  
Dr. F.M. Dekking, Technische Universiteit Delft

Bosveld, Frank

Hierarchical Video Compression using SBC /  
Frank Bosveld - Delft : Technische Universiteit Delft,  
Faculteit der Elektrotechniek. - III.  
Thesis Technische Universiteit Delft. - With ref. - With summary in Dutch.  
ISBN 90-9009-823-02  
Subject headings: video compression / hierarchical coding.

Copyright © 1996 by Frank Bosveld

All rights reserved. No part of this thesis may be reproduced or transmitted in any form or by any means, electronic, mechanical, photocopying, any information storage and retrieval system, or otherwise, without written permission from the copyright owner.

# Contents

<b>Summary</b>	<b>ix</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Interoperability of multimedia applications . . . . .	1
1.2 Scalable video compression . . . . .	2
1.3 Hierarchical coding . . . . .	4
1.4 Scope and outline . . . . .	6
<b>2 Basic techniques</b>	<b>9</b>
2.1 Digital coding . . . . .	9
2.1.1 Memoryless coding . . . . .	11
2.1.2 Predictive coding . . . . .	13
2.1.3 Spectral coding . . . . .	16
2.2 Intraframe subband coding . . . . .	21
2.2.1 Filter bank . . . . .	21
2.2.2 Quantization and entropy coding . . . . .	22
2.2.3 Bit allocation algorithm . . . . .	25
2.3 Implementation of scalability . . . . .	27
2.3.1 Amplitude scalability . . . . .	27
2.3.2 Frequency scalability . . . . .	30
2.3.3 Spatial scalability . . . . .	31
2.3.4 Temporal scalability . . . . .	33
<b>3 Hierarchical coding</b>	<b>35</b>
3.1 Introduction . . . . .	35
3.2 Rate distortion theory . . . . .	36
3.2.1 1-D Gaussian sources . . . . .	36
3.2.2 2-D Gaussian sources . . . . .	38
3.3 Theoretical analysis of SNR scalability . . . . .	38
3.3.1 Description of hierarchical coding . . . . .	39
3.3.2 Conditions for successive refinement . . . . .	39
3.4 Theoretical analysis of spatial scalability . . . . .	41
3.4.1 Definition of successive refinement . . . . .	42
3.4.2 Conditions for successive refinement . . . . .	43
3.4.3 Bit rate and distortion areas . . . . .	46
3.5 Experimental verification . . . . .	48

3.5.1	SNR scalability . . . . .	49
3.5.2	Spatial scalability . . . . .	51
3.6	Discussion . . . . .	55
<b>4</b>	<b>Multirate coding</b>	<b>59</b>
4.1	Introduction . . . . .	59
4.2	Multirate PCM coding . . . . .	60
4.2.1	MR-PCM using multistage quantization . . . . .	61
4.2.2	MR-PCM using embedded quantization . . . . .	65
4.3	Multirate DPCM coding . . . . .	68
4.3.1	MR-DPCM using multistage quantization . . . . .	69
4.3.2	MR-DPCM using embedded quantization . . . . .	72
4.4	Practical multirate coders . . . . .	73
4.4.1	Requirements . . . . .	73
4.4.2	Multistage quantization . . . . .	74
4.4.3	Embedded quantization . . . . .	77
4.5	Experimental results . . . . .	79
4.5.1	Coding scheme . . . . .	79
4.5.2	Coding performances . . . . .	80
4.6	Discussion . . . . .	83
<b>5</b>	<b>Spatially scalable coding for ATM networks</b>	<b>85</b>
5.1	Introduction . . . . .	85
5.2	Video coding for ATM networks . . . . .	86
5.2.1	ATM connections . . . . .	86
5.2.2	ATM traffic contract . . . . .	86
5.2.3	ATM adaptation . . . . .	87
5.2.4	Compatible distribution . . . . .	87
5.3	Intraframe coding scheme . . . . .	88
5.3.1	Encoder and decoder structures . . . . .	89
5.3.2	Adaptive bit/distortion allocations . . . . .	91
5.3.3	Adaptive MR-DPCM and MR-PCM coders . . . . .	92
5.3.4	Synchronization and side information . . . . .	93
5.3.5	Packetization and bit stream shaping . . . . .	95
5.4	Experimental results . . . . .	95
5.4.1	Performance comparison . . . . .	97
5.4.2	Cell-loss resilience . . . . .	98
5.5	Discussion . . . . .	99
<b>6</b>	<b>Spatially scalable coding using MC-SBC</b>	<b>103</b>
6.1	Introduction . . . . .	103
6.2	Subband coding using out-band compensation . . . . .	105
6.3	Subband coding using in-band compensation . . . . .	107
6.3.1	Principle . . . . .	107
6.3.2	Compensation method . . . . .	107
6.3.3	PRMF filter banks . . . . .	109



6.4	Spatial scalability and in-band MC-SBC . . . . .	111
6.5	Interframe coding scheme . . . . .	114
6.5.1	Encoder structure . . . . .	116
6.5.2	Decoder structure . . . . .	116
6.5.3	Coding and prediction methods . . . . .	117
6.6	Experimental results . . . . .	120
6.6.1	Coding performance of low-resolution prediction method . . . . .	120
6.6.2	Coding performances of all prediction methods . . . . .	121
6.7	Discussion . . . . .	123
<b>7</b>	<b>Spatio-temporally scalable coding using 3-D SBC</b>	<b>125</b>
7.1	Introduction . . . . .	125
7.2	Analysis and strategy definitions . . . . .	127
7.2.1	Analysis . . . . .	127
7.2.2	Conversion Strategy . . . . .	129
7.2.3	Direct Strategy . . . . .	130
7.3	Spatio-temporal filter banks . . . . .	131
7.3.1	Non-rectangular filter banks . . . . .	131
7.3.2	Filter bank for Conversion Strategy . . . . .	133
7.3.3	Filter bank for Direct Strategy . . . . .	133
7.4	Spatio-temporal coding schemes . . . . .	137
7.4.1	Encoder structure . . . . .	137
7.4.2	Temporal bit allocation procedure . . . . .	139
7.4.3	Selection Mechanism . . . . .	139
7.5	Experimental results . . . . .	140
7.5.1	Filter bank performances . . . . .	140
7.5.2	Encoding results . . . . .	143
7.6	Discussion . . . . .	148
<b>8</b>	<b>Discussion</b>	<b>149</b>
	<b>Bibliography</b>	<b>151</b>
	<b>List of symbols and abbreviations</b>	<b>161</b>
	<b>Samenvatting</b>	<b>169</b>
	<b>Acknowledgements</b>	<b>175</b>
	<b>Curriculum Vitae</b>	<b>177</b>



# Summary

## Hierarchical Video Compression using SBC

Currently, a new and fascinating area of multimedia applications is emerging. This area embodies the traditionally distinct application areas of telecommunications, broadcasting, computers, and consumer electronics and offers consumers a wide variety of new and enhanced applications. The push behind these applications is the convergence of a number of technologies (digital signal processing, transmission, storage, and integrated circuit design) in combination with a number of standardization activities (JPEG, MPEG, DVB, and DAVIC). The success of these new multimedia applications will depend on the degree of interoperability that will exist between multimedia systems. With interoperability, users can enjoy a large number of applications with a limited number of devices. The degree of interoperability can be increased by compressing and/or representing the video components of multimedia applications in a scalable way. Scalable means here that the video signal can be reproduced at various quality levels by decoding different portions of the compressed data. Such a feature is useful when, for example, a multimedia device has to visualize a high resolution video signal on a low resolution display, or when the delivery of the video signal is error-prone. In the latter case, the scalably compressed video signals can be made to degrade in a graceful way when delivery errors occur.

Video signals are very suitable for scalable compression. This is partly due to the large intelligibility of video signals, which makes them recognizable at a wide variety of signal qualities. For video signals, four basic types of scalability can be defined: namely, amplitude, frequency, spatial, and temporal scalability. Amplitude and frequency scalability are both related to the SNR quality of the signal, while spatial and temporal scalability are related to the spatial resolution and temporal resolution of the video signal, respectively. To implement scalability, multimedia systems may use hierarchical coding schemes. Hierarchical coding schemes can efficiently implement scalability by encoding each reproduction with respect to the previous lower-quality reproduction (except for the lowest-quality reproduction). In particular, hierarchical coding schemes create an ordered set of bit streams, called layers. A layer contains the encoded information that is needed to proceed from one reproduction quality to the next higher reproduction quality. As a result, receivers can reproduce the video signal at various qualities by receiving and decoding the corresponding layers. A drawback of hierarchical coding is that the coding performances of the different reproductions are implicitly dependent on each other.

This thesis focuses on the scalable compression of video signals using hierarchical coding schemes. The main points of interest are the design and implementation of schemes

that achieve an optimal, or nearly optimal, trade-off between bit rate and distortion for each supported reproduction. All proposed schemes use the technique of subband coding (SBC) to compress the input signal as well as to implement scalability. As a result, most chapters in this thesis are dedicated to coding schemes that combine a particular type of scalability with specific compression techniques. The mean square error is used in this thesis as the measure for distortion.

Chapter 3 starts off with a theoretical analysis of hierarchical coding to get insight into the process of hierarchical coding and to get insight into the key issues in the design of such schemes. In particular, it is theoretically analyzed when two elementary hierarchical coding schemes achieve optimal successive refinement of the input signal, i.e. when all reproductions of the schemes are optimally encoded. By assuming a stationary Gaussian distributed input signal, it is shown that i) for SNR scalability (i.e. the combination of amplitude and frequency scalability) optimal successive refinement is always theoretically achievable, and ii) for spatial scalability optimal successive refinement is theoretically only achievable under particular conditions. The theoretical analysis shows that it is essential to encode the coding errors of the low-quality reproduction when encoding the high-quality reproduction. The conditions obtained are verified for both types of scalability using a simple practical hierarchical subband coder. For SNR scalability, the experimental verification shows that in practice optimal successive refinement is not always achieved. In particular, if more than three reproductions are supported, the performance loss compared to a similar non-scalable subband scheme is significant. The cause for this is that the performance of multiple concatenated PCM (or DPCM) coders is less than the performance of a single PCM (or DPCM) encoder used at the equivalent bit rate. For spatial scalability, the experimental verification shows that in practice optimal successive refinement is achieved by the evaluation scheme when two reproductions are supported and when the proper bit rates are used.

Chapter 4 is a follow-up of Chapter 3 and contains an investigation into the performance of multiple concatenated PCM (or DPCM) coders. In particular, specific types of memoryless and predictive coding techniques are investigated, called multirate PCM (MR-PCM) and multirate DPCM (MR-DPCM). MR-PCM and MR-DPCM coders can be used to implement SNR scalability directly and they are also used when other types of scalability are implemented. Multirate coders can be based on either multistage quantization or embedded quantization. With multistage quantization the input signal is quantized by a number of concatenated quantizers which each input the quantization error of the previous quantizer. With embedded quantization the input signal is concurrently quantized by a number of quantizers which have their quantization thresholds aligned. The performances of MR-PCM and MR-DPCM coders based on multistage and embedded quantization are thoroughly investigated using optimally designed dual-rate coders. The analysis of the optimal dual-rate PCM coders shows that coders based on multistage quantization performs less well for the high-quality reproductions than coders based on embedded quantization do. In fact, with embedded quantization the performance for the high-quality reproduction is constant (i.e. it does not depend on the bit rate used to encode the low-quality reproduction) and is competitive with the performance of regular PCM (or DPCM) coders. The performance of the low-quality reproduction is however slightly less than with multistage quantization. As a conclusion of the investigation, novel and competitive MR-PCM and MR-DPCM coders are designed that can be applied in practical hierarchical coding schemes.

Chapter 5 focuses next on spatially scalable coding schemes based on intraframe coding techniques which are designed for ATM networks. ATM can easily support scalable communications, because information is transported via logical connections with negotiable characteristics. ATM transports data through the network in short packets called cells. Unfortunately, cells are occasionally lost in ATM networks because of network congestion and physical layer impairments. Video coding schemes that use ATM networks should therefore be resilient to such errors by using, for example, graceful degradation methods. In Chapter 5 two intraframe transmission schemes are proposed that allow receivers to reconstruct the video signal at multiple spatial resolutions while simultaneously supporting graceful degradation. The coding schemes support 3 reproductions and guarantee the continuity of the application by defining a low-bandwidth CBR layer for the reproduction with the lowest spatial resolutions. In addition, the schemes create two additional high-bandwidth VBR layers to support the other reproductions. One scheme employs multistage quantizers, while the other scheme employs embedded quantizers. For normal network conditions, the simulation results show i) that the performances of the scalable schemes are slightly lower than the performance of a similar non-scalable scheme, and ii) that the scheme based on embedded quantization outperforms the scheme based on multistage quantization. For congested network conditions, the simulations show that the compatible schemes have a much higher error resilience with respect to ATM cell losses than the non-scalable scheme based on the same coding techniques. The error resilience obtained with multistage quantization is slightly higher than the resilience obtained with embedded quantization because the generated layers are less dependent of each other. Based on the experiments of Chapter 5, it is concluded that the expected network performance will influence the architecture of the hierarchical coding scheme.

Chapter 6 concerns spatially scalable coding schemes that are based on interframe coding techniques. For particular applications, e.g. terrestrial TV broadcasting, the use of motion information is imperative to sufficiently compress the video signal. Unfortunately, the integration of motion compensation in spatially scalable subband coding schemes is not straightforward and several difficulties exist. In particular, motion compensation in the subband domain is difficult because subbands of successive frames are not always shifted versions of each other because of aliasing. Further, it is essential that the temporal prediction of a particular reproduction is based on information that is available to the decoder of that reproduction. If this constraint is not fulfilled the reproduction will be distorted because the prediction loops in the encoder and decoder diverge. In Chapter 6 a spatially scalable scheme is designed that is based on interpolative motion compensation in the subband domain (i.e. in-band motion compensation). The scheme supports a number of prediction methods to predict the low-frequency subbands of the high-resolution signal on the basis of low-resolution reference subbands and temporal reference subbands. Depending on the prediction method employed the scheme implements a pure hierarchical coding concept (low-resolution prediction), a simulcast coding concept (temporal prediction), or a concept in between (average prediction and selective prediction). The selective prediction method evaluates a number of candidate predictions and uses the best candidate. From the experimental results it can be concluded that an interframe coding scheme that implements either the low-resolution prediction method or the temporal prediction method can not provide a performance competitive with the performance of a similar non-scalable scheme. When the scheme

employs the selective prediction method it can more successfully minimize the performance loss with respect to a similar non-scalable coding scheme.

Finally, Chapter 7 is focused on two spatio-temporally scalable encoding schemes that support compatibility between progressive-scan video signals and interlaced video signals. To obtain such compatibility, hierarchical encoding schemes are necessary that encode a progressive-scan signal in such a way that it can be reconstructed as an interlaced signal and vice-versa. As compatibility between progressive-scan and interlaced sequences requires a spatio-temporal frequency decomposition, the schemes proposed in Chapter 7 use a three-dimensional (3-D) subband decomposition. The schemes create multiple layers that convey a dynamic collection of encoded 3-D subbands. The first scheme implements the so-called Conversion Strategy, which provides compatibility between progressive-scan signals only. For interlaced signals a lattice conversion is required. The second scheme implements the so-called Direct Strategy and supports compatibility between all signals except for the interlaced signal with the highest spatial resolution. For this signal a lattice conversion is still required to solve the temporal ambiguities inherently associated with the definition of the signals. However, the Direct Strategy filter bank decomposes the input spectrum in such a way that the frequency support of this interlaced signal can exactly be recovered from the subbands, thereby making dedicated coding methods for this signal possible. Experimental results show that the additional complexity of the Direct Strategy scheme does not pay off in terms of coding performance. In particular, the simpler Conversion Strategy scheme outperforms the Direct Strategy scheme by significant values. The results show that the moderate quality of the vertical-temporal filter bank of the Direct Strategy scheme limits its practical applicability. The Conversion Strategy scheme does however outperform an intraframe subband coding scheme for the highest quality reproduction. Hence, temporal subband decompositions can improve the coding performance with respect to subband schemes that use only spatial decompositions. Unfortunately, the subjective quality of the reconstructed progressive-scan and interlaced signals is not very high, because blurred contours and halo images are present in the reconstructed signals. To avoid these artifacts while maintaining the performance gain, the temporal subband filtering should be done along the motion trajectories, as recently proposed in literature.

Based on the results presented in this thesis, it can be concluded that hierarchical coding schemes based on SBC can efficiently implement scalability to support compatibility and graceful degradation in error-prone networks. In general, a loss in coding performance has to be accepted, which can however be minimized by designing the scheme carefully.

# Chapter 1

## Introduction

### 1.1 Interoperability of multimedia applications

Currently, a new and fascinating area of multimedia applications is emerging. This area embodies the traditionally distinct application areas of telecommunications, broadcasting, computers, and consumer electronics and offers consumers a wide variety of new and enhanced applications. Such applications include, for example, high quality audio/video CD, high bandwidth Internet access, multimedia games, Video-on-Demand, Video Telephony, and the broadcasting of digital TV and High Definition TV (HDTV). The push behind these applications is the convergence of a number of technologies (digital signal processing, transmission, storage, and integrated circuit design) in combination with a number of standardization activities (JPEG, MPEG, DVB, and DAVIC). These technologies and standardization activities make it possible to develop and realize new multimedia systems at the required economies of scale [LeGa91, Chia93b, Wood95, Chia95, Schr95].

To stimulate the success of multimedia applications, the Digital Audio-Visual Council (DAVIC) is currently standardizing a set of specifications [DAVI96]. The purpose of these specifications is to achieve interoperability between multimedia systems for interactive applications. Interoperability can be defined as the flawless interworking of these systems and the successful operation of the multimedia applications and services provided. When interoperability is facilitated, users can enjoy a large number of applications with a limited number of devices. Unfortunately, the interoperability of multimedia systems is complicated by several factors. One factor is that multimedia applications will use a wide variety of digital assets to make the applications as attractive as possible. These assets may be for instance compressed audio and video signals, graphics, and text. In addition, applications will use these assets at various qualities and/or resolutions, depending on the application requirements. The diversity in assets complicates interoperability since typical, cost-effective multimedia systems can only support a limited number of assets (both in type and quality). A second factor is that multimedia applications will use an increasing number of delivery systems and media. Types of storage media applied are, for example, CD-ROM and high-density CD, while types of distribution and communication media applied include terrestrial and satellite networks, hybrid-fiber-coax (HFC) networks, and ATM networks. Because each delivery system has its own delivery characteristics (e.g. capacity, latency, and bit error rates), interoperability is not straightforward.

Interoperability between multimedia systems, applications, and services can be facilitated by adopting the concept of open-system architectures [Tane81, Bove92, Chia95]. The concept of open-system architectures is that the different subsystems in the overall system are functionally decoupled from each other by unambiguously defined interfaces. These subsystems could be, for example, the service provider system, the delivery system, and the multimedia devices at the consumer premises. The interfaces between these subsystems specify the functionality of particular layers in the subsystems. For example, the interfaces define which digital assets are supported, how these assets are packetized before transmission, and how the packetized assets are adapted for transmission via a particular delivery channel. In general, open systems can interact with each other as long as they have the same functionalities in corresponding layers. Consequently, the flexibility, extensibility, and interworking capabilities of these systems is enormously increased compared to 'closed' or proprietary systems.

## 1.2 Scalable video compression

The concept of scalability can further increase the interoperability between multimedia systems, applications, and services [Elle90, Gilg91, Lipp91, Bove92]. With scalable open systems, some assets of the application are compressed and/or represented in a scalable way. Scalable means here that the assets can be reproduced at various quality levels by decoding different portions of the compressed data. Such a feature is useful when, for example, a multimedia device has to decode and visualize a high resolution video signal, e.g. HDTV, on a low resolution display [Tzou88, Bosv90, Lipp91, Uz91b, Scha92]. In addition, scalability is also valuable when the delivery of the assets is error-prone, because scalably compressed assets can be made to degrade in a graceful way when delivery errors occur [Bell88, Lee90, Morr91]. Consequently, the continuity of the multimedia application can be guaranteed even if delivery systems with high error rates are used. This feature is generally known as graceful degradation. In general, scalability increases interoperability because the autonomy of multimedia decoding systems is increased. Such autonomous systems are then capable of controlling the reproduction qualities of the application assets, and can thus support applications and services with a limited amount of resources.

Video signals are very suitable for scalable compression. This is partly due to the large intelligibility of video signals, which makes them recognisable at a wide variety of signal qualities. In addition, the processing of video signals in a multimedia system generally requires the majority of the resources. With scalability, the required resources can be reduced for particular applications and services. For video signals, four types of scalability can be defined; namely, amplitude, frequency, spatial, and temporal scalability. Figure 1.1 illustrates these types for either an image or an image sequence. The image or image sequence is shown at five discrete levels of quality for all types of scalability. For amplitude scalability, quality refers to the accuracy with which the amplitudes of the image samples are quantized. As an example, the lowest quality level in Fig. 1.1 has only one possible amplitude level, while the highest quality level has 16 discrete amplitude levels. For frequency scalability, quality refers to the range of frequencies used in the reproduction process of the image. The combination of amplitude and frequency scalability is commonly called SNR (signal-to-noise ratio) scalability because both affect the amount of distortion, or coding noise, in the reproduced image. For spatial scalability, quality refers to the spatial resolution or spatial



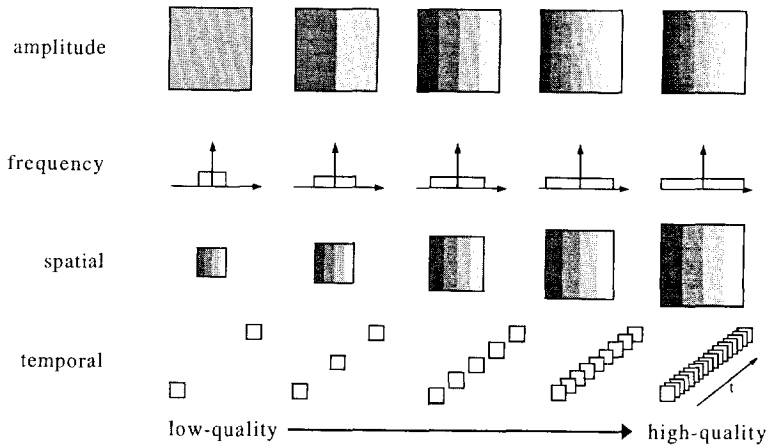


Figure 1.1: Types of scalability for video signals.

sampling rate of the image. Fig. 1.1 illustrates this by varying the size of the reproduced images. For image sequences, temporal scalability involves the temporal resolution, or frame rate, of the image sequence. The combination of spatial and temporal scalability is commonly called spatio-temporal scalability.

Multimedia applications and services may use multiple types of scalability to compress their video signals. The choice of which scalability types to use depends on the application requirements and the expected characteristics of the foreseen delivery systems. In the following, five applications and services are described as an example:

- Compatible broadcasting of TV and HDTV via terrestrial channels [Long92, Lame93]. With the compatible broadcasting of TV and HDTV, receivers should be able to reproduce the compressed HDTV signal at either TV resolution or HDTV resolution. Since TV signals have approximately half the spatial resolution of HDTV signals, spatial scalability can facilitate compatibility. With terrestrial broadcasting, the transmitter may support hand-held portable receivers as well as stationary home-sets. Graceful degradation of the HDTV signal is a requirement here since errors in the terrestrial channel are inevitable; namely, errors will occur due to the limited size of portable antennae, channel fading, multipath interference and varying distances to the transmitter. As a result, the HDTV signal should also be compressed in an SNR-scalable way to provide for sufficient error resilience.
- Multimedia communications via ATM networks [Morr91, Gilg91]. ATM-based networks facilitate the transmission of a variety of signals (audio, video and data) that may have completely different bit stream characteristics (CBR, VBR, and bursty). Multimedia applications that use ATM networks as delivery systems are rapidly emerging. An example of such an application is video telephony via desktop workstations, where people communicate with each other via on-screen windows. Typically, communication in such environment involves size adjustments of the on-

screen window, which implies a change in required spatial resolution from a scalable coding point of view. Spatially scalable compression schemes can support this feature. Graceful degradation (and thus SNR scalability) is of high importance for ATM-based services as well. Although the error probabilities are very low for normal network conditions, large bursts of transmission errors may occur in the case of network congestion [Pryc91, LeBo92].

- Image and video archive systems based on optical storage [Phil94].  
Image and video archive applications such as photo-CD are becoming increasingly popular. When the compressed images and video sequences are stored in multiple spatial resolutions, detailed magnifications of user-selected areas are facilitated. Further, visual search procedures are facilitated by using lower resolution images/sequences as browsing pictures.
- Digital video recording based on magnetic tapes [With92, Frim95].  
Scalable video compression can be used in the digital recording of video signals as well. Since magnetic recording channels are not very reliable, SNR scalability can be applied to provide a robust picture quality. Further, SNR and temporal scalability can be used for particular trick modi such a fast-forward and slow-motion.
- Progressive transmission via slow telephone lines [Tzou87, Burt83].  
Progressive image transmission allows receivers to initially reconstruct an approximate image based upon partially received information and to subsequently receive in several passes additional information about the image details. By doing this, receivers have the opportunity to quickly recognize the image content and to interrupt the transmission at an early stage if desired. Progressive transmission has various potential applications in the area of interactive image retrieval in low speed transmission environments. Typical applications are, for example, teleshopping and telebrowsing for medical and publishing purposes. Because the SNR quality of the transmitted image steadily increases with each transmission pass, progressive transmission can be supported by compressing the image in an SNR-scalable way.

### 1.3 Hierarchical coding

To support scalability, multimedia systems have to use compression schemes that can implement scalability. Such schemes have to compress, or encode, the video signal in such a way that receivers can reproduce the video signal at different qualities. Basically, there are two kinds of coding techniques that implement scalability; namely, simulcast coding and hierarchical coding.

Simulcast coding is a simple way to achieve scalability [Civa92, Morr93b]. In simulcasting the video signal is encoded at all required resolutions and SNR qualities independently and is transmitted in multiple subchannels. Receivers can reproduce the video signal at the desired quality by decoding the appropriate bit stream. An advantage of simulcast coding is that for each reproduction an optimal trade-off between bit rate and distortion can be realized in the encoding scheme without affecting the coding performance of other reproductions. A serious drawback is that simulcasting is a redundant concept since some signal components,

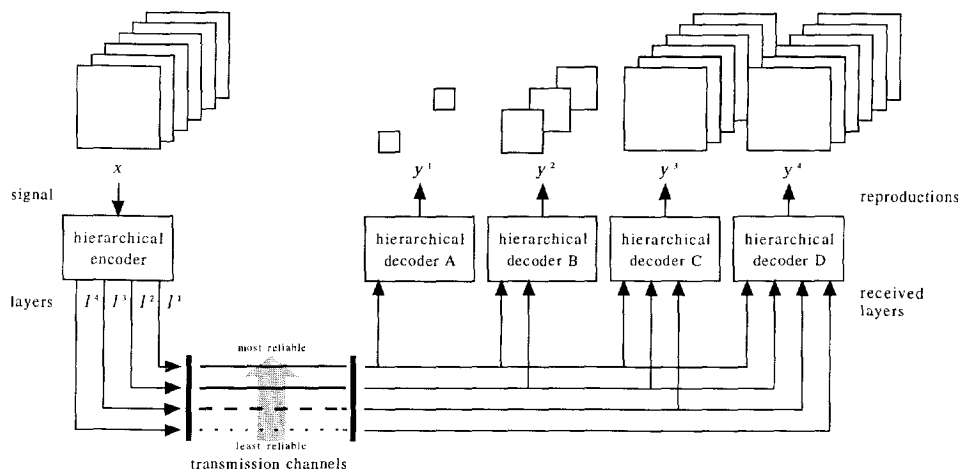


Figure 1.2: Example of hierarchical coding.

such as the low spatial and temporal frequencies, are transmitted several times. Especially when a large number of reproductions is supported, this concept uses the bandwidth of the delivery system rather inefficiently.

Hierarchical coding is a more complex way to achieve scalability but is generally more efficient in bandwidth usage than simulcast coding [Tzou88, Bosv90, Vand91a]. Hierarchical coding schemes facilitate scalability by encoding each reproduction with respect to the previous lower-quality reproduction (except for the lowest-quality reproduction). Hence, duplications in encoded signal components are inherently avoided. To accomplish this, hierarchical coding schemes create an ordered set of bit streams, which are henceforth called layers. A layer contains the encoded information that is needed to proceed from one reproduction quality to the next higher reproduction quality.

The first layer conveys the low spatial and temporal frequencies and facilitates the reproduction of the signal at the lowest quality level. Subsequent layers convey enhancement information and facilitate, in combination with previous layers, the reproduction of the signal at higher quality levels. As a result, receivers can reproduce the video signal at various qualities by receiving and decoding the corresponding layers. The main advantage of hierarchical coding schemes is the efficient usage of bandwidth in the delivery system. In fact, hierarchical coding schemes can transmit a scalably compressed video signal in the same bandwidth normally required to transmit a non-scalably compressed signal. A drawback of hierarchical coding is that the trade-offs between bit rates and distortions of the reproductions have become dependent on each other. This may most affect the coding efficiency of the reproduction with the highest quality level, because it depends on the coding quality of all other reproductions.

As an example, Figure 1.2 shows a hierarchical coding scheme that facilitates the reproduction of a video signal at four quality levels. The video signal is denoted by  $x$ , while the set

of reproductions is denoted by  $y^1, \dots, y^4$ . The scheme creates four layers, which are denoted by  $l^1, \dots, l^4$ . Layer  $l^1$  conveys the compressed data that is necessary to construct  $y^1$ , i.e. the reproduction with the lowest spatial and temporal resolutions. The other reproductions have increased spatio-temporal resolutions compared to  $y^1$ . Signal  $y^4$  has the same spatio-temporal resolution as  $y^3$  but has a higher SNR quality.

Fig. 1.2 also shows how hierarchical coding schemes generally support graceful degradation; namely, the most important layers are transmitted in the most reliable subchannels of the delivery system. Hence, when errors occur in the subchannels, the least important layers are affected first. This will degrade the SNR quality of the reproduction in a graceful way because the most important layers are still received error-free. Many delivery systems support such subchannels with different reliabilities, for example, terrestrial channels based on 64 MR-QAM [Uz91a, Grun92], cell-based communication in ATM [Pryc91, LeBo92, ATM 93], and magnetic tape recording devices [With92, Frim95]. Alternatively, the hierarchical coding schemes may create different subchannels by employing error-correcting codes with different capabilities.

## 1.4 Scope and outline

This thesis focuses on the scalable compression of video signals that are employed by multimedia systems. In particular, this thesis is about hierarchical coding schemes that efficiently facilitate compatibility and robust communications through error-prone networks. The main points of interest are the design and implementation of schemes that achieve an optimal, or nearly optimal, trade-off between bit rate and distortion for each supported reproduction. The proposed schemes use the technique of Subband Coding (SBC) to compress the input signal as well as to implement scalability. SBC has been chosen as the basic data compression technique because it allows for high compression factors [Wood86, West89] and can implement several types of scalability [Tzou87, Tzou88, Bosv90, Bosv92b, Ohm92, Lage96]. As a result, major parts of this thesis concern the coding efficiency of schemes that combine SBC with other kinds of coding techniques which are necessary to implement the required types of scalability.

Throughout this thesis, the Mean Square Error (MSE) is used as the measure for distortion. Although it is recognized that the MSE does not precisely reflect the subjective quality experienced by humans, it is mathematically tractable. As such, it can be used as a guideline to design practical coding schemes from a perceptual point of view. Further, most experiments in this thesis concentrate on the performance loss caused by the introduction of scalability in the compression procedure. For this purpose, it is believed that the MSE measure is sufficient.

Chapter 2 starts out with a review of the basic techniques that are frequently employed by SBC schemes. The review of these techniques consists of three parts, in which the concepts of the employed memoryless coding technique (i.e. PCM), the employed predictive coding techniques (i.e. DPCM, ME/MC), and the employed spectral coding technique (i.e. SBC) are discussed. Next, the various parts of an intraframe subband coding scheme are described. This scheme serves as the basis for all hierarchical coding schemes proposed in this thesis. The parts described include the filter bank, the bit allocation procedure, the quantization section, and the entropy coding section. Finally, Chapter 2 concludes with a survey about how the various types of scalability can be implemented.

A theoretical analysis of hierarchical coding is described in Chapter 3, so that insight can be obtained into the process of hierarchical coding and the key issues in the design of such schemes. The analysis pertains to an elementary scheme that supports two reproductions of the input signal with different spatial resolutions. To keep the analysis mathematically tractable, the input signal is assumed to be a stationary Gaussian source. Although video signals are generally neither Gaussian distributed nor stationary, the analysis provides insight into the requirements of hierarchical coding. As a result, the analysis yields a set of conditions that expresses when both reproductions are optimally encoded and when they are not. The analysis provides valuable guidelines for the design of practical hierarchical coding schemes.

Chapter 4 concerns specific types of memoryless and predictive coding techniques, called multirate PCM (MR-PCM) and multirate DPCM (MR-DPCM). MR-PCM and MR-DPCM coders can be used to implement SNR scalability, but they are also employed in the proposed coding schemes to implement spatio-temporal scalability. The performances of MR-PCM and MR-DPCM coders are thoroughly investigated for different types of quantizers and entropy coders. Consequently, novel MR-PCM and MR-DPCM coders are presented which are competitive with regular (single-rate) PCM and DPCM coders that are based on state-of-the-art quantizers and entropy coding techniques. Chapter 4 concludes with an investigation of the performance of a hierarchical coding scheme for progressive transmission. The scheme implements SNR scalability by combining the intraframe subband coding system of Chapter 2 with MR-PCM and MR-DPCM coders.

Chapter 5 focuses on spatially scalable video coding for ATM networks. ATM can easily support scalable communications, because information is transported via logical connections with negotiable characteristics. In addition, ATM can provide bandwidth on demand, which makes constant quality video communications with variable bit rates possible. In Chapter 5 two intraframe transmission schemes are proposed that allow for the transmission of the video signals at multiple spatial resolutions in ATM-based networks. In particular, the schemes are evaluated for the layered distribution of progressive-scan HDTV. The schemes create a low-resolution constant bit rate layer that guarantees the continuity of service during network congestion. The second and third layers convey data for additional spatial resolution and facilitate the reconstruction of two constant quality reproductions of the input signal (medium and high spatial resolution). Consequently, the bit rates of these layers are variable in time.

Chapter 6 concerns spatially scalable coding schemes that are based on interframe coding techniques. For particular applications, such as the compatible broadcasting via terrestrial channels, the use of motion information is imperative to sufficiently compress the video signal. Unfortunately, the integration of motion compensation in spatially scalable subband coding schemes is not straightforward and several difficulties exist. In Chapter 6 the most prominent problems are discussed and a spatially scalable coding scheme that supports two reproductions is proposed. The scheme employs a so-called Perfect Reconstruction Modulated Filter (PRMF) filter bank to facilitate compatible coding.

Finally, Chapter 7 is focused on two spatio-temporally scalable encoding schemes that support compatibility between progressive-scan video signals and interlaced video signals. As compatibility between progressive-scan and interlaced sequences requires spatio-temporal frequency decompositions, the schemes employ three-dimensional (3-D) subband decompositions. In Chapter 7 two dedicated decompositions are proposed that both consist of two concatenated filter banks: a vertical-temporal filter bank and a horizontal filter bank. The schemes create multiple layers that convey a dynamic collection of encoded 3-D subbands.

A so-called Selection Mechanism determines which subbands are transmitted in which layers. The Selection Mechanism makes this choice such that i) the bit rates of the layers are approximately constant in time, and ii) the distortions in the reproduced signals are minimized.

# Chapter 2

## Basic techniques

This chapter summarizes the basic techniques that are used in this thesis and presents the terminology and notations used as well. Section 2.1 starts out with an overview of the concept of digital coding and describes three specific classes thereof; namely, memoryless coding (specifically PCM), predictive coding (DPCM, ME/MC), and spectral coding based on subband decompositions. In this order, the techniques allow for increasingly higher compression factors but are also of increasing complexity. Next, in Section 2.2 a practical coding scheme is introduced that is based on intraframe subband coding. This scheme serves as the basis for all hierarchical coding schemes proposed in this thesis. The filter bank, the quantization and entropy coding section, and the bit allocation procedure are among the parts discussed. Finally, it is presented in Section 2.3 how the various types of scalability can be implemented. It is shown that amplitude scalability can be implemented by specific types of memoryless and predictive techniques which are known as multirate PCM (MR-PCM) and multirate DPCM (MR-DPCM). In addition, spatial and frequency scalability can be implemented by the spectral technique of subband coding. Temporal scalability can be implemented by both predictive and spectral techniques; the first employs the technique of motion estimation (ME) and motion compensation (MC) while the latter is based on 3-dimensional (3-D) subband coding.

### 2.1 Digital coding

Figure 2.1 shows the principles of digital coding. The prefilter and sampler convert the analog, i.e. continuous-amplitude and continuous-time, input signal  $x(t)$  into the discrete-time signal  $x(kT)$ . Here,  $T$  is the sampling interval and is the reciprocal of the sampling rate  $f_s$ , i.e.  $T = 1/f_s$ . After sampling, an encoder converts the continuous-amplitude samples  $x(kT)$  into a stream of binary digits, or bits. This conversion involves the crucial step of amplitude discretization, or quantization, which largely determines the distortion of the communication system as well as the resulting number of bits required to represent  $x(kT)$ . At the receiving side, a decoder converts the bit stream back into a discrete-time signal  $y(kT)$ . Signal  $y(kT)$  is an approximation of  $x(kT)$  because of the amplitude discretization in the encoder. Finally, a reconstruction filter derives the continuous-time signal  $y(t)$  from the reconstructed discrete-time signal  $y(kT)$  by using an appropriate interpolation procedure. For simplicity, it is henceforth assumed that  $T = 1$ .

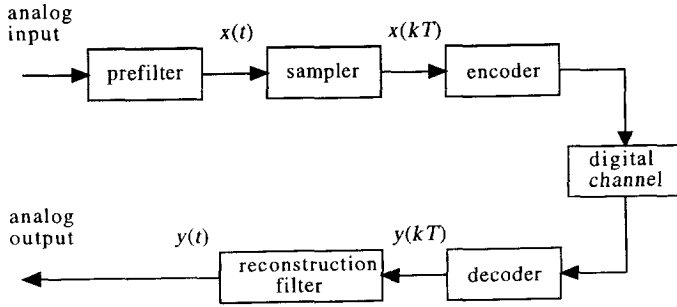


Figure 2.1: Principle of digital coding.

In general, the encoder tries to achieve an optimal trade-off between the number of bits to represent  $x(k)$  and the magnitude of the distortion introduced. The distortion  $D$  that is introduced by the encoder is often measured by the variance of reconstruction error. The reconstruction error variance, denoted by  $\sigma_r^2$ , is given by (assuming  $r(k)$  is zero-mean)

$$D = \sigma_r^2 = E[(r(k))^2] = E[(x(k) - y(k))^2], \quad (2.1)$$

where  $r(k) = x(k) - y(k)$  denotes the discrete-time reconstruction error. If the number of available bits is fixed, the encoder tries to minimize  $\sigma_r^2$ . By doing this, the coder also minimizes the continuous-time reconstruction error  $r(t) = x(t) - y(t)$  because the variance of  $r(k)$  is equal to the variance of  $r(t)$ , provided that  $x(t)$  is properly sampled [Jaya84].

The reconstruction error variance is often referred to as the Mean Square Error (MSE) distortion. A relative measure for the MSE distortion is the Signal-to-Noise Ratio (SNR) measure. The SNR is defined as

$$\text{SNR} = 10 \log_{10} \left( \frac{\sigma_x^2}{\sigma_r^2} \right) \quad (\text{dB}), \quad (2.2)$$

and is expressed in decibels. Here,  $\sigma_x^2$  denotes the variance of  $x(k)$ .

The transmission rate  $I$  of the digital communication system of Fig. 2.1 is the product of the sampling rate  $f_s$  and the mean number of bits per sample  $R$  used by the encoder to represent the amplitudes of the samples:

$$I = f_s R \quad \text{bits per second (bps)}. \quad (2.3)$$

When  $f_s$  is known and fixed, the transmission rate of the system can also be specified by the mean number of bits per sample or per pixel:

$$R \quad \text{bits per sample or pixel (bpp)}. \quad (2.4)$$

In this thesis, the terms transmission rate and bit rate are used to represent  $I$  as well as  $R$ , depending on the context.



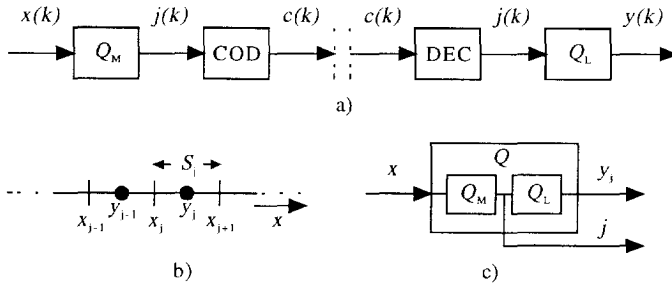


Figure 2.2: Memoryless coding: a) PCM encoder and decoder, b) quantizer decision and representation levels, c) quantizer symbols.

Generally, the operations of prefiltering, sampling, and reconstruction filtering are tacitly assumed when designing digital communications schemes. The design concentrates on the performance, or efficiency, of the coding methods implemented in the encoder and decoder. In this thesis, the term coding performance refers to the trade-off between the bit rate of a coding scheme and the distortion introduced. The next sections concisely describe three classes of coding techniques used, i.e. memoryless coding, predictive coding, and spectral coding based on subband decompositions.

### 2.1.1 Memoryless coding

Memoryless coding schemes employ an instantaneous quantization procedure to discretize the sample amplitude  $x(k)$  at time  $k$ . This implies that the amplitude quantization at time  $k$  is not affected by earlier or later samples. The technique of memoryless coding is commonly referred to as Pulse Code Modulation (PCM). PCM schemes perform best when the input samples are mutually statistically independent.

Figure 2.2a shows the encoder and decoder of a PCM coding scheme in detail. They are interconnected by means of a digital channel, which is henceforth represented by two vertical dashed lines. The PCM encoder transmits a unique codeword  $c(k)$  for each sample  $x(k)$ . The encoder consists of two components to accomplish this; namely, a mapping  $Q_M$ , which maps  $x(k)$  into the index  $j(k)$ , and a lossless encoder COD, which translates the index  $j(k)$  into the appropriate codeword  $c(k)$ . The encoder represents  $x(k)$  by the  $j^{\text{th}}$  codeword if it falls in the interval

$$S_j : \{x_j < x \leq x_{j+1}\}, \quad \text{for } j = 1, 2, \dots, K. \quad (2.5)$$

After transmission, the PCM decoder converts the received codeword  $c(k)$  into  $y(k)$ , which has the value

$$y(k) = y_j. \quad (2.6)$$

The decoder uses reconstruction value  $y_j$  to represent all amplitudes of interval  $S_j$ . To accomplish this, the PCM decoder consists also of two components. These are the lossless decoder DEC, which translates  $c(k)$  back into the corresponding index  $j(k)$ , and the look-up

table  $Q_L$ , which defines the reconstruction value  $y_j$  for each interval  $S_j$ . Fig. 2.2b shows the relation between the decision levels  $x_j$ , the representation levels  $y_j$ , and the decision interval  $S_j$ . The relation between  $x(k)$  and  $y(k)$  is called the quantization characteristic (denoted by  $Q$ ) and is given by

$$y(k) = Q(x(k)) = Q_L(Q_M(x(k))). \quad (2.7)$$

This is shown in Fig. 2.2c, where it is also shown that index  $j$  can be considered as an output of  $Q$ . In the rest of this thesis, the term quantizer is used for both  $Q$  and  $Q_M$ .

The overall distortion of the PCM coding scheme consists of the reconstruction error  $r(k)$ , i.e. the error between the input signal  $x(k)$  and the output signal  $y(k)$ . For PCM,  $r(k)$  is equal to the quantization error  $q(k)$ , which is defined as

$$q(k) = x(k) - Q(x(k)). \quad (2.8)$$

To calculate the distortion introduced by  $Q$ ,  $x(k)$  is regarded as the outcome  $x$  of a real-valued random variable (r.v.)  $X$  with Probability Density Function (PDF)  $p_X(\cdot)$ , mean  $u_x = E[X] = 0$ , and variance  $\sigma_x^2 = E[X^2]$ . Consequently,  $D$  is given by

$$D = \sigma_q^2 = \sum_{j=1}^K \int_{x_j}^{x_{j+1}} (x - y_j)^2 p_X(x) dx, \quad (2.9)$$

with  $x_1 = -x_{K+1} = -\infty$ .

The transmission rate of PCM is given by  $I = f_s R$ . Bit rate  $R$  is determined by the mean length of the codewords used to represent the index  $j$ , i.e.

$$R = \sum_{j=1}^K P_j n_j \quad (\text{bits/sample}). \quad (2.10)$$

Here,  $n_j$  denotes the integer length of the  $j^{\text{th}}$  codeword and  $P_j$  signifies the probability of occurrence of index  $j$ , i.e.

$$P_j = P(X \in S_j) = \int_{x \in S_j} p_X(x) dx. \quad (2.11)$$

The average information content, or entropy, of indices  $j$  constitutes a lower bound for  $R$ . The entropy of the indices, denoted by  $H(J)$  with  $J = Q_M(X)$ , is defined as

$$H(J) = - \sum_{j=1}^K P_j \log_2 P_j, \quad (2.12)$$

so that

$$R \geq H(J). \quad (2.13)$$

To realize bit rates close to  $H(J)$ , the lossless coder generally employs entropy coding techniques [Huff52, Jaya84, Lang84, Witt87]. Entropy coders assign short codewords to indices with high occurrence probabilities and long codewords to indices with low occurrence probabilities. In the case that  $P_j$  and  $n_j$  are related by

$$P_j = 2^{-n_j} \quad (2.14)$$

an entropy code can be constructed that satisfies the lower bound in Eq. 2.13.

From the above it is clear that the quantizer characteristic has to be matched to the input PDF to optimize the coding performance. In the literature, different optimization procedures have been proposed to design quantizers; for example, Lloyd-Max Quantizers (LMQ) [Max60, Lloy82], Uniform Threshold Quantizers (UTQ) [Wood69, Farv84], and rate-constrained quantizers [Farv84]. In the rest of this thesis, two kinds of PCM coders are used. These coders consist of either a Lloyd-Max Quantizer followed by a Huffman coder or a Uniform Threshold Quantizer followed by a lossless Arithmetic Coder (AC).

### 2.1.2 Predictive coding

Predictive coding schemes facilitate the efficient digital transmission of signals whose samples are statistically dependent. Compared to memoryless schemes, they realize an increased coding performance by exploiting the correlation among neighboring samples. The existence of correlation in the input signal allows the prediction of future samples from past samples. Predictive schemes exploit this predictability by removing the predictable part from the input before quantization and by adding the predictable part again afterwards. By doing so, they only have to encode the unpredictable part, which generally has a much lower variance than the input signal. Predictive schemes are also referred to as Differential PCM (DPCM) schemes because frequently PCM is used to encode the difference between the input signal and the prediction. For images and image sequences, predictive schemes can exploit both spatial and temporal correlations. Spatial correlation is usually removed by means of linear prediction. For image sequences, temporal correlation is usually removed by means of a temporally compensated prediction. Temporally predictive schemes are usually referred to as motion compensated (MC) schemes because temporal correlations between subsequent frames are closely related to motion information.

Figure 2.3 illustrates the principle of forward predictive coding. The input to quantizer  $Q_M$  is the prediction error, or difference signal

$$d(k) = x(k) - \hat{x}(k). \quad (2.15)$$

Here,  $\hat{x}(k)$  is the prediction of input signal  $x(k)$ . The prediction  $\hat{x}(k)$  is the result of predictor  $P$ , which uses past reconstructed samples  $y(k-l)$  with  $0 < l \leq L$  to predict  $x(k)$ , i.e.

$$\hat{x}(k) = P(y(k-l)), \quad \text{for } 0 < l \leq L. \quad (2.16)$$

The predictor uses the samples  $y(k-l)$  because these are available in both the encoder and decoder. In fact, sample  $y(k)$  is reconstructed by adding the prediction  $\hat{x}(k)$  to the quantized prediction error  $u(k)$ , i.e.

$$y(k) = u(k) + \hat{x}(k) = Q(d(k)) + \hat{x}(k). \quad (2.17)$$

The performance of DPCM coders can be expressed relative to the performance of PCM. Specifically, when DPCM and PCM encoders have identical bit rates  $R$ , the SNR measures for DPCM and PCM are related by

$$\text{SNR}_{DPCM} \approx 10 \log_{10}(G_P) + \text{SNR}_{PCM}, \quad (2.18)$$

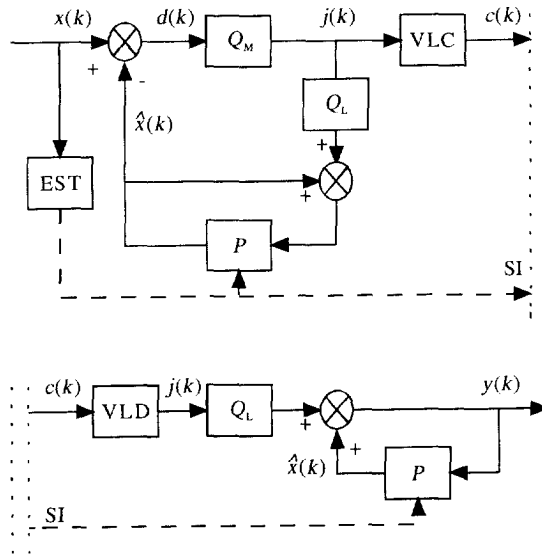


Figure 2.3: Forward predictive coding: a) encoder, b) decoder.

with

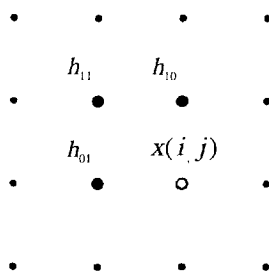
$$G_P = \frac{\sigma_x^2}{\sigma_d^2}, \quad (2.19)$$

the so-called prediction gain, and  $\sigma_d^2$  the variance of the difference signal. Eq. 2.18 shows that DPCM outperforms PCM because the variance of the signal that is quantized is reduced from  $\sigma_x^2$  to  $\sigma_d^2$  by the predictor. The performance of DPCM coders can therefore be optimized by tuning the prediction algorithm such that the prediction error variance is minimized. It should be noted that Eq. 2.18 is not precisely exact because coding performance of the applied PCM coder will vary slightly for different prediction methods. This is because the performance also depends on the shape of the PDF of the input signal, which will vary a little for different prediction methods [Jaya84].

For an optimal prediction, the prediction method requires detailed information about the actual correlations among the input samples  $x(k)$ . For this purpose, the forward predictive encoder shown in Fig. 2.3 contains an estimator (EST) that periodically estimates the sample correlations in a segment of samples. To facilitate the same prediction at the receiver, the parameters derived have to be losslessly transmitted as side information (SI).

### 2.1.2.1 Spatially predictive schemes using linear prediction

A popular predictor to predict images is the linear predictor of order  $N$ . This predictor exploits the existing spatial correlations by predicting the image samples by creating a weighted sum of  $N$  reconstructed samples. As an example, Figure 2.4 shows the samples and their corresponding prediction coefficients  $h_{i,j}$  that may be used in the prediction of sample  $x(i, j)$  by an order 3 predictor.

Figure 2.4: Two-dimensional prediction of  $x(i, j)$ .

Let  $\mathbf{h}$  denote the vector of  $N$  predictor coefficients, i.e.

$$\mathbf{h} = (\dots, h_{i,j}, \dots)^T. \quad (2.20)$$

Then, at high bit rates, the set of optimal prediction coefficients  $\mathbf{h}_{opt}$  is given by the Yule-Walker equations [Jaya84], i.e.

$$\mathbf{h}_{opt} = \mathbf{R}_{xx}^{-1} \mathbf{r}_{xx}. \quad (2.21)$$

Here,  $\mathbf{R}_{xx}$  denotes the autocorrelation matrix of signal  $x$  and  $\mathbf{r}_{xx}$  denotes the vector that contains the autocorrelation values at the positions of the reconstructed samples used.

### 2.1.2.2 Temporally predictive schemes using displacement information

Temporally predictive schemes exploit the temporal correlations that exist in video signals. They realize this by making a prediction based on the samples of the previous reconstructed image. When the samples of the  $k^{th}$  image are denoted by  $x(i, j, k)$ , the temporal prediction is given by

$$\hat{x}(i, j, k) = y(i + \delta_i, j + \delta_j, k - 1), \quad (2.22)$$

where  $\mathbf{d} = (\delta_i, \delta_j)^T$  is an integer displacement vector. For an optimal prediction, the estimator has to determine for each input sample the optimal displacement, or motion, vector, i.e. the vector that points to the maximally correlated samples of the previous reconstructed image. Commonly, the temporally prediction process is known as motion estimation (ME), while the subtraction of the prediction is known as motion compensation (MC).

Schemes that employ temporal prediction have to transmit the motion vectors as side information. To keep the bit rate of the side information at small values, the schemes divide the input frame  $x(i, j, k)$  into blocks and use only one vector  $\mathbf{d}$  per block. This is illustrated in Figure 2.5. The estimation algorithm tries to find the best vector from a finite set of candidate vectors with respect to an appropriate cost function, such as the MSE of the matching error. The candidate vectors are generally integer translations, and the best vector is found by direct search techniques such as brute force search, conjugate directions search, and three-step search (see for example [Musm85, Drie92]). More advanced algorithms incorporate a priori assumptions about spatial and temporal correlations between displacement vectors of neighboring blocks. They impose smoothness by, for example, hierarchical processing

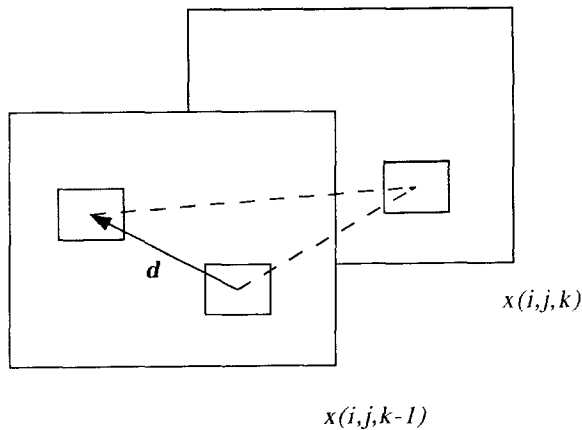


Figure 2.5: Block-based temporal correlations.

techniques [Bier88], recursive processing techniques [Haan92], or assuming an explicit smoothness model [Reut88].

### 2.1.3 Spectral coding

Like predictive schemes, spectral schemes facilitate the efficient compression of correlated signals. Spectral schemes remove linear redundancy in the input signal by decomposing the input signal into a number of separate components. The components are mutually decorrelated and convey particular frequency bands of the input signal. The decorrelated components are separately encoded by means of a dedicated coding method. After transmission, the decoded components are combined into an approximation of the input signal. The advantage of spectral schemes over DPCM and PCM is that the number of bits used to encode each component can be variable. Therefore the most important components can be quantized accurately, while less important components can be quantized coarsely.

The class of spectral coding schemes consists of many subclasses that differ in the way they transform the input signal into the set of decorrelated components. Typical transforms are, for example, subband transforms, orthogonal block transforms (e.g. KLT, DCT, DWT), lapped orthogonal transforms (LOT), short-time Fourier transforms, wavelet transformations, Gabor transformations, and pyramidal schemes (see e.g. [Jaya84, Vett87, Mall89, Daub90, Riou91, Gers92]). All transforms describe the input signal as a weighted sum of basis functions but employ different sets of basis functions. Subband transforms are of particular interest for image coding applications since they provide largely decorrelated components to the encoding stage and facilitate a high coding performance [Vett84, Wood86, West89]. Further, many of the other transforms can be expressed as subband transforms although, in practice, they are implemented in a different way. For example, the frequently used Discrete Cosine Transform (DCT) is traditionally considered as an orthogonal block transform but can be seen as a subband transform as well [Vand91a]. This section describes the structure and coding efficiency of spectral coding schemes based on a subband transform.

### 2.1.3.1 Structure

Figure 2.6 shows the general structure of spectral coding schemes based on subband transforms. The encoder consists of a decomposition stage followed by an encoding stage. The spectral decoder is made up of the inverse stages of decoding and signal reconstruction. The decomposition stage decomposes the input signal  $x$  into  $M$  subbands, which are denoted by  $x_m$ . Typically, each subband has a reduced sampling rate compared to the sampling rate  $f_s$  of the input signal. The downsampling, or decimation, factor of the  $m^{\text{th}}$  subband is denoted by  $V_m$  and defined as

$$V_m = f_s / f_{s,m}, \quad (2.23)$$

where  $f_{s,m}$  is the sampling rate of subband  $m$ . When the sum of the inverse sampling factors equals unity, i.e.

$$\sum_{m=1}^M \frac{1}{V_m} = 1, \quad (2.24)$$

the spectral scheme is said to be critically sampled or maximally decimated [Vett84, Vaid87]. This means that the total number of subband samples is equal to the total number of input samples. The encoding stage encodes the subbands and represents subband  $m$  by a sequence of codewords, denoted by  $c_m$ . The subbands are usually encoded separately using simple coding methods such as PCM and DPCM. At the decoder, the codeword sequences  $c_m$  are transformed into the reconstructed subbands, denoted by  $y_m$ . Subsequently, the subbands  $y_m$  are combined in the reconstruction stage to form the reconstructed signal  $y$ .

### 2.1.3.2 Filter bank

The decomposition stage of the encoder, i.e. the analysis filter bank, consists of a set of bandpass filters  $H_m$  with  $m = 1, 2, \dots, M$  followed by a set of downsamplers. The bandpass filters divide the input spectrum into  $M$  consecutive bands that are subsequently downsampled by a factor  $V_m$  (see Fig. 2.6b). For maximally decimated schemes, the downsampling factors  $V_m$  are related to the bandwidths  $W_m$  of the filters  $H_m$  by  $V_m = \pi / W_m$ , so that each filter output is resampled at its Nyquist rate. The reconstruction stage of the decoder, i.e. the synthesis filter bank, performs the inverse operation of the analysis filter bank. Here, a set of upsamplers increase the rate of the subbands by  $V_m$  by inserting  $V_m - 1$  zeros between two subband samples. The upsampled subbands are subsequently filtered by a set of bandpass filters  $G_m$  with  $m = 1, 2, \dots, M$ , after which they are combined additively to give an approximation  $y$  of the input signal  $x$ .

In the absence of coding, the filter banks of Fig. 2.6a should reproduce the input signal as accurately as possible. This is a complicated problem since the subbands contain aliasing introduced by the downsamplers. This aliasing is unavoidable because the bandpass filters  $H_m$  and  $G_m$ , however, the error can be made very small or even zero, i.e. the input signal is perfectly reconstructed. More specifically, the input-output function of a maximally decimated filter bank with  $M$  uniform bands (i.e.  $V_m = M$ ) is given by [Vett84, Vaid87]

$$Y(\omega) = \frac{1}{M} \left[ \sum_{m=1}^M H_m(\omega) G_m(\omega) \right] X(\omega) +$$

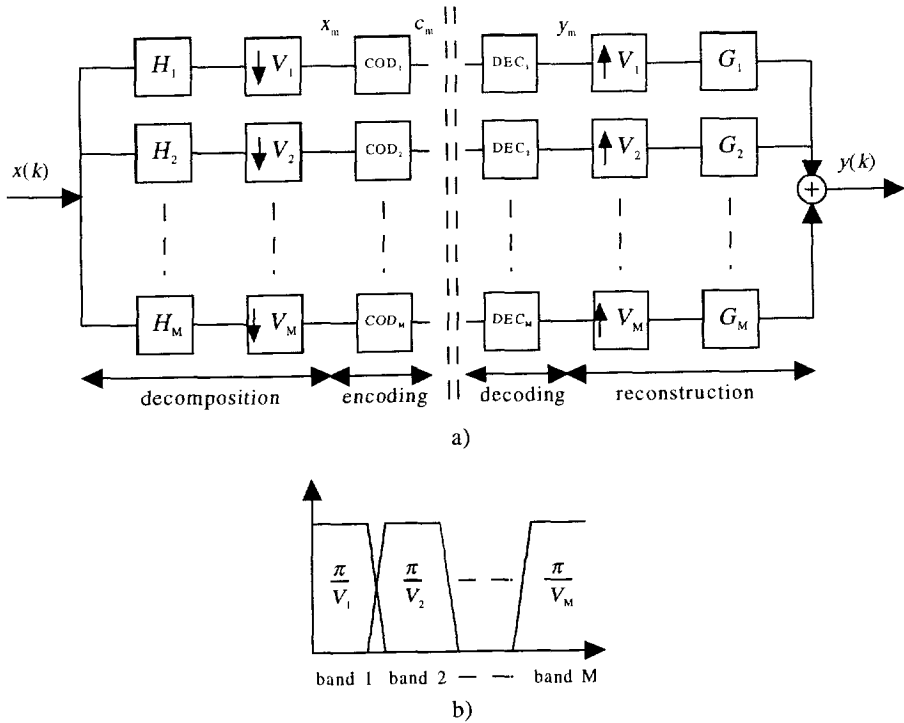


Figure 2.6: Spectral coding scheme based on a subband transform: a) encoder and decoder, b) input spectrum division.



$$\frac{1}{M} \sum_{j=2}^M X\left(\omega - \frac{2\pi(j-1)}{M}\right) \sum_{m=1}^M G_m(\omega) H_m\left(\omega - \frac{2\pi(j-1)}{M}\right). \quad (2.25)$$

Here,  $X(\omega)$  and  $Y(\omega)$  are the Fourier transforms of  $x$  and  $y$ , respectively. Eq. 2.25 demonstrates that, in general, the reconstructed signal suffers from two types of distortion, i.e. aliasing distortion and amplitude/phase distortion. Aliasing distortion is caused by the terms  $X\left(\omega - \frac{2\pi(j-1)}{M}\right)$  with  $(j-1) \neq 0$ . These terms signify the spectral images that are introduced by the downsampler and are the cause of the aliasing in the subbands when they overlap each other. To avoid aliasing in the reconstructed signal, filters  $H_m$  and  $G_m$  have to be designed in such a way that the sum of the products of these spectral images equals zero, i.e.

$$\sum_{m=1}^M G_m(\omega) H_m\left(\omega - \frac{2\pi(j-1)}{M}\right) = 0, \quad (2.26)$$

for  $j = 2, \dots, M$ . This condition implies that the aliasing components of subband  $m$  have to be cancelled out by aliasing components of subbands  $j \neq m$ . Amplitude/phase distortion is related to the weight of the original input spectrum  $X(\omega)$  in Eq. 2.25. To avoid these distortions, the weighting term has to be an integer delay, i.e.

$$\frac{1}{M} \sum_{m=1}^M G_m(\omega) H_m(\omega) = e^{-j\omega p}, \quad (2.27)$$

with  $p$  integer. When the conditions of Eqs. 2.26 and 2.27 are satisfied simultaneously, the input signal is perfectly reconstructed (PR), i.e. the output signal is an exact replica of the input signal except for a shift [Vett84, Vaid87]. To reduce the complexity of design, subband filters are often designed to satisfy the PR condition for a 2-band filter bank only. This does not reduce the applicability of the proposed filter set since all  $M$ -band filter banks can be obtained by cascading 2-band filter banks. Such filter banks are therefore called tree-structured filter banks [Wood86].

Spectral schemes based on subband transforms have attracted a lot of research in the last 10 years and various filter banks have been proposed. Besides the above mentioned design issues of aliasing cancelation and perfect reconstruction, filter bank design efforts have been concentrated on issues like reduced computational complexity using FIR and IIR filters [Gala84, Smit90, Huso90, Horn91, Mau92] and improved coding performance by changing the scales and orientations (in multidimensional signals) of the subbands [Ansa88, Mahe89, Bamb90]. Further, wavelet theory has lead to various new filter types for subband coding [Vett87, Mall89, Anto90, Daub90, Riou91]. In this thesis, linear-phase Quadrature Mirror Filters (QMF) [John80] and non linear-phase Perfect Reconstruction Modulated Filters (PRMF) are used [Mau92].

### 2.1.3.3 Bit rate, distortion, and performance gain

The transmission rate  $I$  of a spectral coding scheme is defined by  $I = f_s R$ , where  $R$  is the mean number of bits used to represent an input sample. If each subband is encoded using  $R_m$  bits per subband sample,  $R$  is given by

$$R = \sum_{m=1}^M \frac{1}{V_m} R_m. \quad (2.28)$$

The reconstruction error  $r(k)$  of a spectral scheme depends on the error made in the filter bank and on the reconstruction errors in the coding stage. For spectral schemes that use PR subband transforms, the variance of reconstruction error  $\sigma_r^2$  is given by [Wood86, West89]

$$\sigma_r^2 = \sum_{m=1}^M \sigma_{r,m}^2, \quad (2.29)$$

where  $\sigma_{r,m}^2$  signifies the variance of the errors made in reconstructing subband  $x_m$ . If PCM and DPCM are used to encode the subbands, then  $\sigma_{r,m}^2 = \sigma_{q,m}^2$ , because for these coding methods the reconstruction errors are identical to the quantization errors. Although Eq. 2.29 is only exact for perfect reconstruction subband transforms, it also provides a good estimate for the variance of the quantization error for near-perfect reconstruction subband transforms [West89].

Also for spectral coding schemes a performance gain over PCM can be defined. For a particular bit rate  $R$ , the performance gain of spectral schemes, denoted by  $G_S$ , is defined by [Jaya84]

$$G_S = \frac{2^{-2R} \sum_{m=1}^M \sigma_{x,m}^2}{\sum_{m=1}^M [2^{-2R_m} \sigma_{x,m}^2]}. \quad (2.30)$$

For a given set of subband signals,  $G_S$  depends thus entirely on the distribution of the available bits among the subbands. To maximize the coding performance, spectral coding schemes use a bit allocation (BA) algorithm to optimally distribute the available bits among the subbands. The bit allocation algorithm minimizes Eq. 2.29 by matching the bit rates  $R_m$  to the estimated variances of subbands  $x_m$ . For a set of subbands with variances  $\sigma_{x,m}^2$  and equal critical sampling rates, i.e.  $V_m = M$ , the theoretical optimal bit allocation is given by [Jaya84]

$$R_{m,opt} = R + \frac{1}{2} \log_2 \frac{\sigma_{x,m}^2}{\left[ \prod_{l=1}^M \sigma_{x,l}^2 \right]^{1/M}}, \quad (2.31)$$

assuming sufficiently high bit rates  $R_m$  and constant quantizer performances. Equation 2.31 shows that important subbands with a variance larger than the geometric mean of variances are accurately quantized with a bit rate  $R_m > R$ . As a result of the optimal bit allocation, the variances of the reconstruction errors of the subbands become equal and have the value

$$D_{m,opt} = \sigma_{r,m}^2 = \epsilon 2^{-2R} \left[ \prod_{l=1}^M \sigma_{x,l}^2 \right]^{1/M}, \quad (2.32)$$

where  $\epsilon$  is a constant (quantizer) performance factor. Consequently, the maximum performance gain equals [Jaya84]

$$G_{S,max} = \frac{\frac{1}{M} \sum_{m=1}^M \sigma_{x,m}^2}{\left[ \prod_{m=1}^M \sigma_{x,m}^2 \right]^{1/M}}. \quad (2.33)$$

It should be noted that spectral decoders require information about the bit allocation to facilitate the proper decoding of the codewords received. Spectral schemes therefore transmit side information (SI) that contains either the labels of the selected quantizers or the estimated variances  $\sigma_{x,m}^2$ . In the latter case, the spectral decoder should perform an identical bit

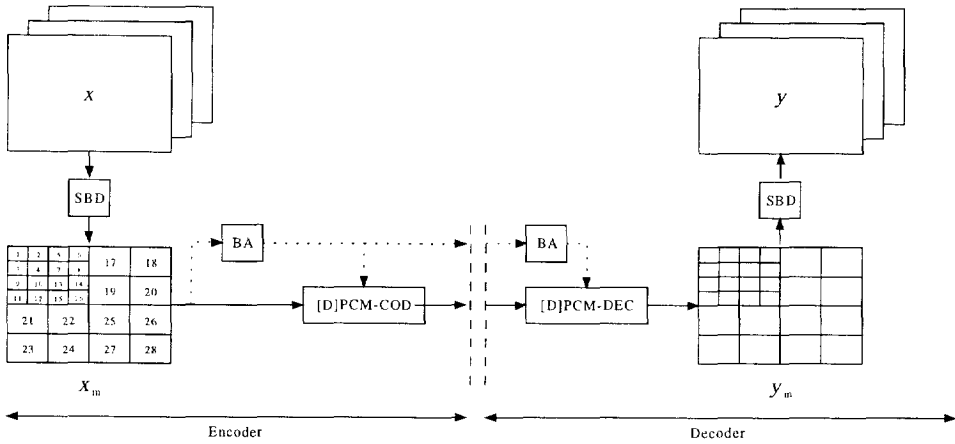


Figure 2.7: Basic intraframe subband coding scheme.

allocation to recover the applied quantizers. Further, note that the theoretical solution to the bit allocation problem is not directly applicable in practical schemes, because the resulting optimal bit rates  $\{R_{m,opt}\}$  are real values and do not necessarily match the bit rate of any available quantizer. In addition, negative bit allocations can occur for subbands with a low variance. Hence, spectral schemes have to use bit allocation procedures that approximate the theoretical solution closely [Wood86, West88].

## 2.2 Intraframe subband coding

This section describes a practical coding scheme that is based on subband coding. This scheme provides the basis for all hierarchical coding schemes presented in this thesis. The subband coding scheme encodes image sequences on a frame-by-frame basis and is henceforth called the basic intraframe subband scheme, or basic scheme for short. Figure 2.7 illustrates the structure of the scheme. Each frame of the image sequence is decomposed into  $M = 28$  subbands that are subsequently encoded using DPCM/PCM under supervision of a bit allocation algorithm (BA). At the receiving side, the subbands are decoded and then reconstructed into an approximation of the input sequence. In the following, three subsections present the subband filter banks, the optimized subband coders, and the bit allocation procedure.

### 2.2.1 Filter bank

The basic scheme uses a tree-structured filter bank to decompose and reconstruct the input signal. The tree-structured filter bank consists of multiple concatenated elementary filter banks that decompose an input signal into 4 rectangular subbands. The elementary filter banks are based on 1-D filters that are used to decompose the 2-D input signal in a separable way (i.e. first decompose the rows, then the columns). The 1-D filters cancel the aliasing

component in the overall transfer function (Eq. 2.25), by choosing the synthesis filters as

$$\begin{aligned} G_0(\omega) &= 2H_1(\omega + \pi), \\ G_1(\omega) &= -2H_0(\omega + \pi). \end{aligned} \quad (2.34)$$

Consequently, the transfer function  $T(\omega)$  becomes

$$T(\omega) = \frac{Y(\omega)}{X(\omega)} = \frac{1}{2} [H_0(\omega)H_1(\omega + \pi) - H_0(\omega + \pi)H_1(\omega)] X(\omega), \quad (2.35)$$

which is indeed free of aliasing. The basic intraframe subband coding scheme uses linear-phase Quadrature Mirror Filters (QMF) [John80], which are widely used for image and video compression using subband transforms. QMF filter banks use half-band low-pass ( $H_0$ ) and highpass ( $H_1$ ) filters which are equal-length, symmetric, and mirror images of each other in the Fourier domain, i.e.

$$H_1(\omega) = H_0(\omega + \pi). \quad (2.36)$$

With this choice, the transfer function becomes

$$T(\omega) = \frac{1}{2} [H_0^2(\omega) - H_0^2(\omega + \pi)]. \quad (2.37)$$

It can be shown that  $T(\omega)$  can not equal a pure delay for even-length linear-phase QMF filters, except for trivial two-tap filters (i.e.  $h_0 = (0.5, 0.5)$ ) and for infinite-length filters [John80]. For other even filter lengths  $T(\omega)$  has a linear-phase response (i.e. no phase distortion), but the amplitude response deviates slightly from unity (i.e. amplitude distortion is present). For odd-length QMF filters  $T(\omega) = 0$  for  $\omega = \frac{\pi}{2}$ , which is a serious amplitude distortion. In [John80] a set of even-length FIR filters is proposed that minimizes the amplitude distortion. In fact, the amplitude distortion of the various filters is so small that it can be neglected, especially in coding environments. The QMF filters used in this thesis are taken from [John80] and are referred to as QMF08 and QMF16b.

## 2.2.2 Quantization and entropy coding

Fig. 2.7 illustrates that the subbands are separately encoded by means of either PCM or DPCM. For both methods, an optimized quantizer is used that is matched to the PDF of the signal to be quantized. Further, both methods use an entropy coder that converts the resulting indices into codewords which have an average bit rate close to the entropy rate. The basic scheme uses one of two combinations of quantizers and entropy coding techniques. These combinations are Lloyd-Max Quantizers with Huffman coders and Uniform Threshold Quantizers with Arithmetic Coders. These two combinations are described in this section after a brief discussion of the statistical properties of the subbands.

### 2.2.2.1 Statistical subband properties

When encoding natural images, the statistical properties of the subband signals are rather similar for a vast range of images. The basic scheme exploits this fact by assuming certain properties for the various subband signals, such as correlations and PDFs of the subbands. In

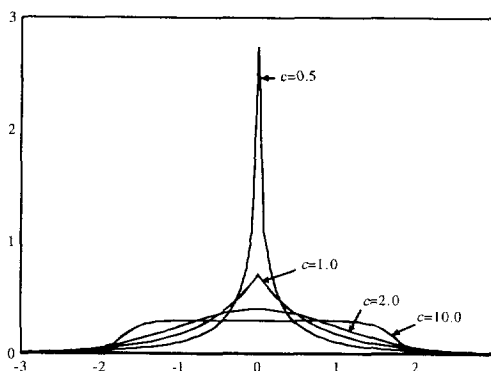


Figure 2.8: Generalized Gaussian PDF (GG-PDF) for  $c = 0.5, 1.0, 2.0, 10.0$  (unity variance).

particular, the spatial correlation among subband samples is generally very small except for the subband containing the lowest frequency band [West89]. Therefore, the basic subband coding scheme encodes this low-pass subband using DPCM, while the other (bandpass) subbands are PCM encoded. The DPCM coder uses an optimal linear predictor that is based on the Yule-Walker equations. Further, the subband and prediction error signals tend to be distributed quite similarly: amplitudes near zero are highly probable, while large amplitudes are less probable. Experimental studies have shown that the distributions of these signals can be reasonably well modelled by a Generalized Gaussian Probability Density Function (GG-PDF) [Mill72] with  $c = 0.5$  for subband signals and  $c = 0.75$  for the prediction error signal [West89]. This GG-PDF is defined as

$$p_X(x) = \frac{bc}{2\Gamma(\frac{1}{c})} \exp(-|bx|^c), \quad \text{with} \quad (2.38)$$

$$b = \frac{1}{\sigma_x} \sqrt{\Gamma\left(\frac{3}{c}\right) / \Gamma\left(\frac{1}{c}\right)}, \quad (2.39)$$

where  $\sigma_x^2$  signifies the variance of the signal to be quantized. In Eq. 2.39, parameter  $c$  determines the shape of the PDF. Figure 2.8 illustrates that for unity variance and  $c = 1.0$ , the GG-PDF equals the Laplacian PDF, while for  $c = 2.0$  the well-known Gaussian PDF is obtained. For  $c = 10.0$ , the PDF approximates a uniform distribution. As a result, the intraframe subband coding scheme uses predefined quantizers which are optimized for unity-variance GG-PDFs with  $c = 0.50$  (subband signals) and  $c = 0.75$  (prediction error signal). Before quantization, these quantizers are scaled by a factor  $\sigma_x$  to match the variance of the quantizer input, i.e. the decision and representation levels of the quantizer are multiplied by  $\sigma_x$ .

### 2.2.2.2 Lloyd-Max quantization and Huffman coding

Lloyd-Max quantizers (LMQ) minimize the distortion of a quantizer with respect to a certain number of representation levels  $K$ . The decision levels  $x_j$  of the Lloyd-Max quantizers are

defined by [Max60, Lloy82]

$$\begin{aligned} x_1 &= -\infty \\ x_j &= \frac{1}{2}(y_j + y_{j-1}), \quad \text{for } j = 2, 3, \dots, K, \\ x_{K+1} &= \infty \end{aligned} \quad (2.40)$$

and representation levels  $y_j$  by

$$y_j = \frac{\int_{x_j}^{x_{j+1}} x p_X(x) dx}{P_j}, \quad \text{for } j = 1, 2, \dots, K. \quad (2.41)$$

Equations 2.40 and 2.41 show that the decision levels are half-way between neighboring reconstruction levels, and that the reconstruction levels are the centroids of the PDF in each decision interval. The most important feature of Lloyd-Max quantizers is that the decision intervals are non-uniform; they are small when PDF  $p_X(x)$  is relatively high, and large when  $p_X(x)$  is relatively low. In fact, this minimizes the distortion for a certain number of levels and causes all quantization intervals to contribute equally to the total distortion. In [Flei64], two iterative design strategies are discussed to calculate the optimum decision and representation levels.

For an LMQ, the entropy rate  $H(J)$  is determined by the set of decision levels  $x_j$  and PDF  $p_X(x)$ . To achieve transmission at a rate close to this entropy rate, the LMQ is followed by a single-symbol Huffman coder. The Huffman coder assigns short codewords to indices with high occurrence probabilities and long codewords to indices with low occurrence probabilities [Huff52]. With properly designed Huffman codes, the codewords of subsequent samples can be concatenated into a single bit stream and can still be flawlessly decoded. It should be noted that the minimum bit rate of a LMQ followed by a Huffman coder is lower bounded at 1 bit per sample. This is not only the minimum entropy rate of a 2-level symmetric LMQ, but also the minimum codeword length of the Huffman code applied.

### 2.2.2.3 Uniform Threshold Quantizers and Arithmetic coding

Uniform Threshold Quantizers (UTQ) are frequently used in image coding applications since these quantizers are known to be among the best scalar entropy-constrained quantizers in the MSE sense [Wood69, Farv84]. UTQs achieve a close to optimal trade-off between the entropy rate of a quantizer and its distortion. UTQs can be described by their stepsize  $\Delta x$  and the number of representation levels  $K$ . For odd  $K$ , the decision intervals  $S_j$  are given by the intervals  $(x_j, x_{j+1}]$ , where

$$x_j = (j - 1 - \frac{K}{2}) \Delta x, \quad 2 \leq j \leq K, \quad (2.42)$$

with  $x_1 = -x_{K+1} = -\infty$ . The representation levels  $y_j$  are given as in Eq. 2.41 and are again the centroids of each decision interval. Another advantage of UTQs is that the entropy rates can be close to zero. To achieve this,  $K$  has to be odd, because the PDF models of the subband data are symmetric [Farv84].

To achieve transmission at a rate close to the entropy rate of the UTQ, the quantizer has to be followed by an Entropy Coder. Since Huffman coders can not achieve rates lower than 1

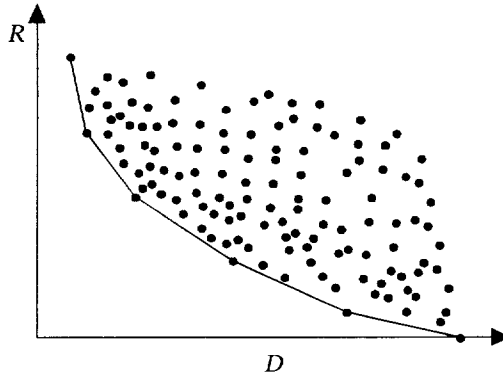


Figure 2.9: Example of bit allocation problem.

bpp, the UTQs are followed by Arithmetic Coders (AC). AC can achieve these rates because they represent a sequence of indices by a single codeword [Witt87, Lang84]. This codeword is actually a binary fraction that indicates a number between zero and one. The number is obtained during the encoding of the indices by multiplying the cumulative probabilities of the indices. In general, Arithmetic Coders allow for a more efficient encoding than Huffman coders do, but are also more complicated to implement.

### 2.2.3 Bit allocation algorithm

The basic scheme encodes the subband data using predefined quantization and entropy coding tables. To maximize the coding performance of the scheme, a bit allocation procedure is used that optimally distributes the available bits  $R$  (per sample) among the subband data. The bit allocation procedure can assign either one quantizer to one subband or one quantizer to one activity class within a subband.

#### 2.2.3.1 Subband-based allocation

The subband-based allocation procedure follows the bit allocation procedure proposed in [West88, Risk91]. It assumes a predefined set of admissible quantizers for each subband and selects for each subband a particular quantizer (and associated entropy code) from this set. The algorithm chooses these quantizers such that the total bit rate closely approximates but does not exceed the available number of bits  $R$ , while the total distortion  $D$  is minimized. Let  $N_m$  denote the number of optimized quantizers that is available for subband  $m$  with  $1 \leq m \leq M$ . Then the set of quantizers that is available for subband  $m$  is given by

$$\{Q_m[l]\} \quad \text{for } l = 1, 2, \dots, N_m. \quad (2.43)$$

The quantizers  $Q_m[l]$  are optimized for the assumed GG-PDFs (unity-variance) of the quantizer inputs. The bit rates of the quantizers are defined as the average codeword length assigned to the quantizer outputs (Huffman coding) or as the entropy rates (Arithmetic coding). When  $\mathbf{Q} = (Q_1[l_1], \dots, Q_M[l_M])$  denotes the vector of the quantizers that are actually

assigned to the subbands and the assigned quantizers have bit rates  $r_m$  and MSE distortions  $d_m$ , then the bit allocation problem can be defined as finding the optimal vector  $\mathbf{Q}^*$  that minimizes the overall distortion

$$D(\mathbf{Q}^*) = d_1\sigma_{d,1}^2 + \sum_{m=2}^M d_m\sigma_{x,m}^2, \quad (2.44)$$

subject to the constraint

$$R(\mathbf{Q}^*) = \sum_{m=1}^M \frac{1}{V_m} r_m \leq R, \quad (2.45)$$

where  $R$  is the average number of bits per sample available to encode the entire frame,  $\sigma_{d,1}^2$  the variance of the DPCM prediction error of subband 1, and  $\sigma_{x,m}^2$  the variance of the subbands. Figure 2.9 illustrates that this is a combinatorial problem of finding  $\mathbf{Q}^*$  from a large number of vectors.

The proposed bit allocation algorithm of [West88, Risk91] finds  $\mathbf{Q}^*$  by tracing the vectors that lie on the lower convex hull of the 'cloud' of possible vectors. The optimal vector must lie on this convex hull because these vectors achieve the best trade-off between bit rate and distortion. The allocation algorithm iteratively traces the convex hull (starting with the vector  $R(\mathbf{Q}) = 0$ ) until it has found the vector  $\mathbf{Q}^*$  that has a bit rate closest to  $R$  but not exceeding  $R$ . It traces the convex hull by finding the next vector that has a minimum increase in bit rate for a maximum decrease in distortion. The computational complexity of the allocation procedure is low since the search process for the next vector can be decomposed into a search for the best rate/distortion trade-off per subband and a search for the minimum of these trade-offs [West88, Risk91].

It should be noted that the bit allocation algorithm distributes the bits among the subbands based on the assumption that the subband signals are distributed according to the expected GG-PDFs. In practice, subband signals will not be in exact agreement with this model. Consequently, the actual bit rate and resulting distortion will deviate slightly from the values obtained by the bit allocation. Fortunately, these deviations are small and can be neglected for regular natural images [West89].

### 2.2.3.2 Activity class-based allocation

The activity class-based allocation procedure is a combination of the bit allocation algorithms proposed in [West88, Wood86, Diab90]. The algorithm divides the subband data of each subband into 4 activity classes with different variances and assigns to each activity class a dedicated quantizer and entropy coder. Consequently, regions in a subband with high variances are encoded more accurately than subband regions with low variances. The activity classes consist of rectangular subband blocks of e.g. 8x8 pixels and are of equal size (i.e. same number of subband blocks). For the PCM-encoded subband, the blocks are classified by means of the estimate of the block variance. For the DPCM-encoded subband, the variance estimate of a block is based on the theoretical optimal prediction error (i.e. no quantization).

After classification, the bit allocation algorithm knows the variances of all activity classes of all subbands. The corresponding allocation algorithm subsequently selects a quantizer from the set of available quantizers for each activity class, based on the variances measured and a set of performance tables. The performance tables indicate for each available quantizer



the expected bit rate and distortion per class after quantization and entropy coding. The final assignment is such that the bit rate is minimized for a certain distortion. More precisely, if  $r_{m,k}$  and  $d_{m,k}$  denote the bit rate and distortion of the assigned quantizer of the  $k^{\text{th}}$  activity class of subband  $m$ , then the distortion of vector  $\mathbf{Q}^*$

$$D(\mathbf{Q}^*) = \frac{1}{4} \sum_{k=1}^4 [d_{1,k} \sigma_{d,1,k}^2] + \frac{1}{4} \sum_{k=1}^4 \sum_{m=2}^M [d_{m,k} \sigma_{x,m,k}^2] \quad (2.46)$$

is minimized under the condition that

$$R(\mathbf{Q}^*) = \frac{1}{4} \sum_{k=1}^4 \sum_{m=1}^M \frac{1}{V_m} r_{m,k} \leq R, \quad (2.47)$$

where  $\sigma_{d,1,k}^2$  denotes the variance of the  $k^{\text{th}}$  activity class of the DPCM prediction error of subband 1, and  $\sigma_{x,m,k}^2$  the variance of the  $k^{\text{th}}$  activity class of the PCM encoded subbands.

## 2.3 Implementation of scalability

Memoryless, predictive, and spectral coding techniques are all of importance for hierarchical coding schemes because these techniques can be used to implement scalability. This section illustrates how the four defined types of scalability (i.e. amplitude, frequency, spatial, and temporal scalability) can be implemented separately by means of the basic techniques presented. The simultaneous implementation of multiple types of scalability is not discussed here, but is studied in the rest of this thesis.

Note that hierarchical coding schemes encode an input signal  $x$  such that decoders can reconstruct multiple signals  $y^n$  from the compressed data (with  $1 \leq n \leq N$ ). The reconstructed signals  $y^n$  have a hierarchical relation to each other with respect to the scalability type. In particular, signal  $y^1$  is a low-quality reproduction of  $x$ , while  $y^N$  is a high-quality reproduction.

### 2.3.1 Amplitude scalability

With amplitude scalability, the reproductions  $y^n$  ( $1 \leq n \leq N$ ) differ in the amount of accuracy with which the amplitudes of the signal are quantized. When  $D^n$  denotes the MSE distortion that is made in the total quantization process for  $y^n$ , the following relation holds for amplitude scalability:

$$D^n > D^{n+1} \quad \text{for } n = 1, 2, \dots, N-1. \quad (2.48)$$

Amplitude scalability is supported by specific variants of memoryless and predictive coding techniques [Good80, Jaya83, Jaya83, Elna86, Tzou86, Wang88], called multirate PCM (MR-PCM) and multirate DPCM (MR-DPCM) in [Taub94, Bosv95]. MR-PCM and MR-DPCM can be used to encode uncorrelated signals and correlated signals, respectively, in an efficient and amplitude-scalable way. The principles of these schemes are presented below, and are further discussed and analyzed in Chapter 4.

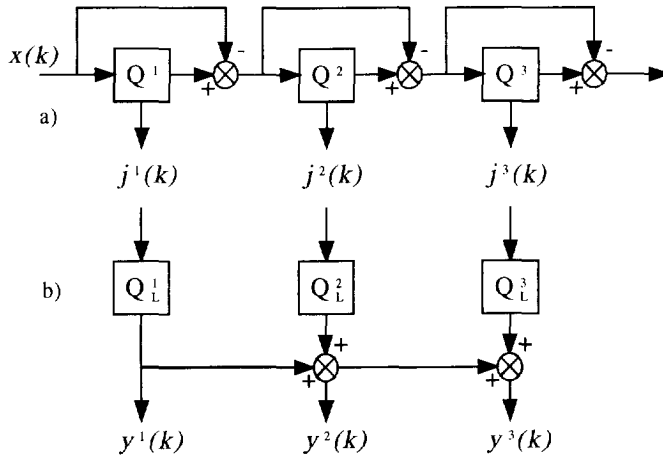


Figure 2.10: Implementation of amplitude scalability by multistage quantization for the case  $N = 3$ : a) encoder, b) decoder.

### 2.3.1.1 Multirate PCM

A MR-PCM coder consists of a multirate quantizer followed by a multirate entropy coding section. A multirate quantizer maps the input amplitude  $x(k)$  into a set of indices  $j^n(k)$  with  $1 \leq n \leq N$  to facilitate  $N$  reproductions of the input amplitude. The value of the  $n^{\text{th}}$  reproduced amplitude  $y^n(k)$  is determined by the indices  $\{j^1(k), \dots, j^n(k)\}$ , with  $n \leq N$ . To facilitate the digital transmission of the indices, the multirate entropy coding section losslessly encodes  $j^n(k)$  into codewords  $c^n(k)$ , which are then conveyed in layer  $l^n$ .

In general, there are two basic ways to implement the multirate quantization process; namely, multistage quantization and embedded quantization. Figure 2.10a shows the block diagram of a multistage quantizer for the case of  $N = 3$ . The input amplitude  $x(k)$  is initially quantized by a coarse quantizer  $Q^1$ , and the obtained quantization error is then quantized by a finer quantizer  $Q^2$ . Subsequently, the new quantization error is quantized again by an even finer quantizer  $Q^3$ , and this refinement process is repeated for  $N$  stages in total. Each quantizer  $Q^n$  produces an index  $j^n$  that is transmitted to the associated decoder (see Figure 2.10b). The decoder obtains  $y^1(k)$  by decoding  $j^1(k)$  only, i.e.  $y^1(k) = Q_L^1(j^1(k))$ . Approximations  $y^2(k)$  and  $y^3(k)$  are obtained by converting  $j^2(k)$  and  $j^3(k)$ , respectively, and combining these additively with  $y^1(k)$ .

The second way to implement multirate quantization is embedded quantization. The main property of embedded quantization is that the decision levels of the involved quantizers are aligned. Figure 2.11 shows the block diagrams of embedded quantization (encoder and decoder) and  $N$  aligned quantizers for the case of  $N = 3$ . At the encoder, the input signal  $x(k)$  is concurrently quantized by a bank of  $N$  aligned quantizers  $Q^n$  that output indices  $i^n(k)$  with  $1 \leq n \leq N$ . The indices  $i^n(k)$  are overcomplete because of the alignment of the quantizer decision levels (see Fig. 2.11c). This implies that index  $i^N(k)$  specifies indices  $i^n(k)$  with  $n < N$  as well. Therefore, indices  $i^n(k)$  are translated into a non-redundant set of

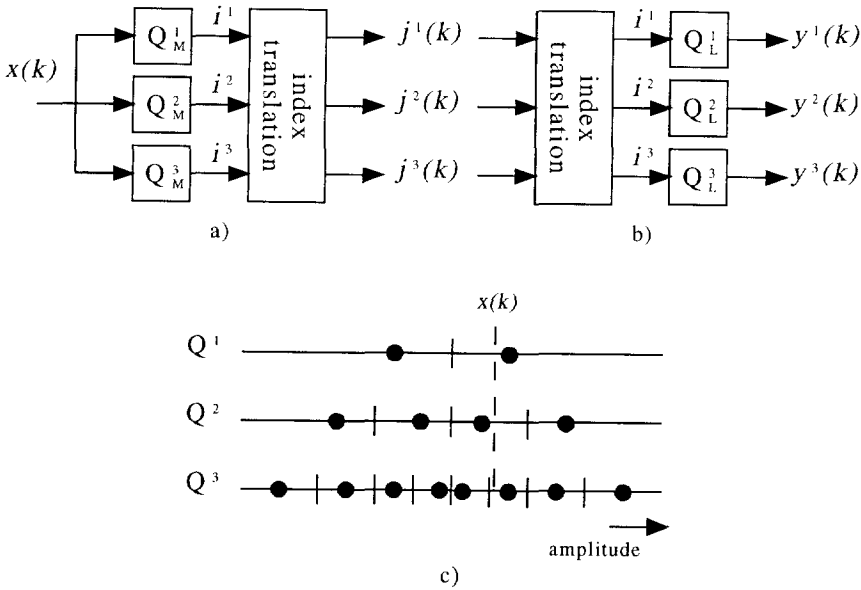


Figure 2.11: Implementation of amplitude scalability by embedded quantization for the case  $N = 3$ : a) encoder, b) decoder, c)  $N$  aligned quantizers.

indices  $j^n(k)$  with  $1 \leq n \leq N$ . At the decoder, the inverse operations are performed. First, indices  $j^n(k)$  are translated into indices  $i^n(k)$ , which are then converted into the reconstructed amplitudes  $y^n(k)$ .

### 2.3.1.2 Multirate DPCM

A MR-DPCM coder consists of a predictive multirate quantizer followed by a multirate entropy coding section. A predictive multirate quantizer is similar to the non-predictive multirate quantizer discussed above except that it incorporates a prediction loop.

If the predictive quantizer is implemented using multistage quantization, a prediction of the input data of each stage is subtracted before quantization and added after reconstruction. Similar to a prediction with DPCM, the prediction is based on the reconstructed signal of each stage to keep the predictions at the encoder equal to those at the decoders. Generally, only the input signal of  $Q^1$  is predicted because the subsequent quantization errors are fairly uncorrelated because of that prediction.

If the predictive quantizer is implemented using embedded quantization, a prediction of the input signal is subtracted before the concurrent quantization and, after reconstruction, added to all the reconstructed signals  $y^n$ . In this way the DPCM coding gain increases the coding performance of all quality levels. However, to keep the prediction at the encoder identical to those at the decoders, the prediction has to be based on  $y^1$ , which is available to all decoders. If the prediction is based on another reproduction, some decoders can not make the same prediction as the encoder and will suffer from accumulating errors in the

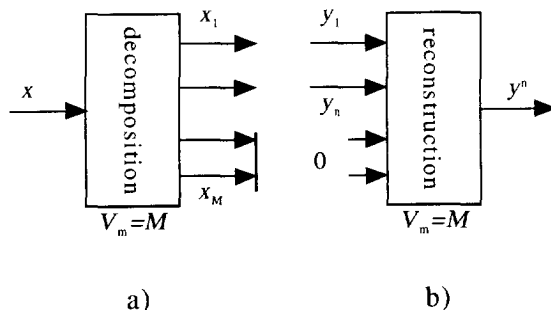


Figure 2.12: Implementation of frequency scalability with a subband filter bank: a) decomposition section, b) reconstruction section.

prediction loop. This phenomenon is called drift to emphasize the diverging predictions of the encoder and decoder. To circumvent drift, the encoder has to make the prediction based on the information available to the lowest-quality decoder.

### 2.3.2 Frequency scalability

With frequency scalability, the reproductions  $y^n$  have increasing frequency supports for higher  $n$ . Reproduction  $y^1$  has a limited frequency support (e.g. only low frequencies) while  $y^N$  has a large frequency support (e.g. the entire input spectrum). For 1-D signals, the low-pass frequency supports of  $y^n$  can be denoted as  $[-\frac{\pi}{\beta^n}, \frac{\pi}{\beta^n}]$ , with  $0 \leq \frac{\pi}{\beta^n} \leq \pi$  the bandwidth of the signal<sup>1</sup>. With frequency scalability, the bandwidths  $\frac{\pi}{\beta^n}$  are related as follows:

$$\frac{\pi}{\beta^n} < \frac{\pi}{\beta^{n+1}} \quad \text{for } n = 1, 2, \dots, N - 1. \quad (2.49)$$

Frequency scalability can be implemented by spectral coding techniques [Vand91a]. In particular, subband schemes can decompose the input spectrum into several subbands, of which some may be used in the reconstruction process. Subbands that are not used in the reconstruction process are set to zero, thereby limiting the bandwidth of the reproduced signal. As an example, Figure 2.12 shows how a subband coding scheme can implement frequency scalability by using a  $M$ -band filter bank. The scheme decomposes the input signal  $x$  into  $M$  subbands  $x_m$  to allow for a maximum of  $N = M$  reproductions. The subbands are maximally decimated and convey a frequency band of the input spectrum with bandwidth  $\frac{\pi}{M}$ . Subband  $m$  is encoded at  $R_m > 0$  and is conveyed in layer  $l^m$  with  $1 \leq m \leq N$ . Decoders reconstruct signal  $y^n$  by decoding the first  $n$  subbands and by using these subbands in the reconstruction process. The subbands  $m > n$  are not decoded but are instead set to zero in the reconstruction filter bank. Consequently, the reproduced signal  $y^n$  has a bandwidth of  $\frac{\pi}{\beta^n} = \frac{n\pi}{M}$ .

The above example shows that by discarding the high-frequency bands in the reconstruction process, the filter banks can be used as a kind of low-pass filter. It should be

<sup>1</sup>Note that  $\beta^n$  is a value associated with reproduction  $n$  and does not mean  $\beta$  to the power  $n$ .

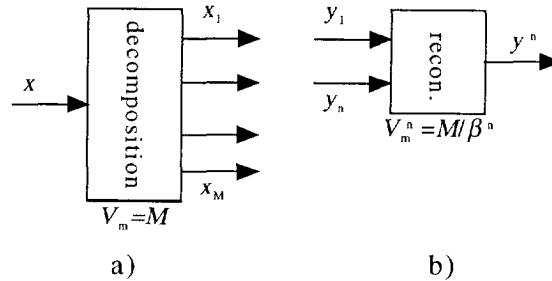


Figure 2.13: Implementation of spatial scalability with a subband filter bank: a) decomposition section, b) reduced reconstruction section.

noted, however, that the system is not equivalent to a linear-phase low-pass filter, because  $y^n$  contains some aliasing components that had to be cancelled by aliasing components in the discarded high-frequency subbands (c.f. Eq. 2.25) [Vand91a]. In addition, the total system response has become (periodically) shift-variant because the aliasing is (periodically) shift-variant [Vand91a].

### 2.3.3 Spatial scalability

With spatial scalability, the reproductions  $y^n$  have increasing spatial resolutions or sampling rates. More precisely, if  $f_s^n$  denotes the sampling rate of  $y^n$ , the sampling rates are related as follows:

$$f_s^n < f_s^{n+1} \quad \text{for } n = 1, 2, \dots, N - 1. \quad (2.50)$$

Like frequency scalability, spatial scalability can be implemented by spectral coding techniques [Bell88, Tzou88, Bosv90, Vand91b]. In particular, subband schemes can decompose the input spectrum into several subbands, of which some may be used in the reconstruction of  $y^n$ . However, in contrast to frequency scalability, spatially scalable decoders employ a modified reconstruction filter bank to obtain  $y^n$  with a reduced sampling rate. As an example, Figure 2.13 shows how a subband coding scheme can implement spatial scalability by using an  $M$ -band decomposition filter bank. The scheme decomposes the input signal  $x$  into  $M$  subbands in a way similar to Fig. 2.12. Decoders reconstruct signal  $y^n$  by decoding the first  $n$  subbands and by using these subbands in the modified reconstruction process. In the modified filter bank, the subbands are upsampled by a factor

$$V_m^n = \frac{M}{\beta^n}, \quad (2.51)$$

and are interpolated with appropriate filters. Consequently, reproduction  $y^n$  has a sampling rate of  $f_s^n$ . The synthesis filters in the modified reconstruction filter bank have to be designed carefully because they have to interpolate the  $\frac{M}{\beta^n}$ -fold upsampled subbands and have to cancel the aliasing components present in the subband as well. A reasonable choice for these filters is to use the  $\frac{1}{\beta^n}$ -fold decimated versions of the original synthesis filters  $G_m$  [Vand91a]. Note that similar to implementation of frequency scalability, the implementation of spatial

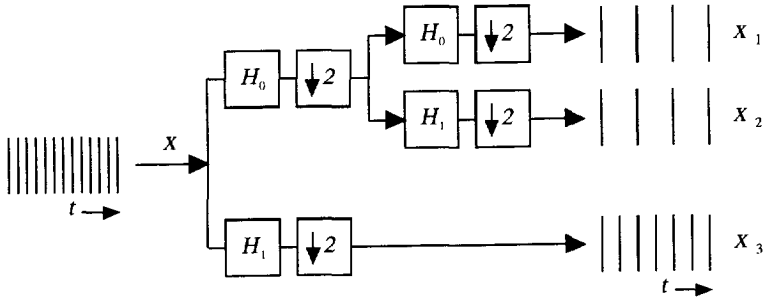


Figure 2.14: Implementation of temporal scalability with a temporal subband decomposition.

scalability by subband filter banks also results in a reproduction  $y^n$  that contains aliasing components which should have been cancelled by the discarded high-frequency bands.

Spatial scalability is of particular interest for the compatible distribution of video signals. For this application, the reproductions must have visually pleasing qualities. The filter bank applied requires therefore attention since filter characteristics like passband ripples (amplitude distortion), stop-band energies (aliasing), non linear-phase characteristics (phase distortion), and step responses (ringing) may have a large influence on the subjective qualities of the reproductions. Because the filters of an  $M$ -band filter bank are heavily constrained (i.e. the filters  $H_m$  and  $G_m$  have to satisfy the aliasing cancelation and PR conditions of Eqs. 2.26 and 2.27) a number of alternative 'filter banks' have been proposed to implement spatial scalability (see e.g. [Tzou88, Vett92, Vand91b, Bosv95]). Such alternatives are, for example, tree-structured subband filter banks [Bosv90, Ghar91a], pyramidal filter banks [Chen88, Uz91b], and filter banks based on the DCT [Tzou88, Vand90, Scha92, Civa92, Gonz93].

In practice, the difference in visual quality of the reproductions obtained by either  $M$ -band filter banks, tree-structured QMF filter banks, wavelet filter banks, or pyramidal schemes with appropriate filters is generally very small. The visual quality only degrades noticeably (and becomes objectionable) when subband filters are used with very strong non-linear-phase characteristics. The visual quality of the low-resolution signals obtained by the DCT is generally considered adequate for coding applications, but is less well compared in respect to the pyramidal and subband schemes. Improvements can be obtained either by modifying the DCT transform [Vand91a] or by using some specific "high frequency" coefficients in the 2-D inverse transform [Scha92].

When tree-structured filter banks are used to implement spatial scalability, the selection of the subbands for the partial reconstruction requires special attention. This is because after one decomposition stage, the frequency orientation in the high-frequency subband is reversed. For a 1-D signal, this implies that the original high frequencies will be contained in subband 3 instead of subband 4 after an additional decomposition. To circumvent this, the filter bank may shift the frequencies in the high frequency subband by  $\pi$  to restore the original frequency orientation. Such modified tree-structured filter banks are used in the rest of this thesis.

### 2.3.4 Temporal scalability

With temporal scalability, the reproductions  $y^n$  have increasing temporal resolutions or frame rates. More precisely, if  $f_r^n$  denotes the frame rate of  $y^n$ , the sampling rates are related as follows:

$$f_r^n < f_r^{n+1} \quad \text{for } n = 1, 2, \dots, N - 1. \quad (2.52)$$

Temporal scalability can be implemented by spectral coding techniques [Bosv92b, Ohm93b, Taub94] as well as by temporal predictive techniques [Uz91b]. When subband coding is used, the image sequence is temporally decomposed into a number of temporal subbands by means of linear filtering. These temporal subbands have reduced sampling rates and convey different temporal frequency bands. As an example, Figure 2.14 illustrates a tree-structured temporal decomposition into 3 temporal subbands  $x_1$ ,  $x_2$ , and  $x_3$ . Here, the input sequence is first decomposed into two temporal subbands by filtering the input sequence with filters  $H_0$  and  $H_1$  and subsequently (temporal) downsampling by a factor of 2. Next, the low-frequency subband is again decomposed into two subbands to obtain three temporal subbands in total. Note that since the filtering is done along the temporal axis, a number of frames have to be stored for computations. The lowest-quality decoder constructs reproduction  $y^1$  by decoding subband  $x_1$  only (not shown). The medium-quality decoder reconstructs reproduction  $y^2$  by decoding and combining the subbands  $x_1$  and  $x_2$ . The highest-quality decoder reconstructs reproduction  $y^3$  by decoding and additively combining all temporal subbands.

Temporal scalability is of particular interest for the temporally compatible distribution of video signals. For this application it is necessary that the reconstructed sequences be of visually pleasing qualities. For temporal subband decompositions, the subjective quality of the reproductions  $y^n$  with  $n < N$  is only adequate for short to very short filter lengths of  $H_0$  and  $H_1$  (i.e. 8 and 2 taps, respectively). For longer filters, the signals  $y^n$  become severely blurred and halo images start to appear. This can be circumvented by using motion information in the decomposition process, i.e. by decomposing the image into temporal subbands by filtering the image sequence along the motion trajectories [Ohm92, Ohm93b, Taub94].

Predictive schemes based on ME/MC can also implement temporal scalability. With such schemes the image sequence is temporally decomposed into three sub-sequences by means of temporal prediction. The three sub-sequences have a reduced frame rate compared to the input sequence and consist of, respectively, intraframe predicted images (I-frames), uni-directional temporally predicted images (P-frames), and bi-directional temporally predicted images (B-frames) [Civa92, ISO/95]. The lowest-quality decoder constructs reproduction  $y^1$  by decoding the I-frames only. The medium-quality decoder reconstructs reproduction  $y^2$  by decoding both the I-frames and P-frames. The highest-quality decoder reconstructs reproduction  $y^3$  by decoding all types of encoded frames. The subjective quality of the reproductions  $y^n$  with  $n < N$  is generally very good. However, the block based nature of the prediction and encoding method may cause blocking artifacts to occur at low bit rates.





# Chapter 3

## Hierarchical coding

### 3.1 Introduction

Multimedia systems can use hierarchical coding schemes to implement scalability in an efficient way. Ideally, these schemes implement scalability in such a way that the coding performances of the  $N$  reproductions are equal to the performances obtained by  $N$  independent coders using similar coding techniques. If this is true then it is said that the coding scheme achieves the successive refinement of the input signal in an optimal way [Equi91, Bosv92c]. Optimal successive refinement implies that features like compatibility and graceful degradation can be supported without any additional costs in bandwidth and/or quality. Optimal successive refinement is however not easily achievable, because the coding performances of the reproductions are dependent on each other. In fact, the optimal coding of a low-quality reproduction may obstruct the optimal coding for all higher quality reproductions.

In this chapter the hierarchical encoding process for SNR scalability<sup>1</sup> and spatial scalability is investigated from a theoretical point of view. The main motivations for such an investigation are to gain insight into the process of hierarchical coding and to identify the key issues in the design of hierarchical coding schemes. The analysis pertains to an elementary scheme that supports  $N$  reproductions of the input signal. It is theoretically determined for this scheme under which conditions optimal successive refinement is achievable for both SNR and spatial scalability. To keep the analysis mathematically tractable, the input signal is assumed to be a realization of a stationary Gaussian source  $X$ . In this case, the rate distortion theory provides an analytical description of the rate distortion function (RDF), i.e.  $R_x(D)$ , which provides the fundamental lower bound  $R_x$  on the bit rate per pixel for a given average distortion of  $D$  [Berg71]. Although visual signals are generally neither Gaussian distributed nor stationary, the analysis provides insight into the requirements of practical hierarchical coding schemes.

The content of this chapter is structured as follows. Section 3.2 starts off with an overview of the rate distortion theory for continuous-amplitude signals. It is followed by Sections 3.3 and 3.4, which describe the theoretical analyses of SNR scalability and spatial scalability, respectively. Subsequently, an experimental verification of the insights and conditions obtained from the theoretical analysis is presented in Section 3.5. To this end, the coding performance

---

<sup>1</sup>In Chapter 1, SNR scalability was introduced as the single term for amplitude and frequency scalability.

of a practical hierarchical coding scheme is evaluated. The scheme is based on the basic subband coding scheme presented in Chapter 2, but uses multiple coding stages to create the layers. To conclude the chapter, Section 3.6 presents some conclusions and generalizes the results obtained.

For the sake of notational simplicity, this chapter uses primarily a one-dimensional (1-D) notation to describe the signals and equations. Two-dimensional (2-D) results can be obtained by replacing single indices and variables by double ones, and by replacing intervals by regions. Whenever the 2-D results are essential to the discussion, they are presented explicitly. Further, the term 'successive refinement' is used as a shorthand notation for the term 'optimal successive refinement'.

## 3.2 Rate distortion theory

For a stationary discrete-time and continuous-amplitude source  $X$  there exists a monotonically non-increasing RDF  $R_x(D)$  that provides a lower bound  $R_x$  on the bit rate for coding with an average MSE distortion of  $D$  [Berg71]. This function serves as an upper bound on the performance of any practical coding system with a MSE  $D$  and a bit rate of  $R'(D)$  bits per pixel, i.e.

$$R_x(D) \leq R'(D). \quad (3.1)$$

The theoretical coding scheme that realizes  $R_x(D)$  is called the optimal coding scheme. This scheme is characterized by a mapping  $x \rightarrow y$  that is equal to the optimal mapping. Here,  $x$  and  $y$  are regarded as realizations of the input random variable (r.v.)  $X$  and the reproduction r.v.  $Y$  [Berg71, Jaya84].

### 3.2.1 1-D Gaussian sources

The MSE RDF of a Gaussian source  $X$  with a Power Spectral Density (PSD) function  $S_{xx}(\omega)$  is given parametrically by [Berg71]

$$R_x(\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \max \left\{ 0, \frac{1}{2} \log_2 \frac{S_{xx}(\omega)}{\theta} \right\} d\omega, \quad (3.2)$$

$$D_x(\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \min \{ S_{xx}(\omega), \theta \} d\omega. \quad (3.3)$$

Here,  $R_x(\theta)$  and  $D_x(\theta)$  denote respectively the rate and distortion of the RDF as a function of coding parameter  $\theta$ . Figure 3.1 illustrates that  $\theta$  acts as a threshold that divides the frequency axis into the sets  $A$  and  $B$ :

$$\begin{aligned} \omega \in A & \quad \text{if } S_{xx}(\omega) \geq \theta, \\ \omega \in B & \quad \text{if } S_{xx}(\omega) < \theta. \end{aligned} \quad (3.4)$$

For a certain level of distortion, the optimal mapping  $x \rightarrow y$  reproduces only the spectral contributions of set  $A$  while the spectral contributions of set  $B$  are discarded. Equation 3.2 shows that the frequencies of set  $A$  contribute to the overall bit rate. These frequency components are each encoded at a rate  $\frac{1}{2} \log_2(S_{xx}(\omega)/\theta)$  with an average distortion of  $\theta$ . The

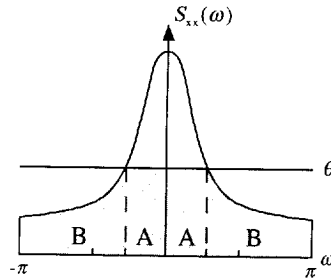


Figure 3.1: Optimal source coding.

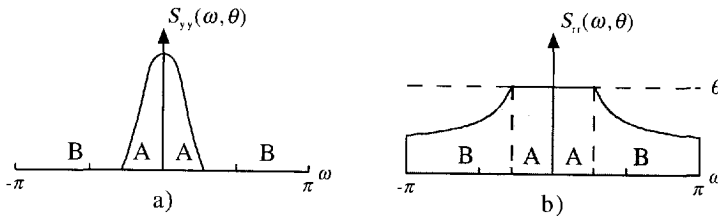


Figure 3.2: Optimal source coding : a) reconstructed spectrum, b) error spectrum.

frequencies of set  $B$  do not contribute to the overall rate. These frequency components are each encoded at a rate equal to zero and have an average distortion of  $S_{xx}(\omega)$ .

The optimal mapping  $x \rightarrow y$  for Gaussian sources is explicitly known [Berg71, Jaya84], as are the characteristics of the reproduced signal  $y$  and the reconstruction error  $r = x - y$ . Specifically, the reproduced signal is Gaussian distributed and has a variance  $(\sigma_x^2 - D)$ . The PSD  $S_{yy}(\omega, \theta)$  of the reproduced signal is given by [Mazo86]

$$S_{yy}(\omega, \theta) = \begin{cases} S_{xx}(\omega) - \theta & \omega \in A, \\ 0 & \omega \in B, \end{cases} \quad (3.5)$$

which is the area above the coding threshold  $\theta$  as illustrated in Figure 3.2a. The reconstruction error is also Gaussian distributed, has variance  $D(\theta)$ , and is independent of the input signal [Mazo86]. The PSD  $S_{rr}(\omega, \theta)$  of the reconstruction error can be expressed as [Jaya84]

$$S_{rr}(\omega, \theta) = \begin{cases} \theta & \omega \in A, \\ S_{xx}(\omega) & \omega \in B, \end{cases} \quad (3.6)$$

as illustrated in Fig. 3.2b. Eq. 3.6 shows that the optimal error spectrum consists of two parts: i) in-band noise that is due to the coding of the frequencies of set  $A$  and ii) the out-band noise that results from the discarded frequencies of set  $B$ . For small distortions, i.e.  $\theta \leq \min\{S_{xx}(\omega)\}$ , the RDF of a colored Gaussian source can be rewritten as [Jaya84]

$$D_x(R) = \gamma_x^2 2^{-2R} \sigma_x^2, \quad (3.7)$$

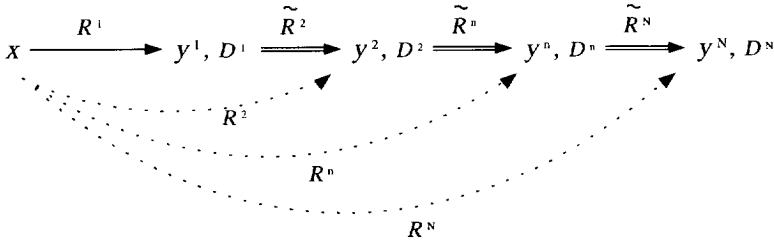


Figure 3.3: Rate distortion view of hierarchical coding.

where  $\gamma_x^2$  is the spectral flatness measure ( $0 \leq \gamma_x^2 \leq 1$ ) and  $D_x(R)$  the inverse of the RDF. In this case, the optimal error spectrum does not contain any out-band noise and it is said that the coding has reached its 'linear' domain. From Eq. 3.7 it can be derived that the MSE is reduced by a factor of 4 for each additional bit spent on transmission, or equivalently, the SNR is increased by approximately 6 dB.

### 3.2.2 2-D Gaussian sources

The MSE RDF of a 2-D Gaussian source with PSD  $S_{xx}(\omega_1, \omega_2)$  is given parametrically by [Berg71]

$$R_x(\theta) = \left(\frac{1}{2\pi}\right)^2 \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \max\left\{0, \frac{1}{2} \log_2 \frac{S_{xx}(\omega_1, \omega_2)}{\theta}\right\} d\omega_1 d\omega_2, \quad (3.8)$$

$$D_x(\theta) = \left(\frac{1}{2\pi}\right)^2 \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \min\{S_{xx}(\omega_1, \omega_2), \theta\} d\omega_1 d\omega_2. \quad (3.9)$$

Clearly, this RDF is a straightforward extension of the RDF for 1-D signals. Similarly, the 2-D spectra of the reproduced signal and the reconstruction error are straightforward extensions of their 1-D versions.

## 3.3 Theoretical analysis of SNR scalability

In the following, it is proven that for SNR scalability successive refinement of a Gaussian input signal  $X$  is always theoretically achievable. This is done by showing that a theoretical coding scheme can be constructed that creates (compound) mappings  $x \rightarrow y^n$  that are equal to the optimal mappings. More formally, if  $R_x(\cdot)$  denotes the RDF of  $X$ , and  $R^n$  and  $D^n$  denote respectively the rate and distortion associated with  $y^n$ , then for all  $y^n$  it can be shown that

$$R^n = R_x(D^n) \quad \text{for } 1 \leq n \leq N, \quad (3.10)$$

when

$$D^n \geq D^{n+1} \quad \text{for } 1 \leq n \leq N-1. \quad (3.11)$$

The latter equation merely formalizes the definition of SNR scalability, which specifies that reproductions  $y^n$  have increasing SNR values for larger values of  $n$ .

The proof consists of two parts. The first part provides a description of hierarchical coding in terms of the rate distortion theory. The second part validates Equation 3.10 by determining the conditions for successive refinement.

### 3.3.1 Description of hierarchical coding

When hierarchical coding schemes are described from the perspective of the rate distortion theory, it can be said that the schemes implement  $N$  mappings  $x \rightarrow y^n$  in an embedded way. Figure 3.3 illustrates this by showing the various mappings necessary to create  $y^n$  with  $1 \leq n \leq N$ . The hierarchical coding scheme produces reproduction  $y^n$  by first mapping  $x$  into a coarse reproduction  $y^1$ , i.e.  $x \rightarrow y^1$ . Mapping  $x \rightarrow y^1$  introduces an average distortion  $D^1$  in  $y^1$  and requires  $R^1$  bits per pixel. Subsequently  $y^{n-1}$  with  $2 \leq n \leq N$  is successively refined into the more accurate reproduction  $y^n$ , i.e.  $y^{n-1} \Rightarrow y^n$ , by creating updates about  $x$  relative to  $y^{n-1}$ . Each update facilitates the reconstruction of  $y^n$  with a distortion  $D^n$  and has a bit rate  $\tilde{R}^n$  (per pixel). The mappings  $x \rightarrow y^1$  and the updates  $y^{n-1} \Rightarrow y^n$  with  $2 \leq n \leq N$  together constitute a compound mapping of  $x$  into  $y^n$ , i.e.  $x \rightarrow y^n$ , with coding threshold  $\theta^n$ . Consequently the total bit rate of  $y^n$  is given by

$$R^n = \sum_{m=1}^n \tilde{R}^m \quad \text{for } 1 \leq n \leq N, \quad (3.12)$$

with  $\tilde{R}^1 = R^1$  by definition. The hierarchical coding scheme achieves successive refinement when all mappings  $x \rightarrow y^n$  with  $1 \leq n \leq N$  achieve the rate distortion bound, i.e.  $R^n = R_x(D^n)$ .

### 3.3.2 Conditions for successive refinement

In this section it is shown that successive refinement is always achievable. In particular, Eq. 3.10 is validated by showing that the optimally encoded signal  $y^{n-1}$  with distortion  $D^{n-1}$  and bit rate  $R^{n-1} = R_x(D^{n-1})$  can be refined into signal  $y^n$  with distortion  $D^n < D^{n-1}$  by an update with bit rate  $\tilde{R}^n$  so that

$$R^n = R^{n-1} + \tilde{R}^n = R_x(D^n), \quad \text{for } 2 \leq n \leq N. \quad (3.13)$$

The update consists of the optimally encoded reconstruction error of mapping  $x \rightarrow y^{n-1}$ , i.e.  $r^{n-1} = x - y^{n-1}$ . Note that by definition reproduction  $y^1$  is optimally encoded so that  $R^1 = R_x(D^1)$ .

Figure 3.4 shows the situation under consideration when a Gaussian signal with PSD  $S_{xx}(\omega)$  is encoded in an SNR-scalable way. Here, the optimal mappings  $x \rightarrow y^{n-1}$  with coding threshold  $\theta^{n-1}$  (Fig. 3.4a) and  $x \rightarrow y^n$  with coding threshold  $\theta^n$  (Fig. 3.4b) have to be embedded into a compound mapping  $x \rightarrow y^{n-1} \Rightarrow y^n$  (Fig. 3.4c). Note that the thresholds are related by  $\theta^n < \theta^{n-1}$  because  $D^n < D^{n-1}$ . The spectra of the reproductions, i.e.  $S_{yy}^{n-1}(\omega)$  and  $S_{yy}^n(\omega)$ , are given by

$$S_{yy}^n(\omega, \theta^n) = \max \{0, S_{xx}(\omega) - \theta^n\}. \quad (3.14)$$

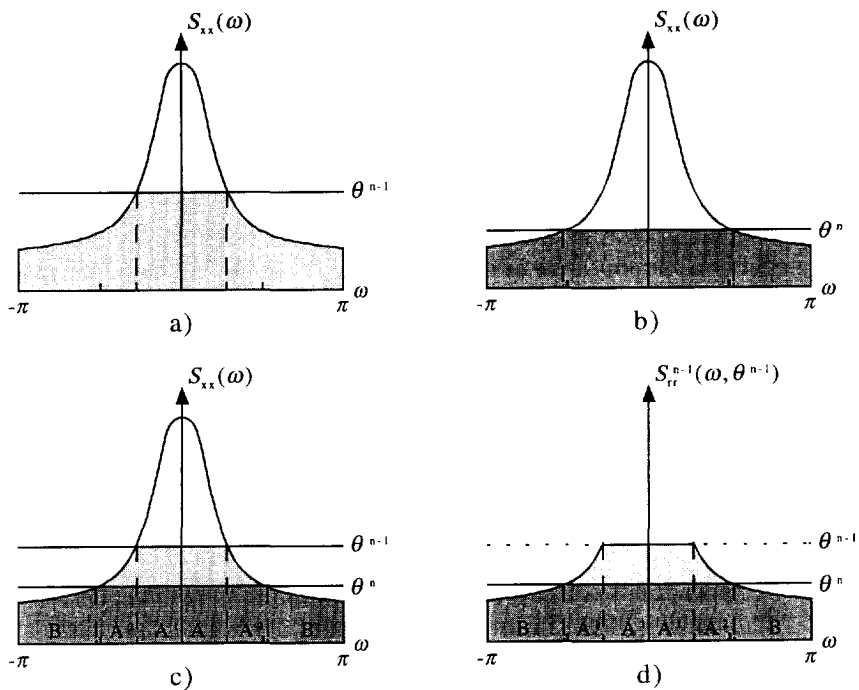


Figure 3.4: Successive refinement for SNR scalability: a) mapping  $x \rightarrow y^{n-1}$ , b) mapping  $x \rightarrow y^n$ , c) compound mapping  $x \rightarrow y^{n-1} \Rightarrow y^n$ , and d) mapping of  $r^{n-1} = x - y^{n-1}$ .

In Fig.3.4c these spectra are the areas above thresholds  $\theta^{n-1}$  and  $\theta^n$ . Hence, from Fig.3.4c, it can be deduced that the update for  $y^n$  has to convey those parts of the input spectrum that are under threshold  $\theta^{n-1}$  but above threshold  $\theta^n$ . Since the area under  $\theta^{n-1}$  represents the error spectrum of mapping  $x \rightarrow y^{n-1}$ , i.e.  $S_{rr}^{n-1}(\omega, \theta^{n-1})$  of  $r^{n-1} = x - y^{n-1}$ , this observation can be rephrased into the statement that the update for  $y^n$  consists of the error spectrum  $S_{rr}^{n-1}(\omega, \theta^{n-1})$  encoded with coding threshold  $\theta^n$ , or equivalently, distortion  $D^n$  (Fig. 3.4d). Because the reconstruction error signal  $r^{n-1}$  is Gaussian distributed, the optimal encoding of the reconstruction error with distortion  $D^n$  requires a bit rate

$$\tilde{R}^n(\theta^n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \max \left\{ 0, \frac{1}{2} \log_2 \frac{S_{rr}^{n-1}(\omega, \theta^{n-1})}{\theta^n} \right\} d\omega, \quad (3.15)$$

with  $\theta^n < \theta^{n-1}$  and  $S_{rr}^{n-1}(\omega, \theta^{n-1})$  defined as (see also Fig.3.4d)

$$S_{rr}^{n-1}(\omega, \theta^{n-1}) = \begin{cases} \theta^{n-1} & \omega \in \{A^1\}, \\ S_{xx}(\omega) & \omega \in \{A^2, B\}. \end{cases} \quad (3.16)$$

Hence, the bit rate  $R^n$  of the compound mapping of  $y^n$  with distortion  $D^n$  is given by

$$\begin{aligned} R^n &= R^{n-1} + \tilde{R}^n \\ &= \frac{1}{2\pi} \int_{\omega \in \{A^1\}} \frac{1}{2} \log_2 \frac{S_{xx}(\omega)}{\theta^{n-1}} d\omega + \frac{1}{2\pi} \int_{\omega \in \{A^1\}} \frac{1}{2} \log_2 \frac{\theta^{n-1}}{\theta^n} d\omega + \\ &\quad \frac{1}{2\pi} \int_{\omega \in \{A^2\}} \frac{1}{2} \log_2 \frac{S_{xx}(\omega)}{\theta^n} d\omega \\ &= \frac{1}{2\pi} \int_{\omega \in \{A^1, A^2\}} \frac{1}{2} \log_2 \frac{S_{xx}(\omega)}{\theta^n} d\omega \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \max \left\{ 0, \frac{1}{2} \log_2 \frac{S_{xx}(\omega)}{\theta^n} \right\} d\omega. \end{aligned} \quad (3.17)$$

Comparison with Eq. 3.2 reveals that the right hand side of Eq. 3.17 is equal to the bit rate of the optimal mapping  $x \rightarrow y^n$  with coding threshold  $\theta^n$ . Hence, Eq. 3.13 has been validated and thus it is shown that successive refinement of a Gaussian input signal is always achievable for SNR scalability. Specifically, the theoretical coding scheme achieves successive refinement by making the update for  $y^n$  equal to the optimally encoded reconstruction spectrum of  $y^{n-1}$ .

### 3.4 Theoretical analysis of spatial scalability

Now it will be proven that for spatial scalability successive refinement of a Gaussian input signal is only achievable for particular combinations of bit rates and distortions. These combinations are dependent on the statistics of the input signal and on the ratios between the sampling rate of  $x$  (i.e.  $f_s$ ) and the sampling rates  $y^n$  (i.e.  $f_s^n = f_s/\beta^n$ ). Therefore, as an example, the conditions obtained for successive refinement are evaluated for a particular two-dimensional model of the input signal and two reproductions  $y^1$  and  $y^2$  that have quarter-size and full-size formats, respectively.

The proof itself consists of two parts. The first part provides a clear definition of successive refinement for spatial scalability. The definition of successive refinement for spatial scalability

differs from the definition used for SNR scalability and involves the definition of separate RDFs for the optimal mappings  $x \rightarrow y^n$ . The second part of the proof determines under which conditions successive refinement is obtained. The result is a set of conditions expressed in the coding thresholds  $\theta^n$  and the decimation factors  $\beta^n$  of the reproductions.

### 3.4.1 Definition of successive refinement

The definition of successive refinement for spatial scalability differs from the definition used for SNR scalability. The reason for this is that the optimal mappings  $x \rightarrow y^n$  are not well described by the RDF  $R_x(\cdot)$ , because  $x$  and reproductions  $y^n$  have different sampling rates, namely,  $f_s$  and  $f_s^n = f_s/\beta^n$ . Hence, the reconstructed spectrum of  $y^n$  can maximally consist of the frequency band  $[-\frac{\pi}{\beta^n}, \frac{\pi}{\beta^n}]$  of the input spectrum  $S_{xx}(\omega)$ , assuming Nyquist sampling. Consequently, when  $R^n \rightarrow \infty$  the RDF  $R_x(\cdot)$  is lower bounded for different reproductions by different amounts of distortion due to discarded high frequency components.

To circumvent these lower bounds for the distortion of  $y^n$ , successive refinement for spatial scalability is defined as

$$R^n = R_x^n(D^n), \quad \text{for } 1 \leq n \leq N. \quad (3.18)$$

Here  $R_x^n(\cdot)$  is a dedicated RDF of the optimal mapping  $x^n \rightarrow y^n$ , where  $x^n$  denotes the input signal resampled at a sampling rate  $f_{s,y}^n$  after it has been perfectly band-limited to frequency band  $[-\frac{\pi}{\beta^n}, \frac{\pi}{\beta^n}]$ . The spectrum of  $x^n$ , denoted by  $S_{xx}^n(\omega)$ , is therefore free of aliasing and is given by [Vaid87]

$$S_{xx}^n(\omega) = \frac{1}{\beta^n} S_{xx}\left(\frac{\omega}{\beta^n}\right), \quad -\pi \leq \omega \leq \pi. \quad (3.19)$$

Since  $x^n$  is Gaussian distributed [Pear91], the RDF of  $x^n \rightarrow y^n$  is given parametrically by Eq. 3.3, i.e.

$$R_x^n(\theta^n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \max \left\{ 0, \frac{1}{2} \log_2 \frac{S_{xx}^n(\omega)}{\theta^n} \right\} d\omega, \quad (3.20)$$

$$D_x^n(\theta^n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \min \{ S_{xx}^n(\omega), \theta^n \} d\omega, \quad (3.21)$$

where  $\theta^n$  is the coding threshold. Note that because of this definition the bit rate  $R_x^n(\theta^n)$  and average distortion  $D_x^n(\theta^n)$  are defined with respect to the sampling frequency of reproduction  $y^n$ . Further note that the distortion  $D^n$  (of  $y^n$  relative to  $x^n$ ) goes indeed to zero when  $R^n \rightarrow \infty$ .

In a hierarchical coding scheme, mapping  $x^{n-1} \rightarrow y^{n-1}$  is embedded in mapping  $x^n \rightarrow y^n$ . To reconstruct  $y^n$ , signal  $y^{n-1}$  is interpolated by a factor  $\beta = \beta^{n-1}/\beta^n$  because  $y^{n-1}$  has a reduced sampling frequency with respect to  $y^n$ . It is illustrative to see how this interpolation affects the optimal mapping  $x^{n-1} \rightarrow y^{n-1}$  as given by Eqs. 3.20 and 3.21. Ideally, the interpolation procedure consists of an upsampler that increases the rate by  $\beta$ , and an interpolation filter that perfectly limits the bandwidth of the upsampled signal to  $[-\frac{\pi}{\beta}, \frac{\pi}{\beta}]$ . Such an interpolation can be expressed as [Vaid87]

$$\hat{S}_{xx}^{n-1}(\omega) = \begin{cases} \beta S_{xx}^{n-1}(\beta\omega) & -\frac{\pi}{\beta} \leq \omega \leq \frac{\pi}{\beta}, \\ 0 & \frac{\pi}{\beta} < |\omega| \leq \pi, \end{cases} \quad (3.22)$$



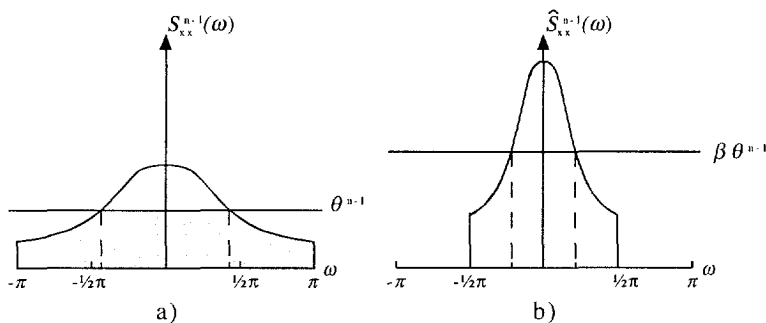


Figure 3.5: Interpolation of optimal mapping  $x^{n-1} \rightarrow y^{n-1}$ ; a) optimal mapping, b) optimal mapping after interpolation by  $\beta = 2$ .

where  $\hat{S}_{xx}^{n-1}(\omega)$  denotes the interpolated spectrum of  $x^{n-1}$ . Using some simple mathematics, Eqs. 3.20 and 3.21 can then be rewritten as

$$R_x^{n-1}(\theta^{n-1})/\beta = \frac{1}{2\pi} \int_{-\pi/\beta}^{\pi/\beta} \max \left\{ 0, \frac{1}{2} \log_2 \frac{S_{xx}^n(\omega)}{\beta \theta^{n-1}} \right\} d\omega, \quad (3.23)$$

$$D_x^{n-1}(\theta^{n-1}) = \frac{1}{2\pi} \int_{-\pi/\beta}^{\pi/\beta} \min \left\{ \beta \theta^{n-1}, S_{xx}^n(\omega) \right\} d\omega, \quad (3.24)$$

where the fact is used that  $\hat{S}_{xx}^{n-1}(\omega) = S_{xx}^n(\omega)$  for  $[-\frac{\pi}{\beta} \leq \omega \leq \frac{\pi}{\beta}]$ . Eqs. 3.23 and 3.24 show that the optimal mapping  $x^{n-1} \rightarrow y^{n-1}$  is transformed by the interpolation procedure into an optimal mapping of the frequency band  $[-\frac{\pi}{\beta} \leq \omega \leq \frac{\pi}{\beta}]$  of  $S_{xx}^n(\omega)$  with a coding threshold  $\beta \theta^{n-1}$ . Figure 3.5 illustrates this transformation of the mapping  $x^{n-1} \rightarrow y^{n-1}$  for the case of  $\beta = 2$ . After interpolation, the bit rate of mapping  $x^{n-1} \rightarrow y^{n-1}$  is equal to  $R_x^{n-1}(\theta^{n-1})/\beta$ , which reflects the change in sampling rate from  $f_s/\beta^{n-1}$  to  $f_s/\beta^n$ . The associated distortion  $D^{n-1}(\theta^{n-1})$  is not modified by the interpolation procedure because the interpolation includes a scaling factor of  $\beta$ .

### 3.4.2 Conditions for successive refinement

Given the definition of successive refinement for spatial scalability, it is now shown that successive refinement is only achieved when the thresholds  $\theta^n$  and relative sampling ratios  $\beta^n$  are related as follows:

$$\beta^n \theta^n \geq \beta^{n+1} \theta^{n+1} \quad \text{for } 1 \leq n \leq N - 1. \quad (3.25)$$

These conditions are obtained by evaluating when an optimally encoded signal  $y^{n-1}$  with distortion  $D^{n-1}$  and rate  $R^{n-1} = R_x^{n-1}(D^{n-1})$  can be refined into signal  $y^n$  with distortion  $D^n$  by an update with bit rate  $\tilde{R}^n$  so that

$$R^n = \frac{\beta^n}{\beta^{n-1}} R^{n-1} + \tilde{R}^n = R_x^n(D^n), \quad \text{for } 2 \leq n \leq N. \quad (3.26)$$

In Eq. 3.26 the fact is used that the bit rate  $R^{n-1}$  contributes only partly to bit rate  $R^n$  because of the different sampling frequencies of  $y^{n-1}$  and  $y^n$ . The update consists of the following two components which are jointly encoded using coding parameter  $\theta^n$ : i.e. the reconstruction error of mapping  $x^{n-1} \rightarrow y^{n-1}$ , i.e.  $r^{n-1} = x^{n-1} - y^{n-1}$ , and the high-frequency components which are part of  $x^n$  but not of the interpolated  $x^{n-1}$  signal. Note that by definition reproduction  $y^1$  is optimally encoded so that  $R^1 = R_x^1(D^1)$ .

For the case of  $\beta = \beta^{n-1}/\beta^n = 2$ , Figure 3.6 shows the situation under consideration when a Gaussian signal with PSD  $S_{xx}(\omega)$  is encoded using spatial scalability. In Fig. 3.6 the optimal mappings  $x \rightarrow y^{n-1}$  with coding threshold  $\theta^{n-1}$  (Fig. 3.6a) and  $x \rightarrow y^n$  with coding threshold  $\theta^n$  (Fig. 3.6b) have to be embedded into a compound mapping  $x \rightarrow y^{n-1} \Rightarrow y^n$  (Fig. 3.6c). Note that the mapping  $x \rightarrow y^{n-1}$  is shown after interpolation by a factor  $\beta$  to make it match the sampling rate of  $x^n$  and  $y^n$ . In Fig. 3.6c the reconstructed parts of the input spectrum for signals  $y^{n-1}$  and  $y^n$  are graphically interpretable as respectively the area above threshold  $\beta\theta^{n-1}$  in the frequency region  $[-\pi/\beta, \pi/\beta]$  and the area above threshold  $\theta^n$  for the region  $[-\pi, \pi]$ . Hence it can be concluded that the update for  $y^n$  has to convey the following spectral contributions:

- the encoded error spectrum of the mapping  $x^{n-1} \rightarrow y^{n-1}$  for  $\omega \in \{A^1, A^2\}$  as illustrated in Fig. 3.6d. This contribution is the encoded reconstruction error signal  $r^{n-1} = x^{n-1} - y^{n-1}$ .
- the encoded input spectrum for  $\omega \in A^3$  as illustrated in Fig. 3.6e. This contribution consists of the encoded (high-)frequency components that are part of  $x^n$  but not of the interpolated  $x^{n-1}$  signal.

Both contributions are encoded with the coding threshold  $\theta^n$  to make the distortion  $D^n$  of the compound mapping  $x \rightarrow y^n$  equal to the distortion of the optimal mapping  $x^n \rightarrow y^n$  (c.f. Eq. 3.24). Since both contributions are Gaussian distributed, the bit rate of the update for  $y^n$  is given by

$$\bar{R}^n(\theta^n) = \frac{1}{2\pi} \int_{\omega \in A^1} \frac{1}{2} \log_2 \frac{\beta\theta^{n-1}}{\theta^n} d\omega + \frac{1}{2\pi} \int_{\omega \in \{A^2, A^3\}} \frac{1}{2} \log_2 \frac{S_{xx}^n(\omega)}{\theta^n} d\omega. \quad (3.27)$$

Consequently, the bit rate  $R^n$  of the compound mapping  $x^n \rightarrow y^n$  with distortion  $D^n$  is given by

$$\begin{aligned} R^n &= \frac{1}{\beta} R^{n-1} + \bar{R}^n \\ &= \frac{1}{2\pi} \int_{\omega \in A^1} \frac{1}{2} \log_2 \frac{S_{xx}^n(\omega)}{\beta\theta^{n-1}} d\omega + \frac{1}{2\pi} \int_{\omega \in A^1} \frac{1}{2} \log_2 \frac{\beta\theta^{n-1}}{\theta^n} d\omega + \\ &\quad \frac{1}{2\pi} \int_{\omega \in \{A^2, A^3\}} \frac{1}{2} \log_2 \frac{S_{xx}^n(\omega)}{\theta^n} d\omega. \end{aligned} \quad (3.28)$$

If and only if  $\beta\theta^{n-1} \geq \theta^n$ , this result transforms into

$$R^n = \frac{1}{2\pi} \int_{-\pi}^{\pi} \max \left\{ 0, \frac{1}{2} \log_2 \frac{S_{xx}^n(\omega)}{\theta^n} \right\} d\omega, \quad (3.29)$$

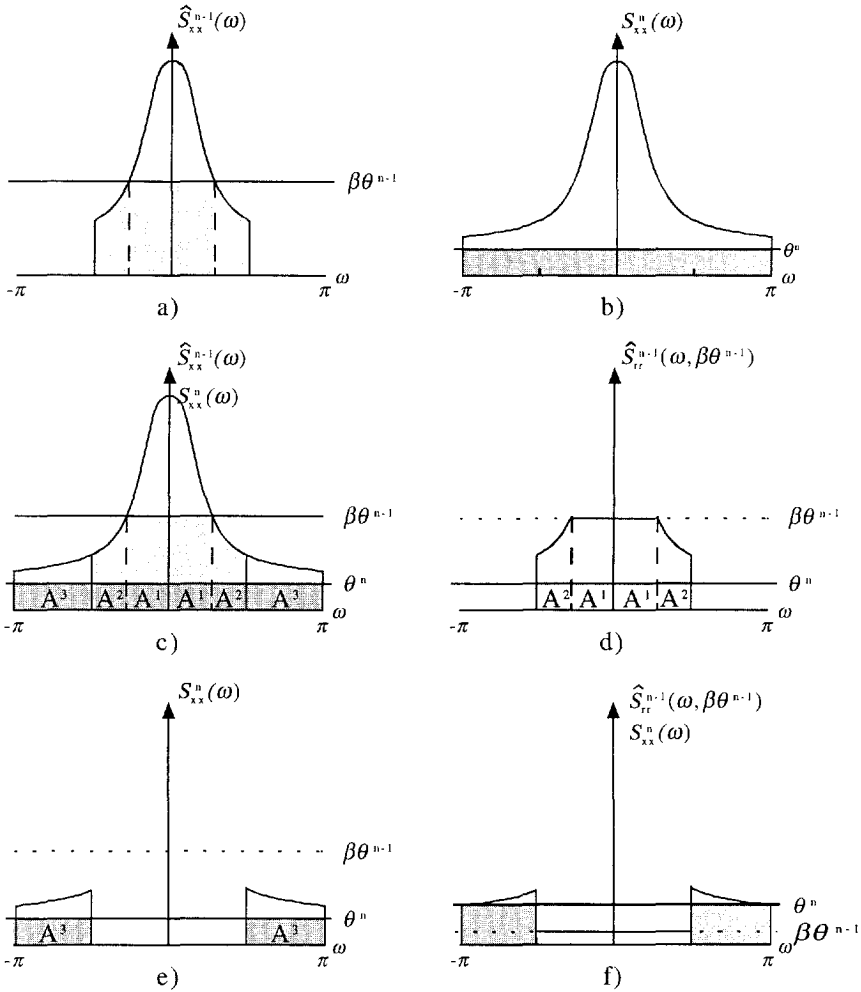


Figure 3.6: Successive refinement for spatial scalability: a) interpolated mapping  $x \rightarrow y^{n-1}$  by  $\beta = 2$ , b) mapping  $x \rightarrow y^n$ , c) compound mapping  $x \rightarrow y^{n-1} \Rightarrow y^n$ , d) mapping of error spectrum of  $y^{n-1}$ , e) mapping of high-frequency components that belong to  $x^n$  but not to the interpolated  $x^{n-1}$  signal, and f) sub-optimal coding of  $y^n$ .

which is equal to the bit rate of the optimal mapping  $x^n \rightarrow y^n$  as described by Eq. 3.23. With this result Eq. 3.25 is validated because condition  $\beta\theta^{n-1} \geq \theta^n$  can easily be expanded for multiple reproductions by replacing  $\beta$  with  $\beta^{n-1}/\beta^n$  and varying  $n$  from 2 to  $N$ .

If  $\beta\theta^{n-1} < \theta^n$  then successive refinement of the input signal is not achievable (see Fig. 3.6f). In this case, the optimal mapping  $x^{n-1} \rightarrow y^{n-1}$  encodes the frequencies in  $A^1$  more accurately than optimal mapping  $x^n \rightarrow y^n$  would do. Hence, the reconstructed spectrum of  $y^n$  contains frequency components that are encoded too accurately (i.e. with threshold  $\beta\theta^{n-1}$  instead of  $\theta^n$ ). Consequently, the compound mapping  $x^n \rightarrow y^n$  can never equal the optimal mapping.

Finally, it is stated here that Eq. 3.25 also holds for 2-D signals. In that case  $\beta^n$  is defined as the ratio of the product of the horizontal and vertical decimation factors and the sampling frequency of  $x$ . This statement is easily verified by replacing single indices and variables in the above equations by double ones, and by replacing intervals by regions.

### 3.4.3 Bit rate and distortion areas

The condition obtained for successive refinement with spatial scalability is not very informative in an image coding environment. In such an environment, coding schemes are described with bit rates and MSE distortions and not by their coding thresholds. Fortunately, for a particular model of the input signal, the bit rate and distortion of mapping  $x^n \rightarrow y^n$  are given by the RDF as a function of the coding threshold. In this case it is thus possible to convert condition  $\beta^{n-1}\theta^{n-1} > \beta^n\theta^n$  via the RDFs into explicit conditions in terms of bit rates  $R^n$  and distortions  $D^n$ .

As an example, the condition for successive refinement is converted for a coding scheme that encodes a Gaussian input image using two levels of spatial scalability (i.e.  $N = 2$ ). The assumed coding scheme supports two reproductions  $y^1$  and  $y^2$ , which are respectively quarter-size ( $\beta^1 = 4$ ) and full-size ( $\beta^2 = 1$ ) images. The input image is assumed to have an isotropic autocorrelation function with the following PSD function [Eric85]

$$S_{xx}(\omega_1, \omega_2) = \frac{2\pi\sigma_x^2}{\alpha_h\alpha_v} \frac{1}{\left(1 + \left(\frac{\omega_1}{\alpha_h}\right)^2 + \left(\frac{\omega_2}{\alpha_v}\right)^2\right)^{\frac{3}{2}}}, \quad \text{for } -\pi \leq \omega_1, \omega_2 \leq \pi. \quad (3.30)$$

Here,  $\alpha_h$  and  $\alpha_v$  are the horizontal and vertical correlation coefficient, respectively.

To obtain the bit rate combinations ( $R^1, R^2$ ) and distortion combinations ( $D^1, D^2$ ) for which successive refinement is achievable, the RDFs  $R_x^1(\cdot)$  and  $R_x^2(\cdot)$  are used. These RDFs are based on Equation 3.30 with  $\alpha_h = \alpha_v = 0.05$  and  $\sigma_x^2 = 1$ . In particular, to determine for which bit rate and distortion combinations  $\beta^1\theta^1 \geq \theta^2$ , the combinations ( $R^1(\theta^1), R^2(\theta^2)$ ) and ( $D^1(\theta^1), D^2(\theta^2)$ ) are calculated that are related by

$$\beta^1\theta^1 = \theta^2. \quad (3.31)$$

By varying  $\theta^2$ , the combinations ( $R^1(\theta^2/\beta^1), R^2(\theta^2)$ ) and ( $D^1(\theta^2/\beta^1), D^2(\theta^2)$ ) make up two lines in the associated bit rate and distortion graphs of Figures 3.7 and 3.8, respectively. These lines are henceforth called the 'equality lines' and form the boundaries between areas 'A' where successive refinement is achievable and areas 'B' where successive refinement is not

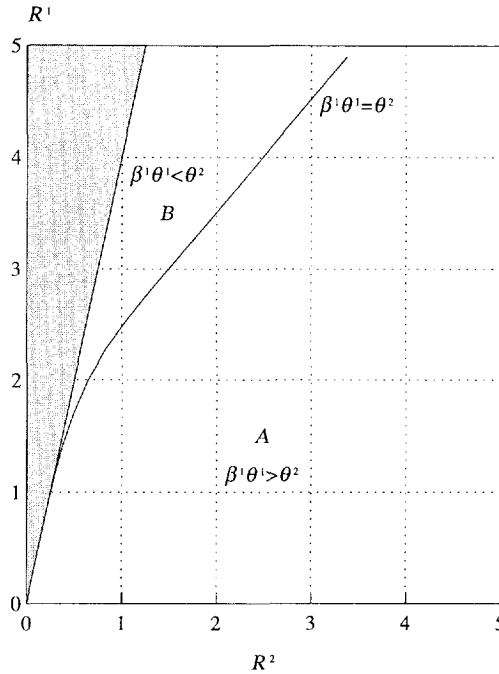


Figure 3.7: Bit rate graph: In area 'A' successive refinement is achievable, while in area 'B' successive refinement is not achievable; the shaded area contains bit rate combinations that are not possible.

achievable. Areas 'A' and 'B' are easily identified by determining the bit rate and distortion combinations where only bits are used to encode  $y^2$ , i.e.  $\theta^2 < \beta^1 \theta^1 = \max\{S_{xx}(\omega_1, \omega_2)\}$ . These combinations are part of the areas 'A', i.e. the areas where successive refinement is achievable.

The shaded areas in Figs. 3.7 and 3.8 contain bit rate and distortion combinations that can not exist for the assumed coding scheme and assumed model of the input signal. For instance, bit rate  $R^2$  can never be smaller than  $R^1 / \beta^1 (= R^1 / 4)$  because it includes the scaled bit rate of  $y^1$  (c.f. Eq. 3.26). Similarly, when  $y^1$  is encoded very accurately so that  $D^1 \rightarrow -\infty$  dB, the distortion of  $y^2$  can never be larger than the variance of the high frequency components that are solely part of  $y^2$ .

From Figs. 3.7 and 3.8 it can be concluded that successive refinement is achievable for most bit rate and distortion combinations. In particular, when the encoding of  $y^1$  is done at a low bit rate and additional bits are available for  $y^2$  then successive refinement is achievable. If  $R^1 = R^2$  then successive refinement is always achievable. Successive refinement is not achievable when too many bits are spent on the encoding of  $y^1$ , thereby implicitly encoding particular frequency components of  $y^2$  too accurately for encoding at bit rate  $R^2$ . From Fig. 3.8 it can be observed that at high bit rates the distortions are approximately related by  $D^2 = D^1 - 6$  dB. This is because both mappings  $x^n \rightarrow y^n$  are in their linear areas. This in

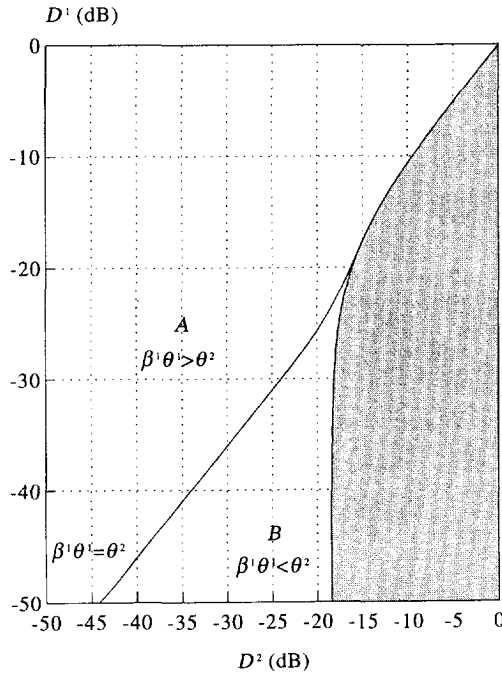


Figure 3.8: Distortion graph: In area 'A' successive refinement is achievable, while in area 'B' successive refinement is not achievable; the shaded area contains distortion combinations that are not possible.

turn implies that the distortion of a reproduction is inversely proportional to the decimation factor of the reproduction.

### 3.5 Experimental verification

The conditions for successive refinement that are obtained from the previous theoretical analyses are verified in this section. To this end, the coding performance of a practical hierarchical coding scheme is evaluated for both SNR scalability and spatial scalability. The hierarchical coding scheme is based on the basic intraframe subband coder described in Section 2.2 and is henceforth referred to as the Refinement System [Bosv90]. To account for the fact that practical schemes have coding performances that are less than the RDF, and that natural images are generally not Gaussian distributed, this section employs the following definition of successive refinement: successive refinement is achieved when the coding performance of reproduction  $y^n$  is approximately equal to the coding performance obtained by the basic subband scheme of Section 2.2. The luminance component of the first frame of the digital TV sequence 'Teeny' is used as a test image. This image is progressively scanned and consists of 720x576 pixels.

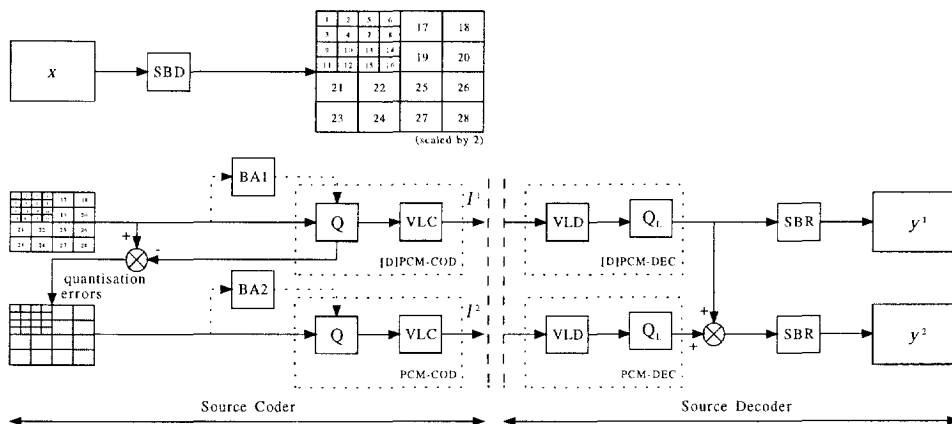


Figure 3.9: Refinement system for SNR-scalability.

### 3.5.1 SNR scalability

#### 3.5.1.1 Refinement system

The Refinement System implements SNR scalability based on the insights obtained by the theoretical analysis of SNR scalability. These insights say that successive refinement is always achievable when the update for reproduction  $y^n$  conveys the encoded reconstruction errors of  $y^{n-1}$ .

Figure 3.9 illustrates the structure of the Refinement system for the case of  $N = 2$ . The encoder consists of a tree-structured subband decomposition and two coding stages. The subband decomposition (*SBD*) decomposes the input signal  $x$  into 28 subbands using the QMF16b filter [John80]. Next, the first coding stage creates layer  $l^1$  by encoding all subbands using DPCM/PCM. The subbands are encoded under supervision of a private forward bit allocation algorithm (*BA1*) that distributes the available bits  $R^1$  optimally among the subbands. It does this by selecting Lloyd-Max Quantizers (*Q*) from a set of 14 available quantizers. Subband 1 is DPCM encoded, while the other subbands are PCM encoded. After coding, the quantized subbands are Huffman encoded (*VLC*) before transmission.

Subsequently, the second coding stage creates layer  $l^2$ . It does this by first calculating per subband the quantization errors made in the first coding stage. Secondly, a forward bit allocation algorithm *BA2* distributes the available bits (i.e.  $\tilde{R}^2 = R^2 - R^1$ ) among the subbands, which are all PCM encoded. The LMQ quantizers of the second coding stage are different from those of the first stage because they are optimized for quantization errors. There are 8 available quantizers for the second coding stage.

The decoders can reconstruct reproduction  $y^1$  by receiving and decoding  $l^1$  and by reconstructing the decoded subbands. To reconstruct  $y^2$ , the decoder also has to receive and decode layer  $l^2$ . The decoded subbands of layer  $l^2$  are then added to the decoded subbands of the first decoding stage. The summed subbands are then reconstructed to form  $y^2$ . Note that the Refinement System can easily support more than 2 reproductions by extending it with additional coding stages.

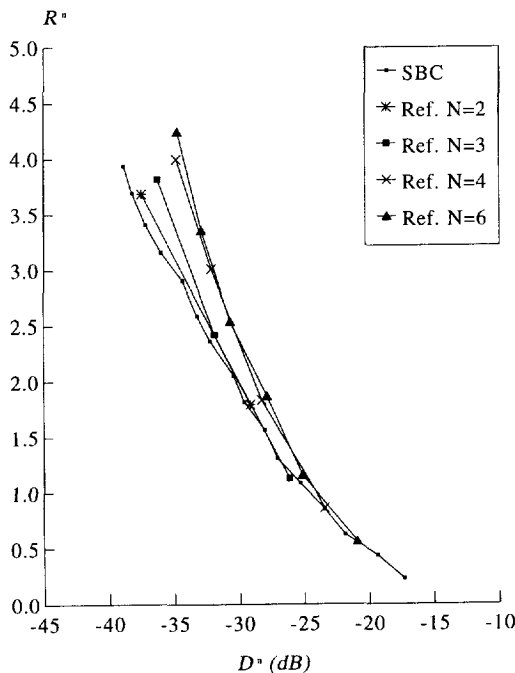


Figure 3.10: SNR scalability: Coding performances of basic subband scheme (SBC) and Refinement System supporting  $N$  reproductions.

### 3.5.1.2 Successive refinement

Figure 3.10 shows the coding performances obtained for the Refinement system and the basic subband scheme. The vertical axis shows the bit rate of the coding scheme, while the horizontal axis shows the obtained distortion using the SNR measure. Fig. 3.10 shows 4 performance curves for the Refinement System that are obtained when the system supports  $N$  reproductions with  $N = 2, 3, 4$  and  $6$ . The Refinement System makes the bit rate of each layer equal to  $\tilde{R}^n = \frac{R}{N}$  so that the bit rate of  $y^N$  is approximately 4 bpp. The curves illustrate the performances obtained (i.e.  $R^n, D^n$ ) for each reproduction. The curve for the basic subband scheme illustrates the SNR obtained when encoding the test image at bit rates  $R$  with  $0.25 \leq R \leq 4.0$ .

From Fig. 3.10 it can be concluded that successive refinement for SNR scalability is not always achieved in practice. Specifically, if more than 3 reproductions are supported, the performances for the higher quality reproductions differ significantly ( $> 2.5$  dB) from the performances obtained by the basic subband scheme. The reason for this is that multiple concatenated [D]PCM encoders are used to encode the subbands. The performance of these concatenated coders is evidently less than the performance of a single [D]PCM encoder used at the equivalent bit rate. Nonetheless, when the number of reproductions supported is low, i.e.  $N = 2$  or  $N = 3$ , successive refinement is achieved, because the coding performances



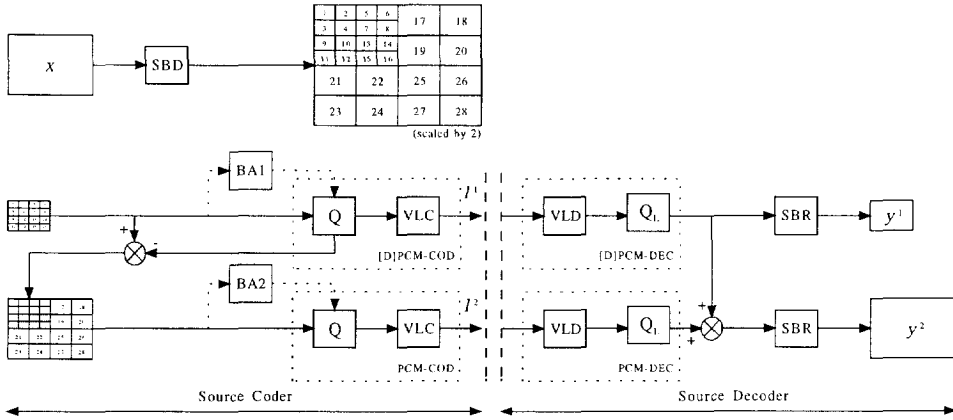


Figure 3.11: Refinement system for spatial scalability.

of  $y^2$  and  $y^3$  are approximately equal to the performance obtained with the basic subband scheme.

### 3.5.2 Spatial scalability

#### 3.5.2.1 Refinement system

The Refinement System also implements spatial scalability based on the insights obtained by the theoretical analysis of spatial scalability. These insights show that successive refinement is only achievable for particular bit rate combinations, and only if the update for reproduction  $y^n$  conveys information that decreases the reconstruction errors of  $y^{n-1}$  as well as information about the high-frequency components that are part of  $y^n$  but not of  $y^{n-1}$ .

Figure 3.11 illustrates the structure of the spatially scalable Refinement System for the case of  $N = 2$ . The two reproductions  $y^1$  and  $y^2$  are quarter-size and full-size images and can be reconstructed using subbands 1-16 and 1-28, respectively. Hence, after the subband decomposition, the first coding stage creates layer  $l^1$  by encoding subbands 1-16 in a way similar to that with the Refinement System for SNR scalability. Subsequently, the second coding stage creates layer  $l^2$ . It does this by calculating the quantization errors made in the first coding stage for subbands 1-16. Secondly, a forward bit allocation algorithm  $BA2$  distributes the available bits (i.e.  $\tilde{R}^2 = R^2 - \frac{R^1}{4}$ ) among the subbands 1-28, which are all PCM encoded. Subbands 1-16 contain the quantization errors of the first coding stage, while subbands 17-28 contain original data from the input image. Note that the LMQ quantizers used to encode subbands 1-16 are different than those used for subbands 17-28 because the characteristics of the input data are different.

The decoders can reconstruct reproduction  $y^1$  by receiving and decoding  $l^1$  and by reconstructing ( $SBR$ ) the decoded subbands. To reconstruct  $y^2$ , the decoder also has to receive and decode layer  $l^2$ . The decoded subbands 1-16 are then added to the decoded subbands of the first decoding stage. The summed subbands 1-16 and decoded subbands

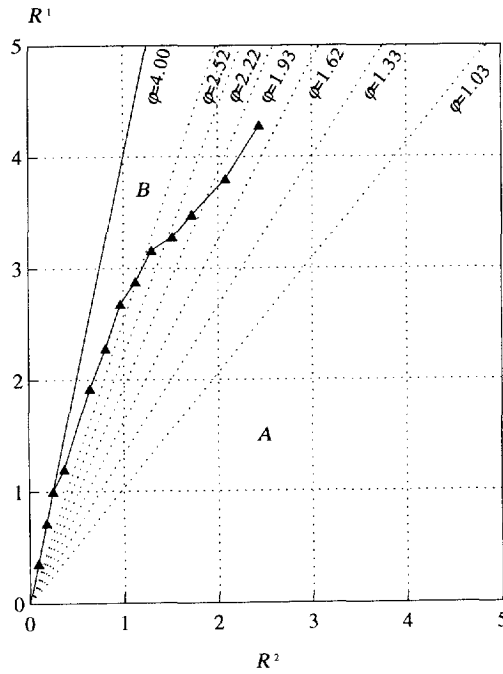


Figure 3.12: Experimental bit rate graph for spatial scalability (with  $\varphi$ -lines).

17-28 are then reconstructed to form  $y^2$ . Note that the Refinement System can easily support  $N$  reproductions with  $N > 2$  by extending it with more coding stages.

### 3.5.2.2 Successive refinement

The verification of the condition for successive refinement for spatial scalability is more complex than the verification for SNR scalability. To limit the complexity of the verification, it is assumed that the Refinement System supports two reproductions that are quarter-size and full-size images. As a first step in the verification, the practical equivalents of the bit rate and distortion areas of Section 3.4.3 are determined. These areas indicate for which bit rate and distortion combinations successive refinement should be achievable and for which combinations this should not be achievable. Subsequently, as a second step, it is then determined whether successive refinement is indeed achieved in the area where it should, and is not achieved in the area where it shouldn't.

The determination of the practical bit rate and distortion areas implies that the practical 'equality lines' in the bit rate and distortion graphs should be resolved. Surprisingly, the basic subband scheme is very suitable for this; the 'equality line' in the bit rate graph indicates those bit rates ( $R^1$ ,  $R^2$ ) for which a hierarchical coding scheme optimally encodes  $y^2$  without encoding any reconstruction errors of  $y^1$ . In other words, the bit rate  $R^1$  used

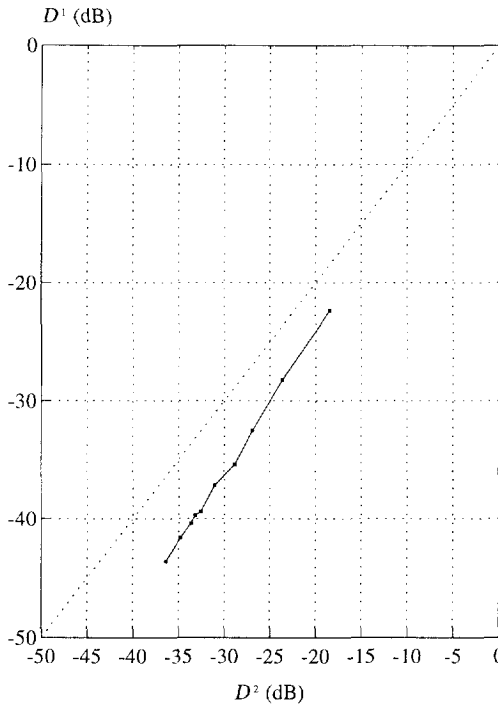


Figure 3.13: Experimental distortion graph for spatial scalability.

to encode subbands 1–16 is exactly equal to the bit rate that has to be allocated to these subbands to optimally encode  $y^2$  with bit rate  $R^2$ . Hence, when  $x$  is optimally encoded by a non-hierarchical coder at bit rate  $R^2$ , the bit rate allocated to subbands 1–16, let us say  $R^*$ , indicates exactly the maximum bit rate  $R^1 = 4R^*$  that a hierarchical coder may use to code  $y^1$  and still achieve successive refinement. Therefore if the basic subband scheme encodes  $x$  at rate  $R^2$ , the combination  $(4R^*, R^2)$  yields a point of the 'equality line'. By repeating this process for various bit rates  $R^2$ , the complete equality line in the bit rate graph is obtained. Similarly, when the distortions  $D^1$  and  $D^2$  are measured at the same time, the equality line in the distortion graph is also obtained.

Figures 3.12 and 3.13 show the practically obtained 'equality lines' in the bit rate and distortion graph, respectively. It is immediately seen that these practical 'equality lines' correspond highly to the theoretical results of Figs. 3.7 and 3.8. It should be noted that this close correspondence does not originate from the fact that the theoretical signal model as defined in Section 3.4.3 is valid for the test images. In addition, it is also neither true that the test images are Gaussian distributed nor that the basic subband scheme implements an optimal mapping. Rather, the close correspondence is due to the hierarchical coding concept itself and the fact that the theoretical and experimental coders support reproductions with the same decimation factors.

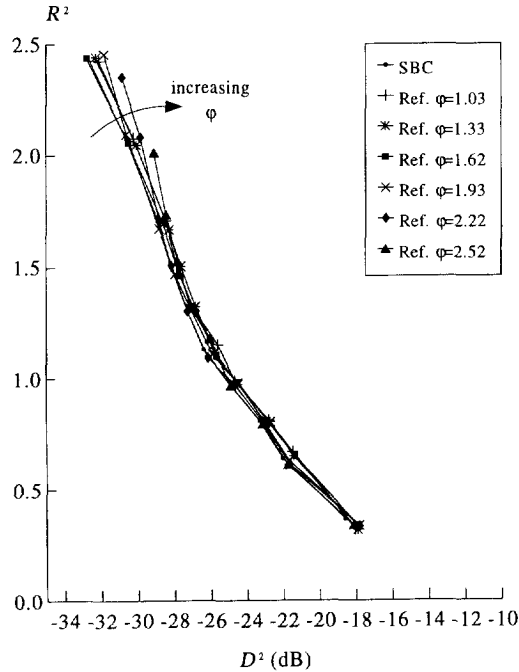


Figure 3.14: Spatial scalability: Coding performance for  $y^2$  for Refinement System and basic subband scheme (SBC).

As a second step in the verification of the condition for successive refinement, it is now evaluated if the Refinement system can obtain the same performance as the basic subband coding scheme. To this end, the coding performance for  $y^2$  is examined for particular bit rate combinations that are defined by  $R^1 = \varphi R^2$ . If  $R^2$  varies with  $\varphi$  constant, this relation results in a straight line in the bit rate graph for each  $\varphi$ , as is illustrated in Figure 3.12. These lines are henceforth referred to as  $\varphi$ -lines.

Figure 3.14 shows the performances of the Refinement System obtained for several bit rate combinations on the  $\varphi$ -lines with  $\varphi > 1.03$ . In addition, it shows the performance obtained with the basic subband scheme as well. For bit rate combinations in area 'A', i.e. at low bit rates or for low values of  $\varphi$ , all curves of the Refinement System are very close to the curve of the basic subband scheme. Hence, it can be said that successive refinement is achieved in this area. For bit rate combinations in area 'B', i.e. at high bit rates and high values of  $\varphi$ , the performance of the Refinement system for  $y^2$  starts to diverge from the performance of the basic subband scheme. Hence, in area 'B', it could be said that successive refinement is not achieved by the Refinement system. These results correspond with the theoretically obtained conditions and therefore verify the theoretical analysis.

Figure 3.15 illustrates the results of an additional experiment, showing the effects on the coding performance for  $y^2$  when the update for  $y^2$  does not contain the encoded reconstruction errors of  $y^1$ . In this case, the quantization errors propagate into  $y^2$  and will thus decrease the

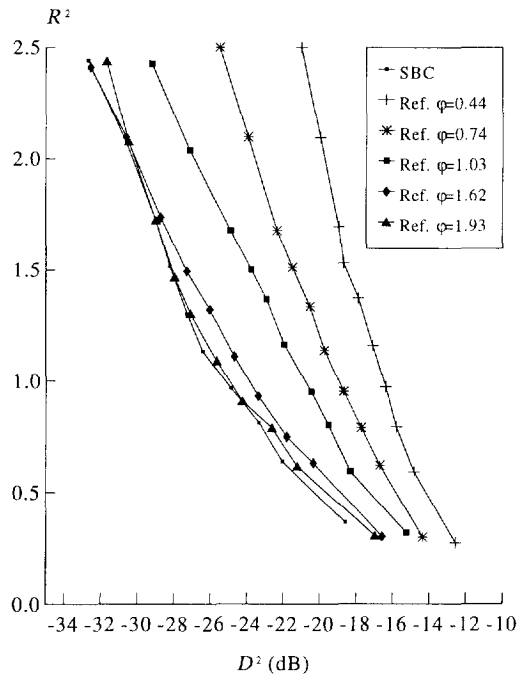


Figure 3.15: Spatial scalability: Coding performance for  $y^2$  for modified Refinement System (no encoding of reconstruction errors of  $y^1$ ) and basic subband scheme (SBC).

coding performance. From Fig. 3.15 it can be observed that the performance of the modified Refinement System is indeed significantly degraded. In particular, successive refinement is no longer achieved in area 'A'. The coding performance for  $y^2$  improves if  $\varphi$  (and thus  $R^1$  for a certain  $R^2$ ) is increased, because the quantization errors of  $y^1$  become smaller. In the case of  $\varphi \approx 1.93$ , the coding performance approximates the coding performance of the basic subband scheme. This is so because the  $\varphi$ -line is sufficiently close to the 'equality line' in the bit rate graph. This experiment clearly shows that the key for successive refinement for spatial scalability is in the encoding of the error made in the encoding of the previous reproduction.

## 3.6 Discussion

In this chapter the process of hierarchical coding for SNR scalability and spatial scalability was studied. By assuming a stationary Gaussian distributed input signal, it was theoretically shown that i) for SNR scalability successive refinement is always achievable and ii) for spatial scalability successive refinement is only achievable under particular conditions. These conditions were verified for both types of scalability using a practical hierarchical coder, i.e. the Refinement System, and a reference coder, i.e. the basic subband scheme of Section 2.2.

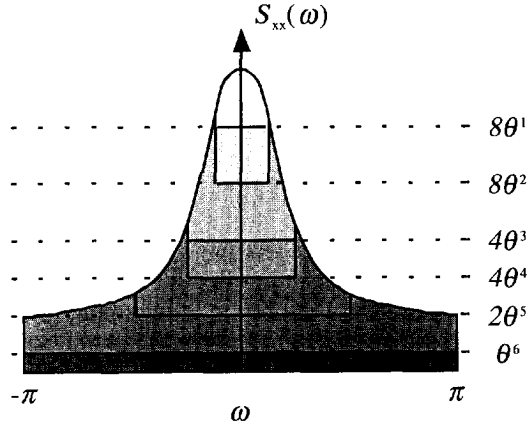


Figure 3.16:  $N$  embedded optimal mappings for combined SNR and spatial scalability.

For SNR scalability, the experimental verification showed that in practice successive refinement is not always achieved by the Refinement System. Specifically, if more than three reproductions are supported, the performance loss compared to the basic subband scheme is significant. The cause for this is that the performance of multiple concatenated [D]PCM coders is less than the performance of a single [D]PCM encoder used at the equivalent bit rate. If the Refinement System supports only two reproductions, successive refinement is approximately achieved.

For spatial scalability, the experimental verification showed that in practice successive refinement is achieved by the Refinement System when two reproductions are supported and when the proper bit rates  $R^1$  and  $R^2$  are used. In addition, the verification showed that the encoding of the quantization errors of  $y^1$  is essential to obtain successive refinement.

Given the insights and results obtained, it is now tempting to address the structure and problems of hierarchical coding schemes that support multiple reproductions with various spatial dimensions and SNR values. The theoretical condition for successive refinement of such a scheme is a special case of Eq. 3.25, where for some reproductions  $2 \leq n \leq N$  the decimation factors are related by  $\beta^n = \beta^{n-1}$ . Figure 3.16 illustrates this condition by showing the embedded mappings  $x \rightarrow y^n$  for the case of  $N = 6$  and  $\beta^1 = \beta^2 = 8$ ,  $\beta^3 = \beta^4 = 4$ ,  $\beta^5 = 2$ , and  $\beta^6 = 1$ . To obtain successive refinement, the hierarchical coding scheme should convey one or two types of information in the update for  $y^n$ ; namely, if the  $y^{n-1}$  and  $y^n$  have the same spatial dimensions, the update should contain information that converts the optimal error spectrum of mapping  $x^n \rightarrow y^{n-1}$  into the optimal error spectrum of  $x^n \rightarrow y^n$ . In Fig. 3.16, this is for instance the case with the updates for  $y^2$  and  $y^4$ . If the  $y^{n-1}$  and  $y^n$  have different spatial dimensions, the update should also contain information about the high frequency components that are part of  $y^n$  but not of  $y^{n-1}$ . In Fig. 3.16, this is for instance the case for the updates for  $y^3$  and  $y^5$ .

Fig. 3.16 prescribes the structure of hierarchical coding schemes to a large extent. Such

schemes should decompose the input spectrum into appropriate frequency bands such that the reproductions with smaller dimensions can be reconstructed. Further, the frequency bands that are part of more than one reproduction have to be encoded in an SNR-scalable way. I.e. for reproductions with higher values of  $n$ , the frequency components have to be reproduced at higher SNR values. Evidently, one crucial component of hierarchical coding schemes is the multirate encoding of the frequency components. To achieve successive refinement, the coding performance of the multirate coders should be equivalent with a non-hierarchical coder operating at the same bit rate. The next chapter examines therefore the structure and performance of multirate coders based on DPCM and PCM techniques.





# Chapter 4

## Multirate coding

### 4.1 Introduction

It was illustrated in Chapter 3 that the performance of a hierarchical coding scheme is directly influenced by the coding techniques used to encode the subbands. The subbands that are part of multiple reproductions have to be encoded using amplitude scalability so that they can be reproduced at different SNR values for different reproductions. Specifically, to obtain a coding performance similar to that of the basic subband scheme for each reproduction, it is necessary that i) the amplitude-scalable subband encoders have coding performances that are competitive with the coding performances of the PCM and DPCM coders used by the basic subband scheme and that ii) the subbands are encoded at the same bit rates as with the basic subband coding scheme.

This chapter explores how well these requirements can be fulfilled when multirate PCM (MR-PCM) and multirate DPCM (MR-DPCM) coders are used to encode the subbands<sup>1</sup>. The performances of the MR-PCM and MR-DPCM coders are evaluated with respect to the performances of regular PCM and DPCM coders based on Uniform Threshold Quantization (UTQ) and Arithmetic Coding (AC). These coders are known to be among the best scalar entropy-constrained quantizers in the MSE sense [Wood69, Farv84]. Based on this exploration, two types of practical MR-PCM and MR-DPCM coders are designed and evaluated. These multirate coders are optimized for use in competitive hierarchical coding schemes.

The content of this chapter is structured as follows. First, the structure and performance of optimal dual-rate PCM and DPCM coders are investigated in Sections 4.2 and 4.3. Optimal dual-rate coders are coders that support two reproductions and use quantizers optimized for particular bit rates. Although these coders are not directly applicable for practical schemes, they are very suitable to investigate the structure and capabilities of different types of coders. In particular, in Section 4.2 a description is given of the two possible structures of a dual-rate PCM coder; namely, MR-PCM based on multistage quantization and MR-PCM based on embedded quantization. The design methodologies of the quantizers are explained for each structure and the obtained performances are discussed. In Section 4.3 the structures of the dual-rate DPCM coder for both multistage quantization and embedded quantization are described. Here, the effectiveness of the different prediction loops and the coding

---

<sup>1</sup>MR-PCM and MR-DPCM were introduced in Chapter 2.

performances of the coders are evaluated for both structures.

Subsequently, in Section 4.4 two types of practical MR-PCM and MR-DPCM coders are designed that can be used in competitive hierarchical coding schemes supporting multiple reproductions (i.e.  $N > 2$ ). One type of coder is based on multistage quantization, while the other type is based on embedded quantization. For each type, a set of quantizers (and associated arithmetic coders) is designed that minimizes the performance loss of hierarchical coding schemes with respect to the basic subband coding scheme. These quantizers are designed after studying the bit allocation requirements of hierarchical coding schemes. In Section 4.5 the performances of the coders designed are evaluated for a hierarchical coding scheme that implements SNR scalability. The performance of the scheme is compared with the performance of the basic subband scheme when it uses UTQs and AC. Finally, Section 4.6 concludes this chapter with a short discussion of the results obtained.

## 4.2 Multirate PCM coding

A MR-PCM coder consists of a multirate quantizer followed by a multirate entropy coding section. The multirate quantizer maps the input amplitude  $x$  into a set of indices  $j^n$  with  $1 \leq n \leq N$  to facilitate  $N$  reproductions of the input amplitude. To facilitate the digital transmission of the indices, the multirate entropy coding section losslessly encodes  $j^n$  into codewords  $c^n$ . The average length of the codewords  $c^n$  is  $\tilde{R}^n$ .

Multirate PCM coding was first discussed in literature in [Elna86, Tzou86], using terms as embedded coding and bit-plane coding. These multirate coders were based on the fact that uniform quantizers with fixed length codewords are embedded quantizers by definition. For instance, a uniform quantizer with a bit rate  $R^{n-1} = n - 1$  bits can be refined into a uniform quantizer with bit rate  $R^n = n$  with one additional bit, i.e.  $\tilde{R}^n = 1$ . The use of embedded Lloyd-Max quantizers with fixed length codewords was proposed in [Tzou86, Elna86]. These quantizers have the property that one of them is a Lloyd-Max quantizer while the other quantizers have their coding thresholds aligned with some coding thresholds of this quantizer. Similar to the uniform quantizers, these quantizers have integer bit rates in the range from 1 to 8 bits and  $\tilde{R}^n = 1$ . More recently, enhanced versions of embedded PCM coders have been proposed that use uniform quantization and AC [Bosv93b, Shap93, Taub94]. Although the MR-PCM coder proposed in [Taub94] is still characterized by  $\tilde{R}^n = 1$ , the so-called pruned Uniform Threshold Quantizers of [Bosv93b] and the successive approximation quantization of [Shap93] are capable of reaching  $\tilde{R}^n < 1$ .

Multirate quantization based on multistage quantization was first proposed in [Wang88]. Here, DCT coefficients were multistage quantized to provide progressive transmission with a high coding performance. In [Biem90, Bosv90], multirate quantizers are proposed that are based on multistage Lloyd-Max quantizers with Huffman coding. In [Barn92], it is shown that appropriately designed multistage quantizers can perform equally well as embedded quantizers, albeit at a reduced complexity. In addition, it is shown that the stage-wise design of multistage quantizers may not always lead to the best approximation of the input amplitude at stage  $N$ . In [Bosv93a], multirate quantizers are proposed that are based on multistage Uniform Threshold Quantizers with Arithmetic Coding.

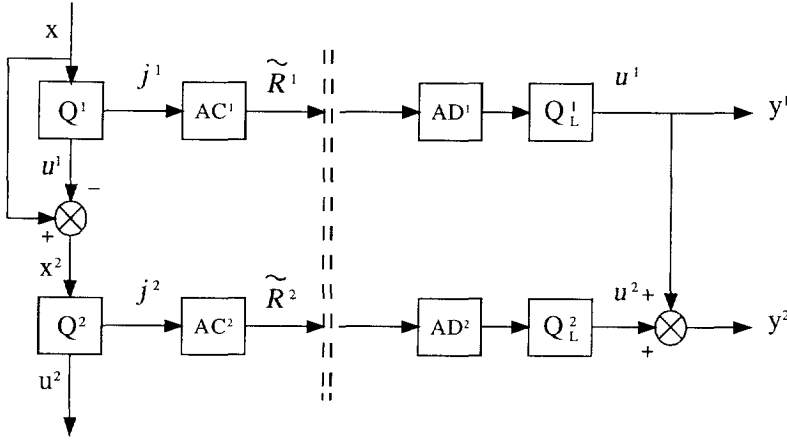


Figure 4.1: Structure of dual-rate PCM coder using multistage quantization.

#### 4.2.1 MR-PCM using multistage quantization

Multistage quantization (or coding) is conceptually the simplest way to implement MR-PCM<sup>2</sup>. The multirate coder consists of a multistage quantizer followed by  $N$  separate entropy coders. Alternatively, one could say that the multirate coder consists of  $N$  concatenated PCM coders, where each PCM coder encodes the quantization errors of the previous coder. Figure 4.1 illustrates the structure for a dual-rate coder. The first quantization stage quantizes the input amplitude  $x$  with quantizer  $Q^1$ , while the second stage quantizes the quantization error of the first stage, i.e.  $x^2$ , with quantizer  $Q^2$ , i.e.

$$\begin{aligned} u^1 &= Q^1(x), \\ u^2 &= Q^2(x^2) = Q^2(x - Q^1(x)). \end{aligned} \quad (4.1)$$

The quantizers put out indices  $j^1$  and  $j^2$ , which are losslessly encoded by the Arithmetic Coders  $AC^1$  and  $AC^2$  before transmission. At the receiving side, the reproductions  $y^1$  and  $y^2$  are obtained by first arithmetically decoding ( $AD^n$ ) the indices  $j^n$  and then combining the reconstructed values  $u^n = Q_L^n(j^n)$ . In particular, reproductions  $y^n$  are defined by

$$\begin{aligned} y^1 &= u^1, \\ y^2 &= u^1 + u^2. \end{aligned} \quad (4.2)$$

The bit rates of  $y^n$  are given by

$$\begin{aligned} R^1 &= \tilde{R}^1 = H(J^1), \\ R^2 &= \tilde{R}^1 + \tilde{R}^2 = H(J^1) + H(J^2), \end{aligned} \quad (4.3)$$

where it is assumed that the output rates of the arithmetic coders are equal to the first-order entropies of their inputs. In Eq. 4.3  $J^1$  and  $J^2$  are the r.v.'s associated with  $j^1$  and  $j^2$ ,

<sup>2</sup>Multistage quantization was also employed by the Refinement System of Chapter 3.

respectively. The MSE distortions of  $y^n$  are given by

$$\begin{aligned} D^1 &= \bar{D}^1 \sigma_x^2, \\ D^2 &= \bar{D}^1 \bar{D}^2 \sigma_x^2, \end{aligned} \quad (4.4)$$

where  $\bar{D}^n$  denotes the distortion introduced by  $Q^n$  relative to the variance of the input signal of  $Q^n$  and  $\sigma_x^2$  the variance of  $x$ . Henceforth, it is assumed that  $\sigma_x^2=1$ .

It is clear from Eqs. 4.3 and 4.4 that the coding performance of the MR-PCM coder is fully defined by the quantizers  $Q^n$ . To make the performance competitive with the performance of regular PCM coders based on UTQs and AC, it is necessary that  $Q^1$  and  $Q^2$  are chosen such that

$$R^n \approx R_{UTQ}(D^n) \quad \text{for } n = 1, 2, \quad (4.5)$$

where  $R_{UTQ}(D)$  is the practical performance curve of PCM coding based on UTQs and AC. A reasonable choice for  $Q^1$  and  $Q^2$  is to make them UTQs [Bosv93a, Bosv95, Nave95]. This choice assures that  $R^1 = R_{UTQ}(D^1)$  (because  $Q^1$  is a UTQ), while the overall quantization characteristic for  $y^2$  will resemble a UTQ to some extent. In addition, it is possible that  $\tilde{R}^2 \leq 1$ , which is an important property if the coder is applied in hierarchical coding schemes.

To investigate the performance of the dual-rate coder, the coder is evaluated for the cases that  $R^2 = 5$  bpp and  $R^1 = 0, 0.25, \dots, 5.00$  bpp. Specifically,  $Q^1$  and  $Q^2$  are sequentially designed according to UTQ design rules for each combination of  $R^1$  and  $R^2$ . The UTQ design procedure requires as inputs the PDF of the input signal of the quantizer and the output entropy desired. For  $Q^1$  these parameters are equal to  $p_X(x)$  and the target bit rate  $\tilde{R}^1 = R^1$ . For  $Q^2$  the PDF of the quantization errors, i.e.  $p_{X^2}(x)$ , is given by [Jaya84]

$$p_{X^2}(x) = \sum_i p_{X|J^1=i}(x) P_i, \quad (4.6)$$

where  $P_i$  is the occurrence probability of index  $i$  of quantizer  $Q^1$ . The target bit rate for  $Q^2$  is given by  $\tilde{R}^2 = R^2 - R^1$ .

The terms  $p_{X|J^1}(x)$  in Eq. 4.6 are called the conditional quantization error PDFs. These conditional PDFs describe the distributions of the quantization errors that result from the various decision intervals of  $Q^1$ . Figure 4.2 shows these conditional PDFs when a GG-PDF is quantized by a UTQ with 5 representation levels (Fig. 4.2a). Figs. 4.2b and c illustrate respectively the 5 conditional PDFs, which are normalized at unit variance, and  $p_{X^2}(x)$ , which is the result of the weighted summation of Eq. 4.6. In general,  $p_{X^2}(x)$  does not resemble any regular PDF model such as the Gaussian and Laplacian distributions. It approaches however the uniform distribution when  $x$  is finely quantized, i.e. at bit rates  $R^1 > 3$  bpp. For quantization at bit rates  $R^1 < 2$  bpp,  $p_{X^2}(x)$  is peaked and has an irregular shape.

The performance of the dual-rate coder is illustrated in Figure 4.3 using the SNR measure. Because  $Q^1$  is a (optimal) UTQ which is not of interest here, only the performance for  $y^2$  is presented. Fig. 4.3 also shows the SNR of a single-rate PCM coder using a UTQ operating at  $R = 5$  bpp. It can be observed that the dual-rate coder performs less well than the single-rate coder. For  $R^1 < 3$  bpp, the performance degradation is approximately 1.0 dB, while for

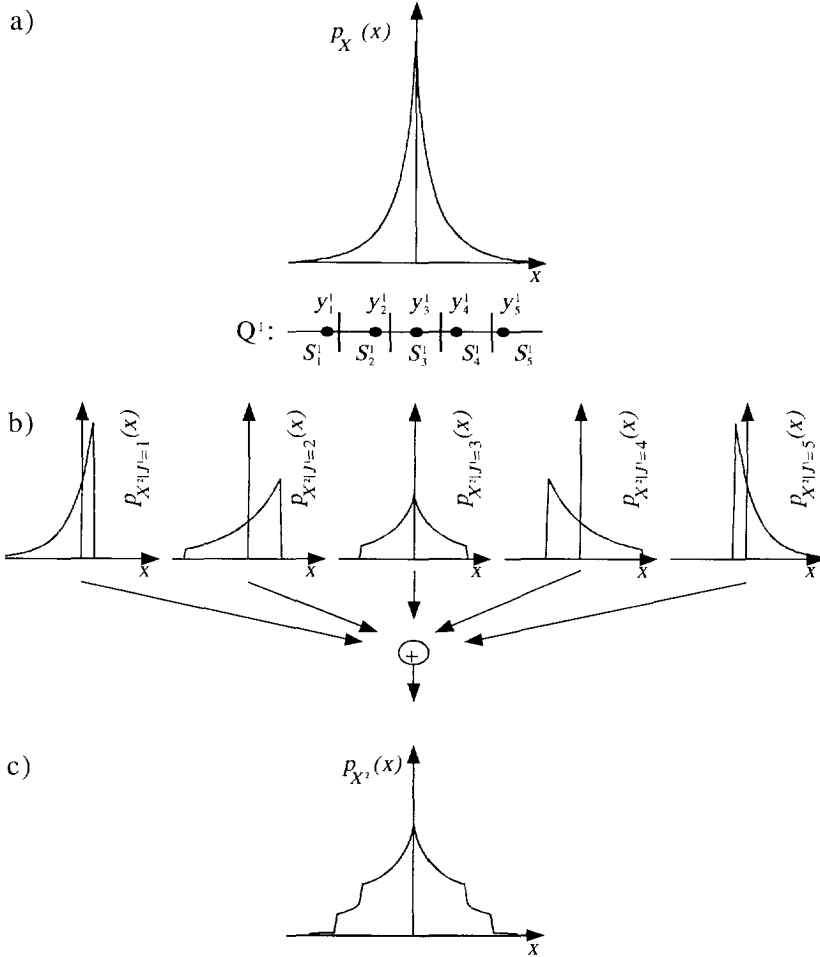


Figure 4.2: Quantization error PDFs: a)  $p_X(x)$ , b)  $p_{X^2|J^i=i}(x)$ , c)  $p_{X^2}(x)$ . Note that all PDFs shown have unit variance.

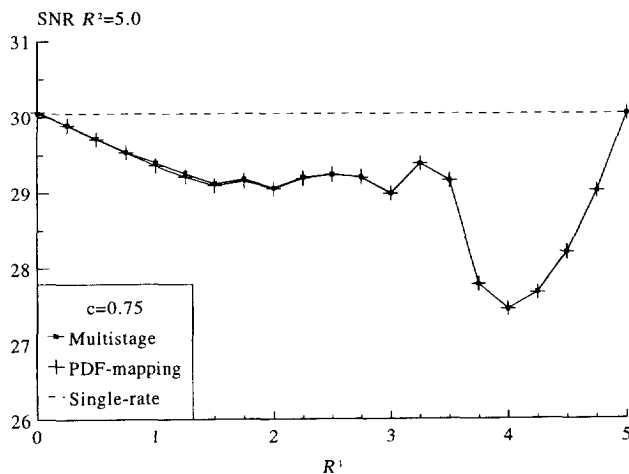


Figure 4.3: Performance of dual-rate PCM coder based on multistage quantization. The performance curve with the label 'PDF-mapping' is discussed in Section 4.4.2.

$R^1 > 3$  bpp, the degradation is maximally 2.5 dB. Other experiments show that the obtained results are typical for a wide range of bit rates and input PDFs.

The performance of the dual-rate coder for  $y^2$  is thoroughly investigated in [Bosv93a, Bosv95]. The performance loss for  $R^1 < 3$  bpp is mainly caused by the fact that the arithmetic coders encode the index pair  $\{j^1, j^2\}$  as two independent indices. By doing this, the total required bit rate is  $R^2 = H(J^1) + H(J^2)$ , which is always larger than or equal to the entropy of the index pair  $\{j^1, j^2\}$ , i.e.

$$R^2 \geq H(J^1, J^2) = H(J^1) + H(J^2|J^1). \quad (4.7)$$

Because  $J^1$  and  $J^2$  are generally dependent on each other, a conditional coding of the indices has a potentially higher coding performance than the current unconditional approach. The dependency between  $J^1$  and  $J^2$  is easily recognized by inspecting the conditional PDFs  $p_{X^2|J^1}(x)$  in Fig. 4.2;  $J^1$  and  $J^2$  are dependent of each other because  $p_{X^2|J^1=s}(x)$  and  $p_{X^2|J^1=t}(x)$  are different for  $s \neq t$ .

The performance loss for  $R^1 > 3.5$  bpp is mainly caused by the fact that the UTQ quantization of  $x^2$  is not very effective at these rates. The reason for this is that  $X^2$ , i.e. the r.v. of  $x^2$ , has a nearly uniform distribution, as do the conditional signals  $X^2|J^1$ . For these distributions and bit rates,  $Q^2$  has three representation levels, of which one is zero and two are situated near the boundaries of the uniform distribution. Such a quantizer yields a bit rate of approximately 1 bpp but also a distortion close to unity ( $\bar{D}^2 \approx 1.0$ ). Consequently, the scalar quantization is largely ineffective, leading to a performance loss compared to single-rate coding.

Both causes of performance loss are intrinsic to the multistage quantization structure employed and are related to the operational independency of the two quantization stages. In [Bosv93a, Bosv95] it is shown how the performance gap of the dual-rate coder can be

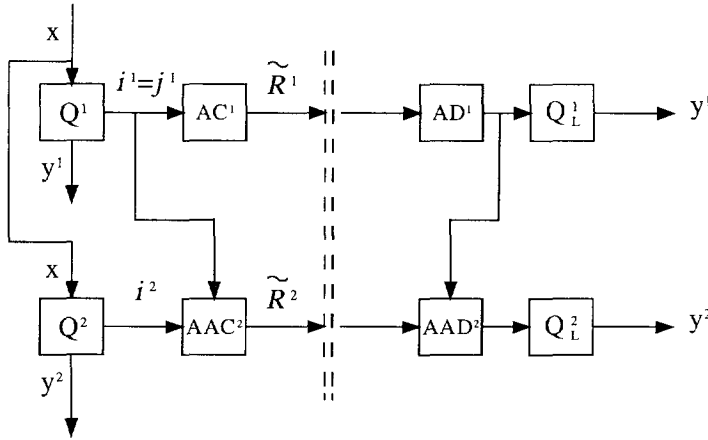


Figure 4.4: Structure of dual-rate PCM coder using embedded quantization.

reduced by making the operation of the second quantization stage dependent on the outcome of the first quantization stage. These modified coders (called conditional entropy coders and conditional quantization coders) are however not used in the rest of this thesis because their complexity and storage requirements are too large for practical use.

### 4.2.2 MR-PCM using embedded quantization

Figure 4.4 illustrates the structure of a dual-rate PCM coder that employs embedded quantization. Embedded quantization means that the input signal is concurrently quantized by multiple quantizers which have their coding thresholds aligned with the coding thresholds of other quantizers. The coder consists of two embedded quantizers  $Q^1$  and  $Q^2$  and two arithmetic coders  $AC^1$  and  $AAC^2$ , of which the latter is adaptive. Quantizers  $Q^1$  and  $Q^2$  map the input amplitude  $x$  into indices  $i^1$  and  $i^2$ , respectively. Since the decision levels of the embedded quantizers are aligned, the relationship between  $i^1$  and  $i^2$  is unambiguous. Consequently,  $AAC^2$  is able to transform the index  $i^2$  into the non-redundant conditional index  $j^2|j^1$  with  $j^1 = i^1$ . After this conversion,  $AAC^2$  encodes  $j^2|j^1$  at a bit rate of  $\tilde{R}^2 = H(J^2|J^1)$  so that

$$\begin{aligned}
 R^2 &= H(J^1) + H(J^2|J^1) \\
 &= H(J^1, J^2) \\
 &= H(I^2).
 \end{aligned} \tag{4.8}$$

Consequently, the bit rates and distortions of the dual-rate coder are given by the bit rates and distortions of  $Q^1$  and  $Q^2$ .

The embedded quantizers are designed using a bottom-up design algorithm. In order to make the overall quantizer characteristics competitive with a UTQ,  $Q^2$  is defined to be a UTQ with bit rate  $R^2$  for PDF  $p_X(x)$ . Subsequently, quantizer  $Q^1$  with  $R^1 < R^2$  is derived from  $Q^2$  by means of a pruning algorithm [deGa92]. The pruning algorithm recursively derives

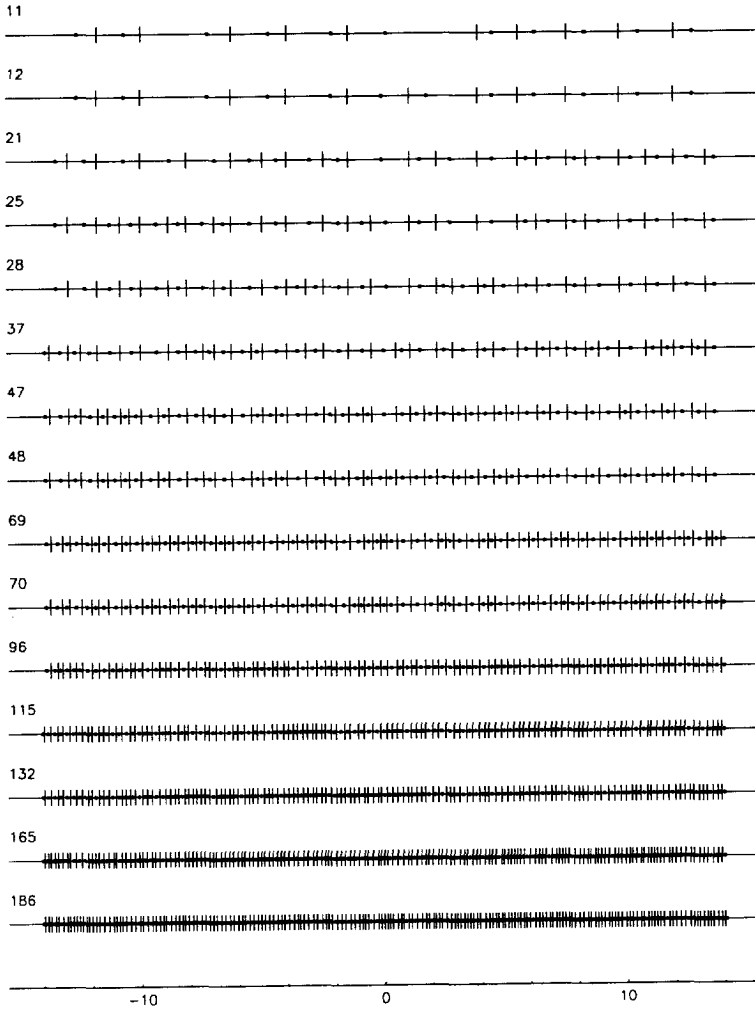


Figure 4.5: Embedded quantizers for GG-PDF with  $c=0.75$ . The number of representation levels is shown at the left-hand side.



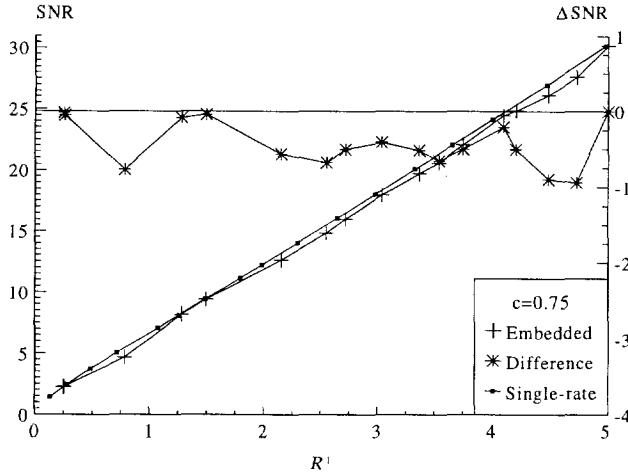


Figure 4.6: Performance for  $y^1$  of dual-rate PCM coder using embedded quantization. The right vertical axis shows the SNR difference values.

an embedded quantizer, let us say  $V$ , from another quantizer, let us say  $W$ , by combining two adjacent decision intervals of  $W$  into one decision interval. In particular, it merges those decision intervals that increase the distortion least for a given decrease in bit rate. The algorithm starts out by setting  $W$  equal to  $Q^2$  and then determines quantizer  $V$ . To minimize the distortion, the new representation level of the combined interval is defined as the centroid of the merged interval. When the bit rate of  $V$  is much larger than  $R^1$ ,  $W$  is set equal to  $V$  and the pruning procedure is repeated. When the bit rate of  $V$  is sufficiently close to  $R^1$ , the pruning algorithm stops and quantizer  $V$  is stored as  $Q^1$ .

The pruning algorithm largely determines the characteristics of  $Q^1$ . For instance, because the pruning algorithm is a convex hull tracing algorithm,  $Q^1$  approximates the UTQ characteristic closely but is generally not a UTQ. In addition,  $Q^1$  will have a bit rate that slightly deviates from the target bit rate  $R^1$ . This is caused by the fact that only a discrete number of bit rates are obtainable by the pruning algorithm. Sometimes the differences between the obtainable bit rates are very small (e.g. when two decision intervals were merged with low probability), but sometimes they are quite large (e.g. when two decision intervals were merged with high occurrence probabilities).

Figure 4.5 illustrates the pruning process by showing some quantizers that are derived from a UTQ with 2043 levels. As can be observed, all quantizers have their decision levels aligned so that the decision intervals of coarse quantizers consist of multiple decision intervals of finer quantizers. Consequently, most quantizers have decision intervals that are of varying sizes.

The performance of the dual-rate embedded coder is shown in Figures 4.6 and 4.7. Fig. 4.6 shows the performance for  $y^1$  as well as the performance of a single-rate PCM coder based on UTQ and AC using the SNR measure. In addition, the difference between these performances

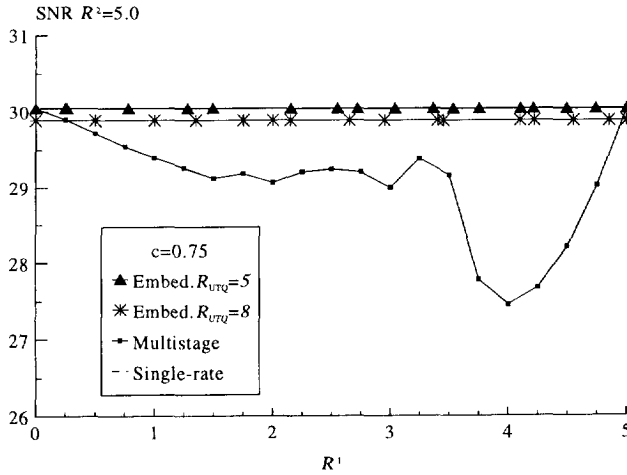


Figure 4.7: Performance for  $y^2$  of dual-rate PCM coder using embedded quantization.

is shown. This curve clearly shows that the performance of  $Q^1$  is slightly less than the performance of the PCM coder, i.e. approximately 1.0 dB at maximum. Fig. 4.7 shows the performance for  $y^2$  for the embedded coder ('Embed.  $R_{UTQ} = 5$ ') as well as the performance of the single-rate PCM coder operating at  $R = 5$  bpp. Because  $Q^2$  is a UTQ, the performance is constant and equal to the performance of the PCM coder. Fig. 4.7 illustrates the influence of the pruning algorithm on  $Q^1$ . From the irregular positioning of the measurement points it can be derived that i) the target bit rate of  $Q^1$ , i.e.  $R^1$ , is not always achieved, and ii) the difference in bit rate of embedded quantizers is generally smaller than 1 bpp.

In practice, a MR-PCM coder based on embedded quantization employs a set of pre-designed quantizers and entropy coding tables instead of designing the quantizers at run-time. To make the coder suitable for a wide range of bit rates, various quantizers are stored that are derived from a UTQ with a high bit rate. Consequently, the quantizer selected for  $Q^2$  will also be a pruned quantizer, which implies that the performance for  $y^2$  will be less than that of the PCM coder. To illustrate this, Fig. 4.7 also shows the performance of the dual-rate coder for  $y^2$  when  $Q^2$  is a pruned quantizer that is derived from a UTQ with  $R = 8$  bpp ('Embed.  $R_{UTQ} = 8$ '). The performance is still independent of  $R^1$  but is slightly less than the performance of the PCM coder. Specifically, the performance of the dual-rate is approximately 0.20 dB less than the UTQ performance. Compared to MR-PCM based in multistage quantization, this is a large performance improvement, especially for the region  $3 \leq R^1 \leq 5$ .

### 4.3 Multirate DPCM coding

The structure of MR-DPCM coders closely resembles the structure of MR-PCM coders. The main difference is of course the use of prediction in the multirate coder. In particular, special constraints exist concerning the prediction methods: to keep the prediction loops

of all decoders synchronized, the encoder has to produce a prediction value for a given reproduction based on the samples that are available to the decoder of that reproduction. If this is not the case, some decoders can not make the same prediction as the encoder, and will suffer from accumulated errors in the prediction loop. This phenomenon is commonly called drift.

Multirate DPCM coding has not received as much attention in literature as multirate PCM coding has. In [Good80, Jaya81], 'embedded' DPCM coders were proposed that had an increased robustness with respect to transmission errors. The simple expedient here was to drop  $R_d$  bits of the encoded input values of the feedback loop where  $R_d$  is equal to the maximum number of bits  $R_e$  that could possibly be corrupted by transmission errors. The use of such a coarse feedback makes the coder perform less well compared to a conventional  $R$ -bit system with an ideal channel  $R_e = 0$ . However, with a non-ideal channel ( $R_e > 0$ ) the performance is very close to that of a conventional  $(R - R_e)$ -bit coder. Later, [Jaya83] proposed an enhanced version of the above scheme. The proposed scheme explicitly encodes the DPCM reconstruction error at an average bit rate of  $R_n$  bits. By varying the bit rate allocated to the quantization error, the scheme achieved a higher quality than a conventional  $(R + R_n)$ -bit DPCM coder.

This section investigates MR-DPCM based on the predictive variants of multistage quantization and embedded quantization. It is analyzed for both variants how the predictability of the signal is employed by a dual-rate encoder implementing amplitude scalability. In addition, it is analyzed how competitive the performances are with respect to the performance of a DPCM coder based on UTQ quantization and a quarter-plane prediction method [West89].

### 4.3.1 MR-DPCM using multistage quantization

Figure 4.8 illustrates the structure of a dual-rate DPCM coder using multistage quantization. In each quantization stage a prediction of the input signal of that stage is subtracted before quantization and added after reconstruction. To keep the predictions at the encoder equal to those at the decoders, the predictions are based on the reconstructed signals of each quantization stage. Consequently, a MR-DPCM encoder can be thought of as multiple concatenated DPCM encoders, each encoding the quantization errors from the previous DPCM encoder. The prediction error signals  $d^n$  are defined as

$$\begin{aligned} d^1 &= x - P^1(y^1), \\ d^2 &= x^2 - P^2(y^2), \end{aligned} \quad (4.9)$$

where  $P^n(y^n)$  are predictions that are based solely on past pixels of  $y^n$ . At the decoders, reproductions  $y^n$  are recovered by

$$\begin{aligned} y^1 &= u^1 + P^1(y^1), \\ y^2 &= u^2 + P^2(y^2), \end{aligned} \quad (4.10)$$

with  $u^n = Q_L^n(j^n)$ . The effectiveness of  $P^1$  and  $P^2$  depends on the amount of correlation that is present in  $x$  and  $x^2$ , respectively, and on the bit rates with which the prediction error signals  $d^n$  are encoded. In practice,  $P^1$  is very effective for regular input signals as long as the bit rate used to encode  $d^1$  is high (i.e.  $R^1 > 3$  bpp).

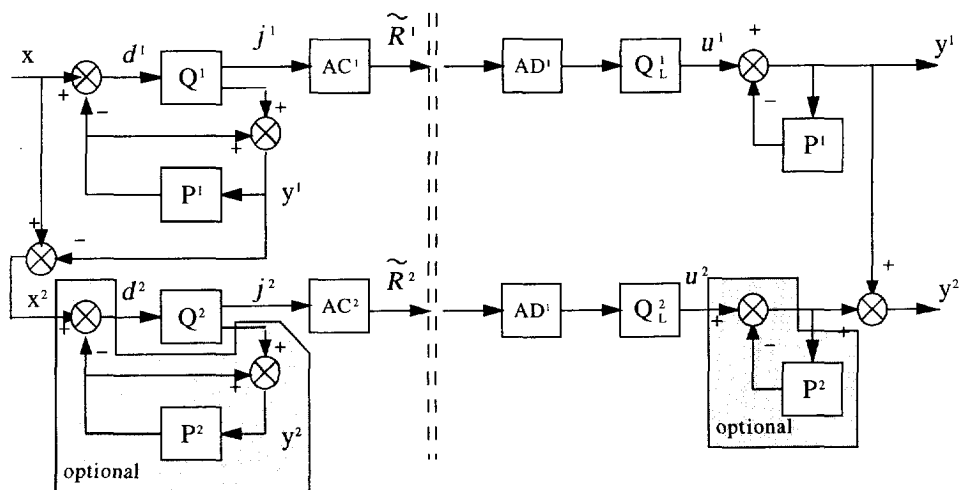


Figure 4.8: Structure of dual-rate DPCM coder using multistage quantization.

Figure 4.9 illustrates this by showing the realized prediction gain for  $x$ , i.e.  $G_{P,x}$ , as well as the maximum prediction gain for  $x$ , i.e.  $G_{P,x,max}$ , obtained by encoding the test image 'Clown' (128x128 pixels). The prediction gain is realized by using UTQs that are optimized for a GG-PDF with  $c=0.75$ . As illustrated, the prediction gain realized is low when coding at low bit rates but reaches the maximum prediction gain value for higher bit rates. In contrast to prediction  $P^1$ , prediction  $P^2$  is generally not very effective since signal  $x^2$  is fairly uncorrelated because of  $P^1$ . Fig. 4.9 also illustrates this by showing the maximum achievable prediction gain for  $x^2$ , i.e.  $G_{P,x^2,max}$ , when encoding the same test image at various values of  $R^1$ . The maximum prediction gain is based on the Yule-Walker equations for the quarter-plane prediction method [West89]. As can be observed, the maximum prediction gain decreases rapidly towards 1 (i.e. no gain) for higher values of  $R^1$ . Consequently,  $P^2$  is usually not implemented. Hence, the dual-rate coder consists of a DPCM coder that is concatenated to a PCM coder.

Figure 4.10 shows the performance of the dual-rate DPCM coder based on multistage quantization together with the performance of a single-rate DPCM encoder (The performance curves shown for the dual-rate DPCM coder based on embedded quantization are discussed in the next section). As can be seen, the shape of the performance curve resembles the performance curve shown in Fig. 4.3 for the dual-rate PCM encoder based on multistage quantization. When  $R^1 > 3$  bpp, the performance of the multistage quantizer is maximally 2.5 dB lower than the single-stage DPCM performance. In the area where  $R^1 < 3$  bpp, the dual-rate coder performs approximately 1.5 db to 3.5 dB less well compared to the single-rate coder. In a relative sense, the dual-rate DPCM coder performs thus less well than its dual-rate PCM counterpart. Inspection reveals that this is a direct consequence of the coarse quantization performed by  $Q^1$  (as illustrated in Fig. 4.9). In particular, the PDF of the prediction error diverges from the (assumed) GG-PDF with  $c = 0.75$ , which implies that applied quantization is less effective.

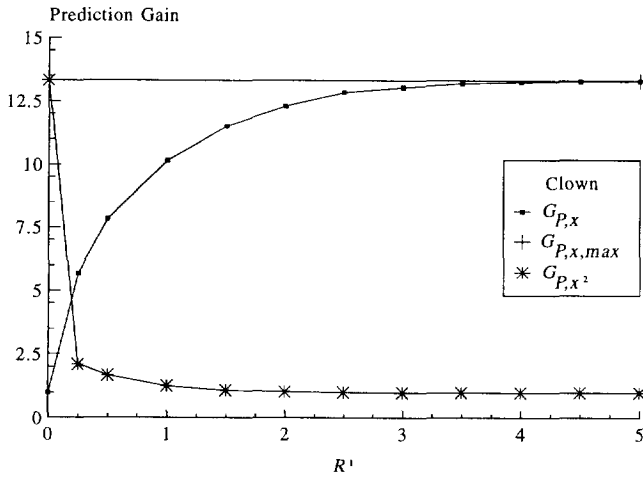


Figure 4.9: Obtained and maximum prediction gains for test image 'Clown' (128x128 pixels).

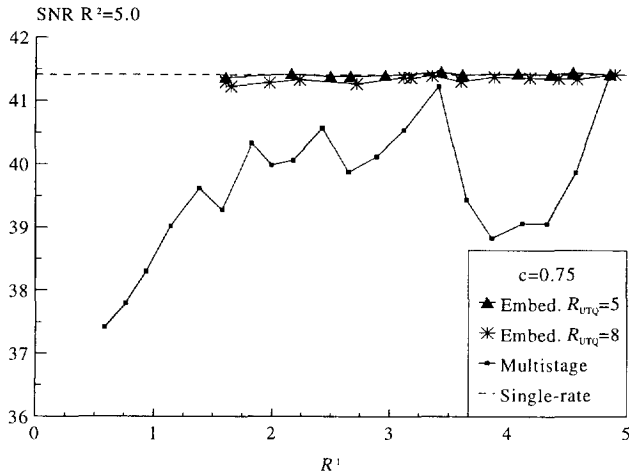


Figure 4.10: Performances of dual-rate DPCM coders for both multistage quantization and embedded quantization.

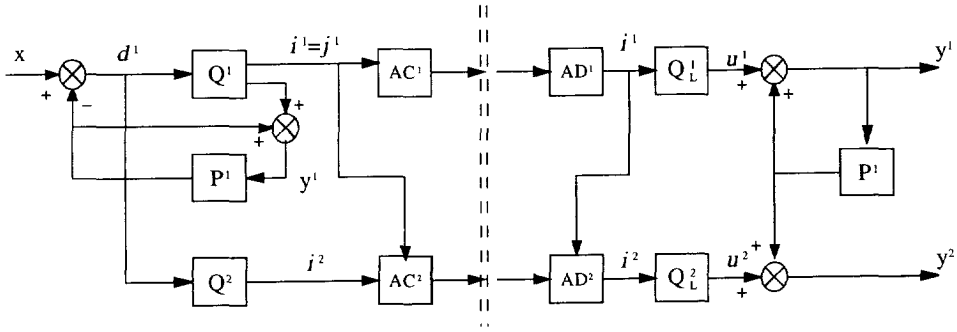


Figure 4.11: Structure of dual-rate DPCM coding using embedded quantization.

### 4.3.2 MR-DPCM using embedded quantization

Figure 4.11 illustrates the structure of the dual-rate DPCM encoder using embedded quantization. Here, a prediction of  $x$  is subtracted from  $x$  before quantization by the two embedded quantizers. More precisely, the prediction error signal  $d^1$  is defined as

$$d^1 = x - P^1(y^1), \quad (4.11)$$

where  $P^1(y^1)$  is a prediction that is based solely on past pixels of  $y^1$ . At the receiving sides, the prediction is added to the quantized versions of the prediction error to form the reconstructed signals  $y^n$ , i.e.

$$\begin{aligned} y^1 &= u^1 + P^1(y^1), \\ y^2 &= u^2 + P^1(y^1), \end{aligned} \quad (4.12)$$

with  $u^n = Q_L^n(i^n)$ . In this way the DPCM coding gain increases the coding performance for all reproductions.

Equation 4.11 implies that drift can not occur in the receivers because all receivers have access to the information carried in  $y^1$ . On the other hand, Eq. 4.11 may seriously affect the reconstruction quality of  $y^2$  because the prediction is based on coarsely quantized data (i.e.  $y^1$  instead of the finely quantized data of  $y^2$ ). Fig. 4.9 already demonstrated that this implies that the bit rate for  $y^1$  should not become too low.

The performances obtained for the dual-rate DPCM coder based on embedded quantization are also shown in Figure 4.10. As with the MR-PCM case, the MR-DPCM coder is evaluated for two cases. In the first case  $Q^2$  is a UTQ with bit rate  $R^2 = 5$  bpp, while in the second case  $Q^2$  is a pruned quantizer derived from a UTQ with bit rate  $R = 8$  bpp. As can be seen, the performance curves when using embedded quantization are situated as with the MR-PCM case for high values of  $R^1$ . If  $Q^2$  is a UTQ the performance loss is negligible. If  $Q^2$  is a pruned quantizer derived from a UTQ with  $R = 8$  bpp a small performance loss can be observed. Unfortunately, no measurement points are available for  $R^1 < 1.5$  bpp, because the PDF of the prediction error diverges too much from the (assumed) GG-PDF with  $c = 0.75$ . This implies that the bit rates obtained for  $y^1$  and  $y^2$  differ too much from the target values.

## 4.4 Practical multirate coders

In the previous two sections, the structures and performances of optimal dual-rate PCM and DPCM coders were investigated. Although these dual-rate coders illustrate the capabilities of the different types of coders well, they are not directly applicable in practical schemes because they design their quantizers at run-time. In this section two sets of quantizers are therefore designed that can be applied by practical MR-PCM and MR-DPCM coders in hierarchical coding schemes. One set of quantizers is for use in multirate coders based on multistage quantization while the other set is for use in coders based on embedded quantization. However before designing the quantizers first the requirements of hierarchical coding schemes regarding the multirate encoding of the subbands are analyzed.

### 4.4.1 Requirements

Hierarchical coding schemes employ MR-PCM and MR-DPCM coders in the same way as the basic subband scheme employs PCM and DPCM coders. The structures of the coders are fixed and at run-time a bit allocation algorithm determines the optimal bit rates for the different subbands for different reproductions. These bit allocation algorithms make these decisions on the basis of performance tables that indicate the expected rate and distortion of the subband when a particular quantizer (and associated arithmetic coding table) is used. In general, hierarchical coding schemes will require multirate coders that can support multiple reproductions, i.e.  $N \geq 2$ . This implies that the multirate coders contain multiple quantization stages for which tables should be stored. In fact, to provide the bit allocation algorithms with enough flexibility, a set of quantization tables has to be stored for each quantization stage.

To get insight in the requirements of the bit allocation algorithms, Figure 4.12 shows various bit allocations that are obtained with the basic subband scheme. The bit allocations show the bit rates allocated to the 28 subbands<sup>3</sup>, denoted by  $R_m$ , when the test image 'Clown' is encoded at the rates  $R = 0.25, 0.50, \dots, 2.50$ . The bit rates obtained would be the target bit rates for a hierarchical coding scheme that implements SNR scalability and supports 10 reproductions (i.e.  $N = 10$ ). The basic subband scheme employs PCM and DPCM coders based on UTQs and AC. For both types of coders two sets of predefined quantization tables exist, namely, one set for the high-frequency subbands (assumed to be distributed as a GG-PDF with  $c = 0.50$ ) and one set for the prediction error of the low-frequency subband (assumed to be distributed as a GG-PDF with  $c = 0.75$ ). Each set has 32 quantizers, which have bit rates from 0 bpp to 8 bpp with a resolution of 0.25 bpp.

From Fig. 4.12 it can be observed that the allocated rates  $R_m$  depend on the subband number and bit rate  $R$ . It shows that the subbands with the lower frequency components are encoded at higher rates than the subbands with the higher frequencies. The maximum bit rate allocated to a subband is approximately 7 bpp. For subsequent values of  $R$ , the bit rates  $R_m$  are incremented with values of less than 1 bpp. For many subbands the smallest increase in bit rate is used, i.e. 0.25 bpp. Roughly speaking, it can therefore be said that the bit allocation algorithm fully employs the range and resolution of the available quantization tables. Hence, the requirements for the quantization tables of the multirate coders can be formulated as follows:

<sup>3</sup>See Chapter 2 for a diagram of the subband numbering.

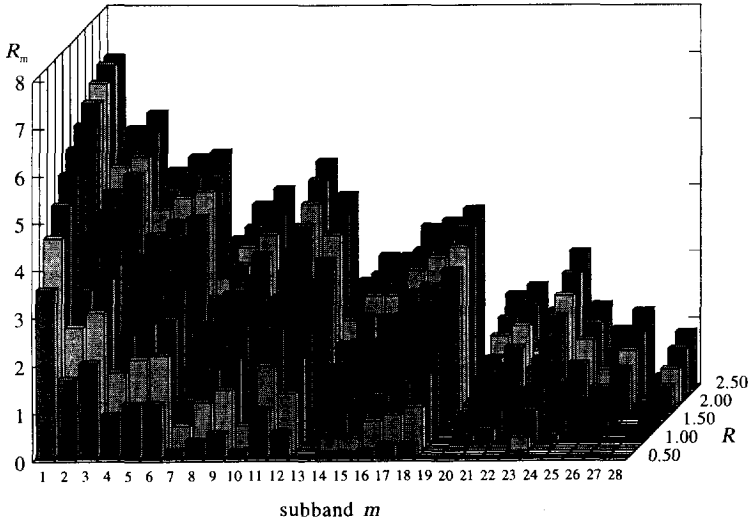


Figure 4.12: Bit rates  $R_m$  of the basic subband coding schemes for bit rates  $R = 0.25, 0.50, \dots, 2.50$ .

- the quantization tables should be applicable for multirate coders that support multiple reproductions,
- quantization tables are necessary to encode subbands that are assumed to be distributed according to GG-PDFs with either  $c = 0.50$  or  $c = 0.75$ , and
- the tables should allow the bit allocation algorithms to encode the subbands for a particular reproduction with (cumulative) bit rates  $0 \leq R^n \leq 8$  bpp with increments of 0.25 bpp.

#### 4.4.2 Multistage quantization

The structures of dual-rate PCM and DPCM coders were discussed in Sections 4.2.1 and 4.3.1. Unfortunately, these structures are not straightforwardly extensible to multiple stages without making the storage requirements excessive. The reason for this is that the design of  $Q^n$  with  $n \geq 2$  depends on the PDF of the input signal of stage  $n - 1$  and the applied quantizer  $Q^{n-1}$ . Because for all  $Q^n$  with  $1 \leq n \leq N$  multiple quantization tables have to be stored, a combinatorial complexity arises with similar storage requirements. For instance, if the multirate coder has 32 quantizers available for  $Q^1$ , it must also have 32 different sets of quantizers for  $Q^2$  available. If each of these sets contains 8 quantizers, then the multirate coder needs  $32 \times 8 = 256$  different sets of quantizers for  $Q^3$ . If these sets also contain 8 quantizers each, a total of 2048 quantizers (plus associated arithmetic coding tables) have to be stored for  $N = 3$ .

Fortunately, the storage requirements of these multirate coders can be significantly reduced by slightly modifying the structure of the multirate coders so that the coders only



support a limited number of PDFs for  $x^n$ . This can be achieved by mapping the (assumed) PDF of  $x^n$  to one of the PDFs supported. The performance tables of this PDF are then used by the bit allocation algorithm to select the appropriate quantizer. In fact, this modification degenerates the multistage quantizer to  $N$  concatenated single-rate coders that all share the same quantization tables. Hence, instead of combinatorial storage requirements, it can be said that this modified multirate coder has linear storage requirements.

Obviously the performance of the multirate coder is dependent on the accuracy of the PDF mapping procedure. In fact, the set of PDFs supported should be chosen carefully to minimize the performance loss. In [Bosv95] it is shown that a set of 12 PDFs is sufficient to represent the full range of possible PDFs. The set consists of the uniform PDF and 11 GG-PDFs with  $c$ -values in the range of  $c = 0.50$  to  $c = 10.00$ . The choice for these PDFs is based on the observations that i) the PDFs of quantization errors resemble the GG-PDFs to a large extent, and ii) the quantization errors are approximately uniformly distributed for quantization at  $R \geq 3.0$  bpp.

Figure 4.13 illustrates these observations by showing the PDFs of 6 quantization error signals (solid lines) as well as the best look-alike GG-PDFs (dashed lines). The quantization errors are obtained by evaluating Eq. 4.6 for a GG-PDF with  $c = 0.75$  and various UTQs with entropy rates  $R = 0, 0.5, 1, 5, 2.5, 3.5,$  and  $4.5$  bpp. The best look-alike PDF is found by comparing the PDF of the quantization error with the set of PDFs supported using the Kolomov-Smirnov fitting test [Cono71, West89]. It can be observed from Fig. 4.13 that the difference between the quantization error PDFs and the look-alike PDFs is generally very small. Hence, the performance loss caused by the PDF mapping will be very small, if not negligible. Figure 4.3 shows the performance obtained for the optimal dual-rate PCM coder of Section 4.2.1 using such a PDF-mapping. As can be seen, the performance loss is almost negligible and occurs only for a very limited range of bit rates.

Since the modified multirate coder can support multiple reproductions with manageable complexity and storage requirements, it is now time to define the characteristics of the quantization tables. With multistage coding, the cumulative bit rate of a particular subband for a particular reproduction is the sum of the bit rates of the quantizers that are used in the various coding stages. Hence, to fulfil the previously defined requirements for the quantization tables, it is necessary that these tables have a bit rate resolution of at least 0.25 bpp where possible. In addition, the bit rate range of the different tables should allow for cumulative rates up to 8 bpp. Consequently, the following characteristics are defined:

- The quantizers are optimized to encode signals distributed according to the set of 12 supported PDFs. These PDFs are 11 GG-PDFs with  $c = 0.5, 0.75, 1.0, 2.0, 2.75, 3.5, 4.50, 6.0, 8.0, 10.0,$  and the uniform PDF.
- The tables for the GG-PDFs each consists of 32 UTQs. These quantizers have entropy rates in the range from 0.0 to 8 bpp with a resolution of 0.25 bpp.
- For the uniform PDF 19 uniform quantizers are designed with an identical bit rate range but different resolution.
- One table contains the PDF mappings.

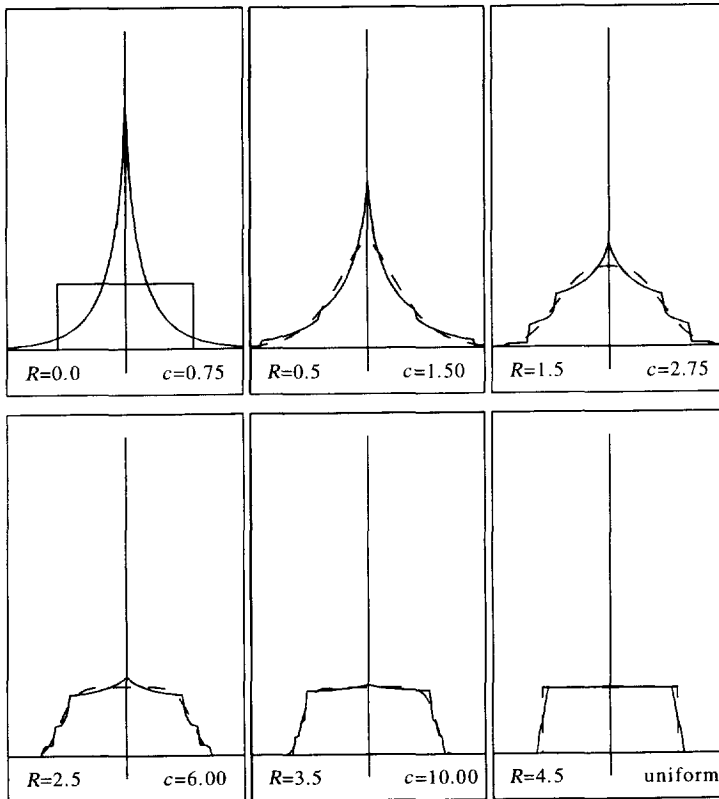


Figure 4.13: Mapping of quantization error PDFs: solid lines depict  $p_{X^2}(x)$  after encoding a signal with GG-PDF with  $c = 0.75$  at the indicated rate  $R$ . The dashed lines depict the best look-alike GG-PDF whose  $c$ -value is also indicated. For the purpose of comparison, the uniform PDF is plotted in the box with  $R = 0.0$ .

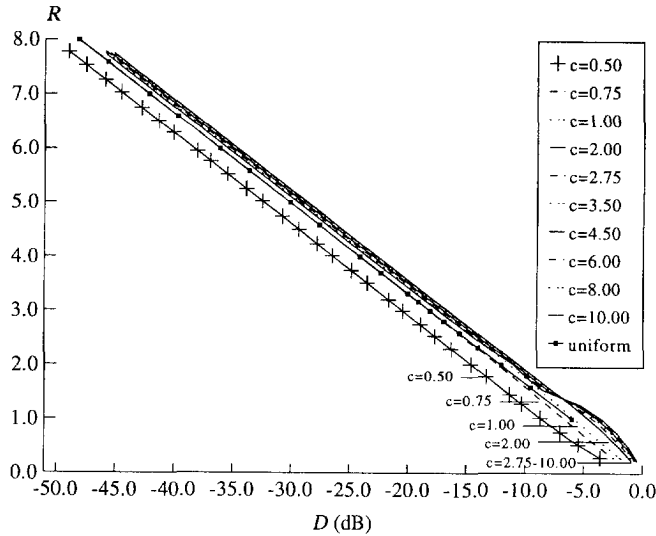


Figure 4.14: Rate-distortion curves of quantizers designed for multirate coders using multi-stage coding.

Figure 4.14 shows the rate-distortion curves of the stored quantization tables. Note that only the quantizers for the GG-PDF with  $c = 0.5$  and the uniform PDF are explicitly marked. As illustrated, most curves are approximately linear and increase by approximately 6 dB per additional bit, except for the quantizers optimized for GG-PDFs with  $c > 6.0$  and  $R < 1.5$  bpp. As outlined before, this is caused by the fact that these PDFs are nearly uniformly distributed. Hence, quantization at  $R < 1.5$  bpp has little effect because one representation level is situated at zero and the other two are situated close to the boundaries of the PDF range. Note that the illustrated curves do not show the performance of the multirate coders. These performances can be derived from the figure, but only when the bit allocations to the different stages are known. In that case, the performance curves can be obtained by stacking the relevant pieces of the performance curves on top of each other.

Table 4.1 shows the mapping table obtained for the 11 supported GG-PDFs. The mapping table indicates to which supported GG-PDF the PDF of the quantization error is mapped after quantization of a signal with a particular GG-PDF at a particular rate  $R$ . It shows that when a signal is encoded at rates higher than 4 bpp, the quantization error is mapped to the uniform distribution. The quantization errors of the uniform distribution are all mapped to the uniform distribution (not shown).

### 4.4.3 Embedded quantization

In Sections 4.2.2 and 4.3.2 the structures of dual-rate PCM and DPCM coders that use embedded quantization were discussed. In contrast to the structures for multistage quantization, these structures are straightforwardly extensible to multiple reproductions because

Table 4.1: Mapping table of the PDFs of the quantization error signals. The entries indicate the  $c$ -value of the best look-alike supported GG-PDF. The label 'uni' refers to the uniform PDF.

$R$	$c$ -value of input GG-PDF										
	0.50	0.75	1.00	1.50	2.00	2.75	3.50	4.50	6.00	8.00	10.00
0.25	0.75	1.00	1.50	2.00	2.00	2.75	3.50	4.50	4.50	6.00	8.00
0.50	1.00	1.50	1.50	2.00	2.75	2.75	3.50	3.50	3.50	3.50	3.50
0.75	1.00	1.50	2.00	2.75	2.75	2.75	2.75	2.75	2.75	2.75	2.75
1.00	1.50	2.00	2.00	2.75	2.75	2.75	2.75	2.75	2.00	2.00	2.00
1.25	1.50	2.00	2.75	3.50	3.50	3.50	2.75	2.75	2.00	2.00	2.00
1.50	2.00	2.75	3.50	3.50	3.50	4.50	4.50	4.50	3.50	3.50	3.50
1.75	2.00	3.50	4.50	4.50	4.50	6.00	6.00	6.00	6.00	8.00	8.00
2.00	2.75	3.50	4.50	6.00	6.00	6.00	6.00	6.00	6.00	6.00	6.00
2.25	2.75	4.50	6.00	6.00	8.00	8.00	8.00	8.00	8.00	6.00	6.00
2.50	3.50	6.00	8.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00
2.75	4.50	6.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00
3.00	4.50	8.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00
3.25	6.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00
3.50	8.00	10.00	10.00	uni	uni	uni	uni	uni	uni	uni	uni
3.75	8.00	10.00	uni	uni	uni	uni	uni	uni	uni	uni	uni
4.00	uni	10.00	uni	uni	uni	uni	uni	uni	uni	uni	uni
>	uni	uni	uni	uni	uni	uni	uni	uni	uni	uni	uni

quantizers  $Q^n$  can share the same quantization tables. Consequently, the requirements for the quantization tables can be directly derived from the previously defined requirements for the multirate coders, i.e.

- The quantizers are optimized to encode signals distributed according to a GG-PDF with either  $c=0.50$  or  $c=0.75$ .
- The tables for the GG-PDFs consist of embedded quantizers that are pruned from a UTQ with  $R = 8$  bpp. The embedded quantizers have entropy rates ranging from 0.0 to 8 bpp with a resolution of 0.25 bpp.

Figure 4.15 shows the rate-distortion characteristics of the quantizers stored for the embedded coding method. Evidently, the number of quantizers is much smaller than the number required for multistage quantization. Both curves have an approximately linear performance that increases by 6 dB per additional bit. Note that the bit rate difference between two embedded quantizers is not always approximately 0.25. Specifically, in the area  $R < 3$  bpp the difference varies between 0 bpp and 0.5 bpp. This characteristic is a direct consequence of the pruning algorithm that in the end has to merge decision intervals with relatively large occurrence probabilities.

It should be noted that the performance curves of the stored quantizers also represent the performance curves of the multirate coders. Therefore, as a reference, Fig. 4.15 also shows the performance curve of a single rate PCM coder based on UTQ and AC for each GG-PDF supported. As illustrated, the performance loss varies but is never larger than approximately 0.75 dB.

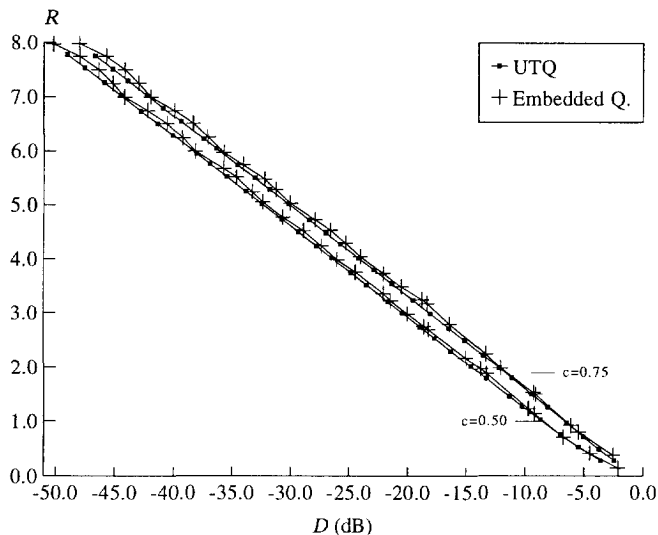


Figure 4.15: Rate-distortion curves of designed quantizers for multirate coders using embedded quantization.

## 4.5 Experimental results

In this section the multirate coders and quantization tables designed in the previous sections are evaluated when used in a hierarchical coding scheme that implements SNR scalability. In particular, it is investigated if these multirate coders and quantization tables can make the hierarchical coding scheme competitive with the basic subband scheme using UTQs and AC. The employed hierarchical coding scheme supports  $N$  reproductions with  $N = 2, 3, 4, 5, 6$ .

### 4.5.1 Coding scheme

Figure 4.16 illustrates the structure of the hierarchical coding scheme. The coding scheme supports  $N$  reproductions by creating  $N$  layers,  $l^n$ , with  $1 \leq n \leq N$ . The encoder consists of a subband decomposition followed by several multirate subband encoders, a multiplexer, and  $N$  bit allocations. The decoder contains the inverse stages of demultiplexing, multirate decoding, and subband reconstruction. The subband decomposition consists of a tree-structured subband decomposition that splits the input image  $x$  into  $M = 28$  subbands (c.f. Fig. 2.7) using separable 16-tap FIR filters (QMF16b) [John80]. The subbands  $x_m$  are separately encoded with multirate encoders that describe the subband samples with  $N$  codewords  $c_m^n$ . Since the samples of subband 1 are highly correlated, this subband is encoded using MR-DPCM. The samples of the other subbands are fairly uncorrelated and are thus encoded using MR-PCM. The multirate encoders are controlled by  $N$  bit allocation procedures  $BAn$ . The bit allocation algorithms make sure that the reproductions  $y^n$  have the desired bit rates  $R^n$ . After subband encoding, the multiplexer multiplexes the codewords into

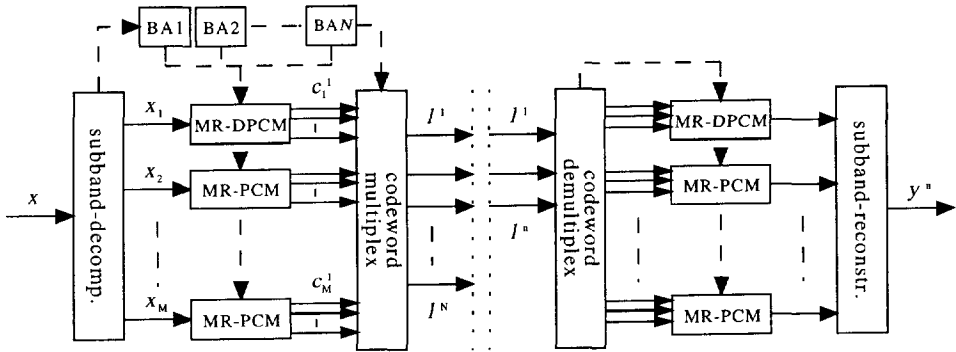


Figure 4.16: Structure of SNR-scalable coding scheme using MR-PCM and MR-DPCM coders.

the layers so that codewords  $c_m^n$  are conveyed in layer  $l^n$ . Note that the information about the bit allocations is also multiplexed in the bit streams to facilitate the proper decoding at the receiver.

### 4.5.2 Coding performances

To evaluate the coding performances of the proposed MR-PCM and MR-DPCM coders, the luminance component of the test image 'Clown' is encoded. The hierarchical coding scheme employs either multirate coders based on multistage quantization or multirate coders based on embedded quantization. The scheme makes the bit rate of each layer equal to  $\tilde{R}^n = \frac{2.5}{N}$  bpp so that the bit rate of reproduction  $y^n$  is approximately given by  $\frac{2.5n}{N}$  bpp.

Figure 4.17 shows 5 performance curves obtained with the hierarchical coding scheme when it employs the multirate coders based on multistage quantization and supports  $N = 2, 3, 4, 5, 6$  reproductions. The curves illustrate the performances obtained (i.e.  $R^n, D^n$ ) for each reproduction. For reference purposes, Fig. 4.17 also shows the obtained performance curve of the basic subband scheme based on UTQs and AC. The curve illustrates the obtained distortions  $D$  when encoding the test image at bit rates  $R$  with  $0.25 \leq R \leq 2.5$ . As can be observed, the SNR value of  $y^N$  clearly depends on the number of reproductions. If more reproductions are supported, the SNR value of  $y^N$  decreases. In fact, for the case of  $N = 6$ , the performance is approximately 1.8 dB less than the performance obtained by the basic subband scheme. An identical relation can be observed for  $y^n$  with  $1 < n < 6$ , although the performance losses are smaller. For the cases of  $N = 2$  and  $N = 3$ , the performances losses for  $y^2$  and  $y^3$  are much smaller. From Fig. 4.17 it can be concluded that the multirate coders based on multistage quantization do not always make the hierarchical coding scheme competitive with the basic subband scheme. For most applications the scheme is only competitive when a small number of reproductions is supported (i.e.  $N = 2, 3$ ).

Figure 4.18 shows the same performance curves as Figure 4.17, but now the hierarchical coding scheme employs the multirate coders based on embedded quantization. As illustrated,

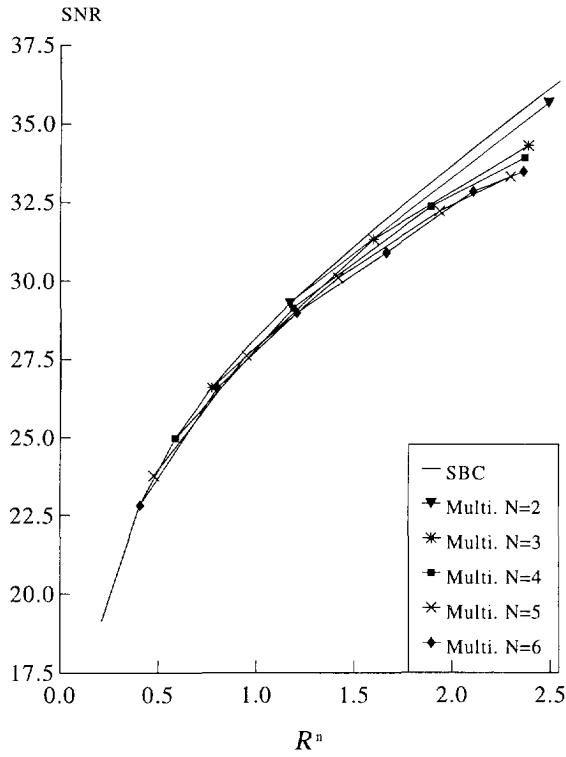


Figure 4.17: Performances of  $y^n$  with  $1 \leq n \leq N$  obtained with SNR-scalable coding scheme using multirate coders based on multistage quantization.

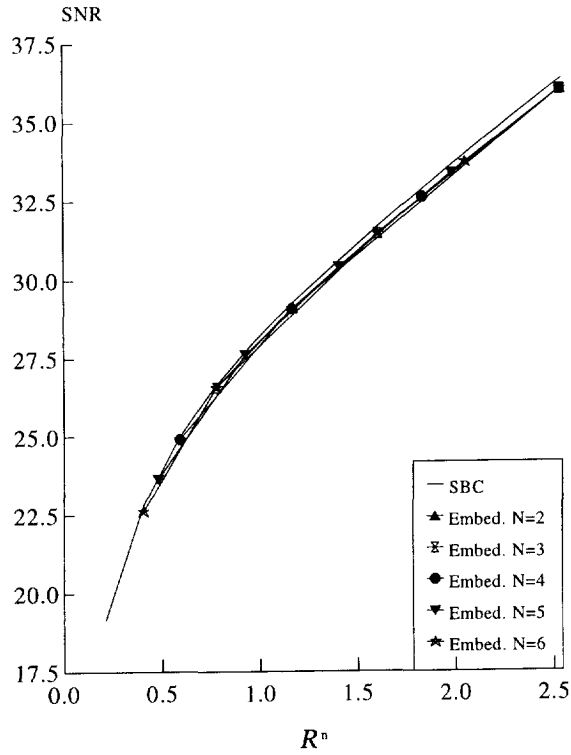


Figure 4.18: Performances of  $y^n$  with  $1 \leq n \leq N$  obtained with SNR-scalable coding scheme using multirate coders based on embedded quantization.



the SNR value of  $y^N$  is independent of the number of reproductions. In fact, for all cases of  $N$ , the performance is only approximately 0.35 dB less than the performance obtained by the basic subband scheme. For  $y^n$  with  $1 < n < N$ , the performance loss compared to the basic subbands scheme is also very small. From Fig. 4.18 it can be concluded that the multirate coders based on embedded quantization make the hierarchical coding scheme fairly competitive. For most applications the scheme is probably competitive irrespective of the number of supported reproductions.

## 4.6 Discussion

The structures and coding performances of multirate PCM (MR-PCM) and multirate DPCM (MR-DPCM) were explored in this chapter. These coders can be used by hierarchical coding schemes to encode the subbands in an amplitude-scalable way. Hierarchical coding schemes employ these coders to limit the performance losses caused by the introduction of scalability. In this chapter PCM and DPCM coders based on UTQ and AC were used as reference coders, as well as the basic subband scheme that employed these coders. Multirate coders can be based on either multistage quantization or embedded quantization. With multistage quantization, the quantizers employed are UTQs. The UTQs are designed to optimally encode the quantization errors of a previous coding stage. With embedded quantization, the quantizers employed are derived from a UTQ by means of a pruning algorithm. The quantizers are designed to optimally encode the input signal.

The analysis of optimal dual-rate PCM coding showed that multistage quantization performs less well for  $y^2$  than embedded quantization does. In fact, for embedded quantization the performance for  $y^2$  is constant and does not depend on the bit rate used to encode  $y^1$ . The performance of  $y^1$  is slightly less with embedded quantization because of the pruning algorithm. The analysis of optimal dual-rate DPCM coding confirmed the above conclusions. In addition, it showed that the prediction loops may have a low prediction gain when the loop quantizer is operating at a low bit rate. This always affects the performances of the coder irrespective of whether it is based on multistage quantization or embedded quantization.

Although the dual-rate coders are useful as study objects, they are not directly applicable in practical schemes. Therefore practical MR-PCM and MR-DPCM coders were designed that can be applied in hierarchical coding schemes. The coders based on multistage quantization had to be modified slightly to limit the storage requirements. The coders based on embedded quantization are direct extensions of the dual-rate coder.

The complexity of the multistage coders is slightly less than the complexity of the embedded coders. This is caused by the fact that the arithmetic coders of the embedded variant are adaptive and have to translate (and encode) the indices of  $Q^2$  into non-redundant indices. The storage space necessary for the tables of the multistage coders is larger than the space required for the embedded coders. In particular, approximately 12 tables have to be stored for the multistage variant, versus 2 tables for the embedded variant. In addition, the special PDF mapping table has to be stored.

The performance of the multistage coders when used in the hierarchical coding scheme is less than the performance of the embedded coders. In particular, if embedded coders are used the performance loss compared to the basic subband scheme is less than 0.35 dB for all

$y^n$  with  $1 \leq n \leq 6$ . With the multistage quantizers the performance loss may increase to 1.8 dB.

Based on the above observations, it can be concluded that multirate coders based on embedded quantization are superior to multirate coders based on multistage quantization, especially when a large number of reproductions has to be supported. For such applications, the designed coders based on embedded quantization can make the performance of hierarchical coding schemes competitive with the performance of non-scalable schemes based on similar coding techniques.

# Chapter 5

## Spatially scalable coding for ATM networks

### 5.1 Introduction

In Chapter 4 it was shown that multirate coders can make the performance of a simple SNR-scalable coding scheme competitive with the performance of a similar non-scalable coding scheme. In this chapter it is analyzed how the multirate coders of Chapter 4 influence the performance of a practical and more complex coding scheme. In particular, the multirate coders are evaluated within a spatially scalable coding scheme that is designed for the compatible distribution of HDTV using ATM.

ATM is a telecommunication concept defined by ANSI and ITU standards for the transport of user traffic ranging from voice and data to audio and video [ITU-88, ATM 93, Prys91, LeBo92]. Video coding schemes can benefit from ATM transmission since bandwidth-on-demand can be provided. Bandwidth-on-demand means here that a video encoder can in principle transmit as many bits as needed to maintain a certain (local or global) reconstruction quality. Hence, constant quality video codecs operating at variable bit rate (VBR) become feasible in addition to existing constant bit rate (CBR) codecs that yield variable quality. ATM transports data through the network in short packets called cells. Unfortunately, cells are occasionally lost in ATM networks because of network congestion and physical layer impairments. Video coding schemes that use ATM networks should therefore be resilient with respect to such errors by using, for example, graceful degradation methods.

The area of scalable video communications for ATM networks has received a lot of attention. In particular, numerous schemes have been proposed for the compatible distribution of HDTV [Tzou88, LeGa88, Bell88, Bosv90, Peco90, Vand90, Ghar91a, Civa92]. The proposed schemes allow receivers to reconstruct the video signal at multiple spatial resolutions while simultaneously supporting graceful degradation. In this chapter the spatially scalable coding scheme of [Bosv92a, Bosv95] is employed to evaluate the multirate coding techniques. The coding scheme supports 3 reproductions with different spatial resolutions and guarantees the continuity of the distribution service by defining a low-bandwidth CBR layer. This layer facilitates the reconstruction of a reproduction with low spatial resolution. In addition, the scheme creates two additional high-bandwidth VBR layers. These layers ensure that two additional reproductions with higher spatial resolutions can be reconstructed

with approximately constant SNR quality. The scheme includes an ATM adaptation section to make the scheme suitable for ATM transmission. The ATM adaptation section consists of bit stream multiplexers and ATM packetizers. The multiplexers are required to insert synchronization points in the bit streams to facilitate the re-synchronization of decoders in the case of transmission errors.

The content of this chapter is structured as follows. Section 5.2 starts off with a description of the basics of ATM as well as the interaction between video coders and the ATM-network. In particular, the mechanisms which are offered by ATM to support scalability, VBR, and graceful degradation are discussed. Subsequently, the spatially scalable coding scheme is presented in Section 5.3. The structures of the encoder and decoders are thoroughly described, as is the mechanism of VBR source coding. In addition, the ATM adaptation of the different layers is also described. Subsequently, in Section 5.4 the coding performance of the scheme is evaluated for both types of multirate coders. Special attention is given to the error resilience of the scheme when ATM cells are lost because of network congestion.

## 5.2 Video coding for ATM networks

### 5.2.1 ATM connections

ATM networks transport data through the network in short cells of 53 bytes. Each cell consists of a header part (5 bytes) and a payload part (48 bytes). The network routes the cells through the network based on routing information in the cell header. The routing information identifies the ATM connection to which the cell belongs. By definition, the network preserves the order of the cells that are transmitted via a particular connection.

The header of each cell contains a number of fields, of which the Cell Loss Priority (CLP) field is of particular interest. The aim of the CLP field is to limit the loss of cells in the network to particular cells. Specifically, in the case of almost full switching buffers, network switching units will first discard the low priority cells (i.e.  $CLP=1$ ), thereby freeing buffer space for high priority cells (i.e.  $CLP=0$ ). The CLP field can improve the error resilience of applications for which retransmission scenarios are impractical (such as real-time video applications) because it allows for a differentiation of cells within one connection. Video coders can use the CLP field to implement graceful degradation by transporting the most important data in high priority cells.

### 5.2.2 ATM traffic contract

ATM networks transport the cells of a particular connection according to a traffic contract [ATM 93]. The traffic contract is negotiated between the network and the ATM user during the set-up of the connection. The traffic contract contains commitments of the ATM network and the ATM user about respectively the quality of the connection and the usage of the connection. The quality of the connection is usually referred to as the Quality of Service (QoS), while the usage of the connection is commonly referred to as the traffic parameters. ATM networks use traffic contracts to reduce the probability of network congestion and to optimize the use of the network resources.

In general, ATM networks support only a limited number of Quality-of-Service (QoS) classes. The available classes meet the requirements of different groups of services such

as CBR video services and VBR video services. The network commits itself to meet the requested QoS as long as the user complies with the traffic parameters. QoS classes are expressed using different performance parameters such as cell transfer delay, cell delay variation, cell loss ratio (CLR), and cell misinsertion rate. Within a specified QoS class, it is possible that two CLRs are defined, namely one for the low priority cells and one for the high priority cells.

The traffic parameters describe the traffic characteristics of the connection. This description includes various rate parameters such as the peak cell rate (PCR), the sustainable (mean) cell rate (SCR), and the Maximum Burst Size (MBS). The PCR and SCR specify upper bounds on respectively the maximum rate and average rate of the traffic that is submitted to the connection. The MBS specifies the maximum size of a burst of cells that will be transmitted at the PCR rate. The SCR and MBS parameters are optional and allow the ATM user to describe the traffic of an ATM connection in greater detail than just with the mandatory PCR parameter. This directly benefits the network provider because it can utilize its network resources more efficiently, but also benefits the ATM user because reduced charges will apply for the ATM connection.

### 5.2.3 ATM adaptation

Before video coding schemes can transport compressed video data through ATM networks they have to fragment the data into smaller segments that fit in the payloads of ATM cells. This function is performed by a so-called ATM adaptation layer (AAL), which isolates the application from specific characteristics of the ATM layer. The ITU has defined a number of AAL protocols that provide particular functions which are required for different service groups [ITU-93, ATM 93]. Examples of such functions include the segmentation and re-assembling of application data, the handling of bit errors, the handling of cell-delay variation, the handling of lost and misinserted cells, and flow and timing control. For video communications, AAL types 1, 2, and 5 are of special interest.

AAL type 1 is designed to transport the data of CBR audio and video services. It provides a 'circuit emulation'-like service to applications. AAL type 2 is intended to transport the data of VBR audio and video services. AAL type 2 is still under study but initial results show a high degree of correspondence with AAL type 1. AAL types 1 and 2 reduce the effective payload of each ATM cell from 48 bytes to 47 bytes. AAL type 5 is a much simpler protocol than AAL types 1 and 2. AAL type 5 is applicable for both CBR and VBR services and is currently widely used because of its low complexity.

### 5.2.4 Compatible distribution

Hierarchical coding schemes that are designed for the compatible distribution of video signals in ATM networks generally use multiple connections to transport the scalably encoded video signal. Specifically, these schemes transport each layer in a separate connection to provide receivers with a means to select and receive only those layers that are really necessary for their reconstruction processes. Receivers can then make a trade-off between the quality of the reproduced video signal and the consumption of the available bandwidth at the User-to-Network interface.

Hierarchical coding schemes that use multiple connections can easily implement graceful degradation by negotiating different QoS classes for different connections (and thus layers). In particular, by specifying low CLR<sub>s</sub> ( $\text{CLR} < 10^{-8}$ ) for the most important layers and high CLR<sub>s</sub> ( $\text{CLR} < 10^{-3}$ ) for the least important layers, the error resilience capabilities are significantly improved [Morr91, LeBo92, Bosv92a, Bosv95]. With these QoS specifications, cell losses are most likely to happen in the least important layers, which affect the subjective quality of the reproductions least. The schemes can further increase their error resilience capabilities by using the CLP bit to differentiate cells within one connection. For example, the schemes could label the cells that carry synchronization information as high priority cells, while other cells could be labelled as low priority cells.

If the hierarchical coding schemes support reproductions with constant SNR qualities, the schemes have to define PCR, SCR, and MBS parameters for each connection. The PCR and SCR parameters of the connection depend on the characteristics of the layers as well as on the applied buffering of the generated data. In particular, the negotiated PCR of the connection may be well below the maximum rate at which the data is generated, as long as the data is sufficiently buffered before transmission. Further, it should be remarked that the SCR is the upper bound on the average rate of the traffic that is submitted to an ATM connection. This implies that the average rate produced by the source coding algorithm may never exceed this value.

From the above it may be concluded that the notion of VBR from a source coding point of view is not necessarily the same as the notion of VBR from an ATM point of view. Specifically, the former uses the term VBR to indicate rate variations in the bit streams generated by the compression algorithm, while the latter uses the term VBR to indicate variations in the ATM cell rate (i.e. varying inter-arrival times of ATM cells). Because these two views have different scales of time, VBR source coding algorithms are not a prerequisite for using connections with VBR traffic contracts. In fact, CBR coding schemes that employ forward bit allocations algorithms may equally well use VBR connections to limit their buffer requirements. For example, such schemes may packetize and transmit the compressed data as soon as possible without buffering the data for bit rate smoothing. This property will be illustrated in Section 5.3.5 by showing the VBR nature of the quantizer output of a CBR layer.

### 5.3 Intraframe coding scheme

This section presents a spatially scalable coding scheme that allows for the compatible distribution of HDTV in ATM networks. The coding scheme employs intraframe subband coding and supports the hierarchy of video signals that is tabulated in Table 5.1. The video signals supported are all progressive-scan video signals with a 50 Hz frame rate and a 16:9 aspect ratio. The HDTV signal, henceforth referred to as HDP (High Definition Progressive), has the highest spatial resolution, i.e. 1440x1152 active luminance pixels. The compatible signals, i.e. EDP (Enhanced Definition Progressive) and VT (Video Telephony), have spatial resolutions of 720x576 and 360x288 luminance pixels, respectively.

The scheme guarantees the continuity of the distribution service by defining a low-bandwidth CBR layer. This layer facilitates the reconstruction of the VT signal. In addition, the scheme creates two additional high-bandwidth VBR layers. These layers ensure that

Table 5.1: Video hierarchy for progressive-scan HDTV.

Video service	Frame rate	Aspect ratio	Active lines	Active samples per line	
				luminance Y	chrominance UV
HDP	50	16:9	1152	1440	720
EDP	50	16:9	576	720	360
VT	50	16:9	288	360	180

Table 5.2: Negotiated layer bit rates  $\tilde{R}^n$  and reproduction bit rates  $R^n$  in Mbit/s.

Layer $n$	Stream type	$\tilde{R}^n$		$R^n$	
		mean	peak	mean	peak
1	CBR	6	6	6	6
2	VBR	11	29	17	35
3	VBR	53	105	70	140

the EDP and HDP signals can be reconstructed with approximately constant SNR values during normal network operation. The scheme uses a separate ATM connection for each layer and negotiates traffic contracts with the ATM network for these connections such that the (maximum) mean bit rates used for the encoding of the HDP, EDP, and VT signals are 70, 17, and 6 Mbit/s, respectively. The peak bit rates for the HDP and EDP signals are 140 and 35 Mbit/s, respectively. Table 5.2 shows negotiated bit rates of the layers, i.e.  $\tilde{R}^n$ , as well as the (cumulative) bit rates of the reproductions, i.e.  $R^n$ , in Mbit/s.

### 5.3.1 Encoder and decoder structures

Figure 5.1 illustrates the basic structure of the spatially scalable encoder. The encoder consists of three parts: a subband decomposition section, a hierarchical source coder, and an ATM adaptation section. The subband decomposition (*SBD*) decomposes the luminance component of the HDP signal, denoted by  $x$ , into 31 subbands, using the QMF16b filter [John80]. Given these subbands, the VT signal, i.e.  $y^1$ , can be obtained by reconstructing subbands 1 to 7. Similarly, the EDP signal, i.e.  $y^2$ , can be obtained by reconstructing subbands 1 to 19, while the HDP signal, i.e.  $y^3$ , is obtained by reconstructing all subbands.

The hierarchical coding scheme encodes the 31 subbands into 3 layers. The scheme encodes the subbands with MR-PCM and MR-DPCM coders in order to optimize the coding performance for each reproduction. In particular, the highly correlated subband 1 is encoded with a MR-DPCM coder, while the other subbands are encoded with MR-PCM coders. The MR-PCM and MR-DPCM coders can use either multistage quantization (as illustrated in Fig. 5.1) or embedded quantization. Note that Fig. 5.1 does not show the prediction loop for the MR-DPCM coder. An adaptive bit allocation algorithm (*BA<sub>n</sub>*) supervises the quantization (Q) and adaptive arithmetic coding (AAC) processes for each layer. For layer

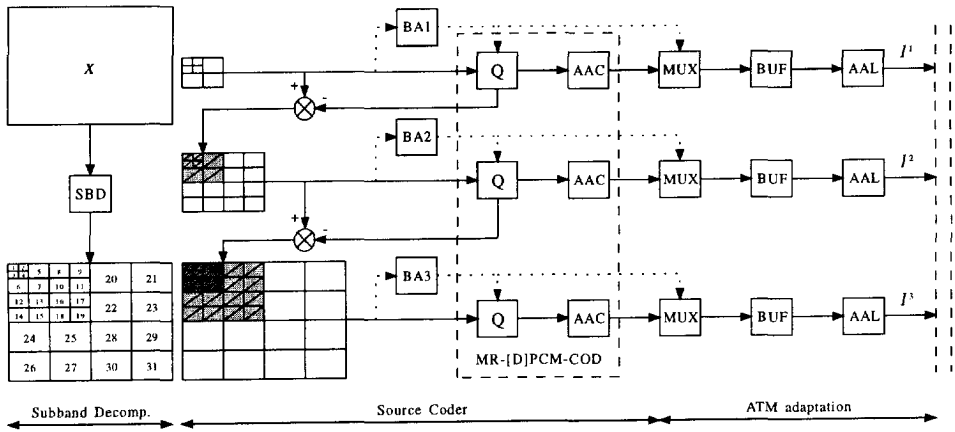


Figure 5.1: Spatially scalable intraframe encoder using multistage quantization.

$I^1$ ,  $BA1$  optimally distributes the available bits over the 7 subbands by selecting quantizers from a set of admissible multirate quantizers. For layers  $I^2$  and  $I^3$ ,  $BA2$  and  $BA3$  allocate the allowed distortion in a similar way to the subbands 1 to 19 and 1 to 32, respectively. Consequently, the layers have bit stream characteristics that can be made compliant with the negotiated traffic parameters.

The ATM adaptation section receives the three layers from the hierarchical source coder and makes these layers suitable for transmission over ATM networks. In particular, each layer subsequently passes, a transport multiplexer (MUX), a buffer (BUF), and an ATM packetizer (AAL). The multiplexer merges the encoded subband data with side information of the source coder and adds synchronization words to the bit stream. The resulting bit stream is subsequently buffered before packetization by the AAL. This buffering makes it possible to shape the layers in such a way that they become compliant to the negotiated traffic parameters. The AAL implements AAL type 1 for layer 1 and AAL type 2 for layers 2 and 3.

Figure 5.2 illustrates the structure of the associated decoders. The decoders can reconstruct the reproductions by receiving and decoding the appropriate layers. After receiving the layers, the ATM adaptation sections of the receivers perform the inverse functionalities of the ATM adaptation sections of the encoder. In particular, the AALs depacketize the multiplexed subband data from the ATM packets and put the data in decoding buffers (BUF). A demultiplexer (DUX) removes the data from each buffer and extracts the side information from the streams. The side information is used to select the quantization and arithmetic coding tables that were used in the encoding process. This is done by performing bit/distortion allocations ( $BA_n$ ) identical to those in the encoder. Subsequently, the subband data is forwarded to the multirate decoding processes, which involves adaptive arithmetic decoding (AAD), table look-up operations ( $Q_L$ ), and, if multistage quantization is used, the addition of the reconstructed subbands. In the last stage, reproductions  $y^1$ ,  $y^2$ , and  $y^3$  are created by reconstructing (SBR) the subbands.

The scalable encoder and decoders process the color components of the video signals in a way similar to the luminance signal. The U and V components are separately decomposed



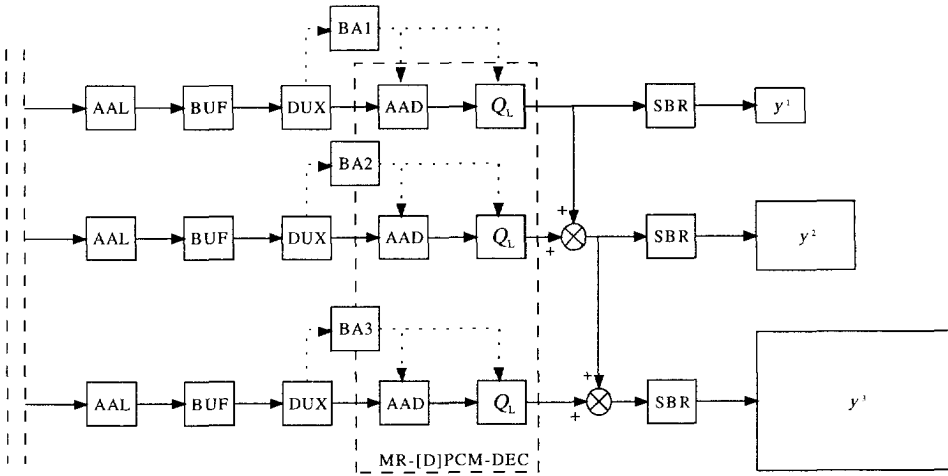


Figure 5.2: Structure of spatially scalable intraframe decoders.

into 31 subbands, using the QMF16b filter [John80]. The subbands containing the low-pass information are MR-DPCM encoded while the other subbands are MR-PCM encoded. The available bit rate for layer 1 is divided between the luminance and chrominance signals on an 80 : 20 % ratio prior to the actual bit allocations. The bit allocation for the chrominance signals selects quantizers for both the U and V subbands, thereby minimizing the joint MSE of these signals. Similarly, the distortion allowed for layers 2 and 3 is divided between the luminance and chrominance signals prior to the distortion allocations. This division is based on the variances of the components. In the rest of this chapter only the processing of the luminance component is discussed and evaluated.

### 5.3.2 Adaptive bit/distortion allocations

The allocation algorithm of layer 1 ( $BA1$ ) allocates a certain amount of bits  $R^1$  for each frame. In general, this results in a varying reconstruction quality per frame because the compressibility of each frame varies. Figure 5.3 illustrates this by showing the rate-distortion characteristics of two frames. Because the rate-distortion characteristics vary for each frame, different distortions will result when both frames are encoded at a constant rate  $R^1$  (i.e.  $D^1(1)$  and  $D^1(2)$ ). The allocation algorithms at layers 2 and 3 allocate a certain amount of distortion  $D^n$  for each frame (with  $D^2 > D^3$ ). Consequently, a different bit rate is obtained for each frame. In Fig. 5.3 these rates are denoted by  $R^n(1)$  for frame 1 and  $R^n(2)$  for frame 2. Since  $R^n(1)$  and  $R^n(2)$  can exceed the negotiated traffic parameters, the resulting bit rates have to be limited to a certain maximum value ( $R^n_{max}$ ). This maximum value is derived from the maximum mean bit rate that is negotiated for that layer (as specified by the SCR). Therefore, the algorithm stops searching when  $R^n_{max}$  is reached and selects a set of quantizers that does not achieve the prescribed minimal reconstruction quality.

The applied allocation algorithms  $BA1 - BA3$  are based on the adaptive bit allocation

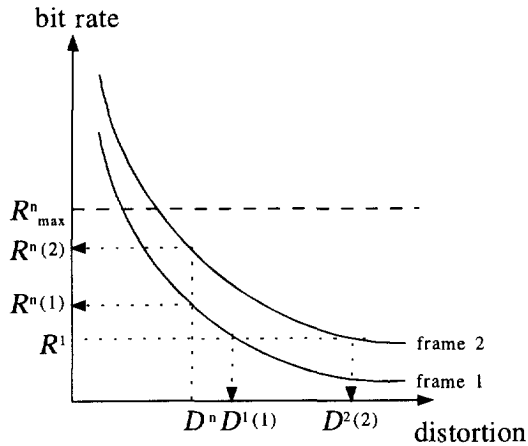


Figure 5.3: Bit and distortion allocation algorithm.

algorithm as discussed in Chapter 2. First, the algorithms divide the data of each subband into four activity classes based on a variance measure. For the MR-DPCM encoded subband the variance estimates are calculated via the Yule-Walker equations. After the classification, the algorithms determine the variances of all activity classes of all subbands. Subsequently, the allocation algorithms select a quantizer for each activity class from the set of available quantizers based on the measured variances and the quantization history (if any). The algorithms use performance tables in the selection procedure that indicate the expected resulting bit rate and distortion per activity class for each available quantizer after quantization and arithmetic coding. The block classification table (BCT) of a subband is inserted into the bit stream the first time the subband is encoded with a non-zero bit rate. This allows decoders to decode the compressed data unambiguously.

### 5.3.3 Adaptive MR-DPCM and MR-PCM coders

For each layer, the applied allocation algorithm assigns a quantizer to each activity class of a subband. Depending on the type of the MR-PCM coder applied, this quantizer quantizes either the quantization error of the previous coding stage (multistage quantization) or the original subband data (embedded quantization). After quantization, the subsequent adaptive arithmetic coder (AAC) encodes all indices of several consecutive subband rows into one arithmetic codeword. By doing this, the AAC obtains an output rate very close to the entropy of the indices, because it encodes a large number of indices simultaneously. Consequently, the arithmetic codewords form the smallest recoverable data units in the bit streams and are henceforth called slices.

The construction of a slice is illustrated in Figure 5.4a where a fictitious subband is shown that is fragmented in blocks. The greyvalue of each block indicates the activity class to which the block belongs. As shown in Fig. 5.4a, the AAC encodes indices of different blocks into one slice, which implies that the AAC has to use different encoding tables for

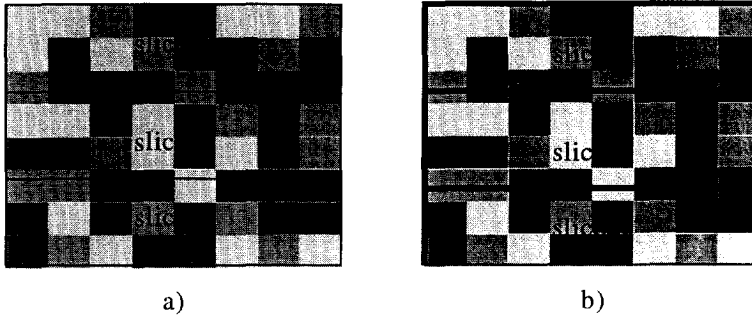


Figure 5.4: Subband blocks belonging to four activity classes (as indicated by greyvalue): a) construction of a slice, b) re-initialization pixels in MR-DPCM subband (thick lines).

different blocks. As can be observed, the slices may have any number of subband rows. If the coding scheme employs multirate coders based on multistage quantization, the AAC has to change the encoding table at block boundaries unless two neighboring blocks belong to the same activity class. If the coding scheme employs multirate coders based on embedded quantization, the AAC has to change the encoding table for every sample because of the (inherent) index translation. Specifically, it creates a new encoding table for each sample within a block based on the applied quantizer at this stage, the applied quantizer at the previous stage (if any), the selected index at this stage, and the selected index at the previous layer (if any).

For the MR-DPCM encoded subband, the prediction loop is re-initialized at the start of a slice to make a partial reconstruction of the subband possible. This procedure is necessary to limit the effects of inevitable cell losses in ATM networks. The re-initialization of the DPCM coding loop implies that particular subband pixels are predicted using fewer reconstructed pixels. In particular, the re-initialization pixels are only predicted using pixels that belong to this slice causing the slices to become self-contained. Fig. 5.4b illustrates these re-initialization points by thick lines.

### 5.3.4 Synchronization and side information

The hierarchical coder transmits each layer in a separate ATM connection. The ATM adaptation of each layer therefore embodies a multiplexer (MUX), a buffer (BUF), and an ATM adaptation layer (AAL) module. The multiplexer interleaves the arithmetic codewords with the required side information and inserts start codes with synchronization capabilities. The resulting bit stream is subsequently buffered before packetization and transmission by the ATM adaptation layer module.

Figure 5.5 shows the syntax of the bit streams that are produced by the three multiplexers. As can be seen, the bit streams are composed of start codes (SOF-Y, SOF-UV, SOB, SOL), side information (SI), and compressed data. A start code indicates the start of a compressed data unit, which may be the luminance information of a frame (SOF-Y), the

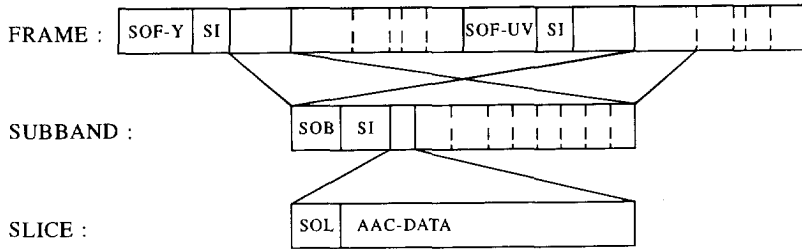


Figure 5.5: Syntax of the layer bit streams after multiplexing.

chrominance information of a frame (SOF-UV), a subband (SOB), or a slice (SOL). Each frame consists of multiple subbands, each of which again consists of multiple slices. The number of slices per subband is determined by the bit allocation algorithm based on the allocated bit rate for that subband.

A start code consists of a synchronization word of 32 bits (i.e. 0x00000001) followed by 8 bits that indicate the type and sequence number of the compressed data unit. Decoders use these start codes to synchronize the decoding of the three separate connections and to recover synchronization after cell loss in a connection. In particular, loss of synchronization usually implies the loss of the remainder of the slice because decoders can not resynchronize in the middle of the arithmetic codeword.

The side information transmitted in a frame contains the variances of the subband activity classes. Decoders require these variances to perform bit/distortion allocations identical to those in the encoder. The outcomes of the bit/distortion algorithms indicate the set of quantizers used, AC tables used, and the number of slices per subband. The mean value of the luminance component or the mean values of the chrominance components are also contained in the side information of a frame. Further, a frame consists of those subbands that are encoded at a non-zero rate.

Subbands may contain side information and multiple slices. When the side information is transmitted it consists of the block classification table (BCT) of the subband. The BCT is transmitted together with the first transmitted data of the subband. The BCT is necessary to decode the different blocks of the subband data with the appropriate tables. For the MR-DPCM encoded (low-frequency) subband of  $l^1$ , the side information also includes the prediction coefficients used by the encoder.

Slices contain single arithmetic codewords. The allocation algorithms dynamically determine the number of slices in a subband for a particular layer in order to limit the effects of ATM cell loss. In particular, they create equally sized slices for each subband of a layer that contain approximately 2500 bits (i.e. approximately 7 ATM cells). The number of slices in a subband therefore depends on the bit rates of the assigned quantizers of the subband activity classes. Hence, large and accurately quantized subbands may be transmitted in as much as 20 slices, while small and coarsely quantized subbands are carried in a single slice. For all subbands, the overhead of slice start codes is approximately 1.5% of the bit rate used.

### 5.3.5 Packetization and bit stream shaping

The bit streams produced by the multiplexers are temporarily buffered and then packetized by the AAL modules. These modules map the encoder-specific bit streams into the payloads of ATM packets according to AAL type 1 (layer 1) and AAL type 2 (layers 2 and 3). In order not to violate the negotiated contracts, the bit streams of all layers have to be controlled, or shaped, with respect to the agreed upon traffic parameters. This shaping is made possible by using a buffer between the multiplexer and the packetizer and by modifying the order in which the subbands are actually encoded.

As an example, consider Figure 5.6a. Here the bit streams of layers 1 and 3 are shown when the subbands are encoded in the order of their subband number (c.f. Figure 5.1), and the output of the multiplexer is packetized instantaneously (i.e. no buffering). The horizontal axis of Fig. 5.6a represents the time and is expressed in units of frame times. As can be observed, the bit rates fluctuate heavily in time and violate the negotiated peak bit rates (dotted line) because of the accurate encoding of the low-resolution subbands. The mean bit rates (per frame time), however, conform to the negotiated contracts, because of the forward bit allocations.

To reduce the peak bit rates to the negotiated values, the AAL modules implement the following elementary shaping procedure: the buffer between the multiplexer and AAL is filled with the output of the multiplexer and is emptied at the negotiated peak bit rate as long as the buffer is not empty. For layer 1 this implies that the output bit stream becomes CBR (although stuffing may be necessary to compensate for the small differences between the assumed subband PDFs and the real subband distributions). Because of the VBR coding, the buffers of layers 2 and 3 are filled with different amounts of bits for different frames. If the buffers of these layers underflow, the packetization is temporarily stopped. Packetization is started again when the buffer fills up again. Figure 5.6b shows the resulting shaped bit streams, while Figure 5.6c shows the fullness of the buffers. The resulting streams now conform to the traffic contracts with respect to the mean and peak bit rates.

To influence the shape of the VBR stream to a larger extent, information from the forward bit allocation procedures may be used in the shaping procedure. For example, after the bit allocation procedure, the expected bit rates of all subbands are known. Consequently, the shape of the output bit stream can therefore be controlled more drastically by changing the encoding order of the subbands.

## 5.4 Experimental results

This section evaluates the coding performances of three coding schemes that are adapted for ATM transmission: i) the compatible distribution scheme using multirate coders based on multistage quantization, ii) the compatible distribution scheme using multirate coders based on embedded quantization, and iii) the basic subband scheme of Chapter 2 using 31 subbands. The compatible distribution schemes are henceforth referred to as the multistage scheme and the embedded scheme, for notational convenience. The basic subband scheme employs adaptive PCM and DPCM coders based on UTQ and AC and has an ATM adaptation section similar to that of the compatible distribution schemes. For simplicity purposes, the schemes employ ATM adaptation sections that incorporates only a subset of the functionalities of the standard AAL types 1 and 2. In particular, they only implement the standardized

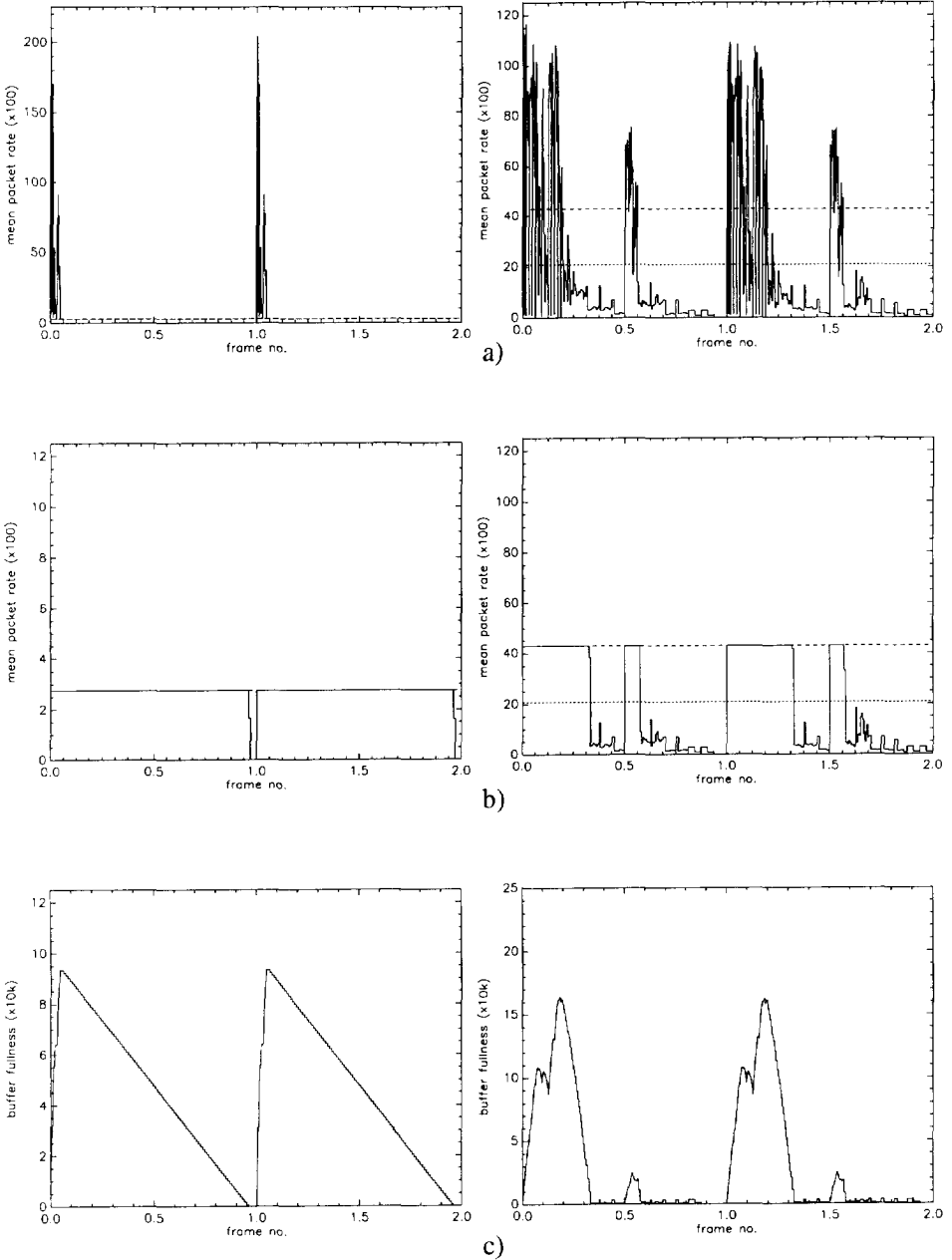


Figure 5.6: Generated bit streams of layer 1 (left column) and layer 3 (right column): a) unshaped bit streams, b) shaped bit streams, c) buffer fullness. The left vertical axes in a) and b) give the number of generated ATM cells (in quantities of 100) per frame time.



Figure 5.7: Coding performances for HDP signal.

fragmentation and reassembly procedures so that decoders can detect lost ATM cells.

The results are obtained using the luminance component of the "Mobile" sequence. This progressive-scan sequence consists of 40 frames (720x576 pixels) and has a frame rate of 50 Hz. Because the "Mobile" sequence does not have HDP dimensions, the actual applied bit rates are down-scaled versions of the ones tabulated in Table 5.2. In addition, 20% of the bit rate is reserved for the encoding of the chrominance components.

#### 5.4.1 Performance comparison

Figure 5.7 shows the obtained SNRs of the HDP signal for each system when the HDP signal is encoded at a constant bit rate of 70 Mbit/s. The basic subband scheme (SBC) outperforms the other two schemes by approximately 0.1 dB for the embedded scheme and by approximately 0.6 dB for the multistage scheme. These performance gaps are in correspondence with the results of Chapter 4 and are apparently due to the reduced coding efficiencies of the multirate DPCM and PCM coders.

Figure 5.8 shows the coding performances of the associated EDP and VT signals as encoded by the spatially scalable schemes. The bit rates of these signals are 17 and 6 Mbit/s, respectively. For these signals, the term 'coding performance' refers only to the quantization process since no corresponding input signals exist. Fig. 5.8 shows that the quality of the reproduced VT signal of the multistage scheme is slightly higher (0.25 dB) than the quality obtained with the embedded scheme. This result could be expected since the quantizers used at the first stage with multistage quantization are (optimal) UTQs, and not the pruned UTQs used with embedded quantization. In contrast, the quality of the reproduced EDP signal of the multistage scheme is slightly lower (0.1 dB) than the quality obtained by the embedded scheme. This result indicates that the better performance of the multirate coders based on embedded quantization already pays off with two reproductions. Indeed, with three stages,

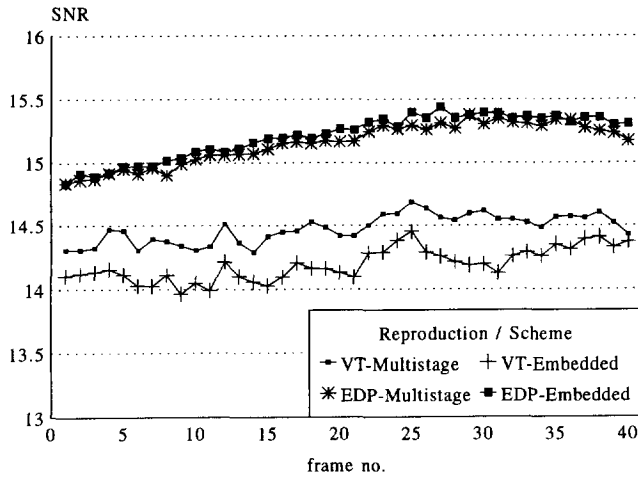


Figure 5.8: Coding performances for EDP and VT signals.

the reproduced HDP signal of the embedded scheme outperforms the multistage scheme by approximately 0.5 dB, as was illustrated in Fig. 5.7.

### 5.4.2 Cell-loss resilience

This section analyzes the resilience of the three schemes with respect to the loss of single ATM cells. None of the three coding schemes uses techniques such as error concealment methods and/or forward error-correcting codes (FEC) that limit the effects of cell losses. If ATM cells are lost, the various decoders will resynchronize at the start code of the next slice and re-start the decoding from that point in the bit stream. The lost subband data will be replaced by zero values in the reconstruction process.

The loss of single ATM cells is simulated for each scheme according to the method described in [ISO/93]. With this method the probability of a lost ATM cell is Poisson distributed with a particular cell-loss ratio (CLR). For each coding scheme the compressed data of the input sequence is decoded four times, while simulating cell-losses using the above model. The basic subband scheme is exposed to CLR of  $10^{-6}$  and  $10^{-3}$ . The compatible schemes are exposed to the following two triples of CLR for  $(l^1, l^2, l^3)$ : i.e.  $(10^{-8}, 10^{-6}, 10^{-6})$  and  $(10^{-8}, 10^{-3}, 10^{-3})$ . Figures 5.9, 5.10, and 5.11 show the results obtained for respectively, the basic subband scheme, the multistage scheme, and the embedded scheme.

Figure 5.9 shows that the HDP reconstruction quality of the basic subband scheme is severely affected by ATM cell-losses for a CLR of  $10^{-3}$ . At frame 155 an especially severe drop in reconstruction quality (8.5 dB) occurs. The performance for the multistage scheme (Fig.5.10) is on the contrary only slightly affected for CLR of  $10^{-3}$  for the higher layers. The performance for the embedded scheme (Fig. 5.11) is moderately affected for CLR of  $10^{-3}$ . Although the drops in SNR value for ATM cell-losses are larger than with the multistage scheme, they are more bounded than the drops that occur with basic subband



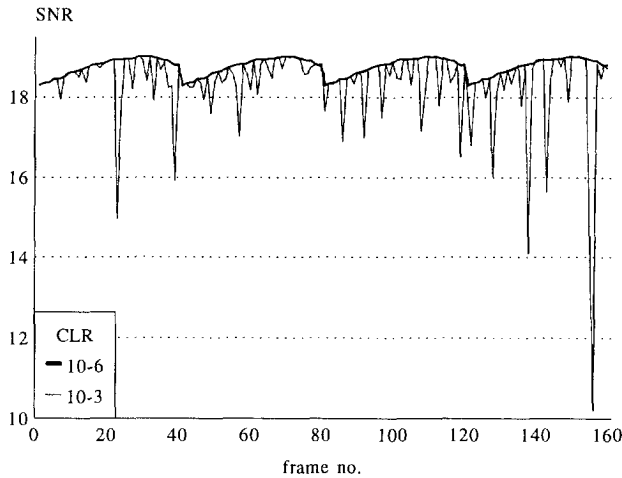


Figure 5.9: Cell-loss resilience of basic subband scheme.

scheme. Subjective evaluations of the reconstructed sequences confirm these numerical results, as errors in the compatible schemes are hardly visible, as opposed to the annoying artifacts caused by cell losses in the basic subband scheme. Note that when the CLR's are decreased to  $10^{-6}$  none of the schemes suffers from severe reconstruction problems.

Figures 5.10 and 5.11 show that a significant difference exists between the cell-loss resilience of the multistage scheme and that of the embedded scheme. This difference can be explained by the fact that with multistage quantization the layers are independent of each other, while with embedded quantization the layers are dependent on each other. Specifically, if with multistage quantization the second layer is lost, the first and third layer can still be decoded and used in the reconstruction process. With the embedded quantization, the loss of a cell in layer 2 automatically implies the loss of corresponding information in the third layer since that layer can not be decoded without the information of layer 2.

## 5.5 Discussion

In this chapter the influence of multirate coders on the performance of a complex hierarchical coding scheme was analyzed. In particular, the multirate coders were evaluated within a spatially scalable coding scheme that is designed for the compatible distribution of HDTV using ATM.

For normal network conditions, the results obtained for the test sequence are similar to the results obtained in the previous chapter for the SNR scalable scheme: the performances obtained with the compatible schemes are slightly lower than the performance of the basic subband scheme. If the compatible scheme uses multirate coders based on embedded quantization, the performance difference is approximately 0.1 dB and is negligible. If the compatible scheme uses multirate coders based on multistage quantization, the performance difference is approximately 0.6 dB. Hence, the scheme based on embedded quantization

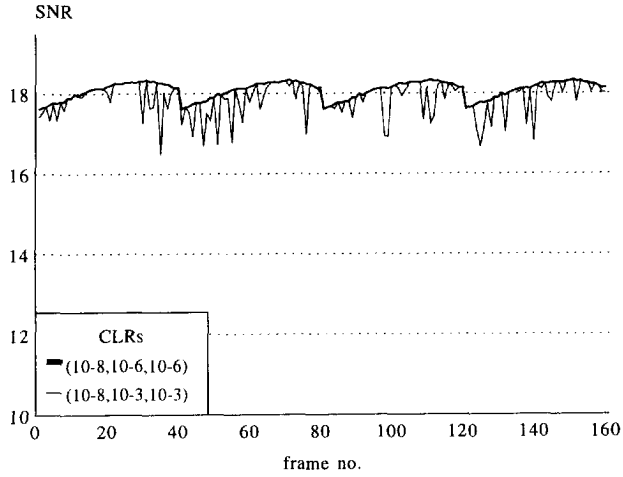


Figure 5.10: Cell-loss resilience of compatible scheme using multistage quantization.

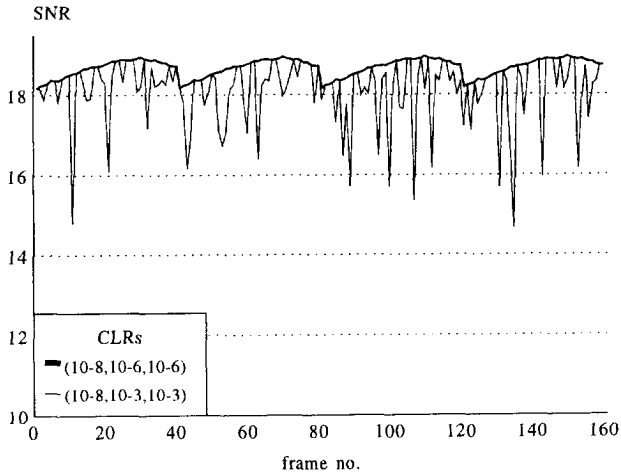


Figure 5.11: Cell-loss resilience of compatible scheme using embedded quantization.

outperforms the scheme based on multistage quantization.

However, for congested network conditions, the relative performances are reversed. The compatible schemes clearly have a much higher error resilience with respect to ATM cell losses than the basic subband scheme does. In fact, with the compatible schemes the cell losses are hardly visible, while with the basic subband scheme the reconstructed images show very annoying artifacts. For the basic subband scheme, the loss of ATM cells can degrade the SNR of the reconstructed signal by almost 8 dB. The compatible schemes limit this loss to approximately 3 dB. The error resilience of the compatible scheme based on multistage quantization is higher than the resilience obtained for the compatible scheme based on embedded quantization. Although subjectively no significant difference can be noticed, the degradation in SNR is less than with embedded quantization. As stated before, this difference is due to the fact that with multistage quantization the layers are independent of each other, while with embedded quantization the layers are dependent on each other.

Based on the above observations, it can thus be concluded that the expected network performance may influence the architecture of the hierarchical coding scheme. In particular, if the expected error rates are high, the hierarchical coding scheme should use multirate coders that are based on multistage quantization to increase the error resilience as much as possible. If the expected error rates are low, the hierarchical coding scheme should use multirate coders based on embedded quantization to increase the coding performances of the different reproductions as much as possible.



# Chapter 6

## Spatially scalable coding using MC-SBC

### 6.1 Introduction

In Chapters 4 and 5 it was shown that scalable coding schemes based on intraframe coding techniques can obtain a performance similar to non-scalable schemes provided that they use the proper multirate coding techniques. In this chapter it is investigated whether the above conclusion also holds for spatially scalable coding schemes that use interframe coding techniques, i.e. motion estimation and motion compensation. For a wide range of applications the use of motion information is imperative for sufficient compression of the input signal. The use of motion information in subband coding schemes is, however, complex and is even more complicated in spatially scalable schemes.

Out-band and in-band motion compensation are two methods that combine the processes of motion compensation and subband coding [Ghar91b, Vial91, Vand91a]. With out-band compensation, the input frame of time  $k$  is temporally compensated prior to a subband decomposition. After coding, the frame is reconstructed from the subbands and is stored as a reference for the input frame of time  $k + 1$ . This method has received a lot of attention and is very similar to the approaches standardized in ITU-T H.261 and MPEG [CCIT89, ISO/92, ISO/95]. With in-band compensation, the subbands themselves are temporally compensated, i.e. the compensation process is performed in the subband domain. After coding, the subbands of input frame of time  $k$  are stored as references for the subbands of time  $k + 1$ . The in-band method has received less attention than the out-band method, probably because motion compensation in the subband domain is not trivial. The reason for this is that subbands of successive frames are not always shifted versions of each other even when the two input frames are.

In literature, both compensation methods have been proposed as the basis for spatially scalable coding schemes [Ghar91b, Vand91b, Vial91, Jeon91, Uz91b, Scha92, Nave92, Ansa92, Vett92, Civa92, Zafa93, Hors93, Gonz93, Lame93, Bosv93b, Nave95]. Such schemes support multiple reproductions by using multiple temporal prediction loops. For high-quality reproductions, it is essential that the temporal prediction of a particular reproduction is based on information that is available to the decoder of that reproduction. If this constraint is not fulfilled the reproduction will be distorted because of drift, i.e. the divergence of the prediction loop in the encoder and the decoder. For hierarchical coding schemes that implement spatial scalability, in-band motion compensation seems to be very

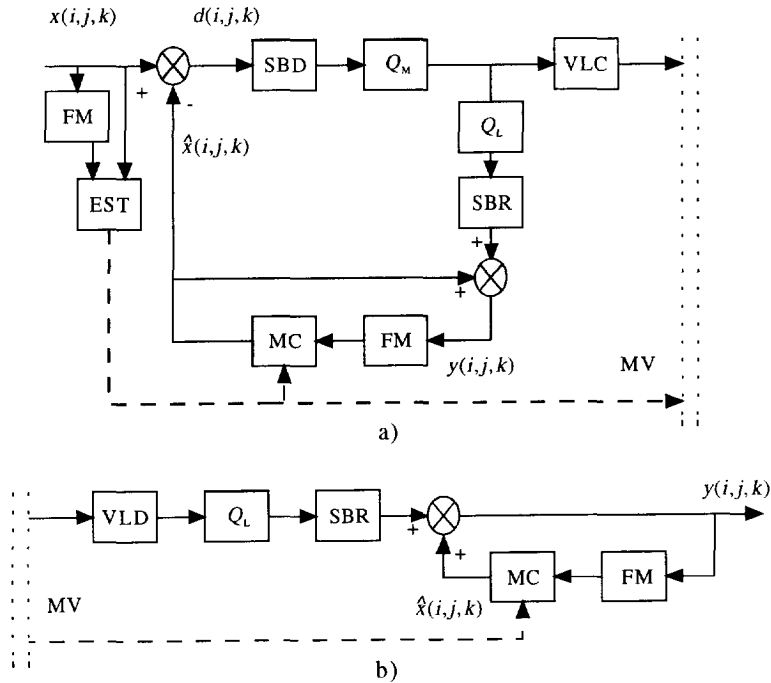


Figure 6.1: Out-band motion compensation: a) encoder, b) decoder.

appropriate; namely, with this method the multirate coders of Chapter 4 can be used to refine the low-frequency subbands for high-resolution reproductions.

This chapter focuses on the structure and performance of spatially scalable schemes that are based on in-band motion compensation. First a more detailed description of the out-band and in-band compensation methods is presented in Sections 6.2 and 6.3. In Section 6.3 it is shown that in-band motion compensation can yield a prediction performance similar to that of out-band motion compensation and that Perfect Reconstruction Modulated Filters (PRMF) filter banks are very appropriate for in-band motion compensation [Mau92, Bouc92]. Subsequently, in Section 6.4 it is analyzed how in-band motion compensation can be integrated with spatial scalability. In this section a number of options are described for the prediction of low-frequency subbands that are part of multiple high-resolution reproductions. These prediction options arise as an intrinsic property of the combination of spatial scalability and motion compensation. Next, a hierarchical coding scheme that supports two reproductions with different spatial resolutions is proposed in Section 6.5. The scheme implements the prediction options of Section 6.4 and employs multirate coders to encode the subbands where possible. Finally, in Section 6.6 some experimental results are presented that illustrate the coding performance of the scheme by comparing the prediction options and the impacts of the multirate coders based on multistage and embedded quantization.

## 6.2 Subband coding using out-band compensation

Figure 6.1 shows how the out-band compensation method combines the processes of subband coding and motion compensation. The structure of the encoder is similar to the structure of a spatial DPCM encoder except that the different building blocks process frames instead of pixels. The coding stage that is used within the prediction loop is a subband coder. The input of the subband decomposition (*SBD*) is a frame that contains the difference  $d(i, j, k)$  between the input frame of time  $k$ , denoted by  $x(i, j, k)$ , and the temporally predicted frame at time  $k$  denoted by  $\hat{x}(i, j, k)$ , i.e.

$$d(i, j, k) = x(i, j, k) - \hat{x}(i, j, k). \quad (6.1)$$

The temporally predicted frame  $\hat{x}(i, j, k)$  is the outcome of the motion compensation procedure (*MC*) that creates  $\hat{x}(i, j, k)$  from the previous reconstructed frame  $y(i, j, k - 1)$  and an estimated motion vector field *MV*, i.e.

$$\hat{x}(i, j, k) = y(i + \delta_i, j + \delta_j, k - 1). \quad (6.2)$$

Here,  $(\delta_i, \delta_j)^T$  denotes the integer motion vector for pixel  $x(i, j, k)$ . The subbands of the prediction error  $d(i, j, k)$  are encoded by the quantizers ( $Q_M$ ) and the variable length coders (*VLC*) before transmission.

To facilitate the temporal prediction of frame  $x(i, j, k + 1)$  the encoder has to reconstruct the encoded frame. To this end the encoder contains an embedded decoder that reconstructs the quantized prediction error frame by means of a quantizer table look-up operation ( $Q_L$ ) and a subband reconstruction (*SBR*). The reconstructed frame at time  $k$ , i.e.  $y(i, j, k)$ , is then obtained by additively combining the reconstructed prediction error frame with the temporally predicted frame, i.e.

$$y(i, j, k) = Q(d(i, j, k)) + \hat{x}(i, j, k). \quad (6.3)$$

The reconstructed frame is subsequently stored in the frame memory (*FM*) for the prediction of the next frame.

For an optimal coding performance, the out-band scheme has to estimate the motion vectors as accurately as possible. However, the motion vectors have to be transmitted as side information to the decoder to facilitate the reconstruction of  $\hat{x}(i, j, k)$  at the decoder side. To keep the bit rate of the side information at sufficiently small values, motion-compensated schemes divide the input frame  $x(i, j, k)$  into small blocks (e.g. 8x8 pixels) and use only one motion vector per block [Bier88, Drie92]. In general, this approach is sufficiently accurate and yields a more efficient encoding than the intraframe coding of the input frame or the encoding of the frame difference between the two successive input frames. The problem of motion estimation is not addressed in this chapter. The interested reader is referred to the abundance of literature in this field, for instance [Tzir94].

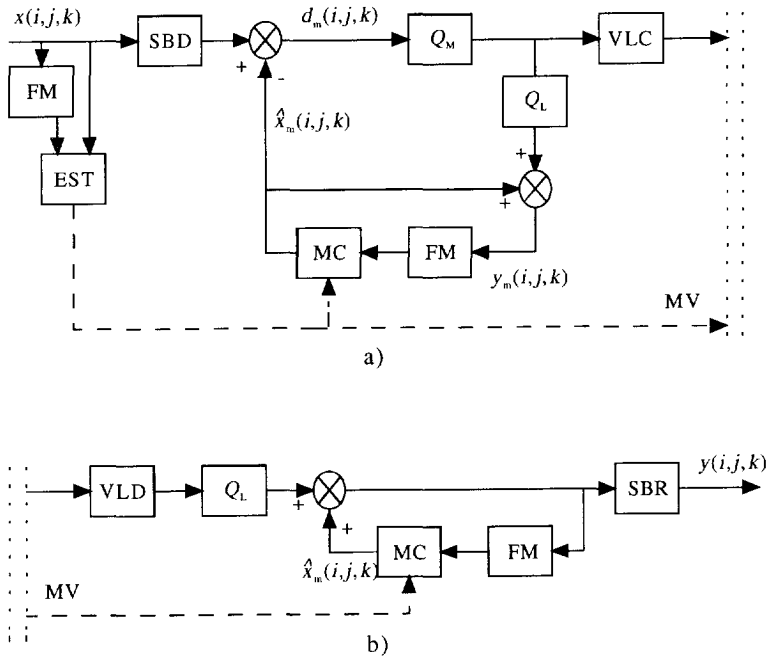


Figure 6.2: In-band motion compensation: a) encoder, b) decoder.



## 6.3 Subband coding using in-band compensation

### 6.3.1 Principle

Figure 6.2 shows how the in-band compensation method integrates the processes of subband coding and motion compensation. With in-band compensation the subbands are temporally predicted and compensated before encoding. Hence, the coder first decomposes the input frame  $x(i, j, k)$  into  $M$  subbands  $x_m(i, j, k)$  with  $1 \leq m \leq M$  by means of a subband decomposition (*SBD*). The differences between the subbands and the temporally predicted subbands  $\hat{x}_m(i, j, k)$  are then calculated as

$$d_m(i, j, k) = x_m(i, j, k) - \hat{x}_m(i, j, k). \quad (6.4)$$

Here,  $d_m(i, j, k)$  denotes the  $m^{\text{th}}$  prediction error subband. The temporally predicted subbands  $\hat{x}_m(i, j, k)$  are the outcome of the motion compensation procedure (*MC*) that creates  $\hat{x}_m(i, j, k)$  from the reconstructed subbands of the previous frame, i.e.  $y_m(i, j, k - 1)$ , and an estimated motion vector field *MV*. The prediction error subbands  $d_m(i, j, k)$  are then encoded by the quantizers ( $Q_M$ ) and the variable length coders (*VLC*) before transmission.

To facilitate the temporal prediction of the subbands of the next input frame, i.e.  $x_m(i, j, k + 1)$ , the encoder reconstructs and stores the quantized subbands of time  $k$ . This is done by first decoding the resulting indices of  $Q_M$  using a quantizer table look-up operation ( $Q_L$ ) and then by combining the decoded subbands additively with the temporally predicted subbands, i.e.

$$y_m(i, j, k) = Q(d_m(i, j, k)) + \hat{x}_m(i, j, k). \quad (6.5)$$

The reconstructed subbands  $y_m(i, j, k)$  are then stored in the frame memory (*FM*). The decoder reconstructs the subbands  $y_m(i, j, k)$  in the same way as the encoder does. The decoder obtains the reconstructed frame of time  $k$ , i.e.  $y(i, j, k)$ , by synthesizing the subbands  $y_m(i, j, k)$  with the subband reconstruction section (*SBR*).

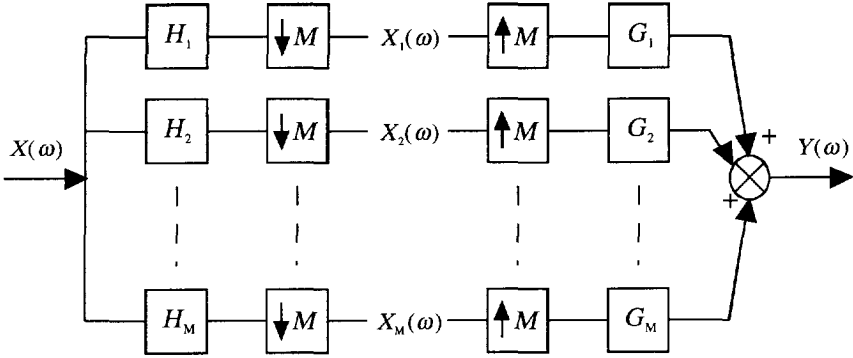
### 6.3.2 Compensation method

The difficulty of the in-band motion compensation method lies in the procedure to predict the subbands of time  $k$ , i.e.  $x_m(i, j, k)$ , from the reconstructed subbands of time  $k - 1$  and a motion vector field *MV*. A simple and straightforward procedure is to use Eq. 6.2 on a subband basis, i.e.

$$\hat{x}_m(i, j, k) = y_m(i + \delta_{m,i}, j + \delta_{m,j}, k - 1). \quad (6.6)$$

The motion vectors  $(\delta_{m,i}, \delta_{m,j})^T$  of Eq. 6.6 could either be estimated from the reconstructed subbands of the previous frame or could be down-scaled versions of motion vectors estimated from the input signal. Unfortunately this procedure does not yield a prediction gain comparable to the prediction gain obtained with the out-band compensation method. The reason for this is that subbands of successive frames are not always shifted versions of each other even when the input frames are.

To illustrate this, let us analyze the temporal relation between subbands in the case that two successive frames are shifted versions of each other [Nave93, Bouc92]. If  $X(\omega, k - 1)$  is the Fourier transform of a 1-dimensional signal at time  $k - 1$  then the subband signals

Figure 6.3:  $M$ -band filter bank.

$X_m(\omega, k - 1)$  of the  $M$ -band filter bank (see Figure 6.3) at time  $k - 1$  are given by:

$$X_m(\omega, k - 1) = \frac{1}{M} \sum_{l=1}^M H_m\left(\frac{\omega}{M} - \frac{2\pi(l-1)}{M}\right) X\left(\frac{\omega}{M} - \frac{2\pi(l-1)}{M}, k - 1\right). \quad (6.7)$$

The aliased signal components in the subbands are represented by  $X\left(\frac{\omega}{M} - \frac{2\pi(l-1)}{M}, k - 1\right)$  with  $(l - 1) > 0$ . Now assume that two frames are related to each other by a translation over  $v_1$  samples:

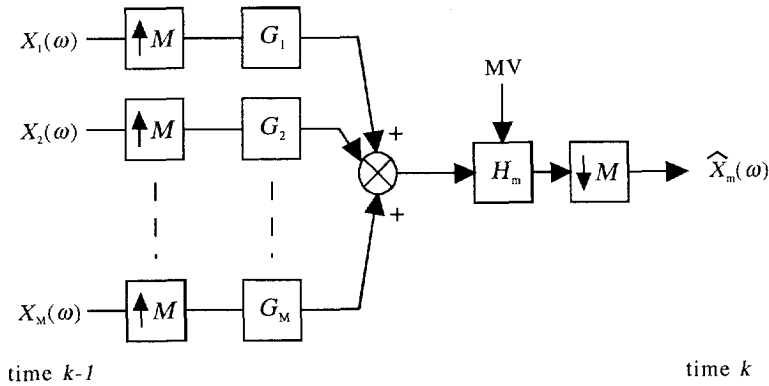
$$X(\omega, k) = X(\omega, k - 1)e^{-j\omega v_1}. \quad (6.8)$$

Then it is also possible to express the subbands signals  $X_m(\omega, k)$  at time  $k$  in terms of signal  $X(\omega, k - 1)$ , i.e.

$$X_m(\omega, k) = \frac{1}{M} \sum_{l=1}^M H_m\left(\frac{\omega}{M} - \frac{2\pi(l-1)}{M}\right) X\left(\frac{\omega}{M} - \frac{2\pi(l-1)}{M}, k - 1\right) e^{-j\left(\frac{\omega}{M} - \frac{2\pi(l-1)}{M}\right)v_1}. \quad (6.9)$$

By comparing Eqs. 6.8 and 6.9 for instance for subband  $m = 0$  it shows that the signal components  $X\left(\frac{\omega}{M} - \frac{2\pi(l-1)}{M}, k - 1\right)$  with  $(l - 1) = 0$  are related by a spatial displacement  $v_1/M$ , whereas the aliasing terms, i.e.  $(l - 1) \neq 0$ , generally are not. Only if the shift is an integer ratio of the decimation factor (i.e.  $v_1/M$  is integer) are the subbands exactly related by a (decimated) displacement.

Given the above example it is now possible to deduce the proper compensation procedure for in-band motion compensation. In particular, in well-designed filter banks, the decimation filters  $H_m(\omega)$  and interpolation filters  $G_m(\omega)$  are designed such that during the reconstruction the aliasing terms  $X\left(\frac{\omega}{M} - \frac{2\pi(l-1)}{M}\right)$  with  $(l - 1) \neq 0$  in one subband are cancelled out by similar terms in other subbands. This implies that in order to predict subband  $m$  at time  $k$  without any aliasing, *all* subbands of time  $k - 1$  are required. More precisely, to predict subband  $m$  at time  $k$ , first all subbands of time  $k - 1$  need to be upsampled by  $M$ , filtered with  $G_m(\omega)$ , and summed to create the reconstruction at time  $k - 1$ . Subsequently, the reconstruction is filtered using subband filter  $H_m(\omega)$  and subsampled subject to a phase shift according to the

Figure 6.4: Prediction of subband  $m$  at time  $k$ .

motion field. This compensation process is illustrated in Figure 6.4, where  $\hat{X}_m(\omega)$  is the predicted subband  $m$  at time  $k$ . Henceforth this compensation procedure is referred to as the in-band compensation method with interpolation.

To illustrate the prediction performances of the different compensation methods, Table 6.1 tabulates the prediction gain for the out-band compensation method, the in-band method without interpolation (i.e. Eq. 6.6), and the in-band method with interpolation. The prediction gains are obtained by predicting a 256x256 subset of the 8th frame of the "Mobile" sequence (no coding). The filter bank employed is a tree-structured QMF filter bank. As can be seen the prediction gain for the out-band compensation method is similar to the in-band method with interpolation. The prediction gain obtained for the in-band method without interpolation is significantly smaller than for the other two.

### 6.3.3 PRMF filter banks

In the previous section it was shown that with in-band motion compensation the subbands have to be interpolated to the sampling frequencies of the input signal before the aliasing components can be cancelled. To limit the complexity of the interpolation procedure it is therefore advantageous to use an efficient filter bank that allows subband data to be interpolated directly to the full-resolution. As a result, the interpolation procedure might be implemented on a block basis: i.e. the (subband) blocks necessary for the motion compensation procedure are interpolated, summed, and decimated independently of neighboring blocks. Unfortunately, the tree-structured QMF filter banks as applied in the previous chapters of this thesis do not allow for such a direct interpolation of the subband data.

In literature, Perfect Reconstruction Modulated Filters (PRMF) filter banks have been proposed as efficient filter banks for in-band motion-compensated schemes [Mau92, Bouc92]. PRMF filter banks directly decompose the input signal into  $M$  subbands using a parallel filter structure (as illustrated in Fig. 6.3). The complexity of the filter bank (and thus the complexity of the in-band prediction procedure) is substantially less than that of the tree-structured QMF filter banks, because fewer additions and multiplications are necessary [Mau92]. Further,

Table 6.1: Obtained prediction gains  $G_P$  for "Mobile" using a QMF filter bank (16 subbands).

Method	$G_P$
Out-band	57.15
In-band without interpolation	8.12
In-band with interpolation	57.04

PRMF filter banks can be implemented in such a way that the  $M \times M$  filter bank contains the embedded  $M/2 \times M/2$  filter bank. It is evident that these properties make PRMF filter banks very suitable for use in spatially scalable schemes based on in-band motion compensation [Mau92].

PRMF filter banks decompose the input signal into  $M$  subbands using filters of moderate length. The analysis filters  $h_m(l)$  and synthesis filters  $g_m(l)$  of a separable  $M \times M$  PRMF filter bank are obtained by cosine modulation of an  $L$ -tap ( $L = 2M - 1$ ) low-pass prototype filter  $p(l)$ , which is defined as [Mau92]:

$$p(l) = \frac{1}{M\sqrt{2}} \sin \frac{\pi(l+1)}{2M}, \quad (6.10)$$

with  $0 \leq l \leq L - 1$ . The filters  $h_m(l)$  and  $g_m(l)$  are subsequently defined by

$$h_m(l) = 2p(l) \cos(\pi(2m-1)(2l-M-L+1)/4M), \quad \text{and} \quad (6.11)$$

$$g_m(l) = 2p(l) \cos(\pi(2m-1)(2l+M-L+1)/4M), \quad (6.12)$$

with  $1 \leq m \leq M$  and  $0 \leq l \leq L - 1$ . Due to this definition the reconstructed signal  $y(k)$  is equal to a scaled version of the input signal (i.e. Perfect Reconstruction):

$$Y(\omega) = \frac{1}{M} \sum_{m=1}^M H_m(\omega) G_m(\omega) X(\omega) = \frac{1}{M} X(\omega). \quad (6.13)$$

To illustrate the capabilities of PRMF filter banks for in-band motion compensation, Table 6.2 shows the in-band prediction gains ( $G_P$ ) that are obtained for a 16-band QMF filter bank (QMF-16) and a 16-band PRMF filter bank (PRMF-16). The prediction gains are obtained by predicting the same 256x256 subset of the 8th frame of the "Mobile" sequence as used for Table 6.1 (no coding). As can be observed, the values obtained for  $G_P$  are very similar to the prediction gain obtained by the out-band compensation method (c.f. Table 6.1), irrespective of the type of the filter bank.

Table 6.2 also shows the in-band prediction gains ( $G_{P,SCAL}$ ) that are obtained when the low-frequency subbands (i.e. subbands 1 to 4) are predicted only on the basis of subbands 1 to 4. This situation is typical for spatially scalable schemes in which high-resolution decoders make the same temporal predictions for the low-frequency subbands as the low-resolution decoders do, in order to avoid drift. The values obtained for  $G_{P,SCAL}$  are less than those of  $G_P$  but still provide a considerable gain. The cause for the reduction in prediction gain is

Table 6.2: Obtained prediction gains for "Mobile" using in-band motion compensation with QMF and PRMF filter banks.

Filter bank	$G_P$	$G_{P,SCAL}$
QMF-16	57.04	46.46
PRMF-16	57.16	41.96

the presence of (non-motion-compensated) aliasing terms in subbands 1 to 4. These aliasing terms are still present in the subbands because subbands 5 to 16 were not used in the interpolation procedure and the aliasing terms in the subbands of subsequent frames are not related by the motion vectors (c.f. Eq. 6.9). Unfortunately, these distortions are inevitable in spatially scalable interframe subband schemes because of the definition of spatial scalability.

## 6.4 Spatial scalability and in-band MC-SBC

To obtain a spatially scalable coding scheme based on in-band motion compensation, the techniques for spatial scalability have to be combined with the technique of in-band motion compensation. The integration of in-band motion compensation and spatial scalability is however not straightforward and can be done in a number of ways [Vial91, Vand91a, Ghar91b, Ansa92, Scha92, Bouc92, Bosv93b]. The main issue with this integration is the prediction of the low-frequency subbands of the high-resolution signal(s).

Figure 6.5 illustrates this issue by showing the subbands of both a low-resolution reproduction  $y^1$  and a high-resolution reproduction  $y^2$  at time  $k - 1$  and  $k$ . The reconstructed and decoded subbands of  $y^n$  at time  $k$  are denoted as  $y_m^n(i, j, k)$ , while the original subbands of  $y^n$  (not encoded) are denoted as  $x_m^n(i, j, k)$ . Fig. 6.5 shows that the low-frequency subbands of  $y^2$ , i.e.  $x_m^2(i, j, k)$  with  $1 \leq m \leq 4$ , can be predicted using the following two types of reference subbands: i.e.

- Low-resolution reference subbands, i.e. the reconstructed subbands of  $y^1$  of time  $k$  are used as references for the prediction, and
- Temporal reference subbands, i.e. the reconstructed and motion-compensated subbands of  $y^2$  of time  $k - 1$  are used as as references for the prediction.

Both types of subbands are suitable to be used as prediction references because both types of subbands are available in the high-resolution decoders. In fact, the low-resolution references are the result of the spatial scalability feature, while the temporal references are the result of the temporal prediction loop. Note that only the low-frequency subbands of  $y^2$  have two kinds of reference subbands available. The high-frequency subbands of  $y^2$ , i.e.  $x_m^2(i, j, k)$  with  $5 \leq m \leq 16$ , have only the temporal reference subbands available.

The existence of two kinds of reference subbands for the prediction of the low-frequency subbands of  $y^2$  has triggered the development of a number of prediction methods

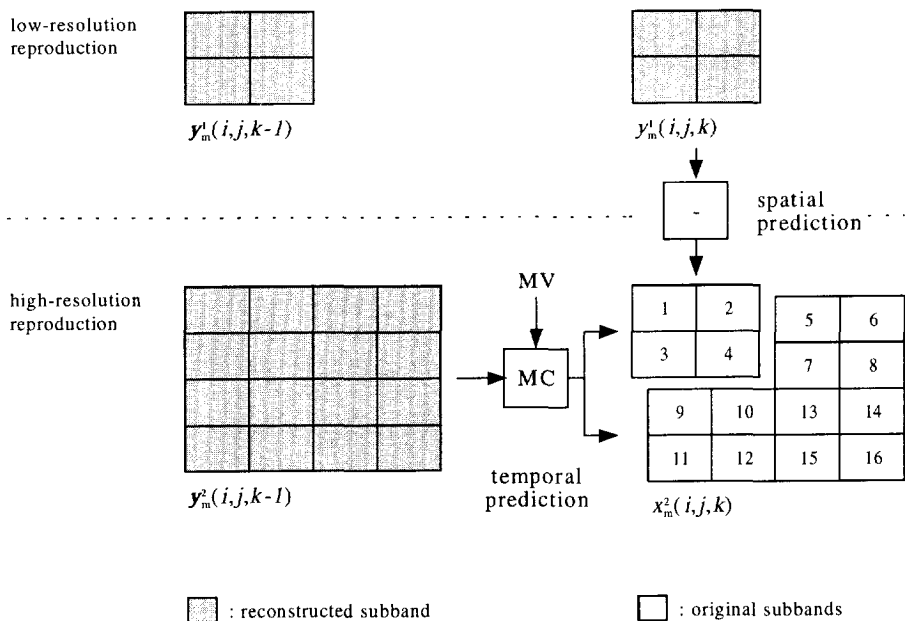


Figure 6.5: Low-resolution and temporal prediction options.

[Vand93, Chia93a]. In the following, four prediction methods are described that are considered as the basic methods proposed in literature. In these descriptions, the predictions for the subbands  $x_m^2(i, j, k)$  are denoted as  $\hat{x}_m^2(i, j, k)$ , while the reconstructed and motion-compensated subbands of  $y^2$  at time  $k - 1$  are denoted as  $\tilde{y}_m^2(i, j, k - 1)$ . With this notation, the four basic methods can now be defined as

1. Low-resolution prediction, i.e.

$$\hat{x}_m^2(i, j, k) = y_m^1(i, j, k). \quad (6.14)$$

The low-resolution prediction method predicts the subbands  $x_m^2(i, j, k)$  from the reconstructed subbands  $y_m^1(i, j, k)$ . As a result, the low-frequency subbands are encoded in an amplitude-scalable way for which the multirate coders of Chapter 4 can be used.

2. Temporal prediction, i.e.

$$\hat{x}_m^2(i, j, k) = \tilde{y}_m^2(i, j, k - 1). \quad (6.15)$$

The temporal prediction method predicts the subbands  $x_m^2(i, j, k)$  from the reconstructed and motion-compensated subbands of  $y^2$  at time  $k - 1$ . As a result, the subbands of  $y^1$  and  $y^2$  are encoded independently of each other. This method makes the encoding scheme equal to a simulcast coding scheme.

3. Average prediction, i.e.

$$\hat{x}_m^2(i, j, k) = \frac{1}{2}y_m^1(i, j, k) + \frac{1}{2}\tilde{y}_m^2(i, j, k - 1). \quad (6.16)$$

The average prediction method combines the low-resolution and temporal references into a single prediction for subbands  $x_m^2(i, j, k)$ . The average prediction method provides a higher prediction gain than the previous two methods when the coding noise in the reconstructed subbands is reduced because of the averaging procedure.

4. Selective prediction, i.e.

$$\hat{x}_m^2(i, j, k) = P_s \in \{P_1, P_2, \dots, P_S\}, \quad (6.17)$$

with, for example,  $S = 3$ , and

$$\begin{aligned} P_1 &= y_m^1(i, j, k), \\ P_2 &= \tilde{y}_m^2(i, j, k - 1), \text{ and,} \\ P_3 &= \frac{1}{2}y_m^1(i, j, k) + \frac{1}{2}\tilde{y}_m^2(i, j, k - 1). \end{aligned} \quad (6.18)$$

The selective prediction method evaluates a number of candidate predictions (i.e.  $P_s$  with  $1 \leq s \leq S$ ) for each subband  $x_m^2(i, j, k)$ . Typical candidates for the prediction would be the predictions produced by the low-resolution, temporal, and average prediction methods. After evaluation, the encoder selects the best prediction (i.e. the one that yields the minimum prediction error) and signals this selection as side information to the decoders. This allows decoders to use the same prediction as the encoder uses. The selective prediction method can be used on a block basis of, for instance, 8x8 pixels. This makes the method more adaptive at the cost of additional side-information.

The low-resolution prediction method is fully in line with the hierarchical coding approach as discussed in this thesis; i.e. the encoded information of the low-resolution reproduction is re-used by the coding stage of the high-resolution reproduction. This could imply that the coding performance for  $y^2$  is comparable with the performance of a non-scalable scheme when optimized multirate coders are used. However, this dependency could also affect the coding performance for  $y^2$ , because the temporal prediction loop for  $y^1$  operates at a lower bit rate and suffers from non-cancelled aliasing components. Consequently, the prediction gain for the subbands  $x_m^1(i, j, k)$  ( $1 \leq m \leq 4$ ) will be less than the prediction gain for  $x_m^2(i, j, k)$ , which will thus reduce the coding efficiency of  $y^2$ .

The temporal prediction method is less in line with the hierarchical coding approach because the coding of  $y^2$  is fully independent from the coding stage of  $y^1$ . This implies that the low-frequency information of the input signal might be transmitted twice to the decoders, depending on the success of the motion compensation procedure. With the temporal prediction method the encoding scheme degenerates to a simulcast scheme (c.f. Chapter 1). On the one hand, this implies that the coding performance for  $y^2$  is less with respect to the performance of a non-scalable scheme because fewer bits are available for the encoding. On the other hand, the independence also implies that coders for  $y^1$  and  $y^2$  can be independently tuned and optimized to maximize the coding performances. The temporal prediction method reduces the applicability for multirate coders, because the (low-resolution) subbands are not encoded in an amplitude-scalable way.

The average prediction method yields an encoding scheme that is half-way between the hierarchical and simulcast coding approaches. The average prediction method will provide a higher prediction gain than the low-resolution and temporal methods do when the coding noise in the different prediction references is small and independent of each other.

The selective prediction method yields an encoding scheme that switches dynamically and locally between the hierarchical coding concept and the simulcast coding concept. This method seems to be the most attractive method because it always finds the best candidate prediction out of a number of candidate predictions. Especially when this method is done on a block basis the gain compared to the previous three methods can be substantial. Consequently, the performance loss with respect to a non-scalable coder for the input signal at the same bit rate can be minimized.

## 6.5 Interframe coding scheme

In this section an interframe coding scheme is presented that allows for the spatially scalable coding of an input signal. The coding scheme supports progressive-scan video signals and is based on a 16-band PRMF filter bank and in-band motion compensation. The scheme supports the four spatio-temporal prediction methods that were described in the previous section. The scheme facilitates the reconstruction of two spatial reproductions by creating two layers  $l^1$  and  $l^2$ . Reproduction  $y^1$  can be reconstructed by decoding the information in layer  $l^1$ , while reproduction  $y^2$  can be reconstructed by decoding the information carried in layers  $l^1$  and  $l^2$ . The scheme assumes that the bit rates available for the layers are constrained by channel properties and are constant in time. The scheme encodes every 20th input frame in an intraframe mode to provide random access points in the layers. The bit rate ratio between the intraframe and interframe encoded frames is fixed to 2:1.



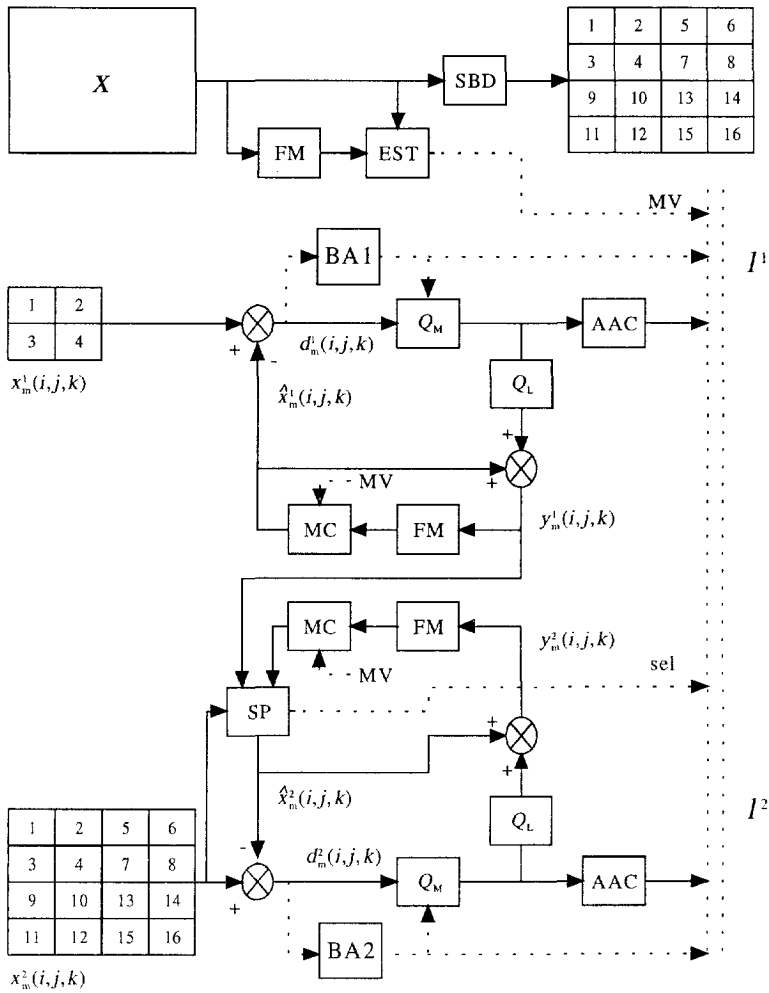


Figure 6.6: Two-layer spatially scalable subband encoder using in-band compensation.

### 6.5.1 Encoder structure

Figure 6.6 illustrates the basic structure of the spatially scalable encoder. The encoder consists of two sections: a motion estimation and subband decomposition section and an encoding section. The first section estimates for each input frame the motion vectors  $MV$  between the current frame and the previous frame, using a motion estimator (EST). The motion estimator is a hierarchical block matcher that estimates one motion vector for each  $8 \times 8$  block using two levels of estimation (max. displacement  $+31/-31$  pixels). The motion vectors are transmitted to the decoders as side information in layer  $l^1$ . Subsequently, each frame is decomposed into 16 subbands using a subband decomposition (SBD) based on separable PRMF filters. Subbands 1 to 4 are necessary to reconstruct reproduction  $y^1$ , while all subbands are necessary to reconstruct  $y^2$ .

The encoding of the subbands 1 to 4 for  $y^1$  is similar to a non-scalable in-band motion-compensated scheme. The interframe coder makes a temporal prediction for each subband based on the reconstructed subbands of the previous frame and the estimated motion vectors  $MV$ . The prediction error subbands are encoded using APCM encoders under supervision of a forward bit allocation algorithm (BA1). This algorithm optimally distributes the available bits of the layer among the activity classes of the subbands by selecting quantizers from a set of available quantizers. If the subbands are not temporally predicted (i.e. intraframe coding) then subband 1 is encoded using ADPCM. The outputs of the quantizers ( $Q_M$ ) are encoded by the adaptive arithmetic coders (AAC) and are transmitted in layer  $l^1$ .

The encoding of the subbands for  $y^2$  is similar to the encoding of  $y^1$ . However, the prediction for each subband is constructed by a Subband Prediction (SP) procedure, which has the following inputs: i) the reconstructed subbands of  $y^1$  at time  $k$ , i.e.  $y_m^1(i, j, k)$ , ii) the motion-compensated subbands of  $y^2$  at time  $k - 1$ , i.e.  $\tilde{y}_m^2(i, j, k - 1)$ , and iii) the original subbands of  $y^2$ , i.e.  $x_m^2(i, j, k)$  with  $1 \leq m \leq 16$ . Based on these inputs the SP procedure creates the prediction  $\hat{x}_m^2(i, j, k)$  that is used to predict subbands  $x_m^2(i, j, k)$  with  $1 \leq m \leq 4$ . The selection of the best prediction method is performed on an activity class basis (2 bits overhead for each class, 32 bits total). The activity classes are determined by the BA1 procedure on the basis of the temporal prediction errors  $d_m^l(i, j, k)$ . The selection information from the SP is transmitted as side-information in layer  $l^2$  to the high-resolution decoders.

The predicted subbands are subsequently encoded under supervision of a second forward bit allocation algorithm (BA2). This algorithm optimally distributes the available bits for  $l^2$  among the activity classes of the subbands by selecting quantizers from a set of available quantizers. If the subbands are not temporally predicted (i.e. intraframe coding) the SP procedure automatically selects the low-resolution reference subbands as predictions. The outputs of the quantizers ( $Q_M$ ) are encoded by the adaptive arithmetic coders (AAC) and transmitted in layer  $l^2$ .

### 6.5.2 Decoder structure

Figure 6.7 shows the basic structure of the spatially scalable decoders. These decoders reconstruct the reproductions by receiving and decoding the appropriate layers. The reconstruction of  $y^1$  involves the selection of the same quantizers as used in the encoder by making

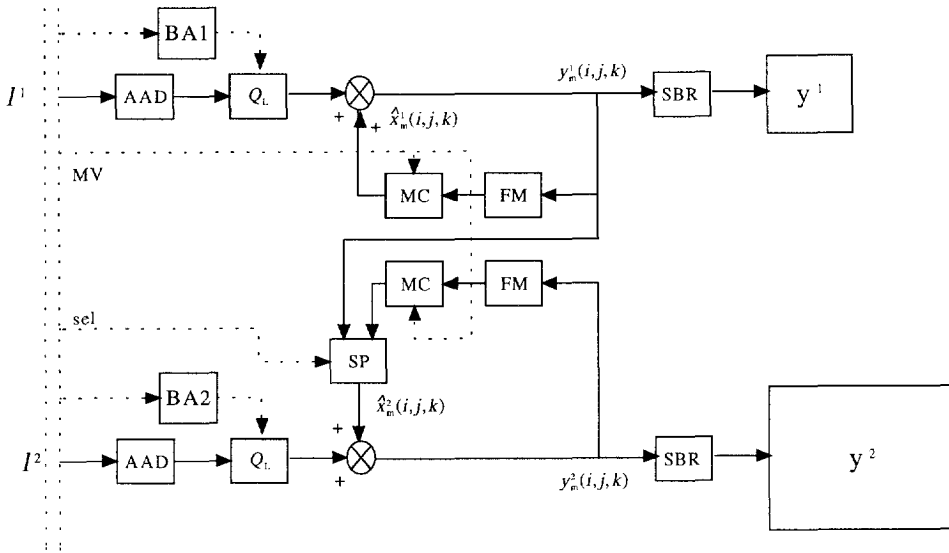


Figure 6.7: Spatially scalable subband decoder using in-band compensation.

the same bit allocation (*BA1*) as the encoder does, based on the transmitted side-information. Subsequently, the decoders decode the codewords in layer  $l^1$  by means of the adaptive arithmetic decoders (*AAD*) and the quantizer look-up procedures ( $Q_L$ ) and construct the temporal predictions based on the transmitted motion vectors (*MV*) and the previously reconstructed subbands. The subbands  $y_m^1(i, j, k)$  are then reconstructed by additively combining the temporal predictions with the decoded prediction error subbands. These subbands are then inserted in the subband reconstruction section (*SBR*), which then reconstructs frame  $k$  of  $y^1$ .

The reconstruction of  $y^2$  is similar to the reconstruction of  $y^1$ . The main difference is that the *SP* procedure makes the same subband predictions as the encoder does, based on the transmitted side-information in  $l^2$ . Next, the decoded prediction error subbands are added to the temporal predictions and are inserted in the subband reconstruction section (*SBR*) to reconstruct frame  $k$  of  $y^2$ .

### 6.5.3 Coding and prediction methods

The coding of the subbands in the interframe encoder and decoders is dependent on the prediction method that is implemented in the *SP* procedure. If the *SP* procedure implements the low-resolution prediction method, the temporally predicted subbands of  $y^1$  ( $d_m^1(i, j, k)$ ) can be encoded using the multirate coding techniques of Chapter 4. The activity classes of each subband are then encoded with quantizers that are either part of a multistage quantizer or part of an embedded quantizer. In the former case UTQs are used, while in the latter case pruned UTQs are used. In both cases, the bit allocation procedure keeps track of the quantization history of each activity class and selects the appropriate quantizer from the set of quantizers that has been optimized for the (expected) GG-PDF of the prediction

Table 6.3: Prediction error variances of subbands  $d_m^2(i, j, k)$  with  $1 \leq m \leq 4$  obtained for various spatio-temporal prediction methods for "Kiel" sequence.

Subband nr.	Activity class	Low-resolution prediction	Temporal prediction	Average prediction	Selective prediction
1	1	19.97	7.02	9.54	7.02
	2	20.39	10.88	11.39	10.88
	3	14.92	16.10	10.88	10.88
	4	16.35	33.50	17.27	16.35
2	1	17.61	7.18	9.12	7.18
	2	15.62	11.22	9.72	9.72
	3	13.66	17.53	10.38	10.38
	4	15.42	43.00	18.37	15.42
3	1	14.63	9.12	8.70	8.70
	2	12.90	16.94	10.04	10.04
	3	14.92	24.17	11.39	11.39
	4	14.00	33.76	13.58	13.58
4	1	11.22	6.01	7.10	6.01
	2	14.75	10.38	9.54	9.54
	3	15.26	13.83	9.88	9.88
	4	12.65	19.86	10.46	10.46
Total variance 1-4		61.11	70.15	44.36	44.08
Total variance 1-16		213.34	222.38	196.59	196.40
$G_P$		12.91	12.38	14.01	14.02

error. Experimental studies have shown that the distributions of these temporally predicted subbands can be reasonably modelled by a GG-PDF with  $c = 2.00$  (subband 1) and  $c = 1.50$  (other subbands).

If the *SP* procedure implements a spatio-temporal prediction method other than the low-resolution prediction method, the applicability of multirate coders is reduced. Specifically, the multirate coders are only applicable to those activity classes for which the selective prediction method selects the low-resolution reference subbands. For other reference subbands and the temporal and average prediction methods, there is no use for multirate coders because the two quantization stages encode different input signals for each reproduction. Nonetheless, even for these prediction modes the framework designed in Chapter 4 for multistage quantization is very useful to encode the resulting prediction errors. In particular, experimental studies have shown that the distributions of the prediction errors can also be reasonably modelled by a GG-PDF with  $c = 2.00$  (subband 1) and  $c = 1.50$  (other subbands). Since UTQs have been designed for these PDFs, the bit allocation algorithm *BA2* merely has to select the appropriate set of quantizers from the multistage framework to encode these prediction errors accurately.

To illustrate the functionality of the *SP* procedure Table 6.3 shows the prediction error variances of the low-frequency subbands of  $y^2$  for the 4 prediction methods. In addition, the prediction gains  $G_P$  for  $y^2$  are shown. The variances and prediction gains are obtained by encoding a 256x256 subset of the 2nd frame of the "Kiel" sequence with  $R^1 = 0.8$  bpp and  $R^2=0.8$  bpp. Activity class 1 contains the subband blocks (8x8 pixels) with the lowest variances.

As can be observed, the activity class variances obtained by the low-resolution prediction method are all approximately equal. This is a direct result of *BA1*, which minimizes the distortion for  $y^1$  by making the different quantization error variances approximately equal. The activity class variances of the temporal prediction are not equal and increase for each class. This is as expected because the temporal prediction errors will be similar to those on which the classification of *BA1* is based. The variances obtained by the average prediction method are mostly between the variances of the low-resolution and temporal prediction methods. Occasionally, however, the variances are lower than any of the other variances (e.g. subband 3, class 1). Finally, the activity class variances of the selective prediction method are practically always equal to the minimum value obtained by the other three methods for the same activity class. This is a consequence of the fact that the overhead in bit rate of the method is negligible (i.e. 0.0005 bpp for 256x256 images).

Table 6.3 also shows the total prediction gain  $G_P$  obtained for  $y^2$  for each prediction method. As expected, the selective prediction method outperforms the other prediction methods, although the difference with the average method is very small. In general, the relative order for the prediction gains of the low-resolution, temporal, and average prediction methods depends on the input signal and on the success of the temporal prediction loops. In particular, when the prediction of the motion in the input signal is difficult, the low-resolution prediction method will give the highest prediction gain. If the prediction of the motion in the input signal is relatively easy and the bit rate of  $y^1$  is sufficiently high, then the average prediction method will give the highest prediction gain since it averages the accurate predictions of both the low-resolution and temporal prediction methods.

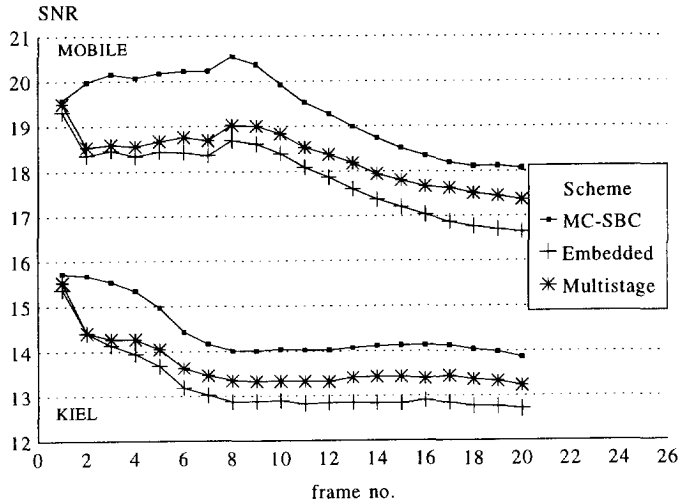


Figure 6.8: Coding performances for  $y^2$  using low-resolution prediction method.

## 6.6 Experimental results

The coding performances of the interframe coding scheme are evaluated in this section. The results are obtained using the first 20 frames (720x576 pixels) of the luminance components of the progressive-scan "Mobile" and "Kiel" sequences (assumed frame rate 25 Hz).

The spatially scalable encoder encodes the low-resolution reproduction  $y^1$  and the high-resolution reproduction  $y^2$  at 1.78 Mbit/s ( $R^1=0.55$  bpp) and 6.48 Mbit/s ( $R^2=0.50$  bpp), respectively. The transmission rates include a 20% overhead for the encoding of the chrominance components.

### 6.6.1 Coding performance of low-resolution prediction method

Figure 6.8 shows the SNRs obtained for  $y^2$  when the  $SP$  procedure implements the low-resolution prediction method. This prediction mode enables the use of multistage and embedded quantizers. For reasons of convenience, these schemes are henceforth called the multistage scheme and the embedded scheme, respectively. In addition to the performances of these schemes, Fig. 6.8 shows also the SNR obtained for a non-scalable scheme based on in-band motion compensation (MC-SBC). This scheme operates at the same bit rate as the scalable schemes.

As illustrated, the non-scalable scheme outperforms both scalable schemes by at least 0.75 dB, while the multistage scheme in turn outperforms the embedded scheme by approximately 0.5 dB. This performance difference indicates that the reduced prediction gain for the subbands of  $y^1$  indeed affects the coding performances for  $y^2$ . Remember that the cause for the reduced prediction gain is that the temporal prediction loop for  $y^1$  operates at a lower bit rate and that the motion-compensated predictions suffer from non-cancelled

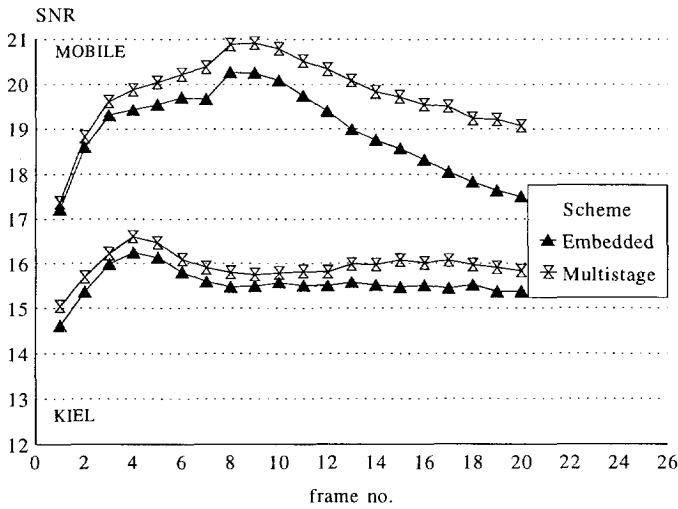


Figure 6.9: Coding performances for  $y^1$  using the low-resolution prediction method.

aliasing components. The performance difference between the multistage scheme and the embedded scheme shows that the UTQ quantizers of the multistage scheme have a better performance at these rates than the pruned UTQs of the embedded scheme.

Figure 6.9 shows the obtained coding performances of the multistage and embedded schemes for  $y^1$ . The term 'coding performance' refers here only to the quantization process since no corresponding input signal exists. Fig. 6.9 shows that the quality of  $y^1$  of the multistage scheme is higher than the quality obtained with the embedded scheme (approximately 0.25 dB to 1.00 dB). This result was expected since the quantizers used at the first stage with multistage quantization are UTQs, as opposed to the pruned UTQs used for embedded quantization. However, the performance difference is much larger than the difference observed for the spatially scalable schemes of Chapter 5 that were based on the same multirate coders. Apparently, the cause for this is the temporal prediction loop, which amplifies the (small) performance difference that by definition exists between UTQs and pruned UTQs.

Based on Figures 6.8 and 6.9 it can be concluded that an interframe coding scheme that implements only the low-resolution prediction method can not really provide a competitive performance compared to a non-scalable scheme. In fact, the performance gap is always larger than 0.75 dB.

### 6.6.2 Coding performances of all prediction methods

Fig. 6.10 shows the SNRs obtained for  $y^2$  when the *SP* procedure implements the four defined prediction methods. For convenience reasons, the four resulting schemes are called the low-resolution scheme, the temporal scheme, the average scheme, and the selective scheme. All schemes use UTQ quantizers, which implies that the low-resolution scheme is equal to the multistage scheme of the previous section. Fig. 6.10 also shows the SNR obtained for

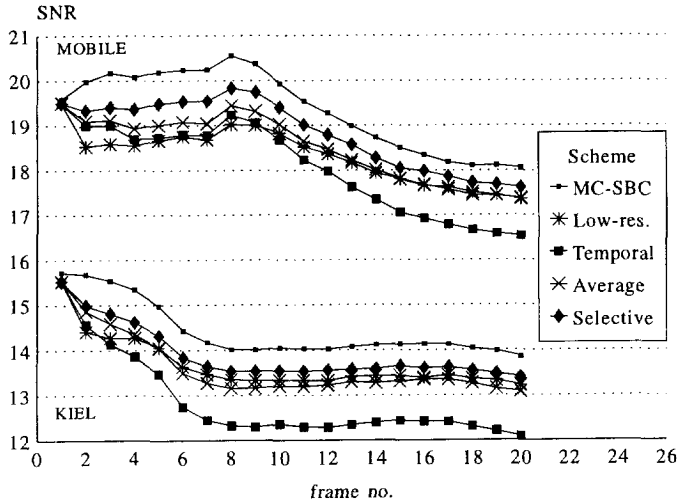


Figure 6.10: Coding performances for  $y^2$  using all subband prediction methods.

the non-scalable scheme based on in-band motion compensation (MC-SBC).

As illustrated, the selective scheme outperforms the other scalable schemes and approaches the performance of the non-scalable scheme to approximately 0.5 dB. The low-resolution and temporal schemes perform less well for most frames in both sequences. For the "Kiel" sequence the temporal scheme clearly has the lowest performance of all schemes. In particular, the performance gap is approximately 1.30 dB between the temporal scheme and the selective scheme and 1.00 dB between the temporal scheme and the other schemes. For the "Mobile" sequence, however, the performance gap is less and is even absent for frames 1 to 9. Hence, the simulcast approach (as implemented by the temporal scheme) may sometimes be a reasonable strategy to provide spatial scalability but mostly it will imply a reduced performance for  $y^2$ . Finally, the average scheme performs either slightly better than the low-resolution and temporal schemes or somewhere in between. Note that the performances of  $y^1$  for the low-resolution, temporal, average, and selective schemes are all identical because the difference prediction methods only affect the coding performance of  $y^2$ . The performance of  $y^1$  is indicated in Fig. 6.9 as the multistage scheme.

Based on Figure 6.10 it can be concluded that a spatially scalable coding scheme based on in-band motion compensation should implement the selective prediction method to minimize the performance loss with respect to a non-scalable scheme. By doing so, the performance of the scalable scheme approaches the performance of the non-scalable scheme up to 0.5 dB for the evaluation sequences. In addition, the selective prediction method offers a better performance than a pure simulcast approach. For the evaluation sequences the performance gain obtained is approximately 0.75 dB.



## 6.7 Discussion

In this chapter the structure and coding performance of a spatially scalable coding scheme based on interframe coding techniques were analyzed. In particular, a spatially scalable scheme was designed that integrates the hierarchical coding techniques for spatial scalability and the technique of in-band motion compensation. The scheme supports a number of prediction methods to predict the low-frequency subbands of the high-resolution signal. These prediction methods predict these subbands on the basis of low-resolution reference subbands and temporal reference subbands. Depending on the prediction method employed the scheme implements a hierarchical coding concept (low-resolution prediction), a simulcast coding concept (temporal prediction), or a concept in between (average and selective prediction). Only if the scheme employs the low-resolution prediction method, multirate coders can be used to encode the subbands in an amplitude-scalable way.

It can be concluded from the experimental results that an interframe coding scheme that implements either the low-resolution prediction method or the temporal prediction method can not provide a performance competitive with the performance of a similar non-scalable scheme. In fact, the performance gap varies between 0.75 dB and 2.00 dB. When the scheme employs the selective prediction method it can minimize the performance loss with respect to the non-scalable scheme. By doing so, the performance of the scalable scheme approaches the performance of the non-scalable scheme up to 0.5 dB for the evaluation sequences.

Based on the above observations, it can be concluded that for applications that require interframe coding, a fixed hierarchical coding concept is not always the best choice (with respect to the coding performance). In fact, coding concepts that switch dynamically between the different prediction options seem to be the most appropriate.



# Chapter 7

## Spatio-temporally scalable coding using 3-D SBC

### 7.1 Introduction

Chapters 5 and 6 included an extensive elaboration on spatially scalable coding schemes for compatible distribution purposes. These schemes were based on intraframe subband coding and on interframe subband coding based on motion estimation and compensation. In this chapter, the focus is directed towards spatio-temporally scalable coding schemes and the compatibility between progressive-scan video signals and interlaced video signals. In particular, it is investigated how well these schemes can support such compatibility based on three dimensional (3-D) subband coding.

Compatibility between progressive-scan video signals and interlaced video signals has become a major issue in recent years because of the rapidly growing variety of video and display standards [CCIR89, Demo92]. To obtain such compatibility, compatible encoding schemes are necessary that encode a progressive-scan signal in such a way that it can be reconstructed as an interlaced signal and vice-versa. Such compatible schemes have to employ a spatio-temporal decomposition of the frequency spectrum because the video signals have different supports in the 3-D frequency spectrum. Consider for example the video hierarchy that is tabulated in Table 7.1. This hierarchy contains, besides the HDP, EDP, and VT<sup>1</sup> signals, interlaced signals called HDI and EDI as well. These signals have the same spatial resolutions as HDP and EDP but have an interlaced scanning pattern.

For each signal in Table 7.1, Figure 7.1 shows the support to the 3-D frequency spectrum. Fig. 7.1a shows the intersection of the support with the spatial (vertical-horizontal) frequency plane, while Fig. 7.1b shows the intersection with the vertical-temporal frequency plane. As can be seen, HDP and HDI have the same support in the spatial frequency plane but have different supports in the vertical-temporal frequency plane. In particular, the progressive-scan HDP signal has a rectangular support while the interlaced HDI signal has a diamond shaped support. EDP and EDI are related in an identical way but have a reduced spatial support with respect to HDP and HDI. The VT signal has the smallest spatial and temporal

---

<sup>1</sup>Note that the VT signal has a 25 Hz frame rate in this chapter as opposed to the 50 Hz frame rate as used in Chapter 5.

Table 7.1: Hierarchy of video signals.

Video service	Frame rate	Scanning pattern	Aspect ratio	Active lines	Active samples per line	
					luminance Y	chrominance UV
HDP	50	1:1	16:9	1152	1440	720
HDI	50	1:2	16:9	1152	1440	720
EDP	50	1:1	16:9	576	720	360
EDI	50	1:2	16:9	576	720	360
VT	25	1:1	16:9	288	360	180

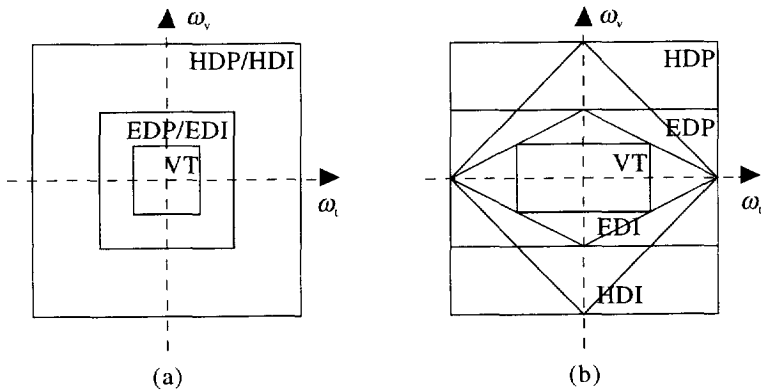


Figure 7.1: Three-dimensional frequency supports of video standards; a) spatial frequency plane, b) vertical-temporal frequency plane.

supports. From Fig. 7.1, it can be concluded that a non-rectangular spatio-temporal filtering and subsampling is required to obtain, for example, a HDI reproduction from an encoded HDP signal.

In recent years different authors have addressed the spatio-temporally scalable encoding of progressive-scan and interlaced video signals using subband decompositions [Vand91a, Peco90, Bree90, Vett90, Bosv91, Bosv92b, Guil92]. The solutions proposed differ primarily in the use of temporal filtering, because of the associated complexity and artifacts like blurring and halo images. The scalable schemes proposed in [Vand91a, Peco90] therefore avoid temporal filtering by converting the interlaced signals into equivalent progressive-scan signals before encoding and vice versa after decoding. Similarly, the schemes proposed in [Bree90, Guil92] avoid temporal filtering by decomposing interlaced signals on a field basis using specially designed vertical filter banks. The schemes proposed in [Vett90, Bosv91] do implement temporal filtering but try to limit the complexity and artifacts by using spatio-temporal polyphase filters [Vett84]. Recently in [Pei94] 3-D subband decompositions based on mathematical morphology are proposed that implement the filter banks of [Bosv91]. These subband decompositions are of moderate complexity and facilitate VLSI implementations.

In this chapter the spatio-temporally scalable schemes of [Bosv91, Bosv92b] are described and investigated. These two schemes support the video hierarchy of Tab. 7.1 and use dedicated 3-D subband decompositions to decompose the 3-D frequency spectrum. Section 7.2 starts out with an analysis of the hierarchy of video signals tabulated in Table 7.1. As will be shown, certain ambiguities exist in this hierarchy that make it impossible to obtain compatibility straightforwardly. Consequently, two strategies are defined that provide some form of compatibility: the Conversion Strategy and the Direct Strategy. A 3-D filter bank is designed for each strategy in Section 7.3 based on the polyphase FIR subband filters. Since the design of these filter banks is based on spatio-temporal filtering and decimation, relevant parts of multirate theory are briefly reviewed as well. Subsequently, the corresponding hierarchical encoding schemes are presented in Section 7.4. These schemes employ regular DPCM and PCM encoders and use a Selection Mechanism to control the bit rate of the layers and the distortion in the reproduced signals. Finally, in Section 7.5 the experimental evaluation of the 3-D subband decompositions and the encoding schemes is described.

## 7.2 Analysis and strategy definitions

### 7.2.1 Analysis

The hierarchy of video signals in Table 7.1 can be separated into a hierarchy in horizontal-vertical plane, and a hierarchy in vertical-temporal plane, as illustrated in Fig. 7.1. Henceforth, these hierarchies are referred to as the spatial hierarchy and the temporal hierarchy, respectively.

The spatial hierarchy is similar to the spatial hierarchies for progressive-scan signals that have been analyzed and used in the previous chapters. In those chapters, it has been illustrated how such a hierarchy can be supported by decomposing the HDP signal into a number of spatial subbands. The VT, EDP, and HDP signals are then reproduced by using some or all subbands in the reconstruction process. This property is actually a fundamental property of the defined signals: i.e. there is no ambiguity in ordering the spatial frequency components or subbands. Denoting spatio-temporal frequencies by  $\omega = (\omega_h, \omega_v, \omega_t)$ , and by indicating the spatial spectrum (i.e.  $\omega_t = 0$ ) of a video signal by the subscript "s", this relation between the spatial spectra of the video standards can be denoted as:

$$VT_s(\omega) \subset EDP_s(\omega) \subset HDP_s(\omega). \quad (7.1)$$

A direct result of this relation is that an increase in the spatial resolution always implies the use of one or more additional subbands in the synthesis process. In addition, the spatial lattices associated with the standards of the spatial hierarchy exhibit a similar relation. Denoting the spatial-temporal coordinates by  $\mathbf{x} = (x_h, x_v, x_t)$ , the following relation between the lattices holds:

$$VT_s(\mathbf{x}) \subset EDP_s(\mathbf{x}) \subset HDP_s(\mathbf{x}). \quad (7.2)$$

This relation states that the samples of the EDP signal coincide with some samples of the HDP (see Fig. 7.2). Since these samples can be retained in a subband decomposition, the

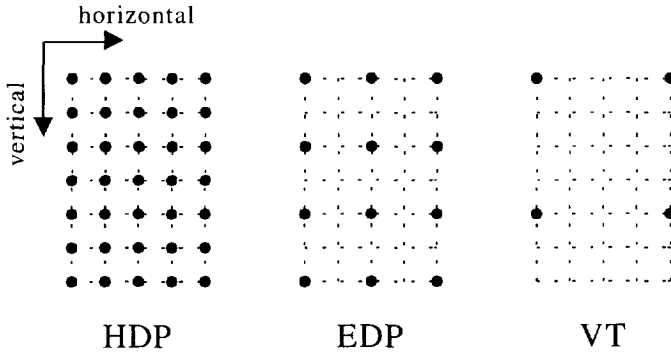


Figure 7.2: Relation between spatial sampling lattices of HDP, EDP, and VT.

hierarchical relation between EDP and HDP is non-ambiguous in both the spectral and sampling domains. Because VT and EDP have the same relation, the entire spatial hierarchy is non-ambiguous.

In contrast with the spatial hierarchy, the temporal hierarchy is more complicated. In fact, in the following it will be shown that the temporal hierarchy is ambiguous in both the spectral and sampling domains. In the spectral domain there is an ambiguity between the supports of HDI and EDP (see Fig. 7.1b). Because the vertical-temporal spectra of HDI and EDP partially overlap each other, neither of the two can be extracted from the other. More formally stated:

$$\begin{aligned} \text{HDI}_t(\omega) &\not\subset \text{EDP}_t(\omega), \quad \text{and} \\ \text{EDP}_t(\omega) &\not\subset \text{HDI}_t(\omega), \end{aligned} \tag{7.3}$$

where the subscript “t” refers to the vertical-temporal spectrum ( $\omega_h = 0$ ). As a consequence, it is not possible to create a regular subband decomposition structure that facilitates the reconstruction of HDI and EDP.

An additional complication in the temporal hierarchy is that vertical-temporal sampling lattices of lower-resolution signals are not all subsets of the sampling lattices of higher-resolution signals. This is, for instance, the case for HDP, HDI, and EDI. The HDI sampling lattice is obtained by a quincunx subsampling of the HDP sampling lattice in the vertical-temporal direction, as illustrated in Figure 7.3. However, the sampling lattice of EDI cannot be obtained from the sampling lattice of HDI by straightforward subsampling. More formally:

$$\text{EDI}_t(\mathbf{x}) \not\subset \text{HDI}_t(\mathbf{x}) \subset \text{HDP}_t(\mathbf{x}). \tag{7.4}$$

Thus, although the EDI spectrum is a subset of the HDI spectrum, simple low-pass filtering and subsampling is not sufficient. To transform the HDI lattice into the EDI lattice (or vice versa), a more complicated vertical-temporal sampling grid conversion is required.

Based on the above observations, it can be concluded that a straightforward decomposition of the spatio-temporal frequency spectrum to support compatibility between all video signals

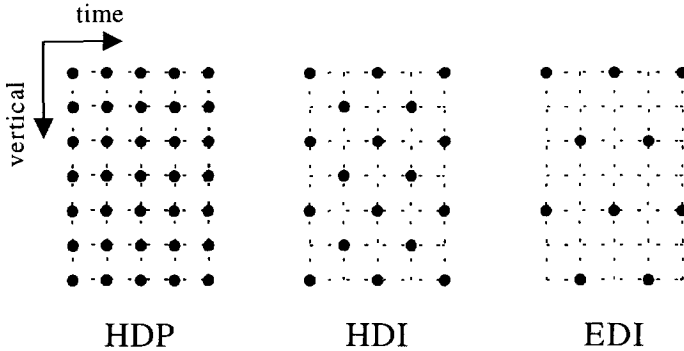


Figure 7.3: Relation between vertical-temporal sampling lattices of HDP, HDI, and EDI.

of Table 7.1 is not possible. Instead, dedicated strategies have to be defined that facilitate compatibility between some signals of Table 7.1. The other video signals of Table 7.1 should then be converted to the supported signals using a sample grid conversion. The following two sections present a discussion of two such strategies: the Conversion Strategy (Section 7.2.2) and the Direct Strategy (Section 7.2.3).

### 7.2.2 Conversion Strategy

One possible way to provide compatibility between the video signals of Table 7.1 is to enforce a certain regularity in the spatio-temporal hierarchy. Such regularity can be enforced by supporting only the progressive-scan signals. The decomposition of the HDP signal would then create a set of subbands from which the VT, EDP, and HDP supports can be reconstructed. The EDI and HDI signals have to be upsampled to EDP and HDP sampling lattices before decomposition, encoding, and transmission (i.e. interlace to progressive-scan conversion). After reception, decoding and reconstruction, these signals are downsampled to the HDI and EDI sampling grids (i.e. progressive-scan to interlace conversion). Figure 7.4 depicts this strategy, which is henceforth called the Conversion Strategy. Note that the upsampling of the interlaced standards refers to the sampling lattices only, and does not involve the recovery of frequency components not present in interlaced signals.

The advantage of the Conversion Strategy is that it has only a moderate complexity and can be realized as a straightforward extension of the spatially scalable schemes. In particular, the only additional requirement is a decomposition along the temporal axis in order to support the VT signal with half the frame rate. Another advantage of the Conversion Strategy is that it can easily be expanded to include other progressive-scan video standards such as, for example, super HDTV [Furu93]. A drawback is, however, the potential loss of efficiency in data compression for interlaced signals; it is not guaranteed that a subsequent encoding stage is able to benefit from the 'empty' vertical-temporal frequency regions in up-converted interlaced signals. In addition, the upsampling and downsampling procedures may also introduce some distortion in the reconstructed signals, independent of the encoding stage.

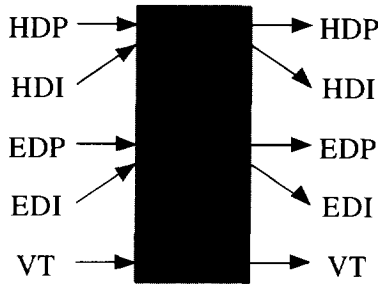


Figure 7.4: Compatibilities and required conversions for Conversion Strategy.

### 7.2.3 Direct Strategy

Another way to alleviate the ambiguity in the spatio-temporal hierarchy is to (initially) exclude either the HDI or EDP standard from the hierarchy. This exclusion provides two regular but incomplete vertical-temporal hierarchies:

$$VT_t(\omega) \subset EDI_t(\omega) \subset EDP_t(\omega) \subset HDP_t(\omega), \quad \text{and} \quad (7.5)$$

$$VT_t(\omega) \subset EDI_t(\omega) \subset HDI_t(\omega) \subset HDP_t(\omega). \quad (7.6)$$

The first hierarchy (Eq. 7.5) is preferable to the second hierarchy (Eq. 7.6) because of the ambiguity in the spatio-temporal sampling lattices of HDI and EDI (see Fig. 7.3). With the first hierarchy, a set of subbands can be defined from which the VT, EDI, EDP, and HDP signals can directly be reconstructed. For this reason, this strategy is henceforth called the 'Direct Strategy'.

One way to facilitate the encoding of the HDI signal is to employ the same approach as with the Conversion Strategy, i.e. i) use upsampling before encoding, ii) ignore the specific frequency support of the HDI signal during the encoding, and iii) use downsampling after decoding. However, this approach has the same potential efficiency problem as with the Conversion Strategy. The Direct Strategy therefore uses a slightly different approach to support HDI, i.e. after the HDI signal is upsampled, the subband decomposition decomposes the spatio-temporal spectrum such that the frequency support of the HDI signal can be recovered exactly. This can be achieved by further decomposing some of the subbands that would suffice for the hierarchy of Eq. 7.5. As a consequence, the subsequent coding scheme can allocate the bits for the HDI signal to the associated subbands and can thus circumvent a loss in coding efficiency.

Figure 7.5 shows the compatibilities and conversions obtained for the 'Direct Strategy'. The VT, EDI, EDP, and HDP signals can be directly inserted in the subband decomposition and can also be directly reconstructed from the decomposition. The HDI signal requires an upsampling of the signal to a HDP lattice before encoding and a downsampling after decoding.



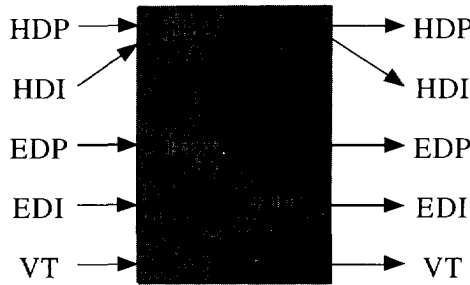


Figure 7.5: Compatibilities and required conversions for Direct Strategy.

### 7.3 Spatio-temporal filter banks

#### 7.3.1 Non-rectangular filter banks

Compatibility between progressive-scan and interlaced signals requires non-rectangular sub-band decompositions and sampling lattices. Therefore the theory and design of subband splitting blocks based on 2-D QMF filters are reviewed concisely in this section. These subband splitting blocks will be the basic building blocks for the spatio-temporal filter banks for the Conversion Strategy and Direct Strategy that will be developed in the following two sections.

The process of decomposing a signal into subbands using a 2-D filter bank is concisely described using multirate theory [Vett84, Ansa88, Bamb90]. As an example, Figure 7.6 shows a basic, critically sampled, 2-band filter bank. It consists of a pair of analysis filters  $H_0(\omega)$  and  $H_1(\omega)$ , a pair of synthesis filters  $G_0(\omega)$  and  $G_1(\omega)$ , and downsampling and upsampling operators  $\downarrow \mu$  and  $\uparrow \mu$  with  $\omega = [\omega_t, \omega_v]^t$  and  $\det(\mu)=1$  (critical sampling). The analysis and synthesis filters serve as anti-aliasing and interpolation filters for the downsamplers and interpolators, respectively. The downsampling operation with input  $x[\mathbf{n}]$  and output  $x_d[\mathbf{n}]$  is defined as [Ansa88]

$$x_d[\mathbf{n}] = x[\mu\mathbf{n}], \tag{7.7}$$

where  $\mathbf{n} = [n_t, n_v]^t$ . The upsampling operation with input  $x[\mathbf{n}]$  and output  $x_u[\mathbf{n}]$  is defined as [Ansa88]

$$x_u[\mathbf{n}] = \begin{cases} x[\mu^{-1}\mathbf{n}], & \mu^{-1}\mathbf{n} \in \mathcal{Z} \\ 0, & \text{otherwise.} \end{cases} \tag{7.8}$$

The analysis and synthesis filters are frequently chosen as modulated versions of a single prototype filter  $H_0(\omega)$ , which simplifies the design of the filters. In particular, the analysis and synthesis filters are chosen as:

$$\begin{aligned} H_1(\omega) &= H_0(\omega - \mu^{-t}\mathbf{k}_1 2\pi), \\ G_0(\omega) &= 2H_0(\omega), \\ G_1(\omega) &= -2H_0(\omega - \mu^{-t}\mathbf{k}_1 2\pi), \end{aligned} \tag{7.9}$$

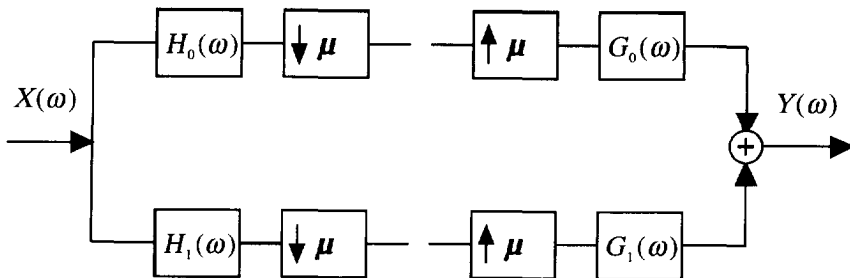


Figure 7.6: Basic 2-band analysis/synthesis bank.

where  $\mathbf{k}_1$  denotes the so-called coset vector; i.e. the vector between the lattice of pixels retained after subsampling and the lattice of the discarded pixels [Ansa88]. Due to the above choice, all aliasing resulting from the non-ideal filter  $H_0(\omega)$  is cancelled in the overall transfer function  $T(\omega)$  of the 2-D filter bank:

$$T(\omega) = H_0(\omega)^2 - H_0(\omega - \mu^{-t}\mathbf{k}_1 2\pi)^2. \tag{7.10}$$

The complete analysis/synthesis bank is now specified except for the filter  $H_0(\omega)$  and the downsampling matrix  $\mu$ .  $H_0(\omega)$  should be chosen in accordance with  $\mu$  because  $H_0(\omega)$  serves as the anti-aliasing filter for the downsampling operation. The relation between the input  $x[\mathbf{n}]$  and output  $x_d[\mathbf{n}]$  of the downsampling operation in the Fourier domain is given by

$$X_d(\omega) = \frac{1}{2} \left\{ X(\mu^{-t}\omega) + X(\mu^{-t}(\omega - \mathbf{k}_1 2\pi)) \right\}. \tag{7.11}$$

The second term of Eq. 7.11 constitutes the aliased signal component, which should be suppressed by  $H_0(\omega)$ . To achieve this  $H_0(\omega)$  should have a nominal passband, which is described by [Bamb90]

$$\{-\pi \leq \mu_{11}\omega_t + \mu_{21}\omega_v \leq \pi\} \cap \{-\pi \leq \mu_{12}\omega_t + \mu_{22}\omega_v \leq \pi\}. \tag{7.12}$$

In practice the prototype filter  $H_0(\omega)$  is not perfect, so  $X_d(\omega)$  always contains some aliasing. Research has shown that a two-dimensional, separable filter based on a 1-D FIR QMF filter is a fairly good choice for  $H_0(\omega)$  [Bamb90]. The construction of  $H_0(\omega)$  from the 1-D QMF filter involves two steps: i) the creation of a separable 2-D filter with a checkerboard frequency response, and ii) the rotation and skewing of the 2-D filter so that the frequency response complies with Eq.7.12. The other filters (i.e.  $H_1(\omega)$ ,  $G_0(\omega)$ , and  $G_1(\omega)$ ) are subsequently derived from  $H_0(\omega)$  according to Eq. 7.9. The filters obtained are polyphase and allow for a very efficient implementation. This reduces the complexity and memory requirements of the resulting filter bank considerably.

In the following two sections spatio-temporal filter banks will be designed for the Conversion Strategy and Direct Strategy. These vertical-temporal subband decompositions are based on a small set of basic filter banks (or splitting blocks) that have different filters and subsampling matrices. Figure 7.7 illustrates this set of basic splitting blocks by showing for each block the (ideal) shape of the  $H(\omega)$  filter together with the associated subsampling

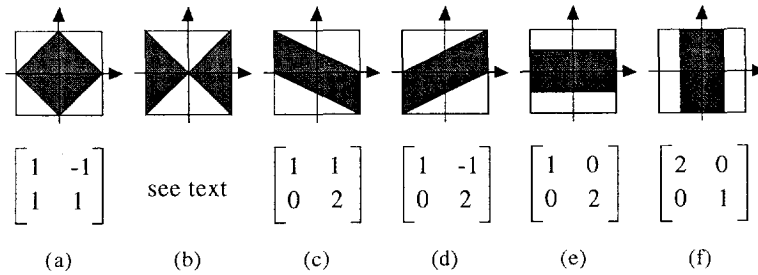


Figure 7.7: Basic 2-D splitting blocks: Filter responses and subsampling matrices.

matrix. The fan filter in Fig. 7.7b is implemented by first pre-modulating the input signal by  $\pi$  in one of the dimensions, and next applying the diamond-shaped filter of Fig. 7.7a.

### 7.3.2 Filter bank for Conversion Strategy

Figure 7.8a shows the minimal analysis filter bank (vertical-temporal plane) for the Conversion Strategy. The filter bank decomposes the vertical-temporal frequency spectrum into 4 subbands using 2 basic splitting blocks. The filter bank derives the spectra of the EDP and VT signals from the HDP signal by a double vertical decimation ( $\mu_{00}$  and  $\mu_{10}$ ) followed by a single temporal decimation ( $\mu_{20}$ ). The low-pass and high-pass filters applied are shown in front of the decimators to which they correspond. Consequently, the vertical-temporal spectrum is decomposed into 3 regions, as illustrated in Figure 7.8b. Subband 4 contains the vertical-temporal information of the VT signal. Subbands 2 and 3 are also required to reconstruct the EDP signal, while for the HDP signal subband 1 also needs to be incorporated in the subband synthesis. Table 7.2 tabulates which subbands are used in the reconstruction process for a particular video signal.

It should be noted that the vertical-temporal filter bank has to be extended with a 1-D filter bank in the horizontal direction to achieve complete compatibility between the HDP, EDP, and VT signals. The additional filter bank should decompose the horizontal frequencies into minimally 4 subbands. Further, it is emphasized that Fig. 7.8 shows only the minimal vertical-temporal decomposition that is necessary for compatibility purposes. In practice, additional vertical-temporal decompositions or a different ordering may be used to obtain a higher coding efficiency. For instance, if the first (vertical) splitting block and the last (temporal) splitting block are exchanged, six vertical-temporal subbands are obtained, while the temporal compatibility is fully preserved.

### 7.3.3 Filter bank for Direct Strategy

Figure 7.9a shows the minimal analysis filter bank (vertical-temporal plane) for the Direct Strategy. The filter bank decomposes the vertical-temporal frequency spectrum into 9 subbands using 8 basic splitting blocks. First, the filter bank decomposes the HDP signal in

Table 7.2: Relation between video signals and subbands for the Conversion Strategy.

Video Signal	Subband			
	4	3	2	1
HDP	✓	✓	✓	✓
HDI	✓	✓	✓	✓
EDP	✓	✓	✓	·
EDI	✓	✓	✓	·
VT	✓	·	·	·

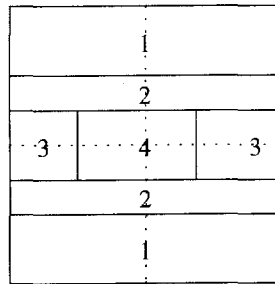
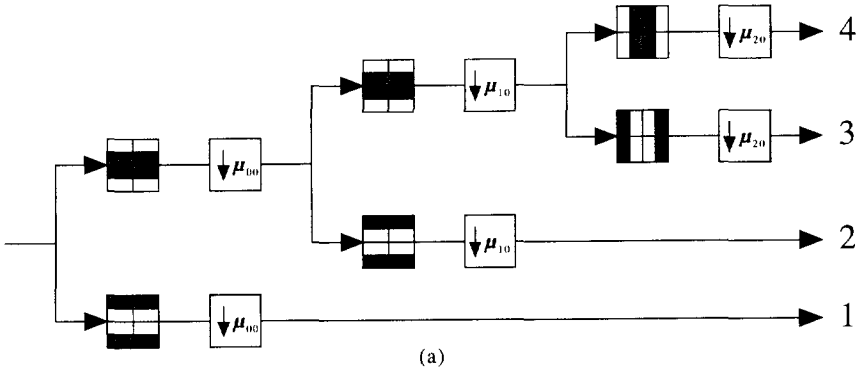


Figure 7.8: Implementation of Conversion Strategy; a) minimal vertical-temporal filter bank; b) resulting decomposition of vertical-temporal spectrum.

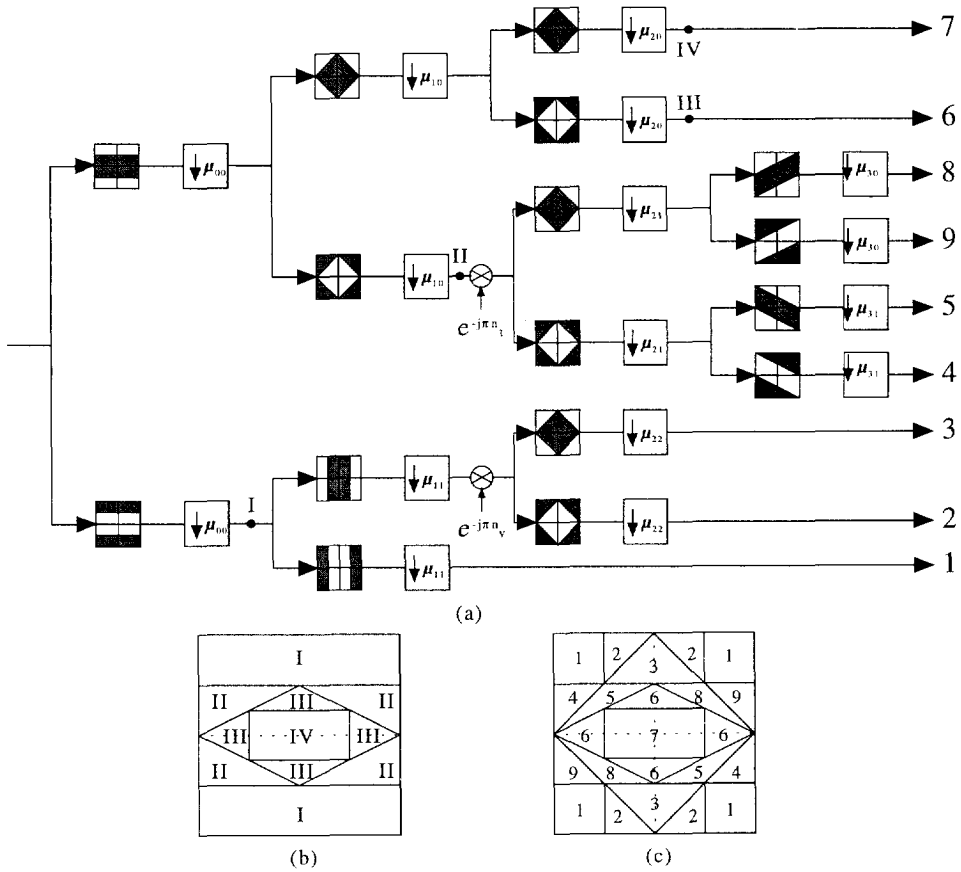


Figure 7.9: Implementation of Direct Strategy; a) minimal vertical-temporal analysis filter bank; b) intermediate stage of vertical-temporal decomposition required for compatibility HDP-EDP-EDI-VT; c) final decomposition of vertical-temporal spectrum.

Table 7.3: Relation between video signals and subbands for the Direct Strategy.

Video Signal	Subband								
	7	6	5	8	4	9	3	2	1
HDP	✓	✓	✓	✓	✓	✓	✓	✓	✓
HDI	✓	✓	✓	✓	·	·	✓	·	·
EDP	✓	✓	✓	✓	✓	✓	·	·	·
EDI	✓	✓	·	·	·	·	·	·	·
VT	✓	·	·	·	·	·	·	·	·

such a way that the hierarchy HDP–EDP–EDI–VT is supported. This requires the filter and decimation operations denoted by  $\mu_{00}$  (vertical decimation), and  $\mu_{10}$  and  $\mu_{20}$  (quincunx decimations). These operations decompose the vertical-temporal frequency spectrum plane into 4 regions as illustrated in Figure 7.9b. The supports of the HDP, EDP, EDI and VT spectra can directly be recognized from this figure.

Next, several vertical-temporal subbands are further decomposed to allow for the incorporation of the HDI standard. Essentially, the filtering and subsampling operations denoted by  $\mu_{11}$  (temporal decimation),  $\mu_{21}$  and  $\mu_{22}$  (quincunx decimations), and  $\mu_{31}$  and  $\mu_{32}$  (directional decimations) are necessary to be able to exactly reconstruct the HDI spectrum from a given HDP spectrum. Fig. 7.9c shows the final decomposition of the vertical-temporal frequency plane. The frequency support of the HDI is clearly recognized in this decomposition; it consists of subbands 7, 6, 5, 8 and 3. Table 7.3 tabulates which subbands are used in the reconstruction processes for the other signals as well.

To exemplify the use of the basic splitting blocks with directional filters, it is now illustrated how vertical-temporal subband 8 is extracted from the HDP spectrum. Figure 7.10 shows the subsequent filter stages that are used to create subband 8. The basis vectors of the input lattices are shown in the lower-left corner for each filter stage, while the basis vectors of the output lattices are shown in the lower-right corner. Fig. 7.10a shows the HDP spectrum with the low-pass frequency shaded to indicate the low-pass filter operation of  $\mu_{00}$ . After decimation, the low-pass frequencies are decomposed further using a diamond-shaped high-pass filter that belongs to  $\mu_{10}$  (Fig. 7.10b). The high-pass frequency band contains, besides subband 8, also subbands 9, 5, and 4. Therefore, after the subband is decimated, the resulting spectrum consists only of frequencies belonging to these subbands (Fig. 7.10c). Subsequently, subbands 8 and 9 are simultaneously extracted by i) shifting the spectrum by  $\pi$  in the temporal direction (i.e multiplying the signal by  $1, -1, 1, -1, \dots$ ) and ii) applying the diamond-shaped low-pass filter of  $\mu_{21}$ . Finally, after decimation, subband 8 is extracted by using the directional low-pass filter of  $\mu_{30}$  and a final decimation.

As mentioned in Section 7.2.3, the HDI signal is upsampled to the HDP lattice before decomposition and encoding. This upsampling does not affect the data compression efficiency because the subsequent coding scheme can exclude subbands 1, 2, 4, and 9 from the HDI encoding. Again, it should be noted that Fig. 7.9c shows only the minimal vertical-temporal filter bank. In practice, this filter bank has to be extended with a 1-D filter bank in the horizontal direction to achieve complete compatibility between the HDP, EDP, EDI, and

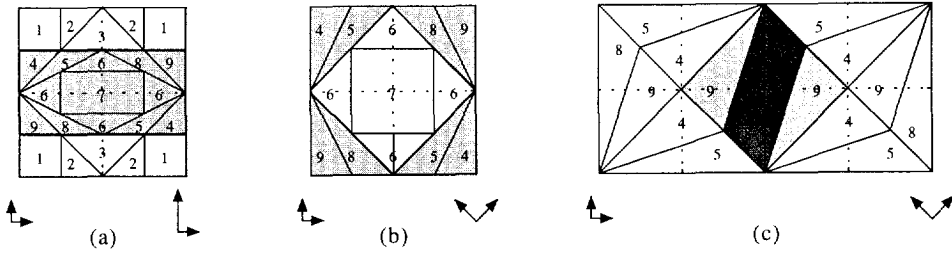


Figure 7.10: Consecutive filter stages required to create vertical-temporal subband 8.

VT signals. Similarly, additional vertical-temporal decompositions may be used to obtain a higher coding efficiency.

### 7.4 Spatio-temporal coding schemes

In this section two hierarchical coding schemes are presented that implement the Conversion Strategy and the Direct Strategy. The schemes have identical structures and differ only in the filter banks applied. The schemes employ regular DPCM and PCM encoders, a temporal bit allocation procedure, and a Selection Mechanism to control the bit rate of the layers and the distortion in the reproduced signals.

The coding scheme that implements the Conversion Strategy creates 5 layers  $l^n$  ( $1 \leq n \leq 5$ ). Decoders can reproduce the signal at VT, EDI, EDP, HDI, and HDP resolutions (and qualities) by receiving and decoding up to 5 layers. Each layer contains encoded subbands that are relevant for the reproduced signal and that fit into the available bit rate for that layer.

The coding scheme that implements the Direct Strategy creates 6 layers  $l^n$  ( $1 \leq n \leq 6$ ). The reason for the extra layer (6 layers for 5 reproductions) is the ambiguity in ordering the subbands for the EDP and HDI signals. Decoders can reproduce the encoded signal at VT and EDI resolutions by receiving and decoding ( $l^1$ ) and ( $l^1, l^2$ ), respectively. To reconstruct the encoded signal at EDP and HDI resolutions, receivers need ( $l^1, l^2, l^3, l^4$ ) and ( $l^1, l^2, l^3, l^5$ ), respectively. Layer  $l^3$  carries subbands 5 and 8, which are common to EDP and HDI. Layers  $l^4$  and  $l^5$  carry the subbands that are specific to EDP and HDI, respectively. To reproduce the signal at HDP resolution, the decoders need all layers.

#### 7.4.1 Encoder structure

Figure 7.11 illustrates the common structure of the spatio-temporal encoding schemes. These schemes consist of a 3-D subband decomposition (SBD), a bit allocation procedure (BA), subband quantization (Q) and variable length coding (VLC), and the Selection Mechanism (SEL). The subband decompositions include the previously designed minimal vertical-temporal decompositions, the required horizontal decomposition, and additional decomposition stages to increase the coding efficiency.

After the subband decomposition, the 3-D subbands are encoded. The schemes encode the 3-D subbands in non-overlapping time segments to limit the complexity, encoding delay,

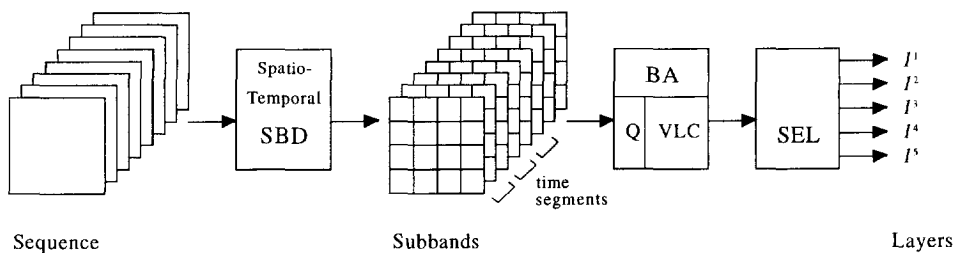


Figure 7.11: Common structure of spatio-temporal encoding schemes.

and memory requirements of the encoder. Each segment consists of a number of subbands that originates from a number of consecutive frames. The schemes choose the size of the time segment such that it can be decomposed into the required spatio-temporal subbands with a minimal temporal size of 1 frame. It will be shown that the minimal segment lengths for the Conversion Strategy and Direct Strategy are 2 and 4 frames, respectively. The segment-wise data compression of the HDP signal is independent of both the structure of the spatio-temporal filter bank and the length of the applied filters. Hence, although the encoding process operates on the basis of temporal segments, truly 3-D subband decompositions are employed.

Within each segment, the schemes use spatial DPCM to encode the subbands with the lowest spatio-temporal frequencies. Similar to spatial subband decompositions, these subbands exhibit a considerable amount of spatial correlation. The schemes employ PCM to encode the other spatio-temporal subbands. The probability density function (PDF) of the prediction error of the low-frequencies subband is fairly well modelled by a Generalized Gaussian (GG) function with a shape-parameter  $c = 0.75$ , while the other subbands are modelled by a GG-PDF with  $c = 0.50$  [Bosv91]. Each subband is quantized by a scalar Lloyd-Max quantizer, followed by a Huffman code.

The distribution of the total number of bits per segment among the subbands is done by the bit allocation procedure. By definition, this means that the HDP signal is optimally encoded. The bit allocation procedure is based on the bit allocation algorithm as described in Chapter 2. Because some subbands in the segments have a temporal size larger than one frame, the bit allocation procedure is temporally adaptive (to a certain extent).

After the subbands within a particular segment are encoded the Selection Mechanism determines which subbands are conveyed in the layers. The Selection Mechanism selects which subbands are transmitted within which layers. It first ranks the encoded subbands according to a particular measure, and then selects for each layer the relevant subbands until the available bit rate for that layer is totally used.

Decoders reproduces the input signal at the desired resolution by reconstructing the encoded time segments on the basis of the relevant layers. Subbands that are not present in a layer, but which are part of the spectrum of the signal, are assumed zero in the subband synthesis process.



### 7.4.2 Temporal bit allocation procedure

The bit allocation procedure operates on the basis of time segments for both schemes. For each time segment, it distributes the available bits among the subband data of that time segment. The minimum sizes of the time-segments are now resolved for both strategies.

With the Conversion Strategy, there is only one temporal decimation involved, namely,  $\mu_{20}$ . Hence, for every two frames that are decomposed, two temporal subband samples become available, i.e. samples of subbands 3 and 4. This implies that each consecutive pair of HDP frames results in one complete set of spatio-temporal subbands after decomposition. These subbands can be encoded under supervision of the bit allocation procedure. Thus the minimal length of a time segment is 2 frames. The temporal size of vertical-temporal subbands 1 and 2 is two frames because these subbands are not temporally decimated. Although it is possible to take into account the 3-D character of these bands in the subsequent data compression, experimental evidence shows that the potential gain in data compression does not compensate for the loss in adaptivity in the temporal direction. Therefore, the two temporal components of subbands 1 and 2 are considered separate frames in the bit allocation and subsequent quantization processes.

The situation is more complicated with the Direct Strategy. Subbands 4, 5, 8, and 9 are subject to three temporal decimations, namely,  $\mu_{10}$ ,  $\mu_{21}$ , and  $\mu_{30}$ . However, as a concatenation of two vertical-temporal quincunx decimations results in a vertical-temporal decimation by a factor of two, a complete set of spatio-temporal subbands is obtained after every four frames. Thus the scheme that implements the Direct Strategy has to operate on time segments that consist of minimally 4 frames. With this segment size, the temporal sizes of the vertical-temporal subbands 2, 3, 4, 5, 8, and 9 are exactly equal to one frame. Subbands 1, 6, and 7 each have a temporal size of two frames.

### 7.4.3 Selection Mechanism

The Selection Mechanism selects which subbands are transmitted within the layers. By doing this, it controls the bit rate of the layers and the distortion in the reproduced signals. In general, the bit rates assigned to the individual subbands are not known beforehand. As a result, the total bit rate that is assigned to the subbands relevant for a particular reproduction is also unknown. Usually, the total assigned bit rate will exceed the available channel capacity that is allocated for that reproduction. As a consequence, only a subset of the relevant subbands can be transmitted in the layers that support this reproduction.

The Selection Mechanism selects subbands based on a measure that includes the trade-off between the bit rate of a layer and MSE distortion in the reproduction [West89, Bosv90]. The MSE distortion of a certain time segment is approximately equal to the summation of all MSE distortions of the associated subband data. If the  $m^{th}$  subband is not selected, the resulting MSE equals the variance  $\sigma_{x,m}^2$  of that particular subband. If subband  $m$  is selected, using  $r_m$  bits of the available bit rate, the distortion equals the quantization error variance  $\sigma_{q,m}^2$  of that subband. The relative decrease in MSE is thus given by:

$$z_m = \frac{\sigma_{x,m}^2 - \sigma_{q,m}^2}{r_m} \quad (7.13)$$

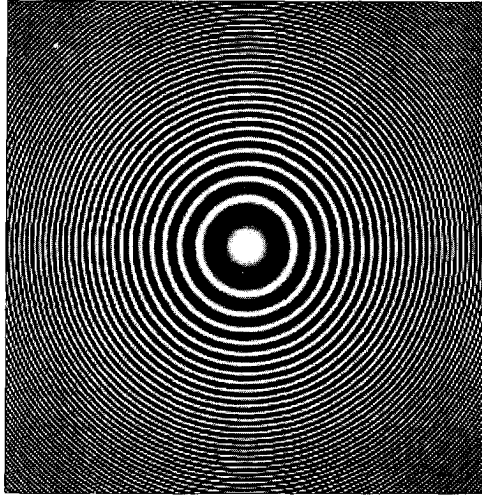


Figure 7.12: Input test image : 2-dimensional frequency sweep.

For a particular reproduction, the Selection Mechanism successively selects those subbands with the largest values of measure  $z_m$ , until the sum of the bit rates of the selected subbands exceeds the channel capacity assigned to that reproduction.

## 7.5 Experimental results

### 7.5.1 Filter bank performances

The performances of the proposed vertical-temporal filter banks of Section 7.5.1 are evaluated in this section. To make the visualization of results easier, the subsequent experiments employ the filter banks in the vertical-horizontal plane. In addition, a 2-dimensional frequency sweep is used as a (HDP) test image. This image is illustrated in Figure 7.12 and contains all 2-dimensional frequencies within the sampling bounds. This image is also used to obtain the other video signals of Table 7.1. These signals are obtained by band-limiting and downsampling the 2-D frequency sweep. In this section, the 2-D filters of the filter banks are all based on the QMF16c filter [John80] unless stated otherwise.

An important property of the proposed filter banks is the frequency discrimination. This property is important because the main difference between the two filter banks lies in the ability to handle the differences in frequency support of HDP and HDI. The frequency discrimination is evaluated by comparing the variances of the subbands obtained for respectively the HDP signal and the HDI signal. Tables 7.4 and 7.5 tabulate the variances obtained for both signals. For the Conversion Strategy, Table 7.4 shows that the reduced frequency support of HDI leads to a rather unlocalized reduction of variance; i.e. the variances of subbands 3, 2, and 1 are all reduced. For the Direct Strategy, Table 7.5 shows that the reduced frequency

Table 7.4: Subband variances for Conversion Strategy.

Standard	Subband			
	4	3	2	1
HDP	1576.6	963.9	2526.3	3114.1
HDI	1576.0	844.2	1932.6	1335.8

Table 7.5: Subband variances for Direct Strategy.

Signal	Subband								
	7	6	8	9	5	4	3	2	1
HDP	1571.6	1400.6	689.6	326.2	717.3	332.0	1351.9	563.4	1184.9
HDI	1566.3	1374.0	559.6	131.1	572.5	154.5	1127.5	175.9	61.5

support of HDI leads to a more localized reduction of the variances. In particular, only the variances of subbands 9, 4, 2, and 1 are significantly reduced. Hence, from this experiment it can be concluded that the 'Direct Strategy' filter bank is indeed better tailored towards the employment of the empty regions in the frequency support of HDI by a subsequent coding stage.

As an illustration, Figures 7.13 and 7.14 show the dimensions of the subbands as well as the subband data actually obtained for the HDP signal for both strategies. Because of the characteristics of the test signal, the frequency supports of the subbands are reflected inside the subbands themselves.

Another important property of the proposed filter banks is the reconstruction quality of the signals in the absence of coding. This quality is an upper bound for any subsequent coding scheme. Table 7.6 shows the SNR values obtained. The tabulated SNR values exclude the image boundaries (8-pixel guard band) to avoid potential boundary effects because of the FIR filtering. For the 'Conversion Strategy' filter bank, the performance for the interlaced signals HDI and EDI is approximately 10 dB less than the performances obtained for the progressive-scan HDP and EDP signals. This is due to the necessary upsampling and downsampling processes, which lead to some additional distortion. For the 'Direct Strategy' filter bank, the SNR obtained decreases when the resolution of the EDI, EDP, and HDP signals increases. This is due to the fact that more near-perfect reconstruction splitting blocks are involved in the analysis/synthesis tree, which increases the distortion in the overall transfer function.

The obtained performance of the 'Direct Strategy' filter bank for the HDI signal is substantially lower than the other values and needs further explication. As explained in Section 7.3 the HDI signal is reconstructed using only subbands 7,6,8,5, and 3. This implies that the transfer functions of the splitting blocks  $\mu_{30}$  and  $\mu_{31}$  degenerate to

$$T(\omega) = H_0(\omega)^2, \quad (7.14)$$

since the high-frequency subbands are missing. As the transfer function of Eq. 7.10 is not

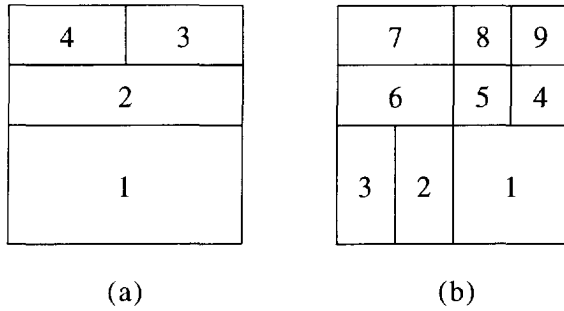


Figure 7.13: Subband positions for a) Conversion Strategy and b) Direct Strategy.

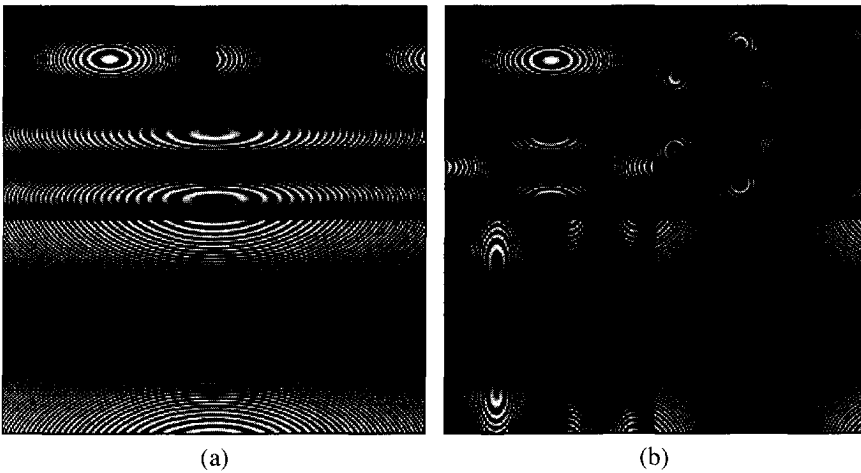


Figure 7.14: Subband data obtained for the a) Conversion Strategy and b) Direct Strategy.

Table 7.6: Performance of the vertical-temporal analysis/synthesis banks for 2-D frequency sweep (in dB, without quantization).

Standard	Conversion Strategy	Direct Strategy
HDP	43.0	37.0
HDI	35.5	13.6
EDP	44.7	37.4
EDI	33.9	43.6

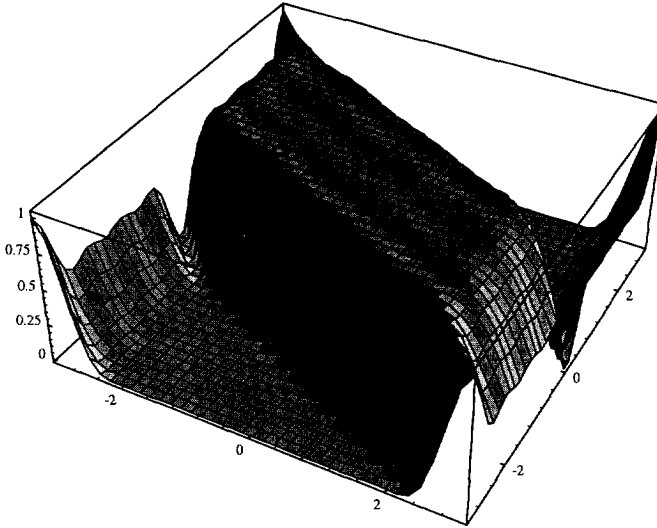


Figure 7.15: Frequency response of the 2D-band diagonal filter.

an ideal 2-D low-pass filter, the 2-D frequencies around the boundaries of the ideal low-pass filter will be distorted. To illustrate this, Figure 7.15 shows the shape of the low-pass filter associated with  $\mu_{31}$ . Normally, these distorted frequencies are not a problem because these frequencies are situated near the boundary of the frequency support of the reconstructed signal. However, because of the two-dimensional nature of these filters, some of these distorted frequencies eventually end up in the middle of the frequency support of the HDI signal and degrade the performance significantly. In particular, for HDI, subbands 9 and 4 are not used in the reconstruction process, so the frequencies at the boundaries of subbands 8 and 5 become distorted. In the reconstruction tree, some of these distorted frequencies are transformed to frequency locations between subbands 3 and 8 and subbands 3 and 5. This transformation can easily be recognized by interpreting Fig. 7.10 from right to left. Figure 7.16 illustrates these distorted frequencies by showing the absolute error between the reconstructed HDI frequency support and the 2-D frequency sweep. The black and white areas indicate small and large errors, respectively. Clearly, two lines of distorted frequencies are visible inside the HDI frequency support.

Although the above results seem to limit the usefulness of the Direct Strategy, it should be remarked that these distortions are dependent on the content of the signal. Because for natural sequences the involved frequencies are not the most important ones, the distortions may be acceptable.

### 7.5.2 Encoding results

In this section, the coding performances of the proposed coding schemes for the HDP and HDI signals are evaluated. The schemes use complete 3-D filter banks, which consist of the

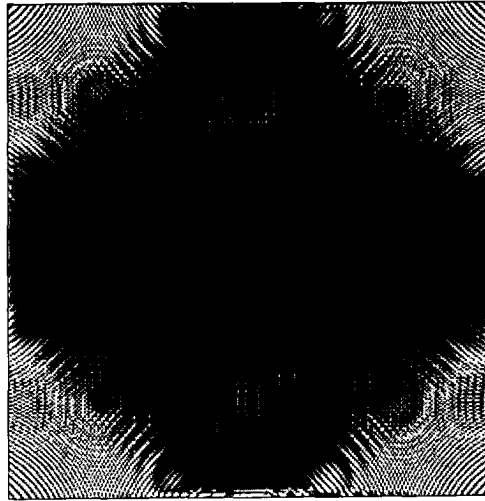


Figure 7.16: Absolute error between 2-D frequency sweep and reconstructed HDI spectrum (black/white = small/large error).

proposed vertical-temporal filter banks, the required horizontal filter bank, and additional decomposition stages to increase the coding efficiency. In particular, for the 'Conversion Strategy' and the 'Direct Strategy' the following spatio-temporal subband decompositions are employed:

- Conversion Strategy: The vertical-temporal subband decomposition of Fig. 7.8 with the vertical decomposition  $\mu_0$  exchanged with temporal decomposition  $\mu_{20}$ . By doing this, the first decomposition results in two temporal subband samples. The temporal subband samples are both decomposed into 16 spatial subbands. Consequently, the total decomposition delivers  $2 \times 16 = 32$  spatio-temporal subbands, which all have a temporal size of one frame.
- Direct Strategy: The vertical-temporal subband decomposition of Fig. 7.5 with subbands 1, 2, and 3 subjected to an additional vertical decomposition (with a factor of two). This yields 12 vertical-temporal subbands, of which 4 bands have a temporal size of 2. As described earlier, these components are all treated independently by the bit allocation, resulting in 16 vertical-temporal subbands. Finally, these vertical-temporal subbands are horizontally decomposed into 4 bands to result in 64 spatio-temporal subbands.

Table 7.7 tabulates the performances of the resulting analysis/synthesis banks using the SNR measure (no quantization). The SNR measures are temporally averaged SNRs computed for the individual frames of the "Mobile" sequence (cut-out of 288x360 pixels). The table lists the quality of the filter banks for three different prototype filters used in the vertical-temporal decomposition, namely, QMF02, QMF08 and QMF16b [John80]. As can be seen

Table 7.7: Overall performance of 3-D analysis/synthesis filter banks for "Mobile" sequence (in dB, no quantization).

Filter	Conversion Strategy		Direct Strategy	
	HDP	HDI	HDP	HDI
QMF08	32.6	25.4	20.1	17.9
QMF16B	44.2	28.6	34.6	23.3
QMF02	47.9	29.3	47.5	17.3

the performances of these banks resemble the results obtained for the minimal vertical-temporal banks. If the prototype filter does not yield a near-unity transfer function (such as the QMF08 filter), the performances obtained are low.

To evaluate the coding performances of the schemes, the same cut-out of the "Mobile" sequence is encoded using both proposed schemes. The schemes encode the HDP and HDI signals at a bit rate of 0.8 bit per luminance pixel. For the HDP signal this would result in an overall bit rate of about 80 Mbit/s including a 20% reserved bandwidth for the chrominance components. Figures 7.17 and 7.18 show the quality (in SNR) of the reconstructed HDP signal for the Conversion Strategy scheme and the Direct Strategy scheme, respectively, as a function of the prototype filter. As a reference, the obtained quality of a basic subband scheme of Chapter 2 (with 16 uniform spatial subbands and a Selection Mechanism) is also shown in both graphs. Two conclusions can be drawn from these plots. First, the Conversion Strategy scheme outperforms the Direct Strategy scheme for all prototype filters evaluated. The cause for this is the higher number of decomposition stages in the latter, which limit the coding efficiency due to their imperfectness. In the second place, it can be observed that the spatio-temporal decompositions yield a more efficient data compression than the spatial decomposition (as used by the basic subband scheme) only. However, this is only true if the overall quality of the spatio-temporal filter bank is sufficient. This requirement is clearly violated for the QMF08 filter in the Direct Strategy scheme, resulting in a coding performance lower than obtained by the standard subband coder. The maximal gain in coding performance is about 1.5 dB for the scheme based on the Conversion Strategy and the QMF16b prototype filter. The subjective quality of the reconstructed HDP signal is for both schemes rather limited because of the blurring of moving contours and the existence of halo images.

Figures 7.19 and 7.20 show the SNR versus frame number of the encoded HDI signal for the spatially scalable schemes at a bit rate of 40 Mbit/sec. It can be observed that although the Direct Strategy has been tailored towards handling the HDI signal efficiently, the performance is less than that of the more straightforward Conversion Strategy due to the large distortions in the overall transfer function. Only in the case when the QMF16b filter is used does the coding performance approach the performances obtained with the Conversion Strategy. From this result it can be concluded that the Conversion Strategy can handle the empty frequency regions sufficiently well in the upsampled HDI signal. For both systems, the subjective qualities of the reconstructed HDI signals is not very high. Similar to the subjective qualities of the HDP signals, blurred contours and halo images are present in the reconstructed signals.

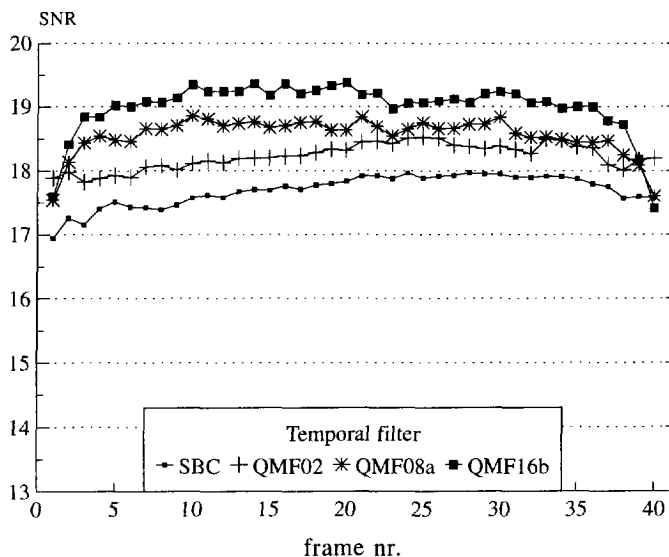


Figure 7.17: HDP coding performance for the Conversion Strategy scheme.

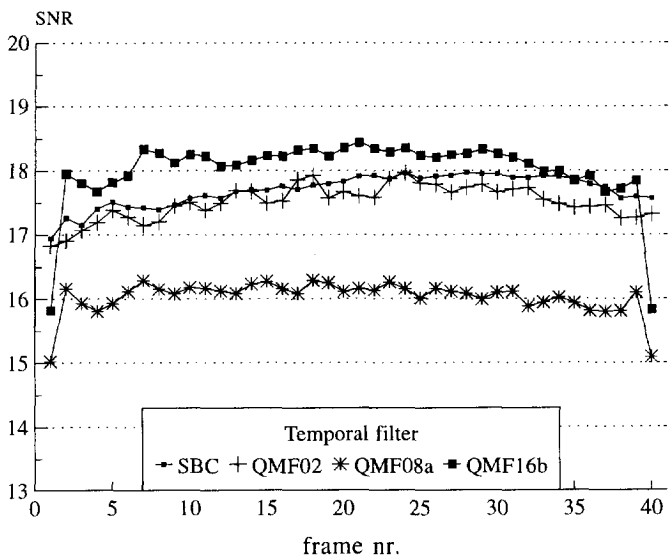


Figure 7.18: HDP coding performance for the Direct Strategy scheme.



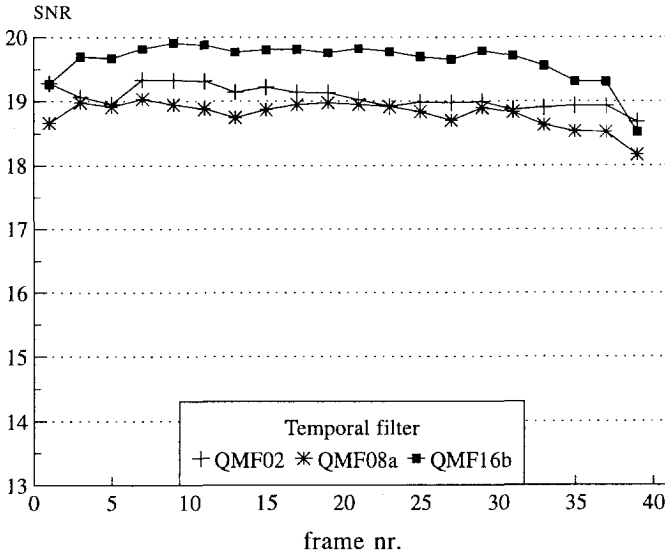


Figure 7.19: HDI coding performance for the Conversion Strategy scheme.

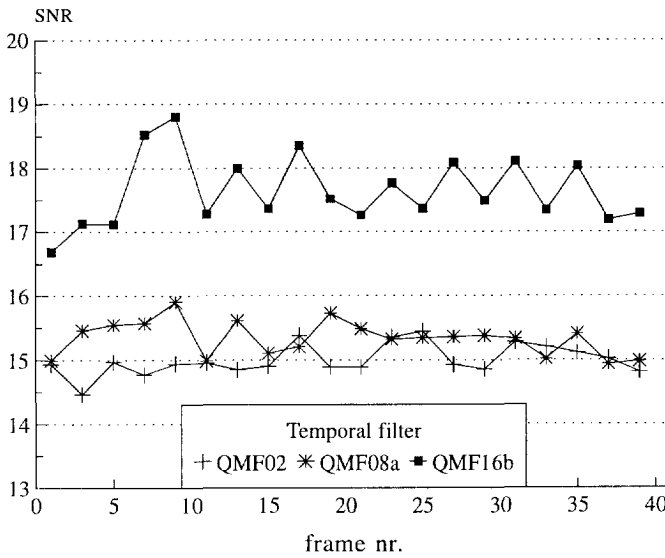


Figure 7.20: HDI coding performance for the Direct Strategy scheme.

## **7.6 Discussion**

In this chapter, two spatio-temporally scalable schemes were investigated that provide limited compatibility between progressive-scan signals and interlaced signals. The schemes are based on a 3-D subband decomposition. The first scheme implements the Conversion Strategy, which provides compatibility between progressive-scan signals only. For interlaced signals a lattice conversion is required. The second scheme implements the Direct Strategy and supports compatibility between all signals except HDI. For this signal a lattice conversion is still required to solve the temporal ambiguities inherently associated with the definition of the signals. However, the Direct Strategy filter bank decomposes the HDP spectrum such that the HDI frequency support can be exactly recovered from the subbands, thereby making dedicated coding methods for the HDI signal possible.

The spatio-temporally scalable schemes are rather complex because of the 3-D subband decomposition. Experimental results show that the complexity does not pay off in terms of coding performance. In particular, the simpler Conversion Strategy scheme outperforms the Direct Strategy scheme by approximately 1 dB for the HDP signal and by approximately 4 dB for the HDI signal (for short temporal filters). This result clearly shows that the moderate quality of the vertical-temporal filter bank of the Direct Strategy scheme limits its practical applicability.

The Conversion Strategy scheme outperforms the basic subband scheme for the HDP signal by approximately 1.5 dB. Hence, the temporal subband decomposition improves the coding performance with respect to subband schemes that use only spatial decompositions. Unfortunately, the subjective quality of the reconstructed HDP and HDI signals is not very high, because blurred contours and halo images are present in the reconstructed signals. To avoid these artifacts while maintaining the performance gain, the temporal subband filtering should be done along the motion trajectories, as recently proposed in [Ohm92, Ohm93a, Choi96].

# Chapter 8

## Discussion

In this thesis hierarchical video coding schemes are investigated that are based on subband coding techniques. These schemes can efficiently implement scalability of compressed video signals to support compatibility and graceful degradation in error-prone networks. The feasibility of implementing scalability of compressed video signals by hierarchical coding schemes has recently been recognised by the Moving Pictures Expert Group (MPEG). In the video part of the MPEG-2 standards (i.e. ISO/IEC 13818-2) tools have been included that support the implementation of SNR scalability, spatial scalability, temporal scalability, and data partitioning [ISO/95]. The MPEG-2 video standard is based on DCT coding and out-band motion compensation and is applicable for a wide range of applications. To stimulate the adaptation of the standard by the market, MPEG has defined a number of conformance points which are expressed using so-called profiles and levels. The main profile is the mainstream MPEG-2 profile that allows for the encoding of standard definition video signals (main level) and high definition video signals (high levels) without scalability tools. The supported bit rates are in the range of 4 to 80 Mbit/s.

The focus of commercial exploitation is currently directed to the main profile, as specified by several emerging application standards [DAVI96, DVB 94, Gran94]. On the one hand, this might imply that the large scale implementation of scalability will not happen in the near future (1-5 years). Especially not because manufacturers will give the highest priorities to the maturation of existing digital compression technology and to get services running to obtain economies of scale. On the other hand, since scalability significantly increases interoperability with features like compatibility and graceful degradation, it can be expected that in the near future other, or revised, application standards will adopt some types of scalability. In particular, it can be expected that techniques for graceful degradation are adopted because graceful degradation is essential when using transmission networks such as ATM-networks, terrestrial networks, and the Internet [Morr93a, Grun92, Scha92]. In fact, the fourth call of proposals of DAVIC already requested protocols and technologies for the improved delivery of audio and video across the Internet [DAVI95].

Within the MPEG-2 framework, the encoding process is not specified and may be used to obtain a competitive advantage. This thesis provides both theoretical and experimental evidence that for SNR scalability the efficient multirate encoding of the subbands is essential. These results indicate that within an MPEG-2 SNR-scalable encoder the multistage encoding of the DCT coefficients should be optimized with respect to the total coding performance of

the multistage quantization process, and only secondly with respect to the performance of the separate layers (see e.g. [Civa92, Tan94]). In particular, the concatenated macroblock encoders should be managed by a single controller that may compromise the accuracy of the first macroblock encoder to optimize the accuracy of the total macroblock encoding. The freedom of such controller within the MPEG-2 framework is however bounded, because the characteristics of macroblock quantizers are specified within the standard.

In this thesis it was illustrated that the performance of in-band motion compensation can be similar to that of out-band motion compensation. The in-band compensation method requires the interpolation of the subbands during the motion compensation process in order to cancel the aliasing components in the subbands. This interpolation might be done locally on a block basis, provided that efficient  $M$ -band filter banks are used [Bouc92]. Both the in-band and out-band motion compensation methods can be used to implement spatial scalability. The in-band compensation method allows for an efficient implementation in which the encoders (and decoders) of the subbands are closely related. This guarantees a high coding performance of the scheme, but reduces the extensibility and flexibility. In particular, new video signals can only be supported when they have spatial resolutions that are in line with the spatial resolutions already supported.

In contrast to the in-band compensation method, the out-band compensation method allows for an implementation of spatial scalability in which the encoders (and decoders) are independent of each other [Morr93a]. By using a pyramidal structure, it is not only easier to add new video signals to the hierarchy supported but also to use different coding schemes within the same hierarchy (e.g. H.261 for the low-resolution signal and MPEG-2 for the high-resolution signal). The cost for the increased flexibility is however a reduced performance because the different encoding schemes are independent of each other.

Experimental results in this thesis showed that the use of 3-D subband decompositions can significantly increase the coding performance of subband schemes (up to 1.5 dB), but that the associated artifacts like blurring and halo images limit the practical applicability of such decompositions. Newly proposed techniques like motion-compensated 3-D subband coding seem therefore very promising because these techniques avoid these artifacts by performing the temporal filtering along the motion trajectories [Ohm93b, Choi96]. High coding performances can especially be expected from these techniques when they are combined with encoding methods that fully exploit the spatio-temporal characteristics of the 3-D subbands. The implementation of spatial and temporal scalability with such schemes seems however to be complicated because the motion compensation is performed *directly on the high-resolution signal*. A possible solution would be to perform first the spatial subband decomposition and secondly perform the motion-compensated temporal decomposition within each subband. Unfortunately, such a solution will be hampered by the same aliasing problems as reported in this thesis for the in-band motion-compensated scheme.

# Bibliography

- [Ansa88] R. Ansari and S. H. Lee, "Two-Dimensional Non-Rectangular Interpolation, Decimation and Filter Banks," in *ICASSP'88*, (New York, U.S.A.), April 1988.
- [Ansa92] R. Ansari and C. Guillemot, "A Hierarchical Scheme for Coding Video at 4-10 Mbps," in *Eurasip European Signal Processing Conference EUSIPCO'92*, (Barcelona, Spain), pp. 227-230, August 1992.
- [Anto90] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image Coding using Vector Quantization in the Wavelet Transform domain," in *ICASSP'90*, (Albuquerque, U.S.A.), pp. 2297-2300, April 1990.
- [ATM 93] ATM Forum, *ATM User-Network Interface Specification, Version 3.0*, September 1993.
- [Bamb90] R. H. Bamberger and M. J. T. Smith, "Efficient 2-D Analysis/Synthesis Filter Banks for Directional Image Component Representation," in *ISCAS'90*, (New Orleans, U.S.A.), pp. 2009-2012, April 1990.
- [Barn92] C. F. Barnes and R. L. Frost, "Residual Vector Quantizers with Jointly Optimized Code Books," *Advances in Electronics and Electron Physics*, vol. 84, pp. 1-59, 1992.
- [Bell88] J. A. Bellisio and K. H. Tzou, "HDTV and the Emerging Broadband ISDN Network," in *SPIE Visual Communications and Image Proc. Conf.*, (Boston, U.S.A.), pp. 772-786, November 1988.
- [Berg71] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*. Englewood Cliffs: Prentice-Hall, 1971.
- [Biem90] J. Biemond, F. Bosveld, and R. L. Lagendijk, "Hierarchical Subband Coding of HDTV in BISDN," in *ICASSP'90*, (Albuquerque, New Mexico), pp. 2113-2116, April 1990.
- [Bier88] M. Bierling, "Displacement Estimation by Hierarchical Block-Matching," in *SPIE Visual Communications and Image Proc. Conf.*, (Boston, U.S.A.), pp. 942-951, November 1988.
- [Bosv90] F. Bosveld, R. L. Lagendijk, and J. Biemond, "A Refinement System for Hierarchical Video Coding," in *SPIE Visual Communications and Image Proc. Conf.*, (Lausanne, Switzerland), pp. 575-586, October 1990.

- [Bosv91] F. Bosveld, R. L. Lagendijk, and J. Biemond, "Three-Dimensional Subband Decompositions for Hierarchical Video Coding," in *SPIE Visual Communications and Image Proc. Conf.*, (Boston, U.S.A.), pp. 769–780, November 1991.
- [Bosv92a] F. Bosveld, R. L. Lagendijk, and J. Biemond, "Compatible HDTV Distribution using Fixed Distortion Subband coding," in *International workshop on HDTV'92*, (Kawasaki, Japan), November 1992.
- [Bosv92b] F. Bosveld, R. L. Lagendijk, and J. Biemond, "Compatible Spatio-Temporal Subband Encoding of HDTV," *Signal Processing*, vol. 28, pp. 271–289, September 1992.
- [Bosv92c] F. Bosveld, R. L. Lagendijk, and J. Biemond, "Hierarchical coding of HDTV," *Signal Processing: Image Communication*, vol. 4, pp. 195–225, June 1992.
- [Bosv93a] F. Bosveld, R. L. Lagendijk, and J. Biemond, "Compatible HDTV Transmission using Conditional Entropy Coding," in *ICASSP'93*, (Minneapolis, U.S.A.), pp. 674–677, April 1993.
- [Bosv93b] F. Bosveld, R. L. Lagendijk, and J. Biemond, "Compatible Video Compression using Subband and Motion-Compensation Techniques," in *International Workshop on HDTV'93*, (Ottawa, Canada), October 1993.
- [Bosv95] F. Bosveld, R. L. Lagendijk, and J. Biemond, "Hierarchical Coding," in *Handbook of Visual Communications* (H. M. Hang and J. W. Woods, eds.), ch. 9, pp. 299–340, Boston: Academic Press, 1995.
- [Bouc92] F. Boucherok and J. F. Vial, "Compatible multi-resolution coding scheme," in *International Workshop on HDTV'92*, (Kawasaki, Japan), November 1992.
- [Bove92] V. M. Bove and A. B. Lippman, "Scalable Open-Architecture Television," *SMPTE Journal*, pp. 2–5, January 1992.
- [Bree90] M. Breeuwer and P. H. N. de With, "Source Coding of HDTV with Compatibility to TV," in *SPIE Visual Communications and Image Proc. Conf.*, (Lausanne, Switzerland), pp. 765–776, October 1990.
- [Burt83] P. J. Burt and E. H. Adelson, "The Laplacian Pyramid as a Compact Image Code," *IEEE Trans. on Commun.*, vol. 31, no. 4, pp. 532–540, 1983.
- [CCIR89] CCIR Study Groups, *Document 11/298-E : Example of a hierarchical family of studio formats towards HDTV*, March 1989.
- [CCIT89] CCITT SG XV WP/1/Q4, *Specialists Group on Coding for Visual Telephony. Description of Reference Model 8 (rm8). Document 525*, June 1989.
- [Chen88] T. C. Chen, K. H. Tzou, and P. E. Fleisher, "A hierarchical HDTV coding system using a DPCM-PCM approach," in *SPIE Visual Communications and Image Proc. Conf.*, (Boston, U.S.A.), pp. 804–811, November 1988.

- [Chia93a] T. Chiang and D. Anastassiou, "HDTV/CDTV Compatible Coding with Improved Prediction," in *International Workshop on HDTV'93*, (Ottawa, Canada), October 1993.
- [Chia93b] L. Chiariglione, "Development of multi-industry information technology standards; The MPEG case," in *International workshop on HDTV'93*, (Ottawa, Canada), October 1993.
- [Chia95] L. Chiariglione, "Digital Audio-Visual Council; Rationale and Goals," *EBU Technical Review*, pp. 51–63, Winter 1995.
- [Choi96] S. J. Choi and J. W. Woods, "Motion-Compensated 3-D Subband Coding of Video," *submitted to IEEE Transactions on Image Processing*, May 1996.
- [Civa92] M. R. Civanlar and A. Puri, "Scalable Video Coding in Frequency Domain," in *SPIE Visual Communications and Image Proc. Conf.*, (Boston, U.S.A.), pp. 1124–1134, November 1992.
- [Cono71] W. J. Conover, *Practical Non-Parametric Statistics*. J. Wiley and Sons Inc., 1971.
- [Daub90] I. Daubechies, "The Wavelet Transform, Time-Frequency Localization, and Signal Analysis," *IEEE Trans. on Inform. Theory*, vol. 36, pp. 961–1005, September 1990.
- [DAVI95] DAVIC, *Fourth call of proposals*, December 1995.
- [DAVI96] DAVIC, *Digital Audio-Visual Council: Specifications 1.0, Revision 4.1*, February 1996.
- [deGa92] D. P. deGarrido and W. Pearlman, "Multirate Codebook Design with the Entropy Constrained Pairwise Nearest Neighbour (ECPNN) Algorithm," in *ICASSP'92*, (San Francisco, U.S.A.), pp. 373–376, March 1992.
- [Demo92] G. Demos, "An Example Hierarchy of Formats for HDTV," *SMPTE Journal*, pp. 609–617, September 1992.
- [Diab90] C. Diab, R. Prost, and R. Goutte, "Block-Adaptive Subband Coding of Images," in *ICASSP'90*, (Albuquerque, U.S.A.), pp. 2093–2096, April 1990.
- [Drie92] J. N. Driessen, *Motion Estimation for Digital Video*. PhD thesis, T.U. Delft, September 1992.
- [DVB 94] DVB Project office (ed. G.T. Waters), *DVB, Television for the third millenium*, August 1994.
- [Elle90] J. C. Ellershaw, M. J. Biggar, and A. W. Johnson, "Interworking of Video Codecs with Different Spatial Resolutions and Aspect Ratios," in *Picture Coding Symposium*, (Cambridge, U.S.A.), 1990.
- [Elna86] S. E. Elnahas, K. T. Tzou, J. R. Cox, R. L. Hill, and R. G. Jost, "Progressive Coding and Transmission of Digital Diagnostic Pictures," *IEEE Transactions on Medical Imaging*, vol. 5, pp. 73–83, June 1986.

- [Equi91] W. H. R. Equitz and T. M. Cover, "Successive Refinement of Information," *IEEE Trans. on Inform. Theory*, vol. 37, pp. 269–275, March 1991.
- [Eric85] S. Ericsson and E. Dubois, "Digital Coding of High Quality TV," in *High Definition Television Colloquium*, (Ottawa, Canada), May 1985.
- [Farv84] N. Farvardin and J. W. Modestino, "Optimum Quantizer Performance for a Class of Non-Gaussian Memoryless Sources," *IEEE Trans. on Inform. Theory*, vol. 30, pp. 485–497, May 1984.
- [Flei64] P. E. Fleisher, "Sufficient conditions for achieving minimum distortion in a quantizer," in *IEEE Int. Convention Rec., pt. 1*, pp. 104–111, 1964.
- [Frim95] E. Frimout, *Fast Playback of Heli-scan Recorded MPEG Video*. PhD thesis, T.U. Delft, November 1995.
- [Furu93] I. Furukawa, M. Nomura, and S. Ono, "Hierarchical sub-band coding of super high definition image with adaptive block-size multistage VQ," *Signal Processing: Image Communication*, vol. 5, pp. 527–538, December 1993.
- [Gala84] C. R. Galand and H. J. Nussbaumer, "New Quadrature Mirror Filter Structures," *IEEE Trans. on Acoust., Speech, Signal Processing*, vol. 32, pp. 522–530, June 1984.
- [Gers92] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Boston/Dordrecht/Londen: Kluwer Academic Publishers, 1992.
- [Ghar91a] H. Gharavi, "Multilayer Subband-Based Video Coding," *IEEE Trans. on Commun.*, vol. 39, pp. 1288–1291, September 1991.
- [Ghar91b] H. Gharavi, "Subband Coding Algorithms for Video Applications: Videophone to HDTV-Conferencing," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 1, pp. 174–183, June 1991.
- [Gilg91] M. Gilge and R. Gusella, "Motion Video Coding for Packet-Switching Networks—An Integrated Approach," in *SPIE Visual Communications and Image Proc. Conf.*, (Boston, U.S.A.), pp. 592–603, November 1991.
- [Gonz93] C. Gonzales and E. Viscito, "Flexibly scalable digital video coding," *Signal Processing: Image Communication*, vol. 5, pp. 5–20, February 1993.
- [Good80] D. J. Goodman, "Embedded DPCM for Variable Bit Rate Transmission," *IEEE Trans. on Commun.*, vol. 28, pp. 1040–1046, July 1980.
- [Gran94] Grand Alliance, *Grand Alliance HDTV System Specification, Version 1.0*, April 1994.
- [Grun92] M. Grundstrom and J. Nieweglowski, "Simulations of a Digital HDTV System with Graceful Degradation," in *International Workshop on HDTV'92*, (Kawasaki, Japan), November 1992.



- [Guil92] C. Guillemot, "Compatible HDTV/TV Hierarchical Scheme for Secondary Distribution of TV and HDTV signals," in *SPIE Visual Communications and Image Proc. Conf.*, (Boston, U.S.A.), pp. 1475–1483, November 1992.
- [Haan92] G. de Haan, *Motion Estimation and Compensation*. PhD thesis, T.U. Delft, September 1992.
- [Horn91] B. R. Horng, H. Samuelli, and A. N. Willson, "The Design of Low-Complexity Linear-Phase FIR Filter Banks using Power-of-Two Coefficients with an Application to Subband Image Coding," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 1, December 1991.
- [Hors93] R. ter Horst, A. Koster, K. Rijkse, E. Fert, G. Nocture, and L. Tranchard, "MUP-COS: A multi-purpose coding scheme," *Signal Processing: Image Communication*, vol. 5, pp. 57–89, February 1993.
- [Huff52] D. Huffman, "A Method for the Construction of Minimum Redundancy Codes," *Proceedings IRE*, pp. 1098–1101, September 1952.
- [Huso90] J. H. Husoy and T. A. Ramstad, "Application of an Efficient Parallel IIR Filter Bank to Image Subband Coding," *Signal Processing*, vol. 20, pp. 279–292, August 1990.
- [ISO/92] ISO/1172-1 IS, *Information Technology Coding of Moving Pictures and associated Audio for Digital Storage Media up to about 1.5 Mbit/s. Part 2:video*, November 1992.
- [ISO/93] ISO/IEC JTC1/SC2/WG11 Test Model Editing Committee, *Test model 4.2, MPEG-2 ed.*, February 1993.
- [ISO/95] ISO/IEC 13818-2 IS, *Information Technology - Generic Coding of Moving Pictures and associated Audio. Part 2:video*, 1995.
- [ITU-88] ITU-T, *Recommendation I.121:Broadband Aspects of ISDN*, 1988.
- [ITU-93] ITU-T, *Recommendation I.363:B-ISDN ATM Adaptation Layer (AAL) Specification*, March 1993.
- [Jaya81] N. S. Jayant and S. W. Christensen, "Effects of Packet Losses in Waveform Coded Speech and Improvements Due to an Odd-Even Sample-Interpolation Procedure," *IEEE Trans. on Commun.*, vol. 29, pp. 101–109, February 1981.
- [Jaya83] N. S. Jayant, "Variable Rate ADPCM based on Explicit Noise Coding," *The Bell System Technical Journal*, vol. 62, pp. 657–677, March 1983.
- [Jaya84] N. S. Jayant and P. Noll, *Digital Coding of Waveforms: Principles and Applications to Speech and Video*. Englewood Cliffs: Prentice-Hall, 1984.
- [Jeon91] J. H. Jeon, C. H. Hahm, and J. K. Kim, "On the Hierarchical Block Motion Estimation for Video Subband Coding," in *SPIE Visual Communications and Image Proc. Conf.*, (Boston, U.S.A.), pp. 954–962, November 1991.

- [John80] J. D. Johnston, "A Filter Family Designed for Use in Quadrature Mirror Filter Banks," in *ICASSP'80*, pp. 291–294, April 1980.
- [Lage96] R. L. Lagendijk, F. Bosveld, and J. Biemond, "Subband Video Coding," in *Subband and Wavelet transforms; Design and Applications* (A. N. Akansu and M. J. T. Smith, eds.), ch. 8, pp. 251–286, Boston: Kluwer Academic Press, 1996.
- [Lame93] J. D. Lameillieure and G. Schamel, "Hierarchical Coding of TV/HDTV within the German HDTV-T Project," in *International Workshop on HDTV'93*, (Ottawa, Canada), October 1993.
- [Lang84] G. G. Langdon, "An Introduction to Arithmetic Coding," *IBM J. Res. Develop.*, vol. 28, pp. 135–149, March 1984.
- [LeBo92] J. Y. LeBoudec, "The Asynchronous Transfer Mode: a tutorial," *Computer Networks and ISDN Systems*, vol. 24, pp. 279–309, 1992.
- [Lee90] D. S. Lee and K. H. Tzou, "Hierarchical DCT coding of HDTV for ATM Networks," in *ICASSP'90*, (Albuquerque, U.S.A.), April 1990.
- [LeGa88] D. J. LeGall, H. Gaggioni, and C. T. Chen, "Transmission of HDTV signals under 140 Mb/s using a Sub-Band Decomposition and Discrete Cosine Transform Coding," in *Signal Processing of HDTV*, (North-Holland, the Netherlands), Elsevier Science Publishers B.V., 1988.
- [LeGa91] D. LeGall, "MPEG: A Video Compression Standard for Multimedia Applications," *Communications of the ACM*, vol. 34, pp. 47–58, April 1991.
- [Lipp91] A. Lippman, "Feature Sets for Interactive Images," *Communications of the ACM*, vol. 34, pp. 93–102, April 1991.
- [Lloy82] S. P. Lloyd, "Least Squares Quantization in PCM," *IEEE Trans. on Inform. Theory*, vol. 28, pp. 129–137, March 1982.
- [Long92] T. Long, "Digital Television Broadcasting Developments in Europe," in *NAB Broadcast Engineering Conference*, pp. 126–135, April 1992.
- [Mahe89] B. Mahesh and W. A. Pearlman, "Hexagonal Sub-Band Coding for Images," in *ICASSP'89*, (Glasgow, Scotland), pp. 1953–1956, May 1989.
- [Mall89] S. G. Mallat, "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, pp. 674–693, July 1989.
- [Mau92] J. Mau, "Perfect Reconstruction Modulated Filter Banks," in *ICASSP'92*, (San Francisco, U.S.A), pp. 273–276, March 1992.
- [Max60] J. Max, "Quantizing for Minimum Distortion," *IRE Transactions on Information Theory*, pp. 7–12, March 1960.

- [Mazo86] B. Mazor and W. A. Pearlman, "A Tree Coding Theorem for Stationary Gaussian Sources and the Squared-Error Distortion Measure," *IEEE Trans. on Inform. Theory*, vol. 32, pp. 156–165, March 1986.
- [Mill72] J. H. Miller and J. B. Thomas, "Detectors for Discrete-Time Signals in Non-Gaussian Noise," *IEEE Trans. on Inform. Theory*, vol. 18, pp. 241–250, March 1972.
- [Morr91] G. Morrison and D. Beaumont, "Two-layer video coding for ATM-networks," *Signal Processing: Image Communication*, vol. 3, pp. 179–195, June 1991.
- [Morr93a] G. Morrison and I. Parke, "COSMIC: A compatible scheme for moving image coding," *Signal Processing: Image Communication*, vol. 5, pp. 91–103, February 1993.
- [Morr93b] G. Morrison and I. Parke, "A spatially layered hierarchical approach to video coding," *Signal Processing: Image Communication*, vol. 5, pp. 445–462, December 1993.
- [Musm85] H. G. Musmann, P. Pirsch, and H. J. Grallert, "Advances in Picture Coding," in *Proc. IEEE*, vol. 73, pp. 523–548, April 1985.
- [Nave92] T. Naveen and J. W. Woods, "Motion Compensated Multiresolution Transmission of High Definition Video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 4, pp. 29–41, February 1992.
- [Nave93] T. Naveen, *Subband Compression of High Definition Video*. PhD thesis, Rensselaer Polytechnic Institute, January 1993.
- [Nave95] T. Naveen, F. Bosveld, J. W. Woods, and R. L. Lagendijk, "Rate Constrained Multiresolution Transmission of Video," *IEEE Trans. Cir. and Sys. for Video Technology*, vol. 5, pp. 193–206, June 1995.
- [Ohm92] J. R. Ohm, "Temporal Domain Sub-Band Video Coding with Motion Compensation," in *ICASSP'92*, (San Francisco, U.S.A), pp. 229–232, March 1992.
- [Ohm93a] J. R. Ohm, "Advanced Packet-Video Coding based on Layered VQ and SBC Techniques," *IEEE Transactions on Circuits and Systems for Video technology*, vol. 3, pp. 208–221, June 1993.
- [Ohm93b] J. R. Ohm, "Layered VQ and SBC Techniques for Packet Video Applications," in *Workshop on Packet Video '93*, (Berlin, B.R.D.), 1993.
- [Pear91] W. A. Pearlman, "Performance Bounds of Subband Coding," in *Subband Image Coding* (J. W. Woods, ed.), ch. 1, pp. 1–41, Boston: Kluwer Academic Publishers, 1991.
- [Peco90] M. Pecot, P. J. Tourtier, and Y. Thomas, "Compatible coding of Television Images, Part 2, Coding algorithm," *Signal Processing: Image Communication*, vol. 2, pp. 245–258, October 1990.

- [Pei94] S. C. Pei and F. C. Chen, "3-D spatiotemporal subband decompositions for hierarchical compatible video coding by mathematical morphology," *Signal Processing: Image Communication*, vol. 6, pp. 83–99, 1994.
- [Phil94] Philips Consumer Electronics B.V., *System Description Photo CD, Version 1.0*, December 1994.
- [Pryc91] M. de Prycker, *Asynchronous Transfer Mode; Solution for Broadband ISDN*. New York: Ellis Horwood, 1991.
- [Reut88] T. Reuter and H. D. Hohne, "Motion Vector Estimation for Improved Standards Conversion," in *Signal Processing of HDTV*, (North-Holland, the Netherlands), pp. 345–353, Elsevier Science Publishers B.V., 1988.
- [Riou91] O. Rioul and M. Vetterli, "Wavelets and Signal Processing," *IEEE SP Magazine*, pp. 14–38, October 1991.
- [Risk91] E. A. Riskin, "Optimal Bit Allocation via the Generalized BFOS Algorithm," *IEEE Trans. on Inform. Theory*, vol. 37, pp. 400–402, March 1991.
- [Scha92] G. Schamel, "Graceful Degradation and Scalability in Digital Coding for Terrestrial Transmission," in *International Workshop on HDTV'92*, (Kawasaki, Japan), November 1992.
- [Schr95] W. F. Schreiber, "A unified approach to moving-image coding for most media and most applications," *Signal Processing*, vol. 43, pp. 305–321, 1995.
- [Shap93] J. M. Shapiro, "Application of the Embedded Wavelet Hierarchical Image Coder to very low bit rate Image Coding," in *ICASSP'93*, (Minneapolis, U.S.A.), pp. 558–561, April 1993.
- [Smit90] M. J. T. Smith and S. L. Eddins, "Analysis/Synthesis Techniques for Subband Image Coding," *IEEE Trans. on Acoust., Speech, Signal Processing*, vol. 38, pp. 1446–1456, August 1990.
- [Tan94] T. K. Tan, K. K. Pang, and K. N. Ngan, "A Frequency Scalable Coding Scheme employing Pyramid and Subband Techniques," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 4, pp. 203–210, April 1994.
- [Tane81] A. S. Tanenbaum, *Computer Networks*. Englewood Cliffs: Prentice-Hall, 1981.
- [Taub94] D. Taubman and A. Zakhor, "Multi-rate 3-D Subband Coding of Video," *IEEE Transactions on Image Processing*, vol. 3, pp. 572–588, September 1994.
- [Tzir94] G. Tziritas and C. Labit, *Motion Analysis for Image Sequence Coding*. Amsterdam: Elsevier, 1994.
- [Tzou86] K. H. Tzou and S. E. Elnahas, "Bit-Sliced Progressive Transmission and Reconstruction of Transformed Images," in *ICASSP'86*, (Tokyo, Japan), pp. 533–536, 1986.

- [Tzou87] K. H. Tzou, "Progressive image transmission: a review and comparison of techniques," *Optical Eng.*, vol. 26, pp. 581–589, July 1987.
- [Tzou88] K. H. Tzou, T. C. Chen, P. E. Fleisher, and M. L. Liou, "Compatible HDTV Coding for Broadband ISDN," in *GLOBECOM'88*, (Florida, U.S.A.), pp. 743–749, December 1988.
- [Uz91a] K. M. Uz, K. Ramchandran, and M. Vetterli, "Multiresolution Source and Channel coding for Digital Broadcast of HDTV," in *Fourth International Workshop on HDTV*, (Turin, Italy), September 1991.
- [Uz91b] K. M. Uz, M. Vetterli, and D. J. LeGall, "Interpolative Multiresolution Coding of Advanced Television with Compatible Subchannels," *IEEE Trans. on Cir. and Sys. for Video Technology*, vol. 1, pp. 86–99, March 1991.
- [Vaid87] P. P. Vaidyanathan, "Quadrature Mirror Filter Banks, M-Band Extensions and Perfect-Reconstruction Techniques," *IEEE ASSP Magazine*, pp. 4–20, July 1987.
- [Vand90] L. Vandendorpe and P. Delogne, "Hierarchical encoding of HDTV by Transform Coefficients Block splitting," in *SPIE Image Processing Algorithms and Techniques*, pp. 343–354, February 1990.
- [Vand91a] L. Vandendorpe, *Hierarchical Coding of Digital Moving Pictures*. PhD thesis, University of Louvain, October 1991.
- [Vand91b] L. Vandendorpe and P. Delogne, "Hierarchical Transform and Subband Coding of HDTV," in *Fourth International Workshop on HDTV*, (Turin, Italy), September 1991.
- [Vand93] L. Vandendorpe, L. Cuvelier, B. Maison, and P. Delogne, "Pyramidal Scheme for the Compatible Coding of Interlaced TV/HDTV," in *International Workshop on HDTV'93*, (Ottawa, Canada), October 1993.
- [Vett84] M. Vetterli, "Multi-Dimensional Subband Coding: Some Theory and Algorithms," *Signal Processing*, vol. 6, pp. 97–112, April 1984.
- [Vett87] M. Vetterli, "A Theory of Multirate Filter Banks," *IEEE Trans. on Acoust., Speech, Signal Processing*, vol. 35, pp. 356–372, March 1987.
- [Vett90] M. Vetterli, J. Kovačević, and D. J. LeGall, "Perfect Reconstruction Filters Banks for HDTV Representation and Coding," *Signal Processing: Image Communication*, vol. 2, pp. 349–364, October 1990.
- [Vett92] M. Vetterli and K. M. Uz, "Multidimensional Coding Techniques for Digital Television: a Review," *Multidimensional Systems and Signal processing*, vol. 3, pp. 161–187, March 1992.
- [Vial91] J. F. Vial, M. Pecot, P. J. Tourtier, and Y. Thomas, "In-Band Interpolation applied to Motion-Compensated Subband Coding," in *Fourth International Workshop on HDTV*, (Turin, Italy), September 1991.

- [Wang88] L. Wang and M. Goldberg, "Progressive Image Transmission by Transform Coefficient Residual Error Quantization," *IEEE Trans. on Commun.*, vol. 36, pp. 75–87, January 1988.
- [West88] P. H. Westerink, J. Biemond, and D. E. Boeke, "An Optimal Bit Allocation Algorithm for Sub-Band Coding," in *ICASSP'88*, (New York, U.S.A.), pp. 757–760, April 1988.
- [West89] P. H. Westerink, *Subband Coding of Images*. PhD thesis, T.U. Delft, October 1989.
- [With92] P. H. N. de With, A. M. A. Rijckaert, J. Kaaden, and H. W. Keesen, "Digital Consumer HDTV Recording based on Motion-Compensated DCT coding of Video Signals," *Signal Processing: Image Communication*, vol. 4, pp. 401–420, August 1992.
- [Witt87] I. H. Witten, R. M. Neal, and J. G. Cleary, "Arithmetic Coding for Data Compression," *Communications of the ACM*, vol. 30, pp. 520–540, June 1987.
- [Wood69] R. C. Wood, "On Optimum Quantization," *IEEE Trans. on Inform. Theory*, vol. 15, pp. 248–252, March 1969.
- [Wood86] J. W. Woods and S. D. O'Neil, "Subband Coding of Images," *IEEE Trans. on Acoust., Speech, Signal Processing*, vol. 34, pp. 1278–1288, October 1986.
- [Wood95] D. Wood, "Satellites, science and success-The DVB story," *EBU Technical Review*, pp. 4–11, Winter 1995.
- [Zafa93] S. Zafar, Y. Q. Zhang, and B. Jabbari, "Multiscale Video Representation Using Multiresolution Motion Compensation and Wavelet Decomposition," *IEEE Journal on Selected Areas in Communications*, vol. 11, pp. 24–35, January 1993.

# List of symbols and abbreviations

## Abbreviations

2-D	Two-Dimensional
3-D	Three-Dimensional
AAC	Adaptive Arithmetic Coder
AAD	Adaptive Arithmetic Decoder
AAL	ATM Adaptation Layer
AC	Arithmetic Coder
AC	Arithmetic Decoder
ADPCM	Adaptive Differential Pulse Code Modulation
ATM	Asynchronous Transfer Mode
BA	Bit Allocation
BCT	Block Classification Table
bpp	Bits per pixel
bps	Bits per second
BUF	Buffer
CBR	Constant Bit Rate
CD	Compact Disc
CD-ROM	Compact Disc-Read Only Memory
COD	Lossless encoder
CLP	Cell Loss Priority
CLR	Cell Loss Ratio
DAVIC	Digital Audio-Visual Council
DCT	Discrete Cosine Transform
DEC	Lossless decoder
DPCM	Differential Pulse Code Modulation
DUX	Demultiplexer
DVB	Digital Video Broadcasting
DWT	Discrete Walsh Hadamard Transform
EDI	Enhanced Definition Interlaced
EDP	Enhanced Definition Progressive
EST	Estimator
FEC	Forward Error Correction
FIR	Finite Impulse Response
FM	Frame Memory
GG	Generalized Gaussian

GG-PDF	Generalized Gaussian Probability Density Function
HDI	High Definition Interlaced
HDP	High Definition Progressive
HDTV	High Definition Television
HFC	Hybrid Fiber Coax
IEC	International Electrotechnical Commission
IIR	Infinite Impulse Response
ISO	International Organisation for Standardization
ITU	International Telecommunication Union
ITU-T	International Telecommunication Union - Telecommunication
JPEG	Joint Photographic Expert Group
KLT	Karhunen-Loeve Transform
LMQ	Lloyd-Max Quantizer
LOT	Lapped Orthogonal Transform
MBS	Maximum Burst Size
MC	Motion Compensation
MC-SBC	Motion Compensated Subband Coding
ME	Motion Estimation
MPEG	Moving Picture Expert Group
MR-DPCM	Multirate-Differential Pulse Codeword Modulation
MR-PCM	Multirate-Pulse Codeword Modulation
MR-QAM	Multirate-Quadrature Amplitude Modulation
MSE	Mean Square Error
MUX	Multiplexer
MV	Motion Vectors
PCM	Pulse Codeword Modulation
PCR	Peak Cell Rate
PDF	Probability Density Function
PR	Perfect Reconstruction
PRMF	Perfect Reconstruction Modulated Filter
PSD	Power Spectral Density
QMF	Quadrature Mirror Filter
QoS	Quality of Service
RDF	Rate Distortion Function
r.v.	Random variable
SBC	Subband Coding
SBD	Subband Decomposition
SBR	Subband Reconstruction
SCR	Sustained Cell Rate
SEL	Selection Mechanism
SI	Side-Information
SNR	Signal to Noise Ratio
SOB	Start of Block
SOF-Y	Start of Frame - Luminance
SOF-UV	Start of Frame - Chrominance
SOL	Start of Slice



SP	Subband Prediction
TV	Television
UTQ	Uniform Threshold Quantizer
VBR	Variable Bit Rate
VLC	Lossless Variable Length Coder
VLD	Lossless Variable Length Decoder
VLSI	Very Large Scale Integration
VT	Video Telephony

## Symbols

Symbols are tabulated according to their first appearance. Symbols with local significance only are not all shown.

### Chapter 1

$x$	input signal
$y^1$	lowest-quality reproduction
$y^n$	$n^{\text{th}}$ -quality reproduction
$y^N$	highest-quality reproduction
$l^n$	layer $n$

### Chapter 2

#### Digital Coding

$x(t)$	continuous-time input signal
$x, x(k), x(kT)$	discrete-time input signal
$y(t)$	continuous-time reconstructed signal
$y, y(k), y(kT)$	discrete-time reconstructed signal
$r(t)$	continuous-time reconstruction error
$r(k)$	discrete-time reconstruction error
$T$	sampling interval
$f_s$	sampling rate
$I$	transmission rate
$R$	bit rate per pixel
$D$	distortion
$\sigma_x^2$	variance of input signal
$\sigma_r^2$	variance of reconstruction error

**Memoryless coding**

$Q$	quantizer characteristic
$Q_M$	quantizer mapping operation
$Q_L$	quantizer look-up operation
$X$	r.v. associated with input signal
$J$	r.v. associated with quantizer index
$K$	number of representation levels of $Q$
$H(\cdot)$	entropy
$p_X(\cdot)$	PDF of input signal
$\mu_x$	mean-value of input signal
$c(k)$	codeword at time $k$
$q(k)$	discrete-time quantization error
$\sigma_q^2$	variance of quantization error
$j, j(k)$	quantizer index (at time $k$ )
$x_j$	quantizer decision level
$S_j$	quantizer decision interval
$y_j$	quantizer reconstruction value
$P_j$	occurrence probability of index $j$
$n_j$	length of $j$ th codeword

**Predictive coding**

$d(k)$	difference value (or prediction error) of time $k$
$\sigma_d^2$	variance of prediction error
$\hat{x}(k)$	prediction signal
$P(\cdot)$	predictor
$G_P$	prediction gain
$\mathbf{h}$	vector of $N$ prediction coefficients
$h_{i,j}$	predictor coefficient
$\mathbf{h}_{opt}$	vector with optimal prediction coefficients
$\mathbf{R}_{xx}$	autocorrelation matrix
$\mathbf{r}_{xx}$	vector with autocorrelation values
$x(i, j, k)$	input frame at time $k$
$\hat{x}(i, j, k)$	prediction of input frame at time $k$
$y(i, j, k)$	reconstructed frame at time $k$
$\mathbf{d}$	displacement vector
$\delta_i, \delta_j$	horizontal and vertical displacement values

**Spectral coding**

$m$	subband number
$M$	number of subbands
$x_m$	subband $m$
$y_m$	reconstructed subband $m$
$c_m$	sequence of codewords of subband $m$
$f_{s,m}$	sampling frequency of subband $m$
$V_m$	decimation factor of subband $m$
$W_m$	bandwidth of subband $m$
$H_m$	analysis filter of subband $m$
$G_m$	synthesis filter of subband $m$
$R_m, r_m$	Bit rate of subband $m$
$D_m, d_m$	Distortion of subband $m$
$R_{m,opt}$	Optimal bit rate of subband $m$
$D_{m,opt}$	Optimal distortion of subband $m$
$X(\omega)$	Fourier transform of $x$
$Y(\omega)$	Fourier transform of $y$
$H_m(\omega)$	Fourier transform of $H_m$
$G_m(\omega)$	Fourier transform of $G_m$
$X_m(\omega)$	Fourier transform of subband $x_m$
$T(\omega)$	Transfer function
$\sigma_{x,m}^2$	variance of subband $m$
$\sigma_{r,m}^2$	variance of reconstruction error in subband $m$
$\sigma_{q,m}^2$	variance of quantization error in subband $m$
$G_S$	spectral prediction gain
$G_{S,max}$	maximum spectral prediction gain

**Intraframe subband coding**

$\sigma_{d,m}^2$	variance of prediction error of subband $m$
$\sigma_{x,m,k}^2$	variance of $k^{\text{th}}$ activity class of subband $m$
$\sigma_{d,m,k}^2$	variance of $k^{\text{th}}$ activity class of prediction error of subband $m$
$r_{m,k}$	bit rate of $k^{\text{th}}$ activity class of subband $m$
$d_{m,k}$	distortion of $k^{\text{th}}$ activity class of subband $m$
$c$ (of GG-PDF)	shape parameter of GG-PDF
$\sigma_x$	standard deviation of input signal
$\Delta x$	stepsize of UTQ

**Implementation of scalability**

$N$	number of reproductions
$R^n$	rate of $y^n$
$D^n$	distortion of $y^n$
$Q^n$	quantizer characteristic of quantizer $n$
$Q_M^n$	mapping operation of $Q^n$
$Q_L^n$	look-up operation of $Q^n$
$\beta^n$	decimation factor of $y^n$
$f_s^n$	spatial sampling rate of $y^n$
$f_r^n$	frame rate of $y^n$
$V_m^n$	subsampling factor of subband $m$ for $y^n$
$j^n, j^n(k)$	index (at time $k$ ) of $n^{\text{th}}$ multirate quantizer
$i^n, i^n(k)$	index (at time $k$ ) of $n^{\text{th}}$ embedded quantizer

**Chapter 3**

$x^n$	filtered and decimated version of $x$
$R_x(D)$	rate distortion function of $x$
$D_x(R)$	distortion rate function of $x$
$S_{xx}(\omega)$	power spectral density function of $x$
$S_{yy}(\omega)$	power spectral density function of $y$
$S_{rr}(\omega)$	power spectral density function of $r$
$\gamma_x^2$	spectral flatness measure of $x$
$\theta$	coding parameter of RDF
$R_x(\theta)$	parametric rate function of $x$
$D_x(\theta)$	parametric distortion function of $x$
$R_x^n(D)$	rate distortion function of $x^n$
$\theta^n$	coding threshold for $y^n$
$R_y^n(\theta)$	parametric rate function of $y^n$
$D_y^n(\theta)$	parametric distortion function of $y^n$
$\tilde{R}^n$	rate of layer $n$
$S_{xx}^n(\omega)$	PSD of $x^n$
$\hat{S}_{xx}^n(\omega)$	PSD of interpolated $x^n$
$S_{yy}^n(\omega)$	PSD of $x^n$

## Chapters 4 and 5

$u^n$	output of $n^{\text{th}}$ multistage quantizer
$\bar{D}^n$	distortion of layer $n$
$R_{UTQ}(D)$	practical performance curve of UTQ
$P^n(y^n)$	prediction based at $y^n$
$BA_n$	bit allocation procedure for $l^n$
$c^n$	codewords for layer $n$
$c_m^n$	codewords of subband $m$ for layer $n$

## Chapter 6

$y^n(i, j, k)$	reconstructed frame of $y^n$ at time $k$
$y_m^n(i, j, k)$	reconstructed subband $m$ of $y^n(i, j, k)$
$x_m^n(i, j, k)$	subband $m$ of $x(i, j, k)$ to be encoded for $y^n$
$d_m^n(i, j, k)$	prediction error subband $m$ of $x(i, j, k)$ to be encoded for $y^n$
$\hat{x}_m^n(i, j, k)$	predicted subband for $x_m^n(i, j, k)$
$\tilde{y}_m^n(i, j, k)$	motion-compensated reconstructed subband $m$ of $y^n(i, j, k)$
$X(\omega, k)$	Fourier transform of $x$ at time $k$
$X_m(\omega, k)$	Fourier transform of subband $m$ at time $k$
$h_m(l)$	$l^{\text{th}}$ filter coefficient of $H_m$
$g_m(l)$	$l^{\text{th}}$ filter coefficient of $G_m$
$\delta_{m,i}, \delta_{m,j}$	horizontal and vertical displacement values in subband $m$

## Chapter 7

$\omega$	vector of spatio-temporal frequencies
$\mu$	upsampling and downsampling matrix
$n$	spatial index
$\mathbf{k}_1$	coset vector
$H_m(\omega)$	two-dimensional decomposition filter
$G_m(\omega)$	two-dimensional reconstruction filter
$\omega_t$	temporal frequency
$\omega_v$	vertical frequency
$\omega_h$	horizontal frequency
$z_m$	selection measure



# Samenvatting

## Hiërarchische Video Compressie met SBC

Een nieuw en fascinerend gebied van multimedia applicaties is momenteel in opkomst. Dit applicatiegebied omvat de traditionele applicatiegebieden van telecommunicatie, TV distributie, computers en consumentenelektronica en biedt consumenten een breed scala aan nieuwe en verbeterde toepassingen. De opkomst van multimedia applicaties is een gevolg van de convergentie van een aantal technologieën (digitale signaal verwerking, digitale transmissie- en opslagtechnieken en de ontwikkeling van geïntegreerde schakelingen) in combinatie met een aantal internationale standaardisatie activiteiten (JPEG, MPEG, DVB, en DAVIC). Het succes van deze nieuwe multimedia toepassingen zal grotendeels afhangen van de mate waarin multimedia systemen kunnen samenwerken. Indien verschillende multimedia systemen kunnen samenwerken dan zal de gebruiker een groot aantal applicaties kunnen gebruiken met behulp van een beperkt aantal apparaten. De mate waarin samenwerking tussen multimedia systemen mogelijk is kan vergroot worden door de videosignalen van multimedia applicaties schaalbaar te comprimeren en/of te representeren. Schaalbaar betekent hier dat het gecomprimeerde videosignaal op verschillende kwaliteitsniveaus gereproduceerd kan worden door meer of minder van de gecomprimeerde data te decomprimeren. Een dergelijke voorziening is zeer bruikbaar wanneer bijvoorbeeld een multimedia systeem een videosignaal met een hoge resolutie moet weergeven op een monitor met een lage resolutie of wanneer het digitale transmissiekanaal gevoelig is voor fouten. In het laatste geval kan een schaalbaar gecomprimeerd videosignaal zorgen voor een geleidelijke vermindering van de beeldkwaliteit wanneer fouten in het transmissiekanaal optreden ('graceful degradation').

Videosignalen zijn zeer geschikt voor schaalbare compressie doordat de inhoud van het signaal op vele kwaliteitsniveaus nog te herkennen is. Voor videosignalen kunnen vier basistypen van schaalbaarheid worden gedefinieerd; namelijk amplitude, frequentie, spatiële en temporele schaalbaarheid. Amplitude en frequentie schaalbaarheid zijn beide gerelateerd aan de hoeveelheid ruis in het signaal (aangegeven door de SNR waarde). Spatiële en temporele schaalbaarheid zijn gerelateerd aan respectievelijk de spatiële resolutie en de temporele resolutie (de beeldherhalingsfrequentie) van het videosignaal. Multimedia systemen kunnen hiërarchische coderingssystemen gebruiken om de videosignalen schaalbaar te comprimeren en te decomprimeren. Hiërarchische coderingssystemen implementeren schaalbaarheid op een efficiënte manier doordat ieder kwaliteitsniveau differentieel gecodeerd wordt ten opzichte van het vorige kwaliteitsniveau (met uitzondering van het laagste kwaliteitsniveau). In het bijzonder, deze systemen genereren een geordende set van bitstromen waarbij iedere bitstroom de informatie bevat die nodig is om van één kwaliteitsniveau naar het volgende hogere kwaliteitsniveau te gaan. Dientengevolge kunnen ontvangende systemen het videosignaal

representeren op verschillende kwaliteitsniveaus door de bijbehorende bitstromen te ontvangen en te decomprimeren. Een nadeel van hiërarchische coderingsystemen is dat de coderingsresultaten voor de verschillende kwaliteitsniveaus van elkaar afhankelijk zijn.

Dit proefschrift gaat over de schaalbare compressie van videosignalen met behulp van hiërarchische coderingsystemen. De belangrijkste aandachtspunten zijn het ontwerp en de implementatie van systemen die een optimaal of bijna optimaal coderingsresultaat bereiken voor ieder van de ondersteunde kwaliteitsniveaus. Alle ontworpen systemen zijn gebaseerd op de techniek van subbandcodering ('subband coding', SBC) welke zowel gebruikt wordt om de videosignalen te comprimeren als ook om de schaalbaarheid te implementeren. Een gevolg hiervan is dat de meeste hoofdstukken in deze verhandeling zich concentreren op een bepaalde klasse van compressiesystemen die een bepaalde type schaalbaarheid combineren met specifieke compressietechnieken. In dit proefschrift wordt de gemiddelde kwadratische fout ('mean square error') tussen het originele en het gereconstrueerde videosignaal gebruikt als maat voor de vervorming van het videosignaal tengevolge van de compressie.

Hoofdstuk 3 bevat een theoretische analyse van hiërarchisch coderen. Het doel van deze analyse is om inzicht te krijgen in het proces van hiërarchisch coderen en om inzicht te krijgen in de benodigde aandachtspunten tijdens het ontwerp van zulke systemen. De analyse concentreert zich op de vraag wanneer twee elementaire hiërarchische coderingsystemen het video signaal optimaal hiërarchisch kunnen coderen ('optimal successive refinement'). Dat wil zeggen wanneer voor beide kwaliteitsniveaus een optimale afweging bestaat tussen het aantal gebruikte bits per seconde (bitsnelheid) en de geïntroduceerde vervorming. Onder de aanname dat het videosignaal stationair en Gaussisch verdeeld is kan er aangetoond worden dat i) voor SNR schaalbaarheid (dat is de combinatie van amplitude en frequentie schaalbaarheid) optimale hiërarchische codering theoretisch altijd mogelijk is, en dat ii) voor spatiële schaalbaarheid optimale hiërarchische codering theoretisch alleen maar mogelijk is onder bepaalde condities. Uit de theoretische analyse blijkt verder dat het essentieel is om de fouten gemaakt tijdens de codering van het lage kwaliteitsniveau te compenseren tijdens de codering voor het hoge kwaliteitsniveau. De gevonden resultaten zijn vervolgens experimenteel geverifieerd met behulp van een eenvoudig en praktisch hiërarchisch coderingssysteem. De experimenten laten zien dat in de praktijk optimale hiërarchische codering niet altijd haalbaar is voor SNR schaalbaarheid. In het bijzonder wanneer meer dan drie kwaliteitsniveaus ondersteund worden door het coderingssysteem dan is het verlies in SNR waarde (bij gelijke bitsnelheid) aanzienlijk ten opzichte van vergelijkbare maar niet-schaalbare compressiesystemen. De oorzaak hiervan ligt in het feit dat totale nauwkeurigheid van meerdere aaneengeschakelde PCM (of DPCM) coders minder is dan de nauwkeurigheid van een enkele PCM (of DPCM) coder met een vergelijkbare bitsnelheid. Voor spatiële schaalbaarheid tonen de experimenten aan dat optimale hiërarchische codering in de praktijk mogelijk is wanneer twee kwaliteitsniveaus ondersteund worden met geschikte bitsnelheden.

Hoofdstuk 4 is een vervolg van hoofdstuk 3 en bevat een analyse van de coderingsresultaten van meerdere aaneengeschakelde PCM en DPCM coders. In het bijzonder worden de coderingsresultaten van specifieke varianten van geheugenvrije en voorspellende coderingstechnieken onderzocht, de zogenaamde 'multirate PCM' (MR-PCM) and 'multirate DPCM' (MR-DPCM) coders. MR-PCM en MR-DPCM coders kunnen worden gebruikt om amplitude schaalbaarheid direct te implementeren en zijn ook toepasbaar wanneer andere vormen van schaalbaarheid geïmplementeerd moeten worden. Multirate coders kunnen gebaseerd zijn op



'multistage' kwantisatie of 'embedded' kwantisatie. Met 'multistage' kwantisatie wordt het binnenkomende signaal gekwantiseerd door een aantal aaneengeschakelde kwantisatoren die ieder de kwantisatiefout van de vorige kwantisator kwantiseren. Met 'embedded' kwantisatie wordt het binnenkomende signaal tegelijkertijd gekwantiseerd door een aantal kwantisatoren die hun kwantisatiedrempels ('decision levels') hebben uitgelijnd. De beschreven analyse onderzoekt de coderingsresultaten van optimaal ontworpen MR-PCM en MR-DPCM coders die twee kwaliteitsniveaus ondersteunen. De analyse toont aan dat het coderingsresultaat voor het hoge kwaliteitsniveau van de coder gebaseerd op 'multistage' kwantisatie minder is dan het coderingsresultaat van de coder gebaseerd op 'embedded' kwantisatie. Bij deze laatste coder is het coderingsresultaat voor het hoge kwaliteitsniveau niet afhankelijk van de bitsnelheid gebruikt voor het lage kwaliteitsniveau en is zelfs vergelijkbaar met het coderingsresultaat van normale PCM (of DPCM) coders. Het coderingsresultaat voor het lage kwaliteitsniveau is echter weer wat beter met 'multistage' kwantisatie in vergelijking met 'embedded' kwantisatie. Hoofdstuk 4 wordt afgesloten met het ontwerp van nieuwe en efficiënte MR-PCM en MR-DPCM coders die gebruikt kunnen worden in praktische hiërarchische coderingssystemen.

Hoofdstuk 5 richt zich op spatieel schaalbare compressiesystemen voor ATM netwerken die gebruik maken van coderingstechnieken voor stilstaande beelden ('intraframe coding'). ATM netwerken kunnen schaalbare multimedia applicaties goed ondersteunen omdat alle informatie via logische verbindingen verstuurd wordt met onderhandelbare karakteristieken. ATM netwerken transporteren de informatie door het netwerk in kleine pakketjes welke van tijd tot tijd verloren raken vanwege congestie in het netwerk en/of fouten in de fysieke verbindingen. Communicatiesystemen voor ATM netwerken moeten hiertegen bestand zijn en kunnen bijvoorbeeld technieken voor 'graceful degradation' implementeren. In dit hoofdstuk worden twee 'intraframe' coderingssystemen ontworpen die het ontvangers mogelijk maken het videosignaal op drie kwaliteitsniveaus te reproduceren terwijl tegelijkertijd 'graceful degradation' ondersteund wordt. De coderingssystemen garanderen de continuïteit van de applicatie door middel van het laagste kwaliteitsniveau wat gereproduceerd kan worden met behulp van een bitstream met een lage constante bitsnelheid. De twee andere kwaliteitsniveaus worden ondersteund met twee bitstromen die een hoge en wisselende bitsnelheid hebben. Eén coderingssysteem maakt gebruik van 'multistage' kwantisatoren, terwijl het andere systeem gebruik maakt van 'embedded' kwantisatoren. De simulatieresultaten laten zien dat met normale netwerkcondities de hiërarchische coderingssystemen iets minder presteren voor het hoge kwaliteitsniveau dan een vergelijkbaar maar niet-schaalbaar coderingssysteem met een gelijke bitsnelheid. Het systeem met 'embedded' kwantisatoren benadert echter voor het hoge kwaliteitsniveau het coderingsresultaat van het niet-schaalbare systeem. Wanneer congestie in het netwerk optreedt dan tonen de simulatieresultaten aan dat de hiërarchische coderingssystemen minder gevoelig zijn voor transmissiefouten dan het niet-schaalbare coderingssysteem. De robuustheid van het coderingssysteem gebaseerd op de 'multistage' kwantisatoren is iets hoger dan de robuustheid van het coderingssysteem gebaseerd op 'embedded' kwantisatoren omdat de gegenereerde bitstromen iets minder afhankelijk van elkaar zijn. Deze resultaten tonen aan dat de verwachte gevoeligheid voor fouten van het transmissiekanaal een invloed zal hebben op de architectuur van het te ontwerpen coderingssysteem.

In hoofdstuk 6 komen spatieel schaalbare compressiesystemen aan de orde die gebaseerd zijn op bewegingsgecompenseerde compressietechnieken ('interframe coding'). Voor be-

paalde applicaties, bijvoorbeeld distributie van TV via aardse transmissiekanalen, is het gebruik van bewegingsinformatie noodzakelijk om het videosignaal voldoende te comprimeren. De integratie van bewegingsgecompenseerde compressietechnieken met spatieel schaalbare subbandcoderingssystemen is echter niet eenvoudig. Zo is bijvoorbeeld bewegingscompensatie in het subbanddomein erg moeilijk omdat objectverplaatsingen binnen de subbanden van opeenvolgende beelden verstoord worden door teruggevouwen frequentiecomponenten ('aliasing'). Verder moeten de compressiesystemen er rekening mee houden dat de temporele voorspelling van een bepaald kwaliteitsniveau alleen mag gebeuren op basis van informatie die beschikbaar is bij de ontvanger van dat kwaliteitsniveau. Indien dit niet gebeurt dan zal de beeldkwaliteit van het gereproduceerde kwaliteitsniveau drastisch verminderen omdat de temporele voorspellingen van de zender en de ontvanger uit elkaar gaan lopen. In hoofdstuk 6 wordt het ontwerp van een spatieel schaalbaar compressiesysteem beschreven dat gebruik maakt van geïnterpoleerde bewegingscompensatie in het subbanddomein ('in-band motion compensation'). Het compressiesysteem is voorzien van een aantal schattingsmethoden om de laag frequente subbanden van het hoge kwaliteitsniveau te voorspellen op basis van de gereconstrueerde subbanden van het lage kwaliteitsniveau en/of op basis van de gereconstrueerde subbanden van het vorige beeld. Afhankelijk van de gebruikte schattingsmethode implementeert het compressiesysteem een zuiver hiërarchisch coderingsconcept (lage kwaliteitsniveau schattingsmethode), een simultaan coderingsconcept (temporele schattingsmethode) of een concept daar tussenin (gemiddelde schattingsmethode en selectieve schattingsmethode). De selectieve schattingsmethode evalueert een aantal kandidaat voorspellingen en gebruikt de beste voorspelling. Uit de experimentele resultaten blijkt dat wanneer het bewegingsgecompenseerde compressiesysteem de lage kwaliteitsniveau of temporele schattingsmethoden gebruikt het de coderingsresultaten van een vergelijkbaar maar niet-schaalbaar compressiesysteem niet kan bereiken. Wanneer de selectieve schattingsmethode wordt gebruikt dan kan het verlies in coderingsresultaat ten opzichte van het niet-schaalbare bewegingsgecompenseerde compressiesysteem met succes verkleind worden.

Tenslotte worden in hoofdstuk 7 twee compressiesystemen beschreven welke het videosignaal met spatio-temporele schaalbaarheid kunnen comprimeren. Deze vorm van schaalbaarheid kan gebruikt worden om compatibiliteit te bewerkstelligen tussen een hiërarchie van progressief gescande videosignalen en geïnterlineerde videosignalen. Het hiërarchische compressiesysteem moet daartoe de progressief gescande videosignalen zo comprimeren dat het signaal gereconstrueerd kan worden als een geïnterlineerde videosignaal en omgekeerd. De beschreven compressiesystemen maken ieder gebruik van een specifieke driedimensionale (3-D) subbanddecompositie omdat de compatibiliteit tussen de voorgestelde hiërarchie van videosignalen een spatio-temporele frequentiedecompositie vereist. De compressiesystemen genereren meerdere bitstromen die ieder een dynamisch geselecteerde set van gecomprimeerde subbanden bevatten. Het eerste compressiesysteem is gebaseerd op de zogenaamde 'Conversion Strategy' welke alleen compatibiliteit biedt tussen de progressief gescande videosignalen. De geïnterlineerde videosignalen moeten bij dit compressiesysteem eerst geconverteerd worden naar een progressief formaat. Het tweede compressiesysteem implementeert de zogenaamde 'Direct Strategy' welke compatibiliteit biedt tussen alle videosignalen behalve het geïnterlineerde videosignaal met de hoogste spatiële en temporele resoluties. Dit videosignaal moet nog steeds geconverteerd worden voordat het schaalbaar gecomprimeerd kan worden vanwege fundamentele compatibiliteitsproblemen ten gevolge van de signaaldefinities. Het compressiesysteem is echter dusdanig ontworpen dat deze con-

versie niet zou moeten leiden tot een verlies in coderingsresultaat. Experimentele resultaten laten echter zien dat het compressiesysteem gebaseerd op de 'Conversion Strategy' veel betere resultaten geeft dan het compressiesysteem gebaseerd op de 'Direct Strategy'. Deze resultaten geven aan dat de beperkte reconstructiekwiteit van de 3-D subbanddecompositie van de 'Direct Strategy' een grote beperking is voor de toepasbaarheid van dit systeem. De coderingsresultaten van het systeem gebaseerd op de 'Conversion Strategy' zijn echter wel beter dan de resultaten van een niet-schaalbaar systeem gebaseerd op 'intraframe' codering. Hieruit kan de conclusie getrokken worden dat temporeel filteren van een beeldsequentie kan leiden tot een verbetering van het coderingsresultaat. Helaas is de subjectieve beeldkwaliteit van deze beeldsequenties niet zo hoog vanwege de vervaagde contouren en de aanwezigheid van echo's. Om deze vervormingen tegen te gaan met behoud van de winst in compressie moet de temporele filtering gebeuren langs de bewegingsvectoren van de objecten binnen het beeld, zoals recentelijk voorgesteld in de literatuur.

Uit het onderzoek beschreven in dit proefschrift kan men concluderen dat hiërarchische coderingssystemen gebaseerd op subbandcodering efficiënt videosignalen schaalbaar kunnen comprimeren. Deze schaalbaarheid kan gebruikt worden om compatibiliteit tussen videosignalen te bewerkstelligen alsmede ook om systemen minder gevoelig te maken voor fouten in het transmissiekanaal. In het algemeen zijn de coderingsresultaten van de resulterende schaalbare compressiesystemen iets minder dan de resultaten van vergelijkbare maar niet-schaalbare compressiesystemen. Het verschil in efficiëntie kan echter tot acceptabele waarden geminimaliseerd worden door het compressiesysteem zorgvuldig te ontwerpen.



# Acknowledgements

At the completion of this thesis, I would like to thank all people who contributed to this thesis. This gratitude goes also to the people who are not explicitly enlisted below but who showed continuous interest in my work.

First of all, I would like to express my gratitude to Jan Biemond for providing me the opportunity to perform this research and for introducing me into the international community of video coding. Special thanks goes to Inald Lagendijk for being a very inspiring co-author of many publications and for providing me the scientific guidance where required.

I also would like to express my gratitude to Prof. Dr. John W. Woods and his former Ph.D. student T. Naveen of the Rensselaer Polytechnic Institute in Troy (NY), USA, for their continuous interest in my work and the joint research we performed regarding multistage quantization.

My gratitude also extends to my former colleagues of the Information Theory Group in Delft for the pleasant teamwork on various projects and for providing me with an amiable working environment. Especially, I would like to thank Ricardo Belfor for the incredible effort we put together into the development and maintenance of the different software packages and toolkits for the HDSP group. Thanks also for my other successive roommates, Rene van der Vleuten, Richard Kleihorst, Hans Driessen, and Ruggero Franich, for their help on various problems and the numerous discussions we had on work and other things.

Further, I like to thank my current colleagues of the Advanced Systems and Application Laboratory in Eindhoven, in particular, Jan Raven, Ton Biemans, and Jef Pijnenburg. By showing their continuous interest in this work they certainly contributed to the completion of this thesis.

Finally, I would like to thank my family for their warm support and Marja for her direct support, her –mostly silent– understanding, and her patience during the past years.



# Curriculum Vitae

Frank Bosveld was born in Rotterdam, the Netherlands, on December 4, 1966. In 1985 he obtained his VWO diploma from the RSG Het Drachtster Lyceum in Drachten. He received his M.Sc. degree in Electrical Engineering cum laude in 1989 from the Delft University of Technology, in Delft, the Netherlands. His Masters thesis 'Hierarchical subband coding of HDTV in BISDN' was awarded with the 'Universiteitsfonds Delft Prijs' for the best Masters thesis in Electrical Engineering in the college-year 1989/1990.

From January 1990 until August 1994 he worked at the Information Theory Group of the Delft University of Technology towards a Ph.D. degree in the area of hierarchical video compression using subband coding. In that area he is the author of several publications in international scientific journals, conference proceedings, and books. During his work towards his Ph.D. degree he has been responsible for the definition of several Masters and pre-Masters projects as well as for the supervision of the students who were working on these projects. In addition, he made considerable contributions to the definition, creation, and maintenance of the infrastructure necessary to perform research in this area. In February 1994 he completed the course 'Management for Young Engineers' of the TSM Business School, in Enschede, the Netherlands.

In August 1994 he joined the Advanced System and Application Laboratory Eindhoven of Philips Sound and Vision in Eindhoven, the Netherlands. In this laboratory he has been working in the development of new multimedia products for video-on-demand and enhanced digital video broadcasting. Since March 1995 he has been participating in the international standardization activities for ISO/IEC 13818-6 (MPEG-2/DSM-CC) and DVB Data Broadcasting Services (DVB/SI-DAT).