Politechnika Poznańska
Wydział Elektryczny
Instytut Elektroniki i Telekomunikacji
ul. Piotrowo 3A, 60-965 Poznań

Sławomir Maćkowiak

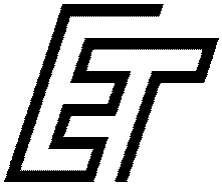
Skalowalne kodowanie cyfrowych sygnałów wizyjnych

Rozprawa Doktorska
Przedłożona Radzie Wydziału Elektrycznego
Politechniki Poznańskiej

Promotor: Prof. dr hab. inż. Marek Domański

Poznań, 2002

Politechnika Poznańska
Wydział Elektryczny
Instytut Elektroniki i Telekomunikacji
ul. Piotrowo 3A, 60-965 Poznań



Sławomir Maćkowiak

Scalable Coding of Digital Video

Doctoral Dissertation

Advisor: Prof. Marek Domański

Poznań, 2002

Contents

Acknowledgements

First and foremost, I would like to express my sincere thanks to my advisor Prof. Marek Domański for the truly memorable research experience I have had with him, for his invaluable support, and for his infinite patience.

Also, I owe many thanks to my colleague Adam Łuczak who helped me to implement some parts of the software, and to create the library for the software simulation of video codecs. I would like to thank Maciej Bartkowiak and Krzysztof Rakowski for discussing the English language phraseology.

I would like to thank Krystyna Ciesielska from Department of Foreign Languages of Poznań University of Technology, for careful reading of the manuscript and correcting and polishing my English.

I would like to thank the Foundation for Polish Science for the material support which allowed me to concentrate on the research.

I owe a huge debt of gratitude to my parents for their love, support, and humor throughout not only the course of my Ph.D. studies, but throughout my entire life. Thanks Mom and Dad.

Last but no way least, Agnieszka has made my life so much more worth it than anyone else, specially in the hard times of dissertation writing. In the last months, she gave me the support I needed and suffered with me as I was trying to juggle dissertation, teaching, sleeping, and didn't have much energy left for anything else.

Sławomir Maćkowiak

Poznań, 27 June, 2002.

Abstract

The dissertation is devoted to scalable coding of digital video. It discusses the topics of scalable video coding based on spatio-temporal decomposition. An original method of achieving multi-layer scalability is introduced. It is based on independent motion compensations in layers and additionally on modified prediction. A multi-loop encoder is introduced. Several versions of the encoder are presented and tested. The proposed encoder is characterized by low bitrate overhead, compared to the MPEG-2 standard solution. A new proposal of fine granularity scalability is included in the dissertation. The proposal solutions are supplemented by the description of software which has been used to construct the verification models of the proposed scalable encoders.

Streszczenie

Rozprawa poświęcona jest skalowalnemu kodowaniu cyfrowych sekwencji wizyjnych opartym na wykorzystaniu skalowalności przestrzenno-czasowej. W szczególności przedstawiona jest struktura wielopętlowego kodera skalowalnego wraz z dokonaniem porównania właściwości różnych zaproponowanych odmian tej struktury. Rozwiązanie to oparte jest o wykorzystanie niezależnych kompensacji ruchu w warstwach oraz zmodyfikowanego sposobu predykcji obrazów. Struktura ta charakteryzuje się niskim kosztem wprowadzenia skalowalności w porównaniu do rozwiązań zawartych w standardzie MPEG-2. W pracy przedstawiono wyniki badań eksperymentalnych zaproponowanego algorytmu kodowania. Jedną z oryginalnych propozycji zawartych w pracy jest koder ze skalowalnością drobnoziarnistą. W pracy przedstawiony jest także oryginalny sposób uzyskania skalowalności wielopoziomowej w oparciu o podział współczynników DCT między warstwy. Uzupełnieniem prezentowanych rozwiązań jest opis oprogramowania, które zostało przygotowane w celu konstrukcji modeli weryfikacyjnych proponowanych koderów skalowalnych.

List of symbols and abbreviations

ΔLL - corrections of the LL subband,

ΔMVx - rounded differences between components *x* of the base and enhancement layer motion vectors,

ΔMVy - rounded differences between components *y* of the base and enhancement layer motions vectors,

2-D - two-dimensional,

3-D - three-dimensional,

4CIF - progressive 4:2:0 704x576 pixels video sequence,

A_H - horizontal activity,

A_V - vertical activity,

AC - DCT coefficient for which the frequency in one or both dimensions is non-zero,

APBIC - Advance Predicted Bitplane Coding,

B_b - base layer bitrate,

$B_{e_1},...,B_{e_{N-1}}$ - bitrates of enhancement layers 1,…,N-1,

B_S - scalable bitrate,

B_{SL} - bitrate achieved with single-layer coding,

B_{SIM} - total bitrate achieved with simulcast coding,

B-frame - bi-directionally interframe encoded frame,

bitrate - number of bits per second,

bitrate (bits per frame) - number of bits in one frame,

$bitrate_{prev}$ - number of bits in the previous I-frame,

BE-frame - B-frame occurring only in the enhancement layer,

BR-frame - B-frame occurring in base and enhancement layers,

CBR - constant bitrate mode,

CIF - progressive 4:2:0 352x288pixels video sequence,

D - motion vectors search range,

$d_e^i(m_1,m_2)$ - distortion of the macroblock in the enhancement layer,

DC - DCT coefficient with zero frequency in both dimensions,

DCT - Discrete Cosine Transform,

dct_b - DCT coefficients in base layer,

dct_e - DCT coefficients in enhancement layer,

dct_m - DCT coefficients in middle layer,

DeMUX - demultiplexer,

DPCM - Difference Pulse Code Modulation,

DWT - Discrete Wavelet Transform,

e_i - difference between corresponding pixels in the original image and distorted image,

e_e^i - prediction error in the enhancement layer,

EOB - End of Block,

EZW - Embedded Zerotree Wavelet,

$f_b^i(m_1, m_2)$ - original value of macroblock (m_1, m_2) in the i^{th} low-resolution frame (the base layer),

$f_e^i(m_1, m_2)$ - original value of macroblock (m_1, m_2) in the i^{th} high-resolution frame (the enhancement layer),

$\hat{f}_b^i(m_1, m_2)$ - macroblock (m_1, m_2) spatially interpolated from the i^{th} low-resolution frame,

$\hat{f}_e^i(m_1, m_2)$ - predicted macroblock (m_1, m_2) in the i^{th} high-resolution frame (the enhancement layer),

$f_{LL}(n_1, n_2)$ - value of pixel (n_1, n_2) in subband LL,

$F(n_1, n_2)$ - DCT coefficient (n_1, n_2),

$F'(n_1, n_2)$ - quantized DCT coefficient (n_1, n_2),

$F_b(n_1, n_2)$ - DCT coefficient (n_1, n_2) from the base layer image,

$\widetilde{F}'_b(n_1, n_2)$ - dequantized DCT coefficient (n_1, n_2) from the base layer image,

FGST - Hybrid temporal-FGS,

FGS - Fine Granularity Scalability,

FIR - Finite Impulse Response,

FM - frame memory,

fps - number of frames per second,

$G(n_1, n_2)$ - element (n_1, n_2) of Intra quantization matrix,

$G(n_1, n_2)_{HH}$ - element (n_1, n_2) of Intra-mode quantization matrix for subband HH,

$G(n_1, n_2)_{HL}$ - element (n_1, n_2) of Intra-mode quantization matrix for subband HL,

$G(n_1, n_2)_{LH}$ - element (n_1, n_2) of Intra-mode quantization matrix for subband LH,

$G(n_1, n_2)_{LL}$ - element (n_1, n_2) of Intra-mode quantization matrix for subband LL,

$g_e^i(m_1, m_2)$ - weighted predicted macroblock,

GOF - Group of Frames,

GOP - Group of Pictures,

$H(z)$ - frequency response of the system,

HDTV - High Definition Television,

HH - high-high-spatial frequency subband,

HL - high-low-spatial frequency subband,

H_t - high-temporal-frequency subband,

i - number of frame,

I-frame - intraframe encoded frame,

IDCT - Inverse Discrete Cosine Transform,

Intra - intraframe,

ITU - International Telecommunication Union,

JM - Joint Model,

JPEG - Joint Photographic Experts Group,

JVT - Joint Video Team,

K - number of macroblocks in horizontal direction,

Kbps - kilobits per second,

L - number of macroblocks in vertical direction,

level - absolute value of a non-zero coefficient,

LH - low-high-spatial frequency subband,

LL - low-low-spatial frequency subband,

LSB - least significant bit,

L_t - low-temporal-frequency subband,

m_1 - position of macroblock in horizontal direction,

m_2 - position of macroblock in vertical direction,

MASCOT - Metadata for Advanced Scalable Video Coding Tools,

MB-PFGS - Macroblock-based Progressive Fine Granularity Scalability,

mb-PFGSS - Macroblock-based Progressive Fine Granularity Spatial Scalability,

Mbps - megabits per second,

MC-FGS - Motion-Compensated Fine Granularity Scalability,

MCP - motion-compensated predictor,

ME - motion estimator,

MOMUSYS - Mobile Multimedia Systems,

MOS - Mean Opinion Score,

MP@ML - Main Profile @ Main Level,

MPEG - Motion Pictures Expert Group,

MSB - most significant bit,

mv_b - base layer motion vectors,

mv_h - high resolution layer motion vectors,

mv_l - low resolution images motion vectors,

mv_m - middle layer motion vectors,

mv_b - base layer motion vectors,

MV_e - full-frame motion vectors,

mv_e - enhancement layer motion vectors,

MVx - motion vector, component *x*,

MVy - motion vector, component *y*,

n_1 - pixel position in a block in horizontal direction,

n_2 - pixel position in a block in vertical direction,

N_1 - size of a frame in horizontal direction,

N_2 - size of a frame in vertical direction,

N_m - control parameter influencing allocation of bits to the base layer and enhancement layers,

NINT - nearest integer value,

NMSE - Normalized Mean Squared Error,

P-frame - interframe encoded frame,

PFGS - Progressive Fine Granularity Scalability,

PFGST - Macroblock-based Progressive Fine Granularity Scalable-Temporal method,

PPVR - personal video recorder,

PSNR - Peak Signal to Noise Ratio,

Q - quantization parameter,

$Q(\ )$ - quantization function,

Q_b - quantization parameter in base layer,

Q_e - quantization parameter in enhancement layer,

Q_{cur} - calculated quantization parameter for the current frame,

Q_{prev} - Q value used in the previous frame,

QMF - quadrature mirror filter,

QoS - Quality of Service,

RC - current reference frame,

RN - next reference frame,

RP - previous reference frame,

RUN, run - number of zero coefficients preceding a non-zero coefficient in the scan order. The absolute value of this non-zero coefficient is called "level".

SDTV - Standard Definition Television,

SIF - progressive 4:2:0 360x288 pixels video sequence,

SMPTE - Society of Motion Picture and Television Engineers,

SNR - Signal to Noise Ratio,

SOT - Spatial Orientation Tree,

SPIHT - Set Partitioning in Hierarchical Trees,

SW - switch,

T - predefined threshold,

TF - Temporal Filtering,

TM - Test Model,

VBR - Variable Bitrate Mode,

VLD - Variable Length Decoder,

VLE - Variable Length Encoder,

x - pixel index in the horizontal direction,

y - pixel index in the vertical direction.

Chapter 1

Introduction

1.1. Scalable video coding

The dissertation deals with special topics in lossy digital video data compression [Doma98d, Skar98]. Lossy compression enables very efficient data representation at the price of some distortion occurring in decoded pictures. Lossy video coding is crucial for video communication.

Video and multimedia communication services have been developing rapidly in the last years. Their availability depends strongly on communication network infrastructure. High bitrates needed for video transmission impose severe requirements on communication networks. The existing networks are very inhomogeneous. Links are characterized by different bitrates, by different error rates, by different levels of Quality of Service (QoS) [IETF97] (Fig. 1.1). Different levels of Quality of Service are often related to different available transmission bitrates.

On the other hand, service providers demand the data to be broadcasted only once to a group of users accessed via heterogeneous links. Different networks are indeed different groups of users who have varying expectations. The same content may reach different group of users. For this purpose, the transmitted bitstream should be partitioned into some layers. The layers represent different quality or different spatial

and/or temporal resolutions of video. The low resolution layers represent a video with reduced quality or reduced resolution. Reduced quality is achieved by progressive coding, while reduced resolution is achieved by hierarchical coding. The higher layers represent a high resolution video. Lower layers can be decoded independently from higher layers. The quality of decoded video depends on the number of decoded layers. This functionality is called scalability [Ohm01b]. Various layers 1,...,N correspond to subsystems N1 to NN of a heterogeneous communication network (Fig. 1.2). In this network, low resolution terminals are able to decode only the low-resolution bitstream. As a result, they display low resolution pictures. High resolution terminals are able to decode all the layers and display high resolution pictures. The low resolution terminals connected to links characterized by full Quality of Service are able to decode, due to their features, only the low resolution bitstreams.

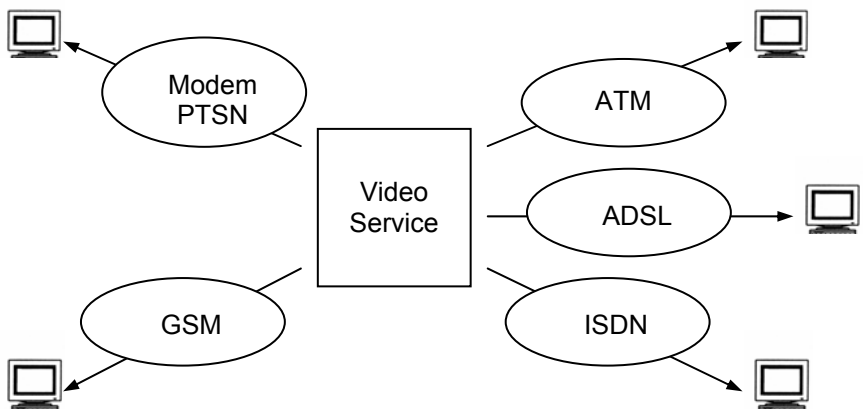


Fig.1.1. Heterogeneous communication network with video services.

The lowest layer is called the base layer. The others layers are called the enhancement layers.

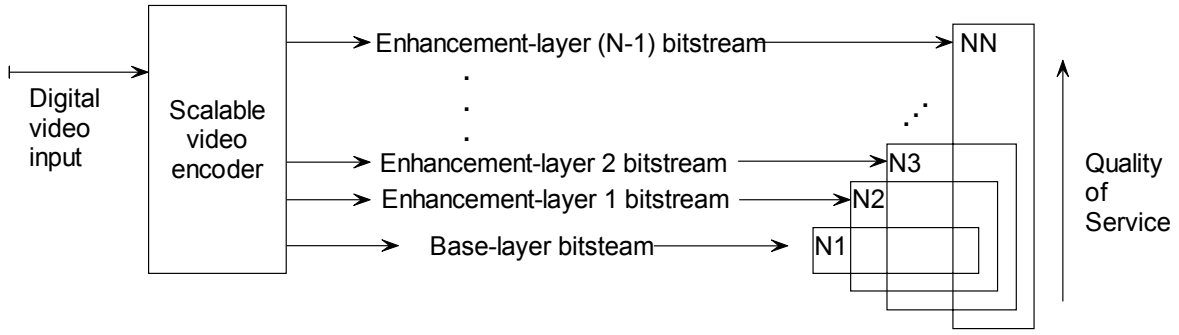


Fig.1.2. Multi-layer scalable video coding system in heterogeneous network that consists of sub-networks N1,…, NN characterized by different throughputs.

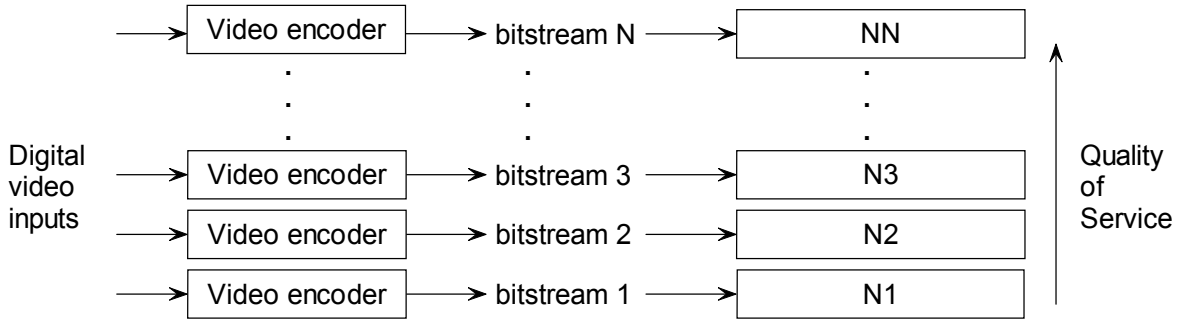


Fig.1.3. Simulcast video coding system.

Scalability of video means the ability to achieve a video of more than one resolution or quality simultaneously. Scalable video coding involves generating a coded representation (bitstream) in a manner that facilitates the derivation of video of more than one resolution or quality from this bitstream.

In scalable video coding, the total bitrate B_S is the sum of layer bitrates

$$B_s = B_b + B_{e_1} + ... + B_{e_{N-1}}, \tag{1.1}$$

where: B_b - base layer bitrate,

$B_{e_1},...,B_{e_{N-1}}$ – bitrates of the enhancement layers 1,…,N-1.

In simulcast coding, each bitstream of video is associated with a certain resolution or quality and is encoded independently (Fig.1.3). Thus, any bitstream can be decoded by

a single-layer decoder. The total bitrate B_{SIM} required for transmission of encoded streams is the sum of bitrates of these streams:

$$B_{SIM} = B_1 + B_2 + ... + B_N ,$$
where N is the number of bitstreams. (1.2)

Therefore, this technique is not efficient.

For a given overall decoded video quality, scalable coding is not acceptable in common applications, if the bitrate is significantly greater than the bitrate achieved in single-layer coding. Likewise, it would be unacceptable if it was in any way less efficient than simulcast coding, i.e. if $B_s > B_{SIM}$.

In practice, bitrate B_S is higher than bitrate B_{SL} required for transmission of data in single layer coding, i.e. $B_s > B_{SL}$.

The difference $B_S - B_{SL}$ is the cost of scalability, called bitrate overhead.

Lossy (irreversible) video coding is described by the rate-distortion curve [Wieg96, Sull98] that defines the relationship between the quality of the decoded image and the bitrate required for transmission of the encoded image. Decreasing quality of the decoded image (i.e. increasing distortion) accompanies increasing compression ratio (i.e. decreasing bitrate) (Fig. 1.4).

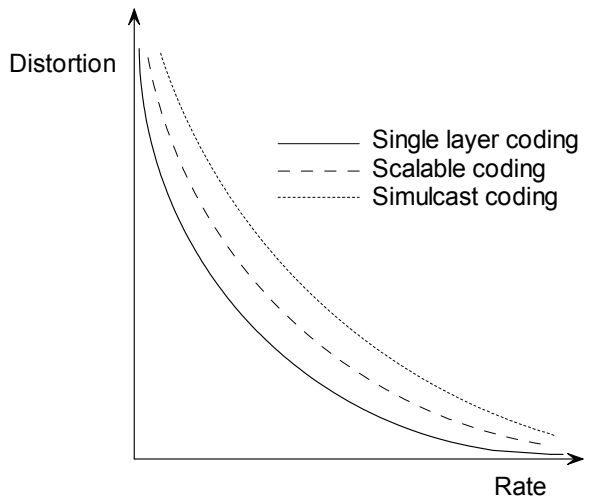


Fig. 1.4. Hypothetical rate-distortion curves.

In order to find the trade-off between bitrate and distortion, it is necessary to measure distortion. However, visual distortion is very difficult to measure, as the human visual system is complex and not well understood. This problem of measuring distortion is aggravated in video coding, because the addition of the temporal domain further complicates the issue.

There exist two ways of evaluating the quality of decompressed images:

- subjective evaluation by assessment of the quality of the decompressed images or by assessment of impairments between the original and decompressed images, made by a panel of viewers,
- objective evaluation by pixel-wise calculation of the error between the original image and the decompressed image.

Subjective evaluation consists in averaging the scores given by a number of viewers. The result is called the mean opinion score (MOS). The assessment of television images is normalized by the International Telecommunication Union (ITU). The respective Recommendation BT.500 [ITUR94] defines experimental conditions such as the minimum number of participants, lighting conditions, timing etc.

In objective evaluation, normalized mean squared error (NMSE) or peak signal to noise ratio (PSNR) [Bert98, Doma98d, Skar98] are used in most comparisons. For monochrome pictures, the peak signal to noise ratio is given by

$$PSNR[dB] = -10 \cdot \log_{10}\left(\frac{\sum_i e_i^2}{N \cdot 255^2} \right) \tag{1.3}$$

where:

- e_i - difference between corresponding pixels in the original image and the distorted image,
- N - number of pixels in an image,
- 255 - maximum possible magnitude of pixel in the original image, which is directly related to the dynamic range.

The results obtained using PSNR do not often coincide with those obtained in subjective tests. It is well known that small errors, randomly distributed over the whole

image, reflect badly in objective measures, i.e. the values of PSNR are small. However, they rarely correspond to visible degradation of image. On the other hand, relatively high values of errors concentrated in a particular part of an image result in relatively high PSNR, while the subjective assessment is very low because of an annoying corruption of a particular portion of the image.

Nevertheless, PSNR is commonly used for calculation of distortion. PSNR is easy to calculate and gives optimistic results. Therefore, this may account for its popularity in reporting the results of compression.

The distortion and quality of images are related to the type of used scalability.

The basic types of scalability are:

- Spatial scalability (also called hierarchical coding) - the encoder produces two or more layers with the same frame rate, but different spatial resolutions.
- Temporal scalability – the encoder produces two or more layers with the same spatial resolution, but different frame rates.
- SNR[1] scalability – the encoder produces two or more layers with the same frame rate and the same spatial resolution, but different quality.

The existing video compression standards MPEG-2 [ISO94, Hask96, Skar98] and MPEG-4 [ISO98] define scalable profiles, which exploit classic Discrete Cosine Transform-(DCT)-based schemes with motion compensation. The MPEG-2 and MPEG-4 standards have defined all the basic types of scalability techniques mentioned above. Unfortunately, spatial scalability as proposed by the MPEG-2 coding standard is inefficient because the bitrate overhead is too large. Additionally, the solutions defined in MPEG-2 do not allow flexible allocation of the bitrate. There exists a great demand for flexible bit allocation to individual layers, i.e. for fine granularity scalability (FGS) [Li01, Li01b], which is also already proposed for MPEG-4, where the fine granular enhancement layers are intraframe encoded. The coding efficiency of this solution is reduced; however, it may be improved by using inter-picture prediction in the enhancement layers [LiWu00].

The scalability is expected to find many applications. The goal of scalable coding is to provide interoperability between different services and to flexibly support receivers

[1] Signal to Noise Ratio (SNR)

characterized by different display capabilities. For example, flexible support of multiple resolutions is of particular importance in interworking between High Definition Television (HDTV) and Standard Definition Television (SDTV), in which case it is important for a HDTV receiver to be compatible with a SDTV application. Compatibility of the receivers can be achieved by means of scalable coding of the HDTV source. Moreover, the transmission of two independent bitstreams to the HDTV and SDTV receivers can be avoided.

Other important applications of scalable coding include video database browsing and multiresolution playback of video in multimedia environments.

The importance of scalability is being more and more recognized as more attention is paid to video transmission in error-prone environments, such as wireless video transmission systems [IEEE99, Giro00, Hanz01]. A video bitstream is error-sensitive due to extensive employment of variable-length coding. A single transmission error may result in a long, undecodable string of bits. It has been shown that an efficient method of improving transmission error resilience is to split the coded video bitstream into a number of separate bitstreams (layers) transmitted via channels with different degrees of error protection. The base layer is better protected, while the enhancement layers exhibit a lower level of protection [Arav96, Gall01, Ghar97, Sted93, Stuh99, TanZa99]. A receiver is able to reproduce at least low-resolution pictures if Quality of Service decreases.

Multicasting of video over the Internet is expected to be an important technology area for many multimedia applications in the 21st century [Horn97]. These applications include the viewing of major television events by millions of Internet users around the globe. A wide range of interactive and non-interactive multimedia Internet applications, such as news on-demand, live TV viewing, and video conferencing rely on end-to-end streaming solutions [Giro97, Giro98, Giro99]. In general, these solutions are required to maintain real-time delivery and presentation of the multimedia content while compensating for the lack of Quality-of-Service (QoS) guarantees [Camp95] over Internet-Protocol-(IP)-based networks [IETF98]. Therefore, any Internet streaming system has to take into consideration crucial network performance parameters such as bandwidth, delay, delay variation, and packet loss rate. Scalable video coding is capable of

supporting a wide range of bandwidth variation scenarios that characterize IP-based networks [Scha01, Scha01b].

Until recently, the cost of non-linear storage of video was an economic constraint to professional applications [Gemm95]. However, with the falling costs of hard disk drives and memories and with the MPEG standard compression becoming more and more widespread, non-linear video storage is a consumer product. For example, a personal video recorder (PPVR) will be able to store an appropriate part of bitstream. It will be possible because new scalable video coding decouples the encoding process from the network and makes if independent from the terminal capabilities. The user will be able to choose resolutions, frame rates, display aspect ratio etc.

1.2. Goals and thesis of the work

To date, different types of scalability have been proposed. Unfortunatelly, existing video compression standards, like MPEG-2 and MPEG-4, define hierarchical video coding which is not effective enough to scalable video coding. Therefore, there exists a great demand for efficient algorithms of scalable video coding.

The goal of the work is to propose modifications of the existing video compression algorithms in order to obtain scalable video coding systems with possibly low bitrate overhead (as defined in Sec. 1.1).

Spatial scalability is already defined by the MPEG-2 video coding standard. Unfortunately, its implementation based on pyramid decomposition is inefficient because the number of pixels to be encoded is increased as compared to a non-scalable scheme. Therefore encoders with spatio-temporal scalability are developed and considered in this dissertation.

Scalability is also offered by the still image JPEG-2000 coding standard [Taub02]. This scalability is, however, beyond the scope of this dissertation.

In the dissertation, the following assumptions are made.

- Motion-compensated interframe prediction coding is used as the basic reference coding technique [ISO94, Hask96].

- High level of compatibility with the MPEG video coding standards should be ensured. Moreover, a scalable encoder should mostly consist of functional blocks present in the standard MPEG-2 encoder.
- Since the algorithm of coding interlaced video is very sophisticated its implementation is beyond the scope of this work. Only the scalable encoding of progressive sequences is considered and tested.

The ***main thesis*** of the dissertation is as follows:

- It is possible to improve the efficiency of scalable video coding using spatio-temporal scalability. It is possible to achieve high coding efficiency, which will make the proposed methods competitive with respect to video encoders standardized to date as well as those being standardized currently.

The ***particular goal*** of the thesis is to extend the coding algorithm already defined by the MPEG-2 standard.

1.3. Overview of the dissertation

The dissertation is organized in seven chapters. Chapter 2 presents to-date knowledge regarding scalable video coding. In particular, the following techniques are discussed:

- the basic scalability techniques used in the MPEG-2 standard,
- the fine granularity scalability technique introduced in the MPEG-4 standard,
- the techniques based on wavelet / subband decompositions,
- other approaches to scalability, used in hybrid encoders.

Spatio-temporal scalability is introduced in Chapter 3. This chapter explains the terms used in the following chapters and describes the general structure of the proposed solutions.

Chapters 4, 5 and 6 describe new scalable video coding techniques based on spatio-temporal scalability developed by the author. Examples of numerous simulation experiments using test sequences are given.

Chapter 4 deals with a new system with a three-dimensional filter bank used for spatio-temporal analysis and a two-dimensional subband encoder with motion compensation, developed by the author.

Chapter 5 proposes spatio-temporal scalability based on the partitioning of data related to B-frames and on exploitation of the interpolated low resolution images from the base layer as reference frames. The scalable encoder proposed in this chapter exploits independent motion estimation and compensation for base and enhancement layer.

Chapter 6 deals with spatio-temporal scalability combined with fine granularity in a scheme based on a slightly modified structure of the MPEG-2 standard.

Since the literature on scalability is enormous, the references have been grouped into the present author's contributions, other cited papers and other works directly related to the topic.

This dissertation has been partially supported by the Polish National Committee for Scientific Research under Grant 8 T11D 009 17 (“Hierarchical video compression techniques for MPEG systems in heterogeneous communication networks”) and by the IST Programme of the EU under contract number IST-2000-26467 (“MASCOT - Metadata for advanced scalable video coding tools”).

Chapter 2

Review of scalable coding techniques

2.1 Introduction

This chapter briefly reviews the existing scalable video coding techniques. The classification of scalability techniques is difficult because the techniques use elements of one another and two techniques can be combined to form a hybrid scalability technique. In this thesis, scalable video coding techniques have been classified into four groups that are not necessarily disjoint. Other classifications are also possible.

The classification is as follows:

- The basic scalability techniques used in MPEG-2 and MPEG-4 standards.
- The Fine Granularity Scalability technique introduced in the MPEG-4 standard and developed in the H.26l standard.
- The techniques based on wavelet / subband decomposition.
- Other approaches to scalability, used in hybrid encoders.

2.2 Basic scalable techniques in MPEG-2

Historically, the MPEG-2 standard [Hask96, ISO94, LeGa92] was the first video coding standard that introduced scalability. It defined four types of scalability, namely: spatial, temporal, SNR and data partitioning [Doma97]. These functionalities are also included into the MPEG-4 standard [ISO98, Koen97, Pere97].

In MPEG-2, many video coding algorithms were integrated into single syntax to meet diverse application requirements. New coding features were added to achieve sufficient functionality and image quality. Thus, prediction modes were developed to support efficient coding of interlaced video. In addition, scalable video coding extensions were introduced to provide additional functionalities, such as embedded coding of digital TV and HDTV, and to ensure graceful quality degradation in the presence of transmission errors. However simultaneous implementation of the full syntax is not necessary for most applications.

Table 2.1. Profiles of the MPEG-2 standard.

Profile	Algorithm
Simple	No scalable coding
Main	No scalable coding
SNR scalable	2 layers, SNR scalable coding
Spatial scalable	2 layers, Spatial scalable coding
High	3 layers, SNR and spatial scalable coding
Multiview	2 layers, temporal scalable coding
4:2:0	No scalable coding

The MPEG-2 standard is organized as a toolkit of algorithmic features, corresponding to syntax elements and tools for decoding semantics, with the goal of supporting many current applications as well as the future ones. Often, several applications may have similar functional requirements that can be served by a set of

common tools. These similar functional requirements are addressed by the concept of profiles [Okub95]. Currently, MPEG-2 includes 4 profiles that support scalability (Table 2.1.), i.e. the SNR profile, the Spatial profile, the High profile and Multiview profile. Non-scalable video coding within MPEG-2 is primarily supported in the Main profile and the Simple profile. The Main profile is aimed at single-layer coding of the 4:2:0 format video.

2.2.1 Spatial scalability

Spatial scalability [Hask96, ISO94] refers to layered coding in the spatial domain. It means that in this type of coding scheme the encoder produces two layers with the same frame rate, but with different spatial resolutions. The spatial resolution of the base layer is lower than the input sequence. The enhancement layer is used to transmit the information needed for restoration of the full spatial resolution. The reconstructed base layer picture is up-sampled to form the prediction of a high resolution picture in the enhancement layer. An important feature of scalable bitstreams is that the base layer bitstream can be decoded independently from other layers. Therefore the base layer may be coded using different video formats, providing interoperability between standards. This functionality is useful for many applications, including embedded coding for HDTV/TV systems, allowing migration from a digital TV service to higher spatial resolution HDTV services.

Figure 2.1 shows a single-loop spatially scalable codec. The input video sequence is down-sampled to a low resolution sequence and constitutes the input sequence for the base layer encoder. The base layer encoder is a motion-compensated DCT-based encoder. The base layer bitstream is fed to the system multiplexer. The locally decoded picture from the base layer is up-sampled to the size of the original sequence and then combined with the temporal prediction signal generated within the enhancement layer encoder. The coded bitstream produced by the enhancement layer encoder is also fed to the system multiplexer. At the decoder, the base layer decoder decodes the base layer bitstream. The locally decoded picture from the base layer decoder is up-sampled and

used for prediction of a full resolution frame in the enhancement layer decoder. The decoder outputs a high resolution video sequence.

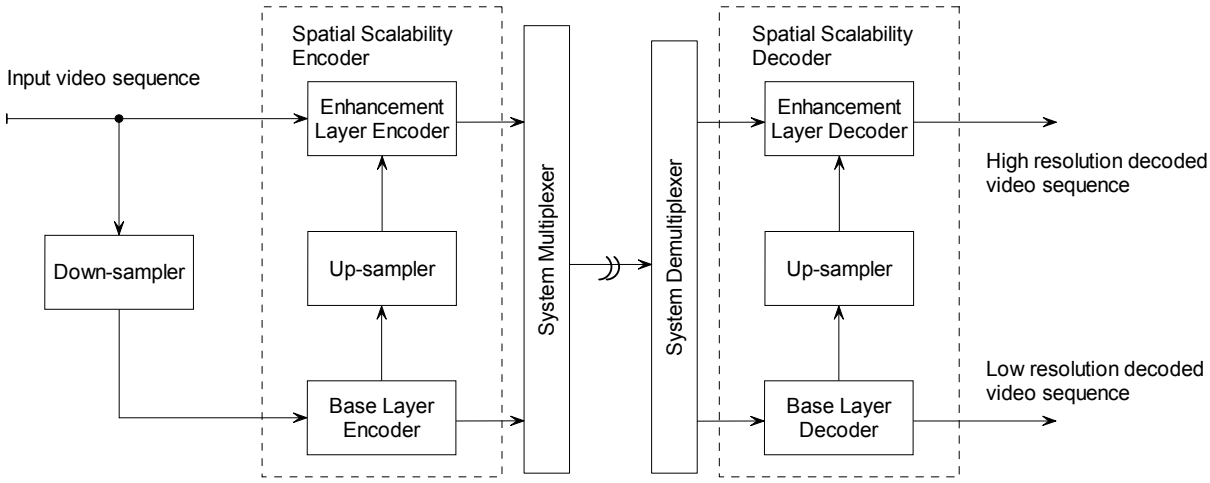


Fig. 2.1. Spatial scalability two-layer codec (based on [Hask96]).

Unfortunately, the bitrate overhead involved in the application of the MPEG-2 spatial scalability is unacceptably high, compared to single-layer MPEG-2 video encoding. There exist some experimental results in which the total bitstream is not much smaller than the sum of bitstreams obtained for simulcast transmission with two different resolutions [Benz96, Benz99].

The author has prepared test software in order to perform a test of scalable encoding [See Annex B]. The software was written in the C++ language. It includes a software implementation of the MPEG-2 Main Profile @ Main Level encoder for the base layer and a verification model of the scalable encoder. The correct operation of the encoder was verified during several tests of selected blocks of the scalable encoder with the MPEG-2 verification model [Mpeg96]. The values of PSNR of selected images and their bitrate were tested. The software also provides an implementation of the MPEG-2 encoder, which has been cross-checked with the MPEG-2 verification model. The encoders process progressive 720 x 576, 50 Hz, 4:2:0 test sequences. The frames are organized into Groups of Pictures (GOPs). The length of a GOP is 12 frames. A GOP consists of one I-frame, two P-frames and nine B-frames. Every three B-frames are located between consecutive I- and/ or P-frames.

The results obtained by the author are presented in Table 2.2. Table 2.2 shows a comparison between spatial scalable coding, non-scalable coding and simulcast coding for two progressive BT.601 test sequences. The conclusion is that spatial scalability coding yields better results than simulcast coding but is worse than non-scalable coding.

Table 2.2. Comparison between spatially scalable coding, non-scalable coding and simulcast coding for two progressive BT.601 test sequences (50 fps) (prepared by the author).

Test sequence		Average PSNR of luminance[dB]	Total bitrate [Mbps]	Overhead [%]
FunFair	Non-scalable coding	32.14	5.01	
	Simulcast coding	32.14	6.75	34.7
	Spatially scalable coding	32.11	5.99	19.5
Cheer	Non-scalable coding	31.85	5.10	
	Simulcast coding	31.85	6.73	32.0
	Spatially scalable coding	31.80	5.91	16.0

Puri and Wong [Puri93] have focused spatially scalable coding and have compared the performance of spatially scalable video coding with the simulcast technique, at two different bitrates. Their results have indicated that spatially scalable video coding provides significant improvement in performance when compared to the simulcast technique.

2.2.2 Temporal scalability

Temporal scalability [Hask96, ISO94, Demo95, Hema99] refers to coding that produces two ore more layers. The layers may have different temporal resolutions. When combined, these layers provide full temporal resolution as available in the input video. The spatial resolution of each layer is the same as that of the input signal. The temporal scalability allows the base layer encoder to process data at a lower frame rate. The

enhancement layer provides the missing frames to form a video sequence with a full frame rate.

Puri, Yan and Haskell [Puri94] introduced the concept of temporal scalability within the context of motion-compensated DCT coding structure defined by MPEG-2. This concept was later included in the MPEG-2 standard. The experiments have proved that temporal scalability provides a significant improvement of 1 to 4 dB over the simulcast technique.

A two-layer temporal scalability codec structure (Fig. 2.2) consists of a temporal demultiplexer, a temporal re-multiplexer, a base layer encoder and decoder, a system multiplexer, a system demultiplexer, and a enhancement layer encoder and decoder. The input video sequence is fed to the temporal demultiplexer. The demultiplexer splits the frames into two video sequences. One video sequence is encoded in the base layer encoder, giving the base layer bitstream. The other sequence is encoded in the enhancement layer encoder, giving the enhancement layer bitstream. These bitstreams are multiplexed by the system multiplexer into a single stream.

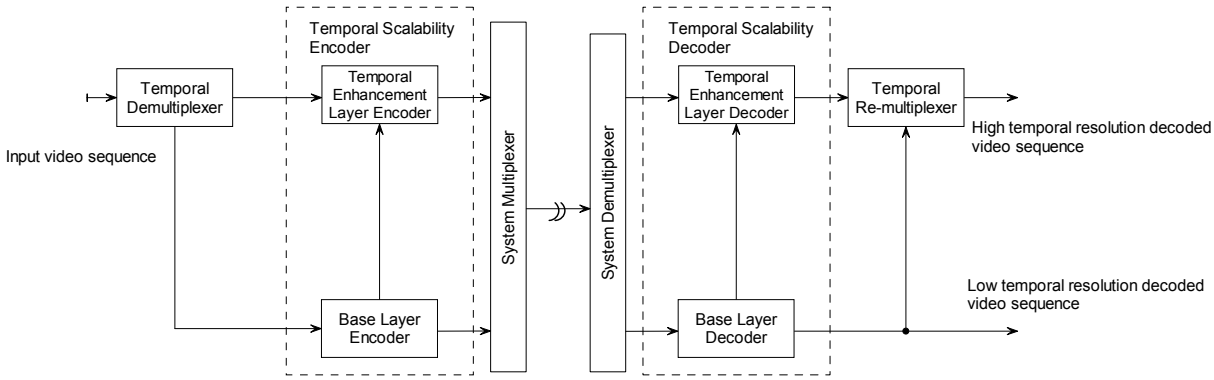


Fig. 2.2. The temporal scalability two-layer codec (based on [Hask96]).

2.2.3 Signal to Noise Ratio (SNR) scalability

Signal to Noise Ratio (SNR) scalability [Hask96, ISO94, Kond98] refers to layered coding that produces more than one layer, each with different picture quality. All layers have the same spatial and temporal resolution. The base layer contains coarsely quantized DCT coefficients. The enhancement layer carries refinement information from which

coefficients at a finer quantization scale can be obtained. In the two-layer case, SNR scalability involves coding the base layer as in the Main profile, whereas the enhancement layer encoder encodes the difference between the enhancement layer signal and the prediction signal, obtained as a weighted combination of the up-sampled decoded base layer signal and motion-compensated prediction obtained on the basis of the previously decoded enhancement layer signal. The base layer signal is essential for the decoder to reconstruct the sequence of pictures. With the additional reception of the enhancement layer signal, the video quality can be improved.

The block diagram of an SNR encoder is shown in Fig. 2.3. This encoder consists of a motion compensation prediction loop and two quantizers, Q_1 and Q_2. In the loop, the intra or inter macroblocks are discrete cosine transformed. The coefficients are then quantized using the first coarse quantizer Q_1. Then the quantized coefficients are variable-length coded (VLC) [Doma98d, Doma98e, Pita00, Skar93] and sent together with the required side information. The enhancement layer is derived by subtracting the base layer quantized coefficients from their original values in the frequency domain. The enhancement layer fine quantizer Q_2 with a fixed step size determines the overall picture quality.

The encoder described above prevents the accumulation of drift in the enhancement layer, since differential refinement coefficients are fed back into the lower layer motion compensation prediction loop in the encoder. However, if there are channel errors, only the base layer is decoded and displayed.

Wilson and Ghanbari [Wils96] have demonstrated that significant correlation exists between the layers of SNR coded video. To reduce this correlation they proposed a different approach to SNR scalability has been proposed by them.

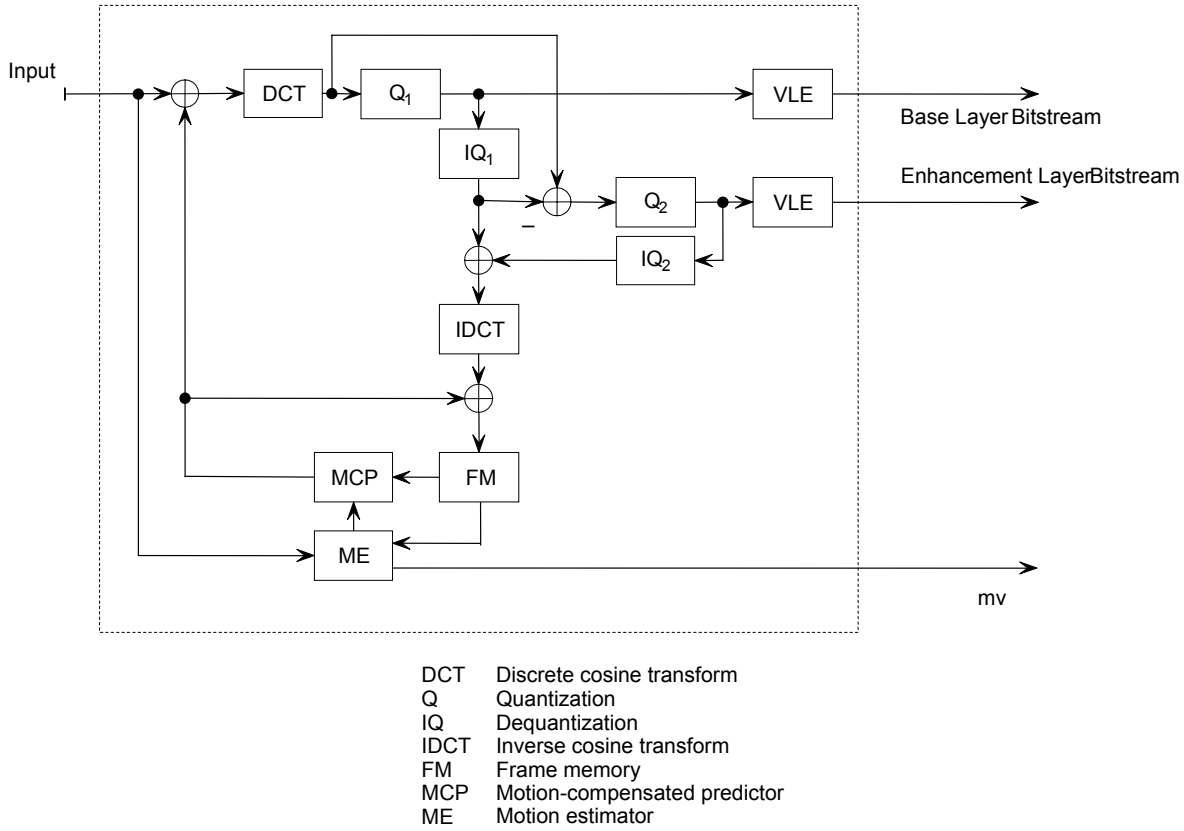


Fig. 2.3. SNR scalability encoder without drift in the enhancement layer.

In their approach, it is the difference between the reconstructed base image and the original input image which is quantized. Fig. 2.4 illustrates the codec structure where the base layer is generated by an independent MPEG-2 encoder using coarse quantization. Next, the reconstructed base image is subtracted from the original to get the spatial domain quantization error, which is then encoded by the enhancement layer encoder, itself another independent MPEG-2 encoder stage. The decoder is complementary to the encoder. Therefore it also consists of two independent MPEG-2 decoders whose reconstructed signals are added just before presentation.

Typically, the enhancement layer adopts a quantizer with a smaller step size to refine the residual information of the base layer. When scalability is introduced, a small degradation generally occurs in the coded picture quality. Hsu et al. [Hsu99] have presented a fine-tuned encoding structure which provides better performance than the previous approach. Hsu has applied an embedded quantizer in the MPEG SNR scalable encoder. The cascade quantization is removed, and the complexity of the encoder is dramatically reduced because only a single quantization device is necessary. The

experimental results have proved that PSNR of the proposed embedded encoder is much closer to the single layer encoder than the standard SNR scalable encoder at any given bitrate.

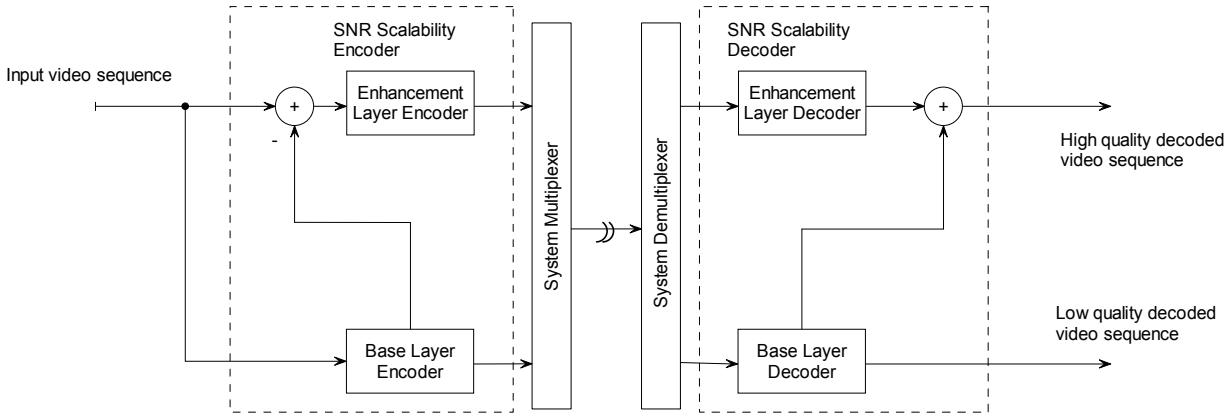


Fig. 2.4. Block diagram of SNR scalability two-layer codec (based on [Hask96]).

2.2.4 Data partitioning

Data partitioning [Hask96, ISO94], although not a true form of scalable coding, provides a means of partitioning coded video data into two priority classes: essential data and additional data. In MPEG-2, a coded video bitstream is divided into two layers, called partitions, such that partition 0 contains address and control information as well as lower order DCT coefficients, while partition 1 contains high-frequency DCT coefficients. Partition 0 is referred to as the base partition (or the high priority partition) whereas partition 1 is called the enhancement partition (or the low priority partition). A codec with data partitioning is shown in Fig. 2.5.

Haskel, Puri and Netravali [Hask96] have compared the base layer for data partitioning to a single layer bitstream, each at 2.25Mbps. The results have shown that the quality achieved for the base layer of data partitioning is quite poor, and is significantly lower than that achieved with non-scalable coding.

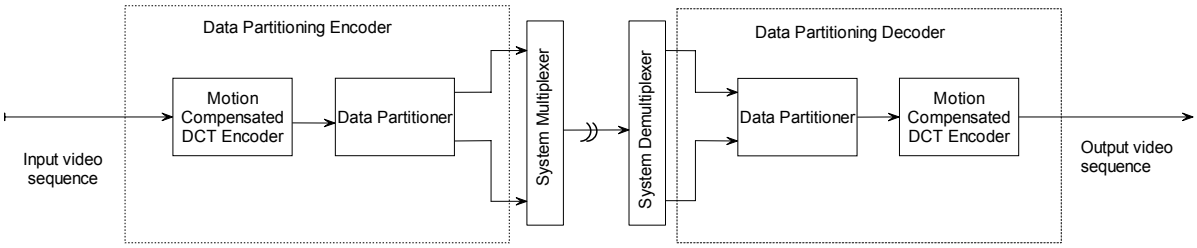


Fig. 2.5. Data partitioning codec (based on [Hask96]).

2.2.5 Hybrid scalability

Beside the above types of scalability, the MPEG-2 standard also supports hybrid scalability, which involves using two different types of scalability from among spatial, temporal and SNR scalabilities. Figs. 2.6 and 2.7 show two examples of hybrid scalability: SNR and spatial, SNR and temporal.

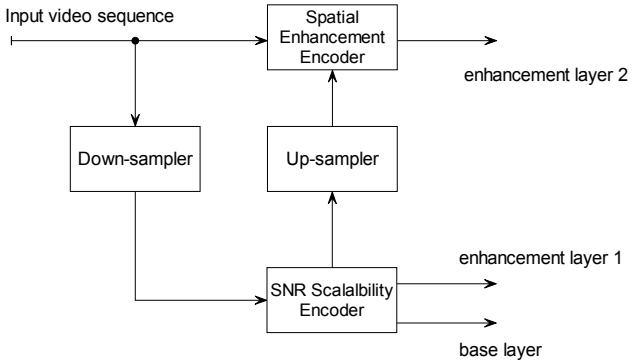


Fig. 2.6. Hybrid scalability: SNR and spatial.

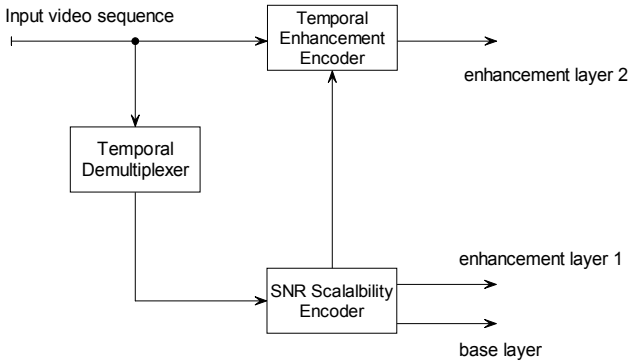


Fig. 2.7. Hybrid scalability: SNR and temporal.

2.3 Fine Granularity Scalability

In a traditional communications system, the encoder compresses the input video signal into a bitrate that is less than, but close to, the channel capacity. The decoder reconstructs the video signal using all bits received from the channel. In such a system, two basic assumptions are made. The first is that channel capacity is known. The second is that the decoder is able to decode all the bits received from the channel fast enough to reconstruct the video. These two basic assumptions are challenged in streaming video applications. Due to varying channel capacity, the encoder no longer knows the channel capacity and does not know at what bitrate video quality should be optimized. Therefore, the objective of video coding for streaming video is changed to optimizing the video quality over a given bitrate range. Fine Granularity Scalability (FGS) enables to produce scalable bitstream depending on varying channel capacity.

The basic idea of FGS is to code a video sequence as a base layer and an enhancement layer. The base layer uses non-scalable coding to reach the lower bound of the bitrate range. The enhancement layer is used to code the difference between the original picture and the reconstructed picture. The bitstream generated by the FGS enhancement layer encoder may be truncated by any number of bits per picture after encoding is completed. The decoder should be able to reconstruct the high resolution video from the base layer and the remaining enhancement layer bitstream. However, the high resolution video quality is proportional to the number of bits decoded by the decoder for each picture.

The MPEG-4 standard offers FGS for scalable video coding [Li00, Li01b, Li02, Ohm01b]. Initially, three types of techniques were proposed for encoding DCT coefficients encoding in FGS, namely, bitplane coding [Li99b, Ling99, Li01, LiLi97] of DCT coefficients, wavelet coding of image residue and matching pursuit coding of image residue [LiJi99, Vlee00, Vlee00b]. After several core experiments, bitplane coding of the DCT coefficients was chosen due to its implementation simplicity while preserving coding efficiency.

As shown in Fig. 2.9, the FGS framework requires two encoders, one for the base layer and the other for the enhancement layer. The base layer can be compressed using

any motion compensation DCT video encoding method. The FGS enhancement layer encoder can be based on a fine granular coding method.

In the bitplane coding, for each 8x8 DCT block 64 absolute values (decimal numbers) are zig-zag ordered into an array. The decimal numbers are then transformed to binary numbers. A bitplane of each block is now defined as an array of 64 bits taken from the same bit position. For every bitplane of each block, RUN and End of Plane symbols are formed and variable-length coded to produce the output bitstream. Experiments show that bitplane coding is always more efficient than run-level coding. Up to 20% of bit savings can be achieved by using bitplane coding [Li01b]. Figs. 2.8 and 2.9 show the FGS encoder and decoder structures, respectively.

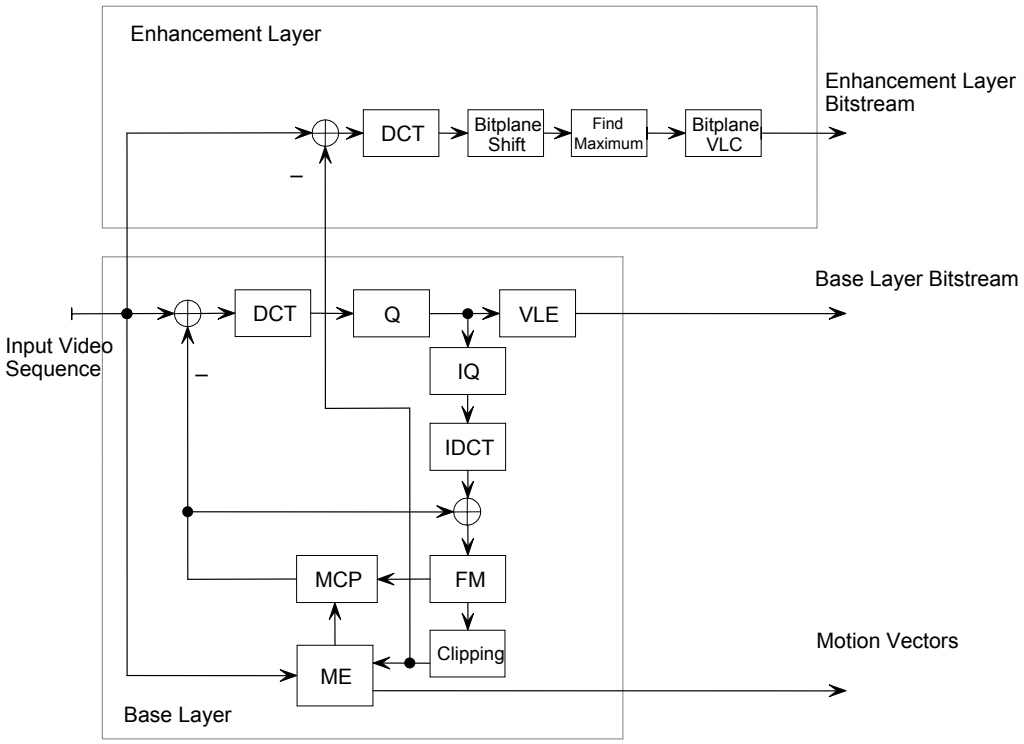


Fig. 2.8. FGS encoder structure (based on [Li01b]).

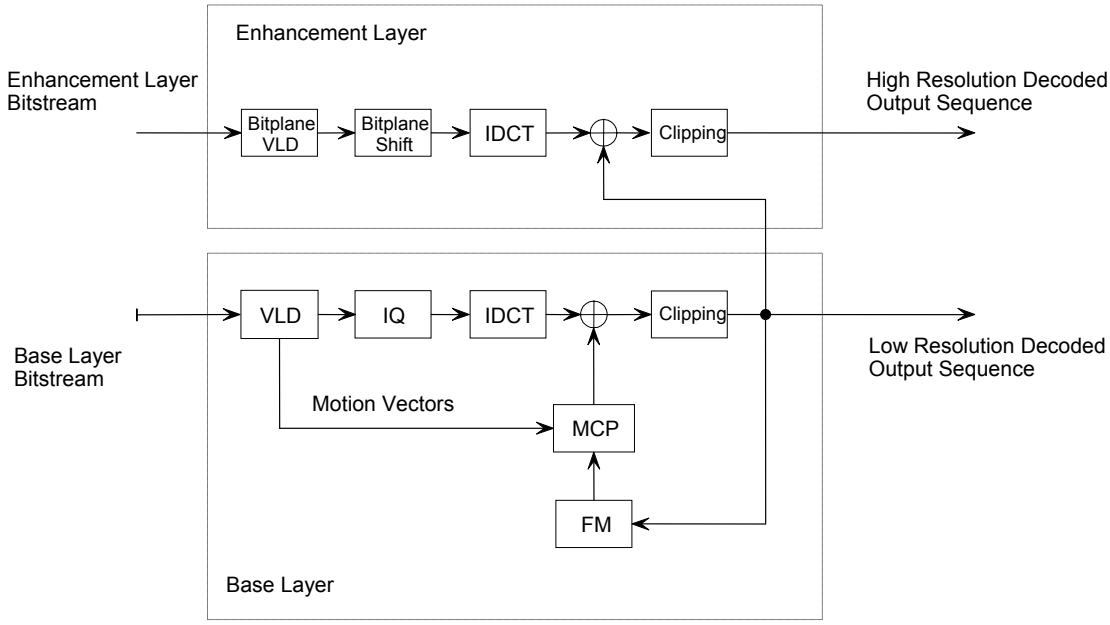


Fig. 2.9. FGS decoder structure (based on [Li01b]).

The FGS encoder has only one motion compensation loop in the base layer. The reference in the motion compensation loop is always reconstructed from the low-quality base layer to avoid possible drifting errors caused by truncations of enhancement bitstream. But low-quality reference makes motion compensation inefficient in FGS coding. Compared with non-scalable coding, the FGS coding may lose 2.0dB or more at the same bitrate [HeYa01].

Li [Li99] has compared the FGS coding technique with multi-level SNR scalability and simulcast coding. The experimental results have proved that FGS reference technique is much better than the multi-level SNR scalability, but about 2dB worse than non-scalable single-layer coding.

Table 2.3. Comparison between FGS reference technique, multi-level SNR scalability and simulcast coding (based on [Li99]).

		Multi-Level SNR		FGS		Simulcast	
Sequence	Level	bitrate (kbps)	PSNR Y (dB)	bitrate (kbps)	PSNR Y (dB)	bitrate (kbps)	PSNR Y (dB)
Carphone	Base layer	17.07	29.95	17.59	30.04	20.00	29.56
	Enhancement layer	119.85	35.16	120.37	37.18	121.64	33.18
Coastguard	Base layer	16.35	26.41	15.68	26.22	16.03	26.04
	Enhancement layer	98.90	29.91	98.23	31.86	128.12	27.22
Foreman	Base layer	17.93	27.39	17.98	27.35	18.48	27.22
	Enhancement layer	100.17	31.89	100.23	34.07	127.81	30.32

2.3.1 Hybrid temporal-FGS (FGST)

The limitation of the FGS is that the frame rate is the same as the input video frame rate, regardless of the available bandwidth. To ensure optimal viewing conditions, and to take into account the user's preferences, it is necessary to combine FGS and temporal scalability. Van der Schaar and Radha [Scha99, Scha00c, Scha01] have proposed single layer hybrid temporal-SNR scalability scheme called FGST.

The proposed hybrid temporal-SNR scalability provides total flexibility in supporting SNR scalability and the frame rate. Fig. 2.10 shows a hybrid temporal-SNR scalability structure. In the presented implementation, the base layer contains only I and P frames.

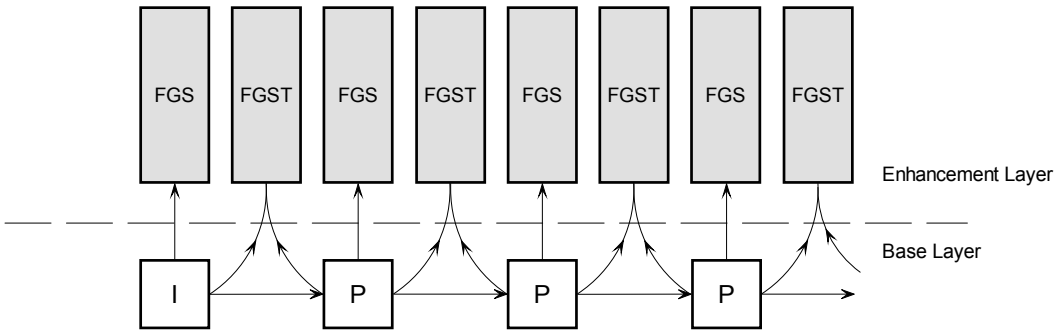


Fig. 2.10. FGS hybrid temporal-SNR scalability structure. FGS denotes a frame which is encoded with fine granularity scalability; FGST denotes a frame which is encoded with hybrid temporal-SNR scalability (based on [Scha01]).

As shown in Fig. 2.10, each frame which is hybrid temporal-SNR encoded in the enhancement layer is predicted from two base layer frames. These two frames do not coincide temporally with the encoded frame.

The enhancement layer includes two types of information:

1) motion vectors, which are computed with respect to the temporally adjacent base layer frames,

2) bitplane coded residual signals.

These two sets of information are coded and transmitted using a data partitioning strategy. Unlike the base layer, where the motion vectors and corresponding texture data are sent macroblock by macroblock, all motion vectors for an FGST frame are clustered and transmitted first. Subsequently, the coded representation of DCT bitplanes residual signal is sent [Chen99a, Chen99b].

The encoder is able to encode both the temporal and FGS enhancement layers and the server can choose which enhancement layer and which portion of layers should be transmitted, depending on the available channel bandwidth. For some types of content, it is desirable to transmit the base layer only, while at higher bitrates, improving the temporal resolution leads to better visual results.

Fig. 2.11 shows the architecture of a hybrid temporal-SNR FGS encoder. The SNR FGS residual is computed directly in the DCT domain. The FGST residual is computed in the pixel domain because motion compensation takes place there. Therefore, the FGST residual frames have to be DCT transformed before they are

bitplane-based-entropy encoded. In addition, and as mentioned above, the FGST residual is computed using a motion-compensated frame from the base layer. The functional blocks available for the encoding of the base layer can be reused for FGST. This can be achieved through data-flow control within the FGS encoder. The sharing of resources is feasible because the encoder never compresses a base layer frame and an FGST frame at the same instant. Similarly, the SNR FGS entropy encoder can be shared between the SNR FGS and FGST frames, since these picture types are never compressed at the same instance of time.

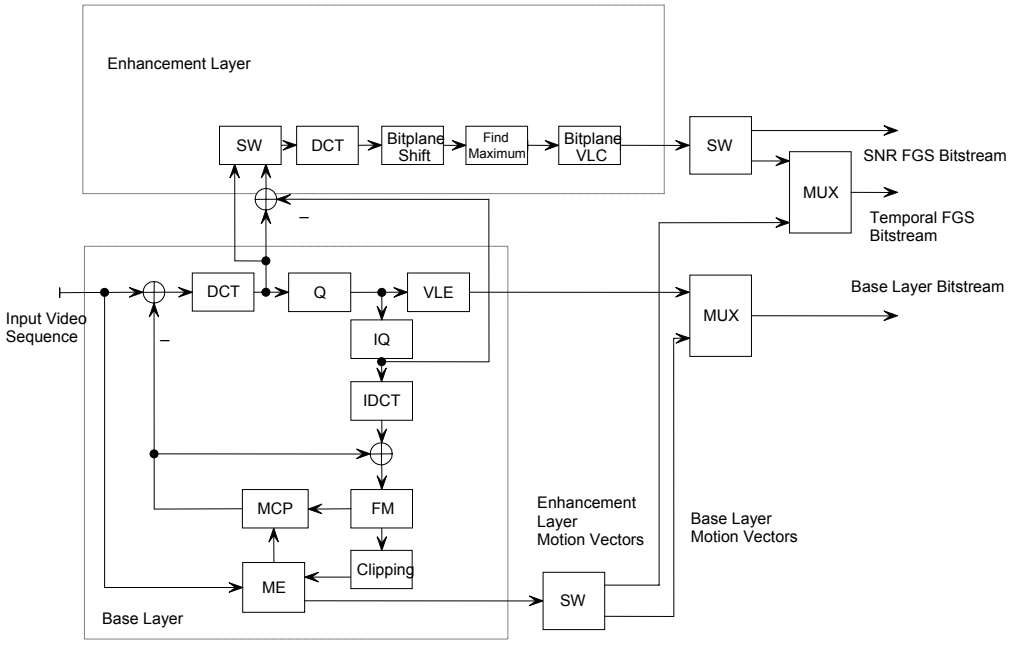


Fig. 2.11. Hybrid FGS temporal-SNR encoder: SW- switch FGST/FGS (based on [Scha01]).

Fig. 2.12 shows the architecture of a corresponding hybrid temporal-SNR FGS decoder. The IDCT of the FGST residual signals can be computed using the IDCT block of either the enhancement layer decoder or the base layer decoder. This is possible because the base layer IDCT block is not in use when the receiver is decoding a FGST frame. However, both IDCT blocks are in use when decoding a FGS SNR frame (one block is used for computing the IDCT of the FGS residual and the other for computing the IDCT of the corresponding base-layer frame). Therefore, they cannot be used to decode the FGST residual at this moment.

The motion estimator in the encoder produces two sets of motion vectors, one set for the base layer and the other for the enhancement layer. Motion vectors associated with FGST frames are multiplexed with the enhancement layer bitstream. In the decoder, the temporal FGS bitstream is demultiplexed to separate motion vectors from other information. The motion vectors are used by the motion compensation block to create a FGST predicted frame. In the enhancement layer, the residual signal is decoded and the IDCT signal is computed. The two signals are added together to generate a FGST frame.

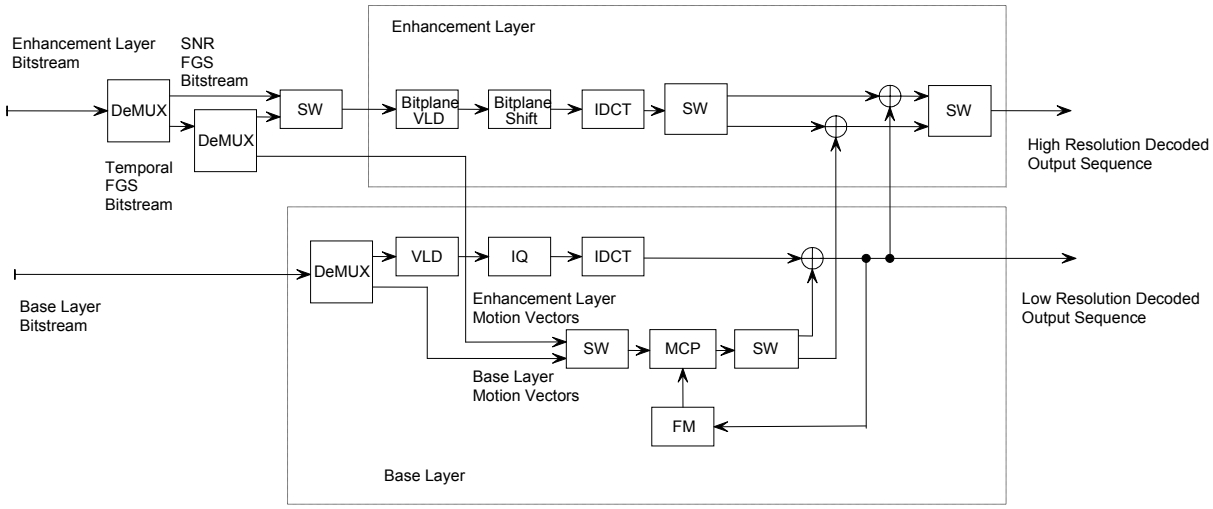


Fig. 2.12. Hybrid FGS temporal-SNR decoder: SW- switch FGST/FGS (based on [Scha01]).

In 1999, van der Schaar and Radha proposed a multi-layer FGST scheme [Scha99, Scha01, Scha01b]. In this scheme, the FGS layer is coded in addition to the base and temporal enhancement layers and therefore enhances the SNR quality of all frames. The temporal enhancement layer contains only B-frames, which are predicted from the corresponding I- and P-frames in the base layer. This multi-layer FGST structure is shown in Fig. 2.13. The experimental results have proved that the FGST and the multilayer FGS-temporal scalability achieve the same efficiency.

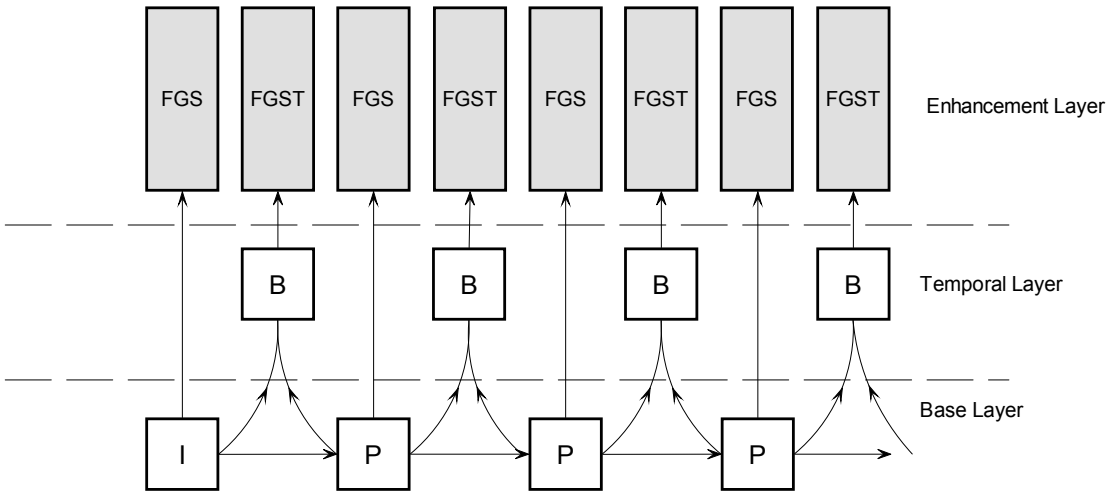


Fig. 2.13. Multilayer FGST scalability structure (based on [Scha99]).

2.3.2 Progressive FGS (PFGS)

Li, Wu and Zhang have proposed an improvement of the FGS solution called Progressive FGS (PFGS) [LiWu99, LiWu00], using the experimental results from [Macn99]. In the traditional FGS scheme, all the enhancement layers in the prediction frames are coded with reference to the base layer frames (Fig. 2.14). Since frame prediction is always based on the low-quality base layer, the coding efficiency of FGS coding is not as good as, and sometimes much worse than, the traditional SNR scalability schemes. The traditional schemes use accurate prediction of the same layer in the reference frames, and they normally provide better coding efficiency. An error in enhancement layers propagates to the end of a GOP and causes drift in higher layers in the next prediction frames. Even though there is sufficient bandwidth available later on, the decoder cannot recover the highest quality until another GOP starts.

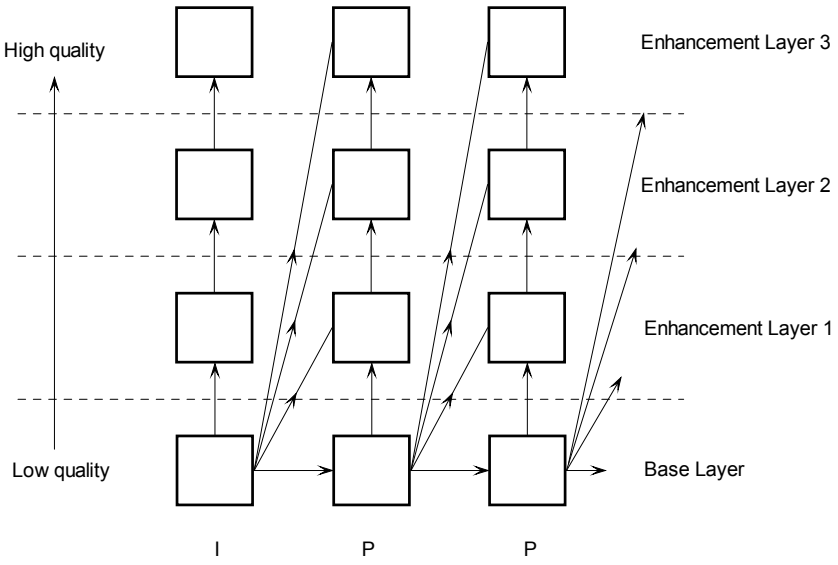


Fig. 2.14. Framework of the FGS scheme (based on [LiWu99]).

Li, Wu and Zhang have used as many predictions in one layer as traditional SNR scalability schemes do (see Fig. 2.15). They have observed that the changes in channel conditions are not abrupt since there is always a stream buffer to smooth them. Therefore channel adaptation in their scheme can span several frames. They introduced two key solutions in their scheme. The first one is to use as many predictions from the same layer as possible (for improving coding efficiency of the higher layers). The second one is to establish a prediction path which always uses prediction from a lower layer in the reference frames (for error recovery and channel adaptation). The first solution makes sure that the motion prediction is as accurate as possible to maintain coding efficiency. The second point makes sure that there is no drift in case of channel congestion, packet loss or packet error. And with this coding structure, there is never a need to retransmit lost/error packets since the higher video layer can be automatically reconstructed over a few frames.

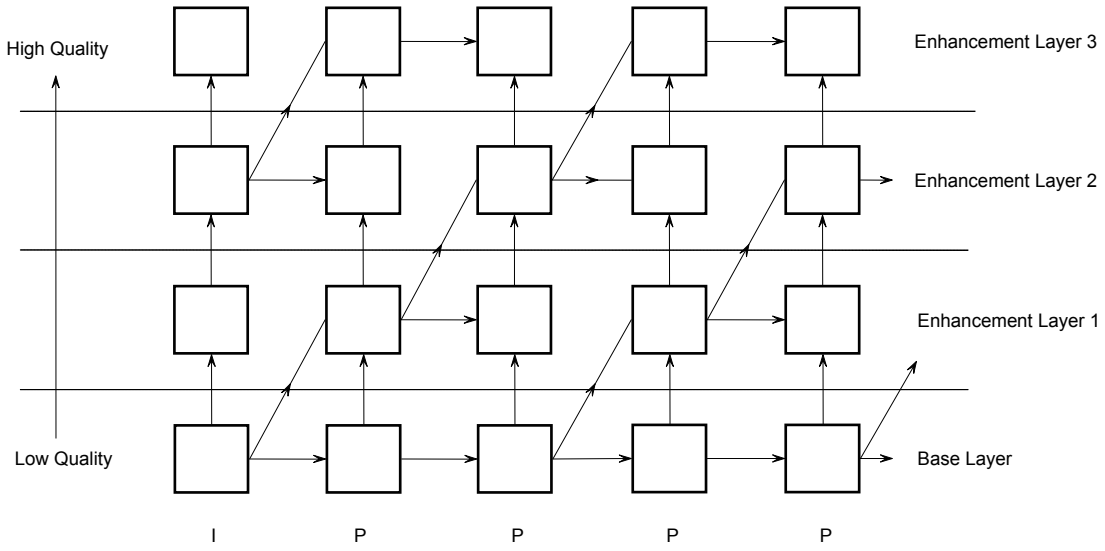


Fig. 2.15. Framework of the PFGS scheme (based on [LiWu99]).

Another feature of the PFGS scheme is that the coding of higher layers, the prediction of some layers in the prediction frame (P-frame) is based on the reconstructed reference frame from the lower layer. The correlation between the lower layer and the current layer in the prediction frame can also be exploited to raise coding efficiency, as presented in Fig. 2.15. In the PFGS scheme, each frame in the base layer is always predicted from the previous frame in the base layer and each frame in an enhancement layer is predicted from the previous frame in the enhancement layer. Because the quality of a frame is higher in the enhancement layers than in the base layer, PFGS coding provides more accurate motion prediction than FGS coding.

The experimental results [LiWu99] have shown that the PFGS achieves a bitrate reduction of up to 15% at the same PSNR, or up to 1dB PSNR gain at the same bitrate in comparison to the standard FGS scheme.

Li, Wu and Zhang have achieved better coding efficiency in higher enhancement layers by using the Advance Predicted Bitplane Coding (APBIC). In this scheme, the lower enhancement layer prediction can be obtained using higher enhancement layer reference. The unique feature of the APBIC scheme is that in frame prediction, the reference used for prediction may be different from the reference used for reconstruction. A more detailed description of APBIC can be found in [LiWu00].

2.3.3 Macroblock-based PFGS (MB-PFGS)

In 2001, Wu, Li, Sun, Yan and Zhang proposed further improvements of the PFGS scheme [WuLi01]. In their solution, called Macroblock-based PFGS, the PFGS scheme is extended from the frame level to the macroblock level. Three prediction modes of the enhancement layer macroblock coding are presented in Fig. 2.16. Gray rectangular boxes in Fig. 2.16 denote frames in layers to be reconstructed as references. Solid arrows with solid lines represent temporal predictions, hollow arrows with solid lines represent reconstruction of high quality references, and solid arrows with dotted lines represent predictions in the DCT domain. The three modes use different references for prediction and reconstruction of the enhancement layer. The coding mode of each macroblock indicates which temporal prediction should be used to reconstruct the high quality reference and the display frame in the enhancement layer.

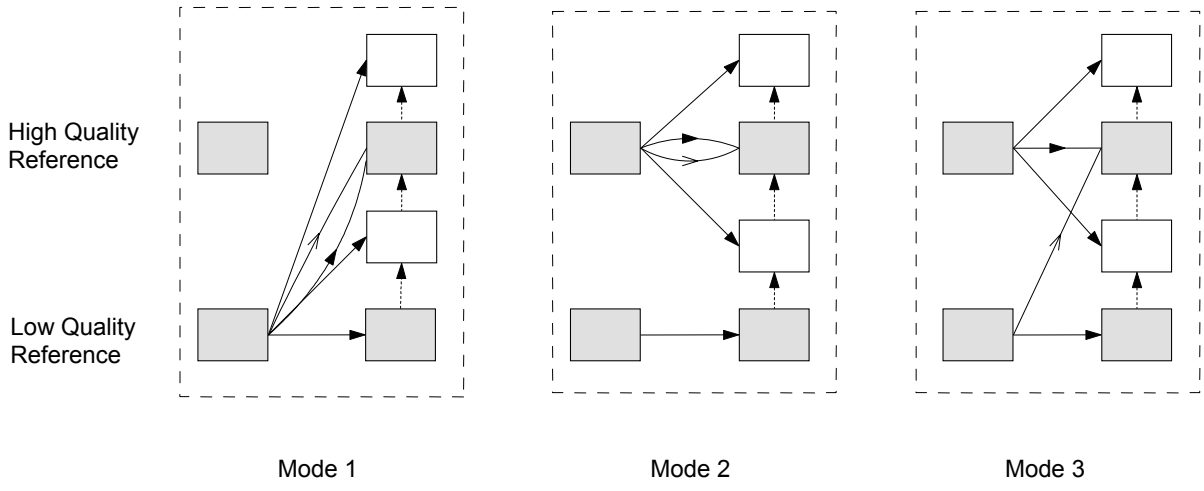


Fig. 2.16. Prediction modes of the enhancement macroblock. Gray rectangular boxes denote the layers which will be reconstructed as references (based on [WuLi01]).

Table. 2.4. Characteristic of prediction modes in PFGS.

Mode	
1	- low coding efficiency - the enhancement macroblock is predicted from the previous low quality reference - the high quality reference for the next frame is reconstructed from the current high quality frame - no drift (low quality references are always available in the decoder) Note 1 If all enhancement macroblocks are encoded with this mode, the MB-PFGS scheme is exactly the same as the FGS scheme
2	- high coding efficiency - the enhancement macroblock is predicted from the previous high quality reference and reconstructed from the same reference - if all enhancement macroblocks are encoded with this mode, the PFGS scheme provides higher coding efficiency at higher bitrates than FGS - drift (if the high quality reference in the previous frame is not available due to low network bandwidth or transmission errors in the previous frames, the decoder has to use the low quality reference instead)
3	- the extension of the error reduction method to the macroblock level - the enhancement macroblock is predicted from the previous high quality reference - the high quality reference is reconstructed from the previous low quality reference in both the encoder and decoder - since the encoder and decoder can always obtain the same temporal prediction, the error propagated from the previous frames can be effectively eliminated in this mode

Macroblock-based PFGS scheme achieves a coding efficiency gain of up to 2dB over FGS for all bitrates [Ling01].

2.3.4 Macroblock-based Progressive Fine Granularity Scalable-Temporal Method (PFGST)

In 2001 Sun, Wu, Li, Gao and Zhang presented other improvements of the FGS scheme [Sun01b]. This time they combined the FGST scheme with the macroblock-based PFGS, and called this coding method macroblock-based Progressive Fine Granularity Scalable-Temporal Method (PFGST).

Since there are two reconstructed layers in each P-frame and I-frame, the temporal frame in the PFGST scheme can be predicted either from the reconstructed base layer or from the reconstructed enhancement layer. Which reference should be used in the prediction of temporal frame is determined by the sum of absolute errors of the predicted residues in the DCT domain. Gray boxes indicate the reconstructed frames in layers used as references. If the sum of absolute errors of the DCT residues predicted from the base layer is lower than that predicted from the reconstructed enhancement layer, the reference of the temporal frame is the base layer; otherwise the reference of the temporal frame is the reconstructed enhancement layer. In the PFGST scheme, it is possible to choose the temporal reference flexibly for every enhancement macroblock, according to the above criterion. Accordingly, there are two prediction modes for enhancement macroblocks in the temporal frames, as shown in Fig. 2.17.

The experiments have shown that the macroblock-based PFGST scheme can give the coding efficiency gain of up to 2.8dB, compared with the FGST scheme.

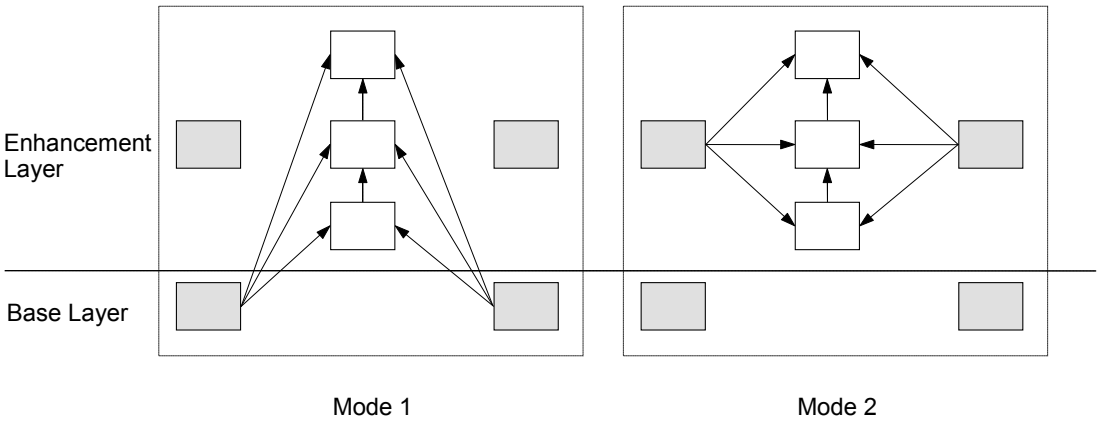


Fig. 2.17. Prediction modes of the enhancement macroblock in B-frames (based on [Sun01]).

2.3.5 Macroblock-based Progressive Fine Granularity Spatial Scalability (mb-PFGSS)

In 2001, Yan, Wu, Li and Zhang [Yan01] added spatial scalability to macroblock-based PFGS framework and called this solution the Macroblock-based Progressive Fine Granularity Spatial Scalability (mb-PFGSS). The basic idea is that the base layer and the low resolution enhancement layers still utilize the whole coding scheme of macroblock-based Progressive Fine Granularity Scalable coding (mb-PFGS). At the same time, each frame generates an additional high resolution bitstream from a low resolution enhancement layer. In the first frame of each Group of Pictures (GOPs), the high resolution enhancement layer is encoded as a P-frame, and the reference is derived from the low resolution enhancement layer. The high resolution enhancement layers in other frames are encoded with bi-directional prediction, i.e. from the previously reconstructed high resolution or low resolution references and the currently reconstructed low resolution reference. Because most macroblocks are encoded from the high resolution reference, the high resolution layer has high coding efficiency. At the same time, the low resolution reference is used to eliminate drift, just as in the PFGS scheme. Fig. 2.18 illustrates the architecture of the macroblock-based Progressive Fine Granularity Spatial Scalability (mb-PFGSS) coding scheme. Gray rectangular boxes in Fig. 2.18 denote the frames in particular layers, which will be reconstructed as references.

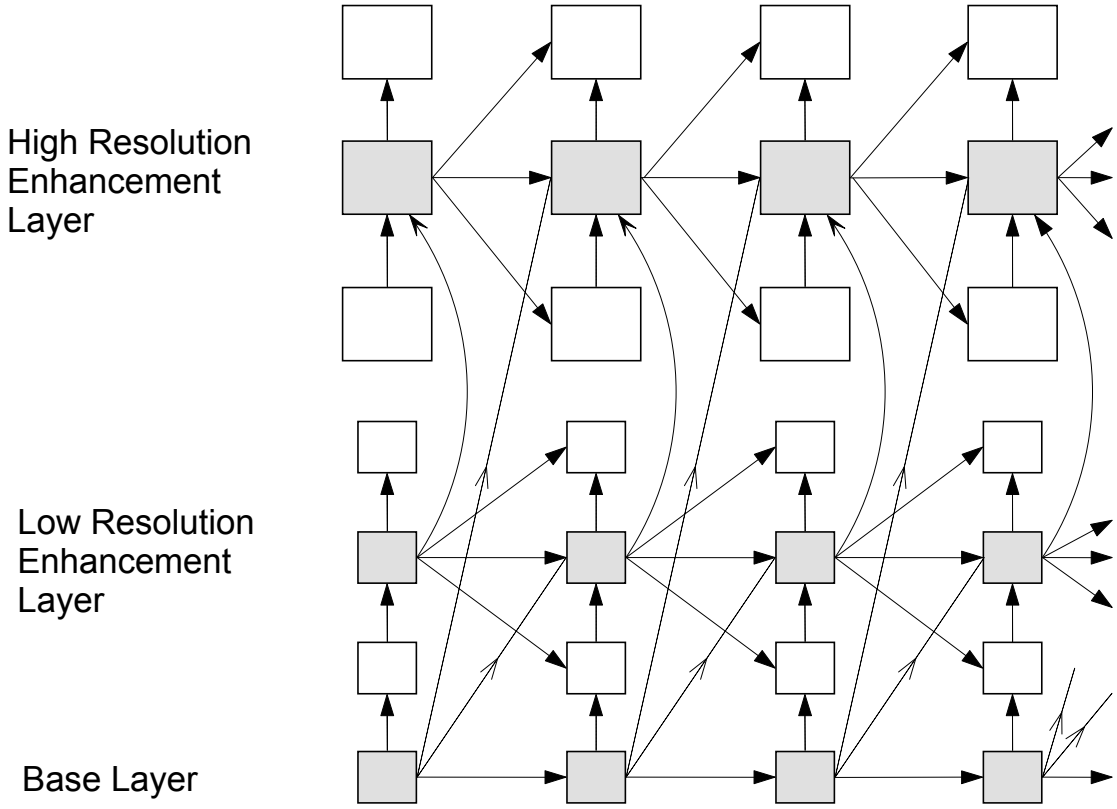


Fig. 2.18. Architecture of the macroblock-based Progressive Fine Granularity Spatial Scalability coding scheme (based on [Yan01]).

Yan, Wu, Li and Zhang have compared the macroblock-based PFGSS scheme with the traditional spatially scalable MPEG-4 Advanced Simple Profile scheme [Yan01]. Experiments show that the traditional spatially scalable scheme provides only two different levels of video quality (Fig. 2.19). In this scheme, the bitrate of the enhancement layers is 128kbps or 256kbps. In the mb-PFGSS scheme, the bitrate of enhancement layers is not limited. Experiments show that the coding efficiency of the traditional spatially scalable scheme is higher than that of the proposed mb-PFGSS scheme only within a small bitrate range after the video sequence is switched to high resolution. On the other hand, the bitrate in the mb-PFGSS scheme can be arbitrarily selected from a wide range. Video quality in the mb-PFGSS scheme improves if the bitrate increases.

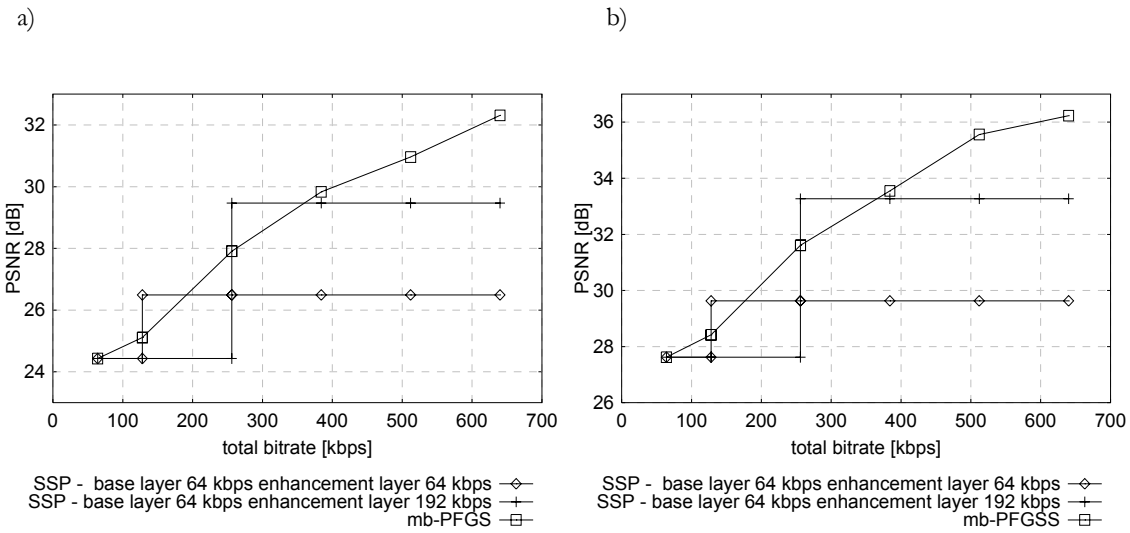


Fig. 2.19. PSNR versus bitrate for mb-PFGSS and the traditional spatially scalable coding SSP for luminance. Test sequences: a) *Coastguard*, b) *Foreman* (based on [Yan01]).

2.3.6 Motion-Compensated FGS (MC-FGS)

In 2000 van der Schaar and Radha introduced a motion-compensation loop in the FGS enhancement layer [Scha00b, Kall01], using earlier experiment results [Benz99b]. They called this scheme MC-FGS. In this solution (see Fig. 2.20), the enhancement layer is coded by a fine-granular encoder but the enhancement layer residue is no longer computed only by subtracting the decoded base layer image from the original image. The motion-compensated residual of the enhancement layer reference image is also used to predict the FGS enhancement layer residual of B-frames. Although it is possible to use motion compensation for P-frames in the FGS enhancement layer, restricting motion compensation to B-frames provides a relatively high gain in quality while preserving many of the benefits of the basic FGS structure. In addition, in this approach, the encoder can utilize any desired portion of the FGS enhancement layer I- and P-frames for predicting the FGS B-frames. Therefore, I- and P-frames do not suffer from drift, and the propagation of any possible drift is significantly reduced since it is confined to the B enhancement frames only. In MC-FGS, the base layer structure remains unchanged. The encoder and decoder have an additional motion compensation loop in the enhancement layer for the B-frames. The enhancement layer motion-compensation loop for the B-frames reuses the base layer motion vectors and prediction modes.

The experiments have shown that the MC-FGS scheme achieves a larger gain in coding efficiency (up to 2dB improvement in quality) than the standard FGS scheme. Karande in [Kara01] reports cross-checking results of the experiments with MC-FGS. The MC-FGS approach was implemented both in the Microsoft encoder and decoder. The encoder was used to generate MC-FGS compressed bitstreams coded over a very wide range of bitrates, for several test sequences. The resulting bitstreams were used as input to the MC-FGS decoder, and they were decoded successfully. This shows that it is possible to implement MC-FGS in the Microsoft software. The cross-check with the Momusys [MOMU] code has also been performed.

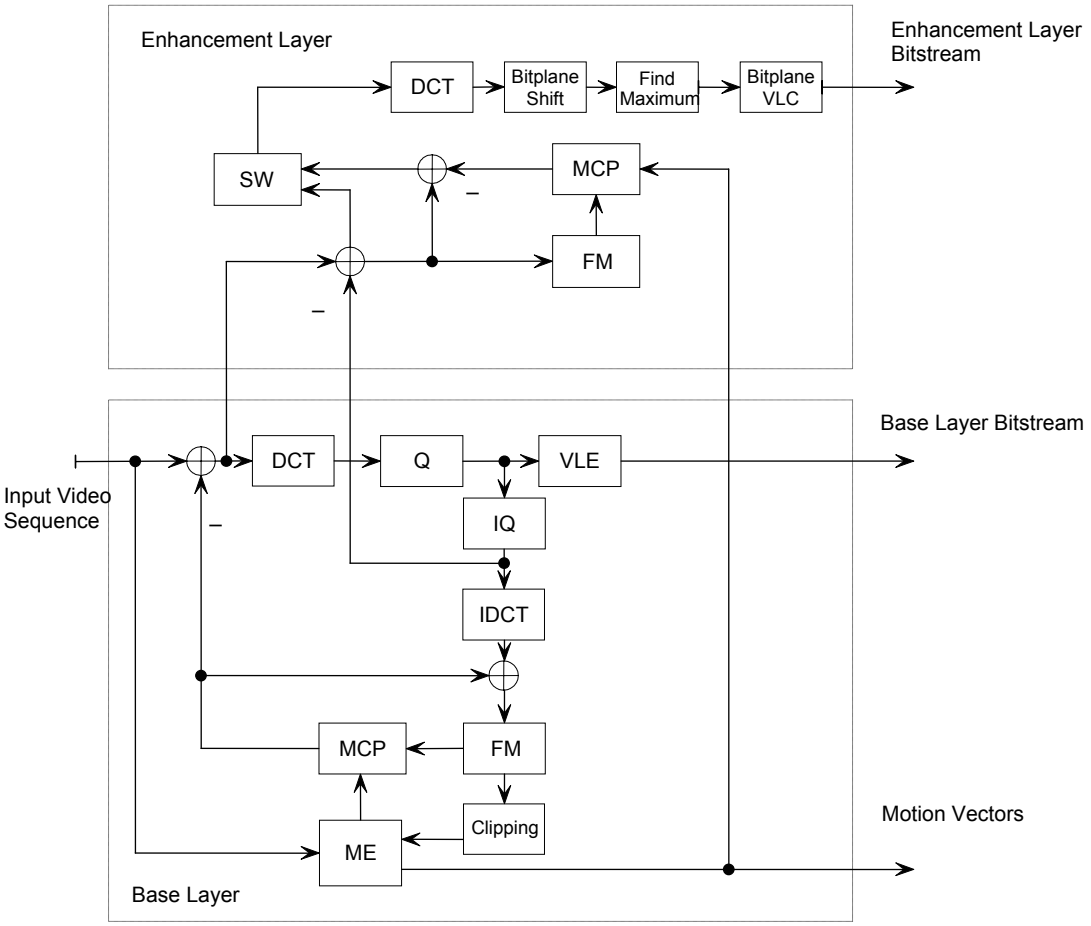


Fig. 2.20. FGS encoder structure with motion compensation in the enhancement layer (based on [Scha00b]).

2.4 Other approaches to scalability used in hybrid encoders

The fourth group includes other approaches to scalability, used in hybrid encoders. Hybrid denotes scalable systems based on a hybrid motion-compensated DCT structure similar to the existing MPEG standards and on a combination of two or more different types of scalability (see Section 2.2.5).

The hybrid motion-compensated DCT structure appears in nearly all existing video coding schemes. Scalable video coding uses a variation of this structure. In many solutions, the block DCT residual encoder is replaced with another coding method, such as subband coding [Topi98] or matching pursuit [Neff97, AlSh99].

A SNR scalability with matching pursuits has been proposed by de Vleeschouwer and Macq [Vlee00, Vlee00b]. Matching pursuit was originally developed for signal analysis by Mallat and Zhang in 1993 [Mall93]. Instead of expanding the motion residual signal using the complete DCT base, an expansion using a larger, more flexible base has been proposed. Neff and Zakhor [Neff97] have chosen an over complete set of separable Gabor functions, which do not contain artificial block edges. Neff and Zakhor have applied the matching pursuit technique to code the motion prediction error signal. The matching pursuit encoder divides each motion residual into blocks and measures the energy of each block. The center of the block with the largest energy value is adopted as an initial estimate for an inner product search. A dictionary of Gabor basis vectors is then exhaustively matched to an S x S size window around the initial estimate. The location of the window, basis vector index, and the value of the largest quantized inner product are then coded together. This procedure is applied recursively until either the bit budget is exhausted or the distortion goes below a prespecified threshold.

A combined subband-DCT approach to scalable multi-resolution video coding has been proposed by Benzler [Benz96, Benz99, Benz00b, Seeg98]. In this technique each image is decomposed into four subbands by analysis filtering, where the low frequency subband represents the base layer. Together with the high frequency subbands the enhancement layer is reconstructed by synthesis filtering. The low frequency subband and the three high frequency subbands are separately coded using DCT-based techniques similar to MPEG-2. Next, in order to adjust the bitrate distribution between the base and

enhancement layers, SNR scalability for the base layer is used. Compared with non-scalable MPEG-2 coding, the additional overhead for scalability is only 5-9% of the total bitrate, whereas the MPEG-2 Spatial Scalable Profile (SSP) leads to an overhead of 66-70%.

Another approach to SNR scalability has been presented by Robers, Kondi and Katsaggelos [Robe97, Robe98]. The proposed scalable video coding applies concepts of the progressive JPEG image coding technique to a video sequence. The quantized DCT coefficients are partitioned for both inter and intra blocks to enable several scans of increasing quality. Robers, Kondi and Katsaggelos have developed scan-dependent variable length codes which use the characteristics and properties of each scan. They have also implemented a rate control mechanism that modifies the scan definitions to meet prespecified bitrate constraints. Robers, Kondi and Katsaggelos have compared their scalable encoder with the non-scalable H.263 encoder. The scalable encoder gives the results about 1,5dB worse than the non-scalable H.263 encoder for 28.8, 56 and 128kbps.

Another scalable video coding technique uses spatio-temporal scalability. This technique, based on spatial and temporal scalability, is introduced by the author. The author has considered two basic approaches to spatio-temporal scalability. The first approach exploits three-dimensional subband analysis. In the second approach, the technique employs data structures already designed for standard MPEG-2 coding. The author has obtained very promising results for spatio-temporal scalability. Spatio-temporal scalability is presented in detail in next sections.

2.5 Scalable techniques based on subband / wavelet decompositions

In recent years, there has been an impressive development in wavelet image coding. Its success is mainly attributed to innovative strategies of data organization and representation of wavelet-transformed images, which exploit the statistical properties of a wavelet pyramid. The success of wavelet image coding [Topi98, Wood91, Taub02] was followed by various proposals of subband decomposition for scalable video coding [Lage96].

After analysis of the proposed scalable subband techniques, encoders can be divided into two classes, i.e. 2-D and 3-D subband decompositions [Ohm94, Ohm95, Taub94]. 2-D subband decomposition video encoders are further divided into intraframe and interframe encoders. An intraframe encoder encodes all frames in a video sequence independently. Thus an intraframe subband video encoder is applied to each individual frame in a sequence, without taking into account temporal redundancies.

The 3-D wavelet encoders are fully scalable by definition. Wavelets offer a natural multiscale representation for still images and high efficiency of progressive encoding. The multiscale representation can be extended to video data, by a 3-D (or 2-D+t) wavelet analysis, which includes the temporal dimension in the decomposition. 3D wavelet decomposition leads to a spatio-temporal multiresolution description of the original group of frames (GOF). Thanks to multiscale representation of video, lower spatial resolutions and/or lower frame rates can be obtained. The temporal filtering (TF) and the spatial filtering modules perform the spatio-temporal analysis. However, a simple TF does not allow efficient compression, as the temporal details can be very important. The movement within the GOF has to be estimated and compensated, in order to get homogenous frames.

2.5.1 EZW and SPIHT

In image compression, quantizers are designed specifically for each band. The problem of optimal quantizers is presented in Section 4.4.4. The quantized coefficients are binary coded using either Huffman coding or arithmetic coding [Held96]. Effective encoding of significant coefficients is the most important objective. It has been observed that coefficients which are quantized to zero in a certain pass have structural similarity in the wavelet subbands in the same spatial orientation. Spatial Orientation Trees (SOT) are groups of wavelet transform coefficients organized into trees rooted in the lowest frequency or the coarsest scale subband, with several generations of offspring along the same spatial orientation in higher frequency (resolution) subbands. Thus spatial orientation trees can be used to quantize large areas of insignificant coefficients efficiently.

Data organization plays a key role in the recent success of wavelet image coding algorithms. Two such high performance wavelet image encoders have been developed, i.e. Shapiro's embedded zerotree wavelet encoder (EZW) and Said and Pearlman's set partitioning in hierarchical trees (SPIHT) [Said96, Topi98]. The EZW and SPIHT algorithms are used in intraframe coding. Both EZW and SPIHT exploit cross-subband dependence of insignificant wavelet coefficients. The major difference between these two algorithms lies in the fact that they use different strategies to scan the transformed pixels. The SOT used by Said and Pearlman is more efficient than this of Shapiros'.

The embedded zerotree wavelet (EZW) algorithm (invented by Shapiro in 1993) uses the similarity of coefficients in different wavelet bands to achieve significant compression ratio. The EZW encoder is based on an important observation, i.e. that natural images in general have a low-pass spectrum. When an image is wavelet transformed, the energy in the subbands decreases as the scale decreases, so the wavelet coefficients will, on average, be smaller in the higher subbands than in the lower subbands. This shows that progressive encoding is a very natural choice for compressing wavelet transformed images, since the higher subbands only add the details. Large wavelet coefficients are more important than small wavelet coefficients. This observation is exploited by encoding the wavelet coefficients in decreasing order, in several passes. For every pass a threshold is chosen, against which all the wavelet coefficients are measured. If a wavelet coefficient is larger than the threshold, it is encoded and removed from the image. If it is smaller, it is left for the next pass. When all the wavelet coefficients have been measured, the threshold is lowered and the image is scanned again to add more details to the already encoded image. This process is repeated until all the wavelet coefficients have been encoded or another criterion has been satisfied (e.g. maximum bitrate).

A simplified version of the algorithm is as follows: starting with a threshold *T*, each coefficient may be either parent of a zerotree root, or a significant coefficient, or an isolated zero. A coefficient is a zerotree root if the coefficient of the parent and all coefficients of the children are lower than *T*. A coefficient is significant if its amplitude is greater than *T*. The isolated zero symbol signifies a coefficient below T, but with at least one child not below *T*. During each pass, *T* is reduced and all coefficients are scanned again. Smaller *T* results in quantization of coefficients that is much more accurate

(Successive Approximation Quantization is similar to binary representation of real numbers).

The SPIHT algorithm [Cho01] utilizes three basic concepts: 1) searching for sets in spatial orientation trees in wavelet transform; 2) partitioning the wavelet transform coefficients in these trees into sets defined by the level of the highest significant bit in a bitplane representation of their magnitudes; and 3) coding and transmiting bits associated with the highest remaining bitplanes first. The SPIHT consists of two main stages, sorting and refinement. At the sorting stage, SPIHT sorts pixels by magnitude with respect to a threshold, (called the level of significance) which is a power of two. However, this sorting is a partial ordering, as there is no prescribed ordering among the coefficients with the same level of significance or highest significant bit. The sorting is based on a significance test of pixels along the SOTs rooted from the highest level of the pyramid in a 2-D wavelet transformed image. SOTs were introduced to test the significance of groups of pixels for efficient compression by exploiting similarity and magnitude localization. Fig. 2.26 shows parent-offspring relationship of coefficients for a 2-D wavelet transform with two levels of decomposition. The details are presented in [Said96, Bani01, Cho01].

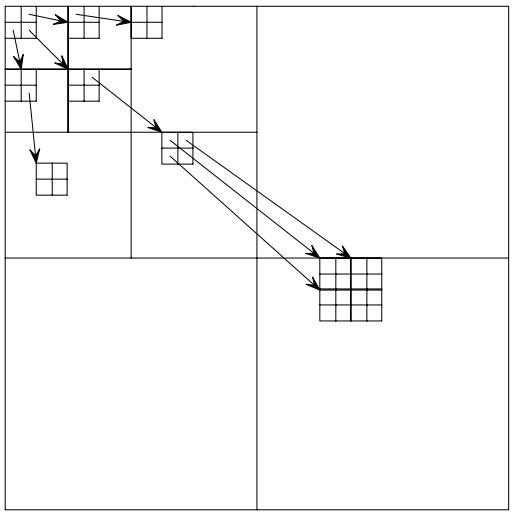


Fig. 2.26. Quadtree organization of subband pixels in SPIHT algorithm.

The new image compression JPEG-2000 standard [Marc00, Taub02] puts strong emphasis on scalability. The JPEG-2000 bitstreams are highly scalable. The Discrete

Wavelet Transform (DWT) provides a natural framework for scalable image compression. The JPEG-2000 offers tools for spatial and SNR scalable coding.

2.5.2 Out-band and In-band motion compensation

Intraframe video encoders do not give the highest achievable compression because temporal redundancies are ignored. Motion compensation is very effective in reducing temporal redundancy, and is commonly used in video coding. In recent years, many subband encoders with motion compensation have been proposed e.g. [Tsun94, Fuld96]. The embedding of subband decomposition into a motion-compensated encoder leads to in-band or out-band motion compensation performed on individual subbands or on the whole image, respectively [Bosv93, Bosv96].

In out-band compensation, video frames are compensated for motion. The result is then spatially decomposed into subbands. The out-band compensation cannot be used directly for scalable video [Akans96]. This is because the subband analysis and synthesis are placed inside the motion compensation prediction loop. Motion-compensated prediction in the encoder is based on full spatial resolution of the images. Because low resolution decoders cannot carry out the same motion-compensated prediction as the encoder, the reconstruction of video will suffer from drift.

In the in-band compensation, motion compensation is carried out directly in the subband domain.

The experiments show that the out-band compensation is more efficient [Bosv96]. That is because, although the correlation between consecutive frames is high, the temporal correlation between the subbands is generally much lower. Even if one frame is a perfect shift of another, the subbands of these frames are not identical, because of the aliasing components in the subbands. Motion compensation carried out directly on subbands is therefore not successful in compensating the subband aliasing term, leading to severe performance losses. The only way to compensate correctly the aliasing is to reconstruct the subbands prior to motion compensation. This is, however, nothing else than out-band compensation.

2.5.3 3-D Subband Coding (3-D SBC)

In 3-D subband coding, the video signal is not only decomposed into subbands in the spatial domain, but also in the direction of the temporal axis. Fig. 2.28 illustrates the basic scheme of a 3-D subband encoder. The spatio-temporal filter banks are assumed to be separable. After filtering, temporal decimation by a factor of two follows. The low-pass L_t and high-pass H_t temporal subband are computed. This means that two new sequences of half the original frame rate are created, where a frame in the first sequence is the average of successive frame pairs, and a frame in the second sequence is the difference between successive frame pairs. Each of these sequences consists of a temporal subband. To achieve higher compression, the low-pass temporal band can be decomposed further by concatenating two-band analysis filter banks in a tree-structure.

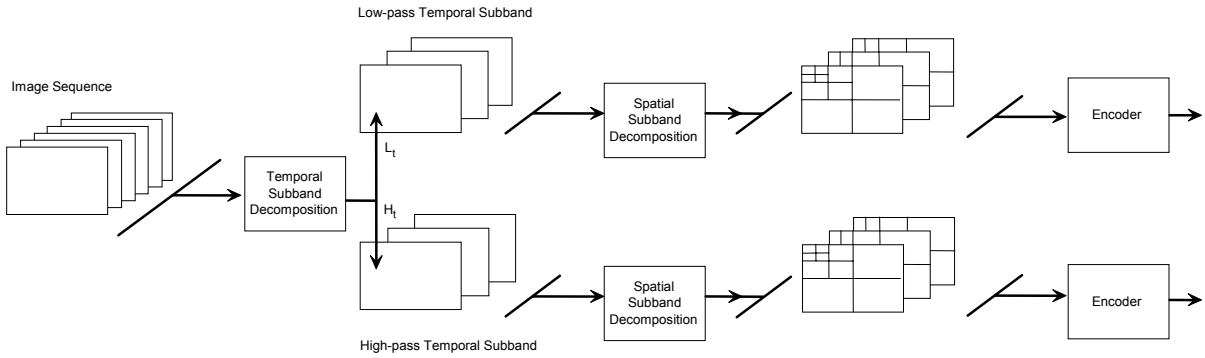


Fig. 2.28. Basic scheme of 3-D subband encoder (based on [Akan96]).

Karlsson and Veterli took the first step toward 3-D subband coding, using a simple 2-tap Haar filter for temporal filtering. Kronander first presented motion-compensated temporal filtering within the 3-D SBC framework. Ohm [Ohm93, Ohm93b, Ohm94] and later Choi and Woods refined the method and utilized different quantization techniques in subbands produced by perfect reconstruction filter banks. The literature proposes many solutions based on the significance tree quantization. This 2-D embedded zerotree method has been extended to 3-D EZW by Chen and Pearlman [Chen96]. Kim and Pearlman have employed SPIHT to wavelets and have proposed a 3-D wavelet coding system with multiresolution scalability [Kim97, Kim00]. Choi and

Woods have discussed the problem of motion-compensated 3-D subband coding of video [Choi97, Choi99].

Domański and Świerczyński have proposed a video coding technique based on the 3-D subband analysis with recursive spatial filter banks [Doma95, Doma96]. In this technique, in order to avoid annoying artifacts at edges and thin lines, the filter banks are switched adaptively. Flat areas are processed with recursive filters, exhibiting long impulse responses and good selectivity, while object edges and other details are processed with recursive filters with highly attenuated impulse responses and poorer selectivity. For a very simple encoding scheme, good visual quality has been obtained for real test video sequences in the CIF format for the bitrate of about 150kbps.

The experiments presented in [Chen96] have shown that a 3-D zero-tree algorithm gives results comparable to the MPEG-2 standard encoder with motion compensation. As for the visual quality, the reconstructed images obtained from the 3-D zero-tree scheme are blurred in some regions, while the images obtained from the MPEG-2 have blocking effects. Kim, Xiong and Pearlman have compared the efficiency of their 3-D SPIHT with the H.263. 3-D SPIHT is about 0.89-1.39dB worse than the H.263 for the bitrate of 30kbps. Delp [Shen99, Asbu99] has tested the scalable adaptive motion-compensated wavelet algorithm. The results show that the proposed algorithm is worse than the H.263 for all tested bitrates.

In 2001, Ad hoc Group on Exploration of Interframe Wavelet Technology in Video was formed to explore new tracks in video coding in the area of wavelet technology, including exploitation of interframe redundancy [ISO01b, ISO02, Ohm02b]. For video coding, motion compensation seems to be the key to achieve high coding efficiency, in particular for high-compression coding. Since modern technology allows to combine wavelet decomposition with motion compensation, this group also studies the potential merits of this approach.

In particular motion-compensated in-band prediction and motion-compensated temporal filtering have been proposed by this group [Ohm02, Tura02, Wood02]. In terms of coding efficiency, motion-compensated temporal filtering leading to a 3D wavelet decomposition was identified as the most promising technique, achieving coding efficiency which is competitive with non-scalable encoders on the market today.

2.6 H.26l

In recent months, a new part of MPEG-4 standard (MPEG-4 part 10 and ITU-T H.264) has been launched jointly by MPEG and ITU-T (JVT Joint Video Team) [Mpeg01, MPEG01d, MPEG01e, Sull02a, Sull02b]. This part will use a new algorithm, H.26l as the baseline [Dovs00, HeWu02, ITUT01, ITUT98, Sull01]. H.26l is not compatible with existing MPEG standards. H.26l aims at higher coding efficiency. It is being developed by ITU-T as a long-term standard.

Many efforts on switching among single bitstreams and scalable coding have been made on H.26l [Blät00, Kurc01]. H.26l does not include any scalable coding techniques so far. But bitrate scalability has been listed on the work plan as an important functionality that should be supported by the joint standard and the integration of the basic PFGS scalable coding scheme has been proposed into the H.26l [Mpeg01, HeYa01].

At the end of 2001, Ad hoc Group on Fine Granularity Scalability in JVT was formed to adapt the FGS into the Joint Model (JM) framework and to explore further possible optimizations of the FGS encoder [ISO01c, ISO01d, ISO01e, Sun02].

Chapter 3

Spatio-temporal scalability

3.1. Introduction

Most scalability proposals are based on one type of scalability. The universal scalable coding has to include different types of scalability [Ohm01]. A single scalable video technique cannot serve a broad range of bitrates in networks (e.g. from a few kbps to several Mbps) or a wide selection of terminals with different characteristics. Moreover, there are limitations in the number of layers. Optimized solutions are usually tied to specific applications.

Among various possibilities, the combination of spatial and temporal scalability called spatio-temporal scalability seems very promising. Spatio-temporal decomposition allows to encode the base layer with a smaller number of bits because the base layer corresponds to reduced information. Here, the term of spatio-temporal scalability describes a functionality of video compression systems where the base layer (low resolution layer) corresponds to frames with reduced spatial and temporal resolution. An enhancement layer (high resolution layer) is used to transmit the information needed for restoration of full spatial and temporal resolution. Fig. 3.1 shows an example of a video sequence structure obtained after spatio-temporal decomposition, for three layers.

Spatio-temporal scalability has been proposed in several versions, in particular:

- with 3-D spatio-temporal subband decomposition [Taub94, Kim00, Doma98],
- with 2-D spatial subband decomposition and partitioning of B-frames data [Doma98, Doma00, Doma00b],
- based on DCT coding, on exploiting the partitioning of B-frames data and on exploiting I- frame subband decomposition,
- exploiting as reference frames the interpolated low resolution images from the base layer [Doma00c, Doma01].

Except the first approach, the rest are original proposals of the author of this dissertation. These proposals will be discussed in detail in the next section.

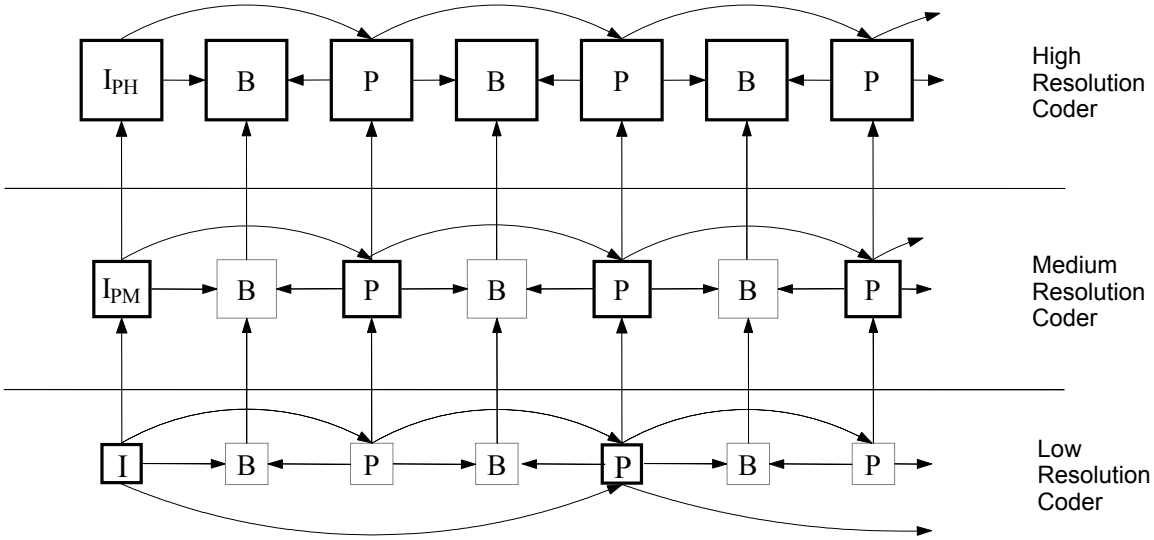


Fig. 3.1. Exemplary structure of the video sequence.

For the sake of simplicity, spatio-temporal scalability will be reviewed for the simplest case of a two-layer encoder, i.e. a system that produces one low resolution bitstream and one high resolution bitstream. Fig. 3.2 shows the block diagram of an encoder with the functionality of spatio-temporal scalability. The following basic video sequences are processed in the presented encoder:

- The low resolution sequence with reduced picture frequency and reduced horizontal and vertical resolution.
- The high resolution sequence with original resolutions in time and space.

Therefore, the encoder consists of a low resolution part and a high resolution part (Fig. 3.2). The advantage of this solution is that the low resolution layer is an independently coded layer and does not use any information from the other layer. That is why the base layer can be decoded in the decoder independently, without any information from the high resolution layer.

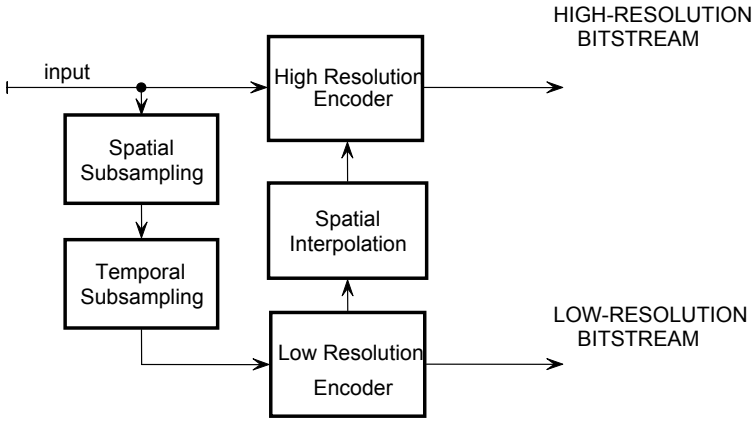


Fig. 3.2. Basic block diagram of a hybrid encoder with the functionality of spatio-temporal scalability.

3.2. Temporal decomposition

As explained in Chapter 2, temporal scalability is achieved using bi-directionally predicted frames, or B-frames. B-frames are disposable, since they are not used as reference frames for the prediction of any other frames. This property allows B-frames to be discarded without destroying the ability of the decoder to decode the sequence and without adversely affecting the quality of any subsequent frames, thus providing temporal scalability.

In this thesis, temporal resolution reduction is achieved by partitioning the stream of B-frames: every second frame is skipped in the low resolution layer. Therefore two kinds of B-frames are obtained: BE-frames that exist in the high resolution sequence only and BR-frames that exist in both sequences (Fig. 3.3). For the sake of simplicity, let us

assume that in the high resolution video sequence the number of B-frames between two consecutive I- or P-frames is even. A typical GOP structure is as follows:

I–BE–BR–BE–P–BE–BR–BE– P–BE–BR–BE– P–BE–BR–BE.

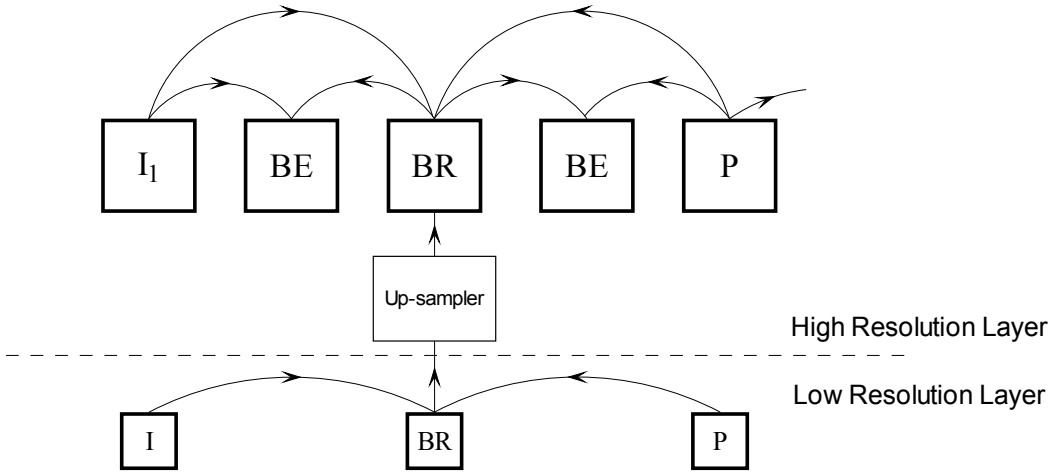


Fig. 3.3. Partitioning the stream of B-frames.

3.3. Spatial resolution reduction

In spatio-temporal scalability, the low resolution coder uses a sequence whose spatial resolution is lower than the resolution of the input sequence. The input video sequence is down-sampled to the low resolution sequence and applied as an input sequence to the low resolution encoder. The spatial resolution reduction is achieved by spatial filtering. In the proposed solutions (Fig. 3.4), spatial filtering is achieved in two different ways:

- by subband / wavelet decomposition (see Chapter 4),
- by pyramid decomposition, using an appropriate down-sampling filter (see Chapters 5 and 6).

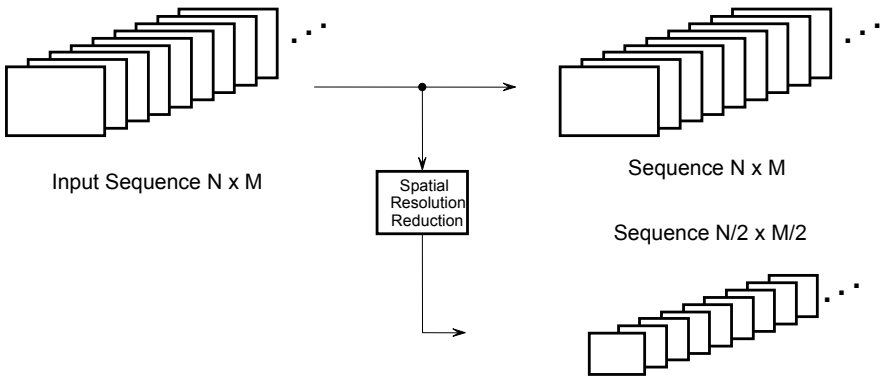


Fig. 3.4. Spatial resolution reduction.

In subband decomposition, the consecutive images are decomposed into four spatial subbands, using a pair of low- and high-pass filters. The details concerning analysis filters are presented in Chapter 4. The lowest frequency subband LL constitutes the base layer.

In pyramid decomposition, the low spatial resolution input sequence is achieved by using only one down-sampling filter. Low resolution images are derived from the original high resolution image sequence using windowing, low-pass filtering and down-sampling operations. In MPEG-2 the 2:1 horizontal and vertical decimation (both for luminance and chrominance) is done using a 7-tap low-pass filter in order to remove high frequency components and hence reduce the horizontal aliasing. In this thesis the top layer is a 4:2:0 format 4CIF progressive sequence and the bottom layer is a CIF format.

Up-conversion from the low resolution layer CIF format is spatial up-sampling operation as recommended in the MPEG-2 standard. It is a linear approximation proposed in the MPEG-2 standard.

Chapter 4

Spatio-temporal scalability with subband decomposition

4.1. Introduction

In this chapter the author presents his original proposals of system with 3-D filter banks for spatio-temporal analysis and a 2-D subband encoder with motion compensation. The author aims to ensure a high level of compatibility of scalable coding with the MPEG video coding standards. Therefore the MPEG-2 standard algorithm is the base algorithm in the presented solution.

It is generally accepted to use subband decomposition in order to obtain spatial scalability. As mentioned earlier, a number of such solutions have been already proposed by other authors. Subband decomposition runs as follows. The video sequence is decomposed frame by frame by FIR filter bank. This analysis results in the decomposition of the video sequence into four spatial subbands (Fig. 3.4.), which are assigned to the base layer and the enhancement layers.

The problem of motion-compensated scalable coding is closely related to subband coding of video. It has been already found that in-band motion estimation and compensation usually leads to poor coding efficiency.

Irrespective of the type of motion compensation, the above solution exhibits a substantial disadvantage. It is caused by the fact that the bitstream does not decrease linearly with the number of pixels in a frame. The experiments show that the base layer bitstream is usually much larger than the enhancement layer bitstream [Doma96]. Such a solution is not viable if the enhancement layer bitstream is much smaller than the base layer bitstream. The solution seems to be the application of spatio-temporal scalability [Doma99c, Doma00]. Another proposal is to use systems with three-dimensional subband analysis [Doma98c, Doma99, Doma99b]. The input video sequence is analyzed in a three-dimensional separable filter bank. Both variants will be further discussed in the next chapters.

4.2. Spatio-temporal scalability in 3-D wavelet/subband encoders

4.2.1. Introduction

In this section, the author presents his proposal of scalable encoder from earlier works [Doma98, Doma98c, Doma99, Doma99b Doma99c]. This solution uses three-dimensional subband analysis (Fig. 4.1).

The input video sequence is analyzed in a three-dimensional separable filter bank, i.e. there are three consecutive steps of analysis: temporal, horizontal and vertical. Temporal analysis results in two subbands, L_t and H_t (Fig. 4.2), which are partitioned into four spatial subbands (LL, LH, HL and HH) each. For spatial analysis, both horizontal and vertical, separable nonrecursive half-band linear-phase filters in a polyphase implementation are used. The linear phase is assumed here because motion vectors estimated in the LL subband of the L_t temporal subband are used for motion compensation in the LL subband of the H_t temporal subband.

The three-dimensional analysis results in eight spatio-temporal subbands. Three high-spatial-frequency subbands (LH, HL and HH) in the high-temporal-frequency subband H_t are discarded as they correspond to the information which is less relevant for the human visual system. Although the discarding of subbands reduces PSNR, e.g. in the

intraframe mode to about 32-33 dB, it has little influence on the subjective quality of the decoded video.

Therefore only five subbands are encoded:

- in the base layer - the spatial subband LL of the temporal subband L_t.
- in the enhancement layer - the spatial subbands LH, HL and HH of the temporal subband L_t and the spatial subband LL of the temporal subband H_t.

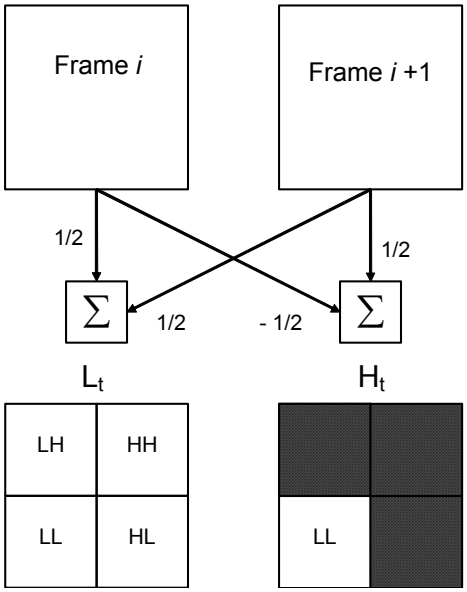


Fig. 4.1. Three-dimensional subband analysis.

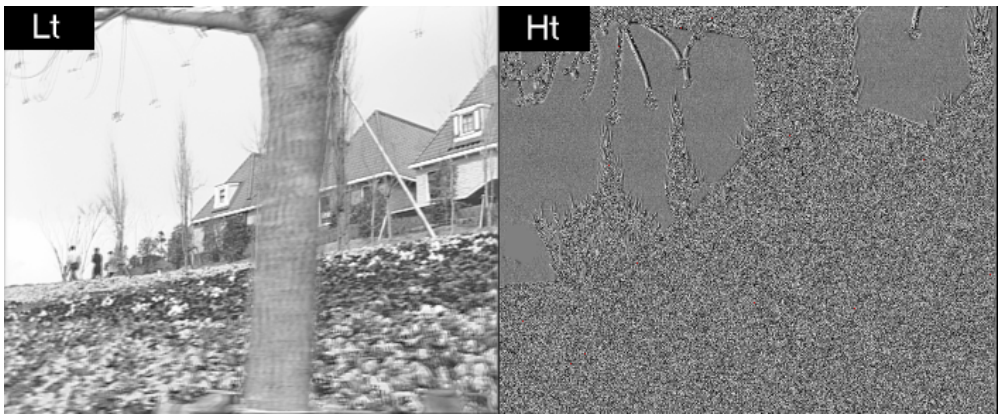


Fig. 4.2. Temporal subbands L_t and H_t in a frame from the test sequence *Flower Garden* (50Hz, progressive).

4.2.2. 3D spatio-temporal filter banks

The input video sequence is analyzed in a three-dimensional separable filter bank, i.e. there are three consecutive steps of analysis: temporal, horizontal and vertical. For temporal analysis, simple linear-phase two-tap filters are used, similarly as in pure three-dimensional subband coding [Ohm94, Doma96, Podl95]:

$$H(z) = 0.5 \cdot \left(1 \pm z^{-1}\right), \tag{4.1}$$

where "+" and "-" correspond to low- and high-pass filters, respectively. This filter bank has a very simple implementation and exhibits small group delay (half of the sampling period), resulting in short system response times. Short response times are highly critical for many applications in interactive multimedia.

4.2.3. Encoder structure

The encoder structure has been chosen after experiments with several variants of the coding algorithm. The encoder structure is presented in Fig. 4.3. The author assumed a high level of compatibility of scalable coding with MPEG video coding standards. Therefore a scalable encoder should consist mostly of functional blocks used in the standard MPEG-2 encoder.

Two subbands are encoded using MPEG-compatible DCT-based encoders. The base layer is produced by an MPEG-2–compatible encoder which processes the subband LL from the temporal subband L_t. Motion estimation, computationally the most demanding task, is performed in this subband. The motion vectors obtained are also used for the LL subband from the temporal subband H_t, which is encoded using an MPEG-2–compatible motion compensation encoder. Nevertheless, the latter encoder does not include a motion estimator.

Motion compensation in the LL subband of the temporal subband H_t is not very efficient. The experiments conducted by the author result in a conclusion that the bitrate reduction caused by motion compensation usually does not exceed 20%.

As mentioned above, the LL/L_t subband is encoded using an MPEG-2 encoder. In order to obtain high compression ratio for the base layer, rough quantization of the DCT coefficients is used, e.g. with the quantization coefficient of about 8-16. Since such quantization decreases the quality of the full resolution image too much, some corrections for non-zero valued DCT coefficients are additionally transmitted in the enhancement layer. These corrections are encoded using Huffman codes and their locations are defined by the locations of the non-zero coefficients from the base layer.

The experiments prove that motion compensation in three high-spatial-frequency subbands (LH, HL and HH) in the low-temporal-frequency subband L_t is inefficient. This conclusion is similar to that from [Uz91]. Therefore these subbands are encoded without motion compensation. In order to exploit mutual dependencies between subbands, a technique previously proposed for still images [Doma94] is used. The pixels which do not correspond to the "active" portions of the LL/L_t subband are discarded (see section 4.4.5). The remaining pixels are quantized by a nonlinear quantizer and encoded using DPCM and variable-length coding.

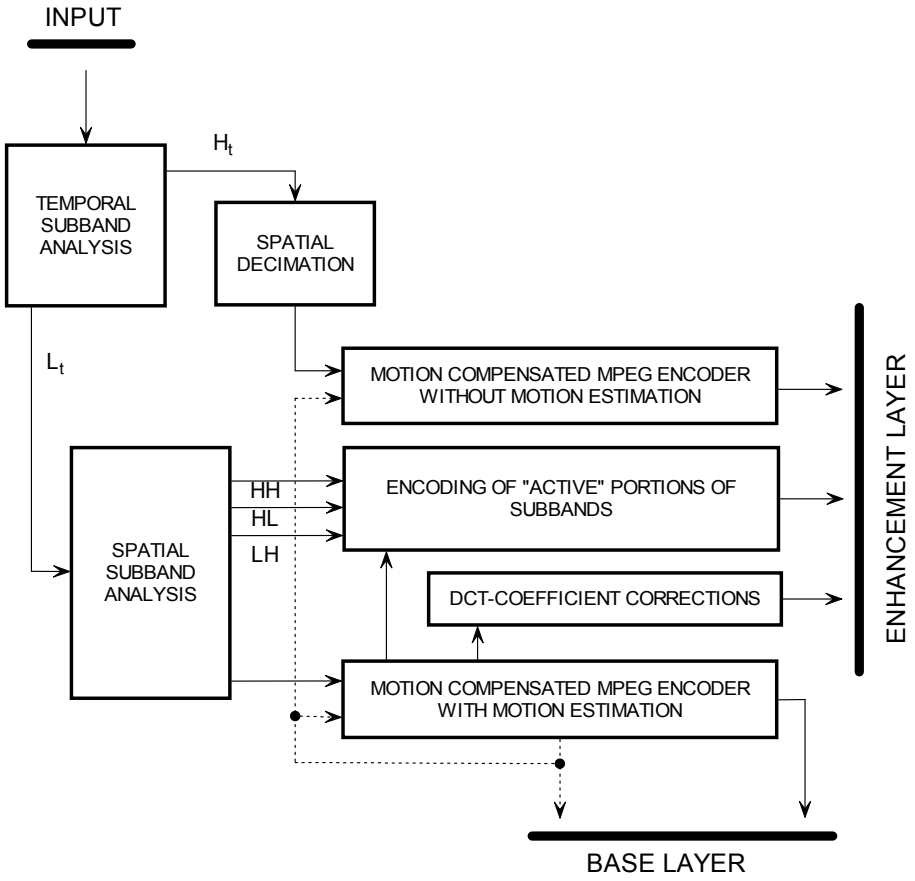


Fig. 4.3. 3D SBC encoder structure.

4.2.4. Results of experiments

The purpose of the experiments is to find the best encoder structure and its properties. Therefore the software is written in C++ language and is not optimized for run time. The most important feature is its flexibility, allowing tests of different variants of the coding algorithm. The software runs on Sun 20 workstations under the Solaris operational system.

Using the software mentioned above, the structures described in 4.2.3 have been examined with standard test sequences. Their luminance format is the progressive BT.601 [ITUR], i.e. 720 x 576 pixels and 50 frames per second. The chrominance subsampling scheme is 4:2:0. It means that the base layer provides SIF frames.

In order to evaluate the encoder efficiency, a verification model of the scalable encoder and an implementation of the MPEG-2 encoder have been used (see Annex B).

The experimental results for the test sequences ”Mobile & calendar” and ”Flower garden” are shown in Table 4.1. The results show that the bitstream of the base layer is less than half the total bitstream.

Table 4.1: Experimental results for progressive BT.601 test sequences.

Sequences	Bitstream Mbps	Base layer as percentage of total	PSNR for full resolution [dB]			PSNR for base layer [dB]		
			Y	Cr	Cb	Y	Cr	Cb
Flower garden	3.0	37.09	27.5	31.8	33.8	31.2	33.8	35.6
	3.5	37.95	28.2	32.4	34.2	32.3	34.7	36.2
	4.0	39.95	28.9	33.1	34.7	33.4	35.6	36.9
Mobile & calendar	3.3	35.46	26.9	31.8	32.8	29.4	32.8	33.5
	3.6	38.92	27.6	32.4	33.5	30.5	33.6	34.4
	4.0	43.32	28.3	33.2	34.2	31.8	34.6	35.3

4.2.5 Conclusions

Although the base layer bitstream is less than half of the total bitstream, the author has given up the proposed 3-D subband encoder for the time being since the results achieved for the 2-D subband encoder have been more promising. Moreover, the

proposed solution with 2-D subband decomposition is characterized by higher level of compatibility with the MPEG video coding standards.

However, there appears to be a growing interest in developing 3-D subband scalable encoders [Andr02b, Andr02c, Bott01, Brün00, Felt00, Pesq00, Pesq01, Pesq02]. A wavelet encoder has already been presented in MPEG and attracted a lot of attention. A group within MPEG is exploring wavelet-based video encoders for possible standardization. MASCOT members are very active in this group [Andr02a, ISO02b, ISO02c]. A new standardization process is starting shortly. It will include new technologies developed in MASCOT.

4.3. 2-D subband encoder with motion compensation

4.3.1. Introduction

A 2-D subband encoder with motion compensation incorporates B-frames into the enhancement layer. It is assumed that video sequences are encoded using a Group of Pictures (GOP) structure that consists of one or three B-pictures between two consecutive I- or P-pictures. Discarding B-frames and decoding only I- and P-frames is equivalent to temporal decimation. Some kind of poor low-pass filtering is made due to quantization performed in the encoder. In the proposal, each odd-numbered frame is included into the enhancement layer. These frames are obviously B-frames.

Reduction of temporal resolution is obtained by removing every second frame from the video sequence. The technique employs data structures already designed for standard MPEG-2 coding. It is assumed that Groups of Pictures consist of an even number of frames. Moreover, it is assumed that each second frame is a B-frame that can be removed from a sequence without affecting decodability of the remaining frames.

Spatial resolution is reduced using subband decomposition. Proper design of the filter bank results in negligible spatial aliasing in the LL subband, which is used as the base layer. Unfortunately, this technique does not allow to suppress temporal aliasing. The effects of temporal aliasing are similar to those related to frame skipping in hybrid encoders.

The standard order of frames in the base and enhancement layers is given in Table 4.2.

Table 4.2. GOP structure in both layers

Base layer (subband LL only)	Enhancement layer
I	I (without LL subband)
skipped	B
B	B (without LL subband)
skipped	B
P	P (without LL subband)
skipped	B
B	B (without LL subband)
skipped	B
P	P (without LL subband)
skipped	B
B	B (without LL subband)
skipped	B
P	P (without LL subband)
skipped	B
B	B (without LL subband)
skipped	B

The base layer is a low-quality channel. Therefore it is reasonable to perform coarser quantization in this layer than in the enhancement layer. After quantization, the base layer represents a low-quality sequence. On the other hand, the quality of the subband LL strongly influences the quality of the full-sized picture. Low quality of subband LL restricts the full-sized picture quality to a relatively low level, despite the amount of information in the remaining subbands. Therefore it is essential to transmit additional information ΔLL in the enhancement layer (see Section 4.3.5). This information is used to improve the quality of the subband LL when used to synthetize a full-sized image in the enhancement layer.

4.3.2. Encoder structure

The base layer bitstream is produced in a standard MPEG-2 encoder, which processes every second frame with reduced spatial resolution.

The enhancement layer encoder is not MPEG-2 compatible but it mostly consists of the building blocks that are used in an MPEG-2 MP@ML standard encoder. The main feature of the proposed encoder is full-frame motion estimation and compensation. In order to ensure it, spatial synthesis must be performed, i.e. four subbands (including the LL subband that constitutes the base layer) are synthetized into one signal. Therefore there are two subband analysis stages and one subband synthesis stage in the encoder. The analysis and synthesis filter banks are Johnston 12th order FIR QMFs [John80] and Daubechies (9,3) wavelets [Daub92]. Then motion estimation and compensation is performed. In order to speed up motion estimation, the base layer motion vectors are used as an initial guess. Note that their accuracy is full-pel in a full-size image. After motion compensation the pictures are split into four spatial subbands and encoded individually. The prediction errors are calculated and encoded for three subbands (HL, LH, and HH) of I- and P-frames.

In the enhancement layer encoder, the subband LL used for frame synthesis is more finely quantized than the subband LL transmitted in the base layer. It corresponds to the sum of information contained in the base layer and ΔLL transmitted in the enhancement layer.

The enhancement layer encoder produces several bitstreams:

1. LH subband,
2. HL subband,
3. HH subband,
4. ΔLL – corrections of the LL subband,
5. MV_e – full-frame motion vectors.

Many different compression techniques are applied to encode LH, HL and HH subbands. Some experiments have exploited inter-subband correlation, i.e. mutual dependencies between subbands, or more exactly, between prediction errors in individual subbands (see Fig. 4.4). The data obtained are then encoded using Huffman encoders. The tables of codes are roughly estimated. Another possibility considered is context-based geometric vector quantization.

The second set of the experiments has used using Discrete Cosine Transform coding applied to the subbands. The implementation of the encoder structure with DCT coding of the subbands is presented in Fig. 4.5.

ΔLL is the most difficult signal to encode. This signal is needed to improve the quality of the LL subband because full frame image quality depends on the LL image quality [Wood91]. This fact is well-known from subband coding of images. Unfortunately, *ΔLL* bitrate grows rapidly as the bitstream in the base layer decreases. The bitstream *ΔLL* contains bitplanes correcting the transform coefficients transmitted in the base layer.

MV_e is a bitstream which encodes motion vector refinements. In the decoder, motion vectors are restored using the base layer motion vectors and the refinements MV_e.

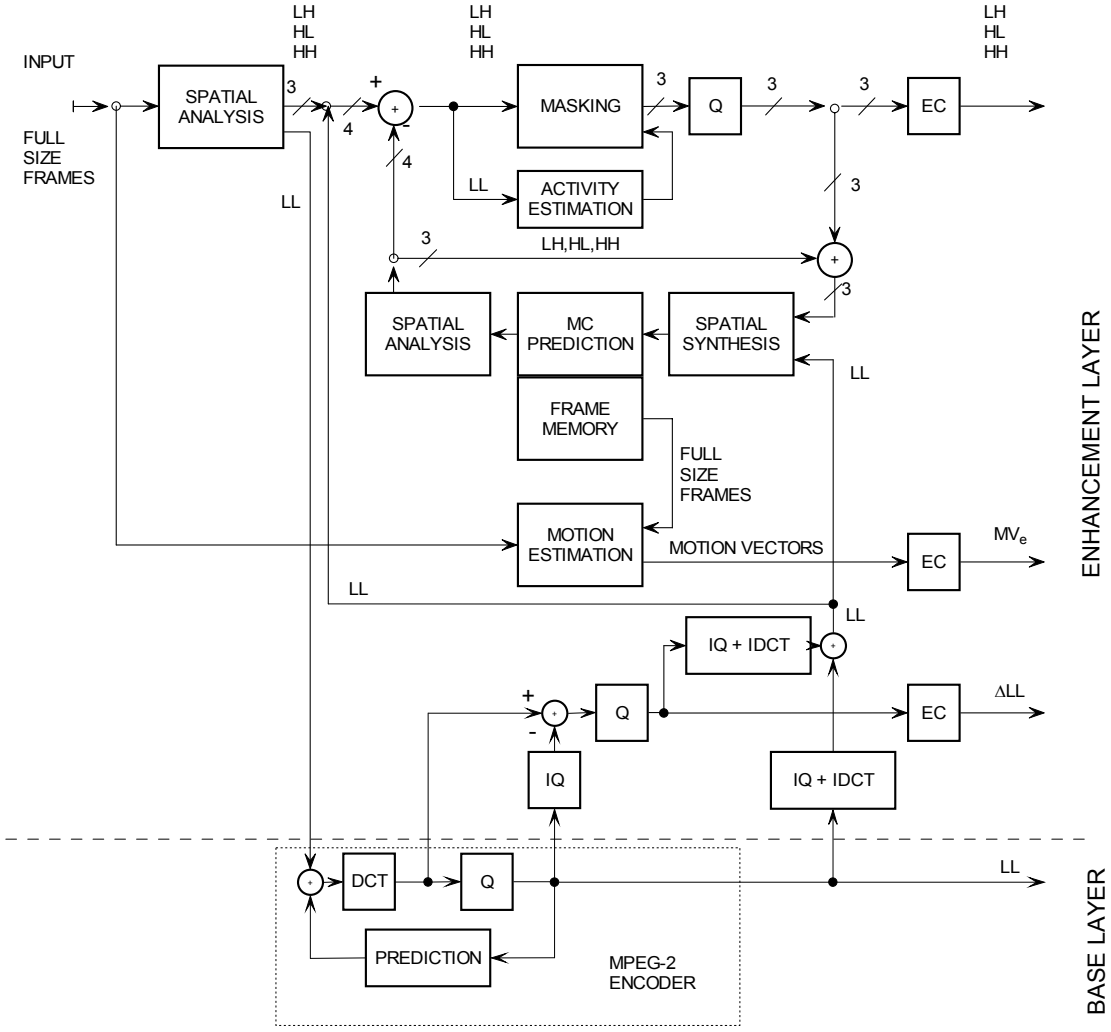


Fig. 4.4. Encoder structure which exploits inter-subband correlation.

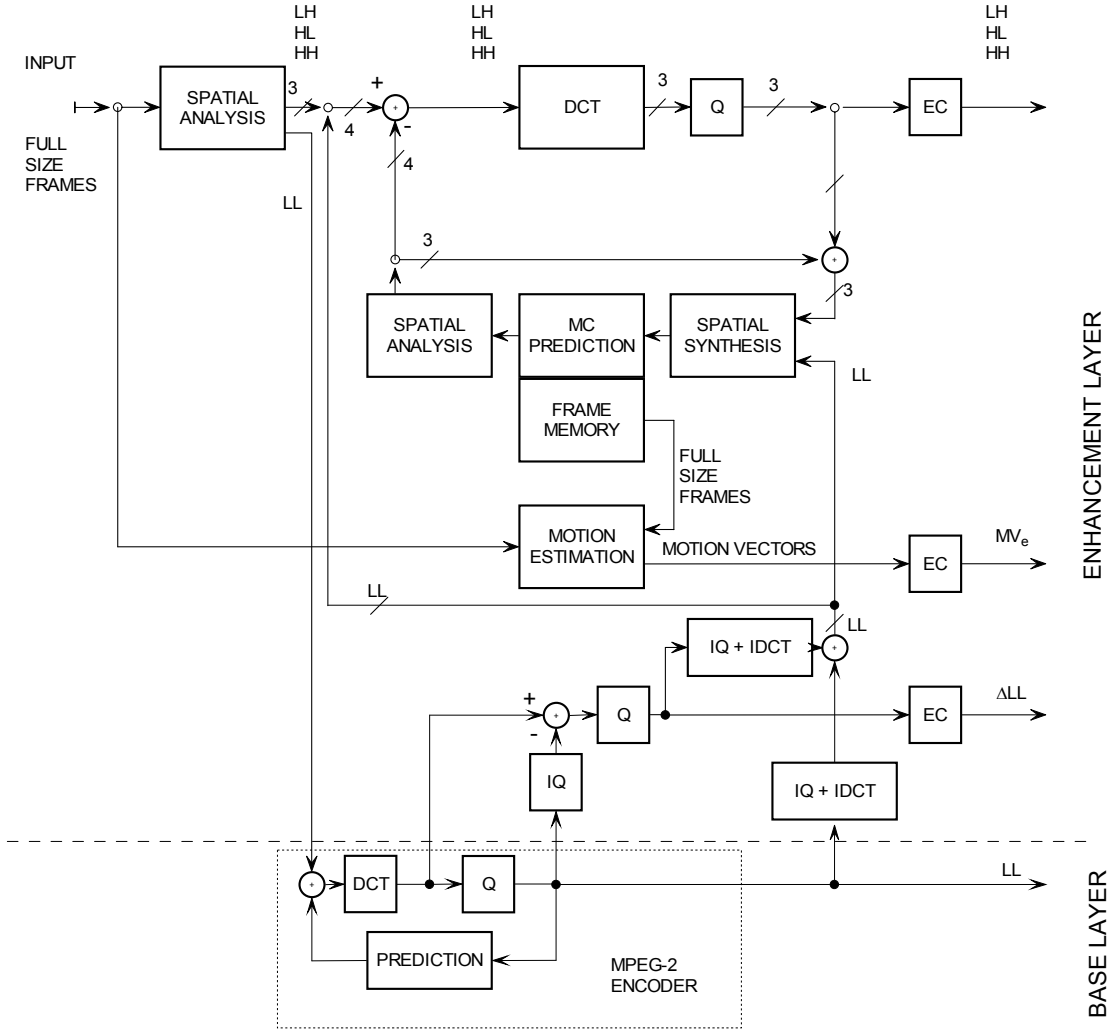


Fig. 4.5. Encoder structure with DCT encoding of subbands.

4.3.3. Adjusting quantization matrices for subbands

Transform coding has been studied extensively in the last two decades and has become a very popular compression method for still image coding and video coding. The idea of transform coding is to de-correlate the picture content and to encode transform coefficients rather than the original pixels of the images. In order to achieve this aim, input images are split into disjoint blocks of pixels (i.e., 8 x 8 pixels). The transformation can be represented as a matrix operation using an 8 x 8 transform matrix to obtain 8 x 8 transform coefficients. The MPEG-2 standard uses Discrete Cosine Transform (DCT) [Doma98d, Doma98e, Pita00].

A major objective of transform coding is to obtain as many small coefficients as possible. These coefficients then can be neglected without significant degradation of image quality (in terms of statistical and subjective measures) and consequently are not coded for transmission. At the same time, it is desirable to minimize statistical dependencies between coefficients with the aim to reduce the amount of bits needed to encode the remaining coefficients. On average, only a small number of DCT coefficients must be transmitted to the receiver to obtain a sufficient reconstruction of image blocks. Moreover, the most significant DCT coefficients are concentrated around the upper left corner (low-frequency DCT coefficients) and the significance of the coefficients decreases with increasing distance. This implies that higher DCT coefficients are less important for reconstruction than lower DCT coefficients.

The DCT coefficients are quantized prior to encoding. Depending on the quantization step-size, quantization results in a significant number of zero valued coefficients. In all coding standards, a run-length coding procedure is employed to efficiently encode the DCT coefficients which remain nonzero after quantization. In order to achieve this, all the nonzero coefficients are detected along a scan line and one code word is encoded for each (run, level)-pair. "Level" indicates the quantization level of the particular nonzero coefficient to be encoded and "run" indicates the distance to the previously encoded nonzero coefficient along the scan line.

In intra-frame coding the pixels are quantized as follows:

$$F'(n_1,n_2) = NINT\left(\frac{F(n_1,n_2)\cdot 16}{G(n_1,n_2)\cdot Q}\right) \tag{4.2}$$

The process of inverse quantization of coefficients is defined by:

$$F(n_1,n_2) = NINT\left(\frac{F'(n_1,n_2)\cdot G(n_1,n_2)\cdot Q}{16}\right) \tag{4.3}$$

where:

$F(n_1,n_2)$ - DCT coefficient,

$F'(n_1, n_2)$ - quantized DCT coefficient,

$G(n_1, n_2)$ - intra quant matrix,

Q - quantization parameter,

NINT – the nearest integer value.

The MPEG-2 standard defines the intra quantization matrix $G(n_1, n_2)$ in the following way.

$$G(n_1, n_2) = \begin{bmatrix} 8 & 16 & 19 & 22 & 26 & 27 & 29 & 34 \\ 16 & 16 & 22 & 24 & 27 & 29 & 34 & 37 \\ 19 & 22 & 26 & 27 & 29 & 34 & 34 & 38 \\ 22 & 22 & 26 & 27 & 29 & 34 & 37 & 40 \\ 22 & 26 & 27 & 29 & 32 & 35 & 40 & 48 \\ 26 & 27 & 29 & 32 & 35 & 40 & 48 & 58 \\ 26 & 27 & 29 & 34 & 38 & 46 & 56 & 69 \\ 27 & 29 & 35 & 38 & 46 & 56 & 69 & 83 \end{bmatrix}$$

This matrix has been defined for SDTV sequences. After Discrete Cosine Transform of a block of 8 x 8 pixels, most energy in the block is typically concentrated in a few low-frequency coefficients. Therefore high frequencies are quantized more roughly than low frequencies. The standard MPEG-2 intra-quantization matrix is used for quantization of the DCT coefficients of an input 4CIF picture. A graphic representation of the MPEG-2 standard intra-quantization matrix is given in Fig. 4.6. Unfortunately, this matrix is not appropriate for quantization of the DCT coefficients of the subbands. The resolution of the DCT coefficients of the subbands LH, HL and HH is CIF. The characteristics of the DCT coefficients of these subband are different from the characteristics of the DCT coefficients of a natural picture. The MPEG-2 intra-quantization matrix used for quantization of the DCT coefficients of the subbands LH, HL and HH causes aliasing between subbands (Fig. 4.7). Kronacher in 1996 [Kron96] proposed two types of intra-quantization matrix for subbands:

- Matrices with linearly interpolated values of the MPEG-2 standard quantization matrix.
- Matrices with modified values.

These quantization matrices have been used in scalable encoders to quantize the DCT coefficients of the subbands LH, HL and HH.

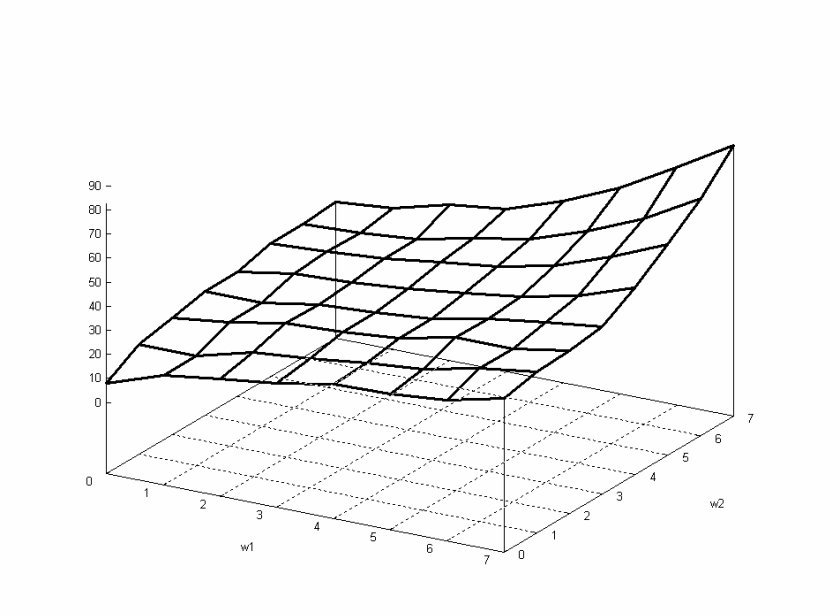


Fig. 4.6. Graphic representation of the MPEG-2 standard intra quantization matrix.

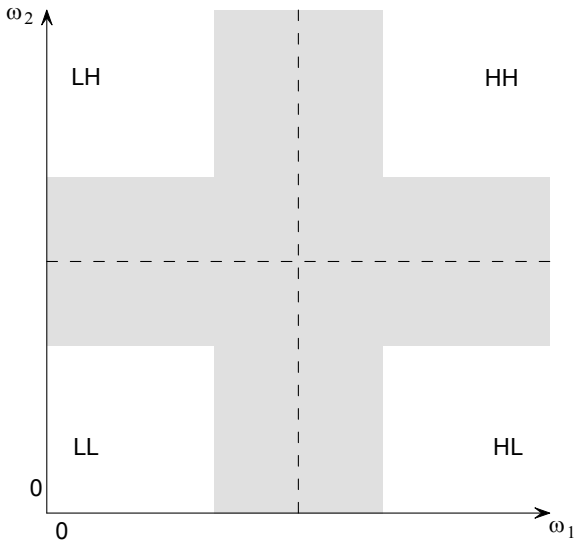


Fig. 4.7. Schematic illustration of the aliasing between subbands.

In the first type, the matrices are created as a linear interpolation of the MPEG-2 standard quantization tables. The standard quantization matrix is extended to 16 x 16 size and the empty places are completed with interpolated values. This matrix is divided into

four 8 x 8 size matrices $G(n_1,n_2)_{LL}$, $G(n_1,n_2)_{LH}$, $G(n_1,n_2)_{HL}$ and $G(n_1,n_2)_{HH}$. These matrices are presented in Fig. 4.9.

The modified values have been achieved in the following way. Let us assume that MPEG-2 coding is reference coding. Let us assume that error in MPEG-2 coding is defined by $E_{MPEG}(n_1,n_2) = Spectrum1(n_1,n_2) - Spectrum2(n_1,n_2)$ (Fig.4.8). This error is accepted by the coding process. The error of the subband coding is defined by $E_{subband}(n_1,n_2) = Spectrum1(n_1,n_2) - Spectrum3(n_1,n_2)$. It is changed by modifications of the quantization matrices $G(n_1,n_2)_{LL}$, $G(n_1,n_2)_{LH}$, $G(n_1,n_2)_{HL}$ and $G(n_1,n_2)_{HH}$. If error $E_{subband}(n_1,n_2)$ is the same or similar to error $E_{MPEG}(n_1,n_2)$, the adjusting of the quantization matrices is completed. Matrices $G(n_1,n_2)_{LL}$, $G(n_1,n_2)_{LH}$, $G(n_1,n_2)_{HL}$ and $G(n_1,n_2)_{HH}$ in Fig. 4.10 have been achieved in this way.

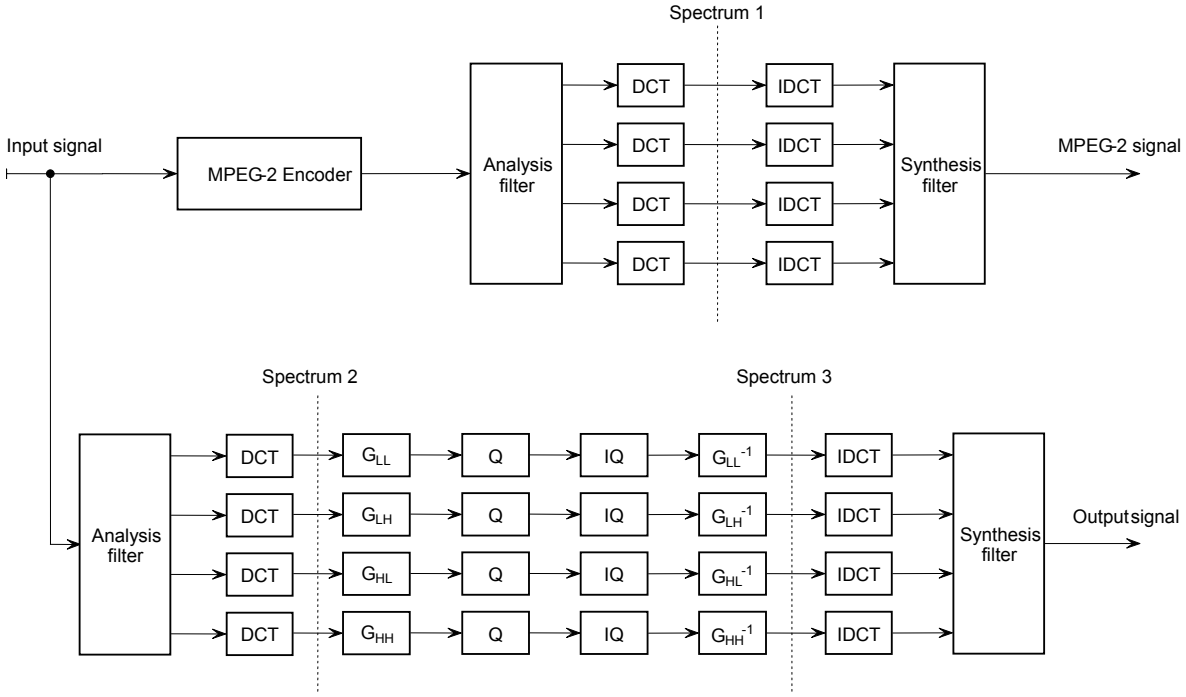


Fig. 4.8. Construction of subbands $G(n_1,n_2)_{LL}$, $G(n_1,n_2)_{LH}$, $G(n_1,n_2)_{HL}$ and $G(n_1,n_2)_{HH}$.

$$G(n_1,n_2)_{LL}=\begin{bmatrix} 8 & 12 & 16 & 18 & 19 & 21 & 22 & 24 \\ 12 & 14 & 16 & 19 & 21 & 22 & 23 & 25 \\ 16 & 16 & 16 & 19 & 22 & 23 & 24 & 26 \\ 18 & 19 & 19 & 22 & 24 & 25 & 26 & 27 \\ 19 & 21 & 22 & 24 & 26 & 27 & 27 & 28 \\ 21 & 22 & 22 & 24 & 26 & 27 & 27 & 28 \\ 22 & 22 & 22 & 24 & 26 & 27 & 27 & 28 \\ 22 & 23 & 24 & 26 & 27 & 28 & 28 & 30 \end{bmatrix} \quad G(n_1,n_2)_{HL}=\begin{bmatrix} 26 & 27 & 27 & 28 & 29 & 32 & 34 & 36 \\ 27 & 28 & 28 & 30 & 32 & 34 & 36 & 37 \\ 27 & 28 & 29 & 32 & 34 & 36 & 37 & 37 \\ 28 & 30 & 32 & 33 & 34 & 36 & 38 & 38 \\ 29 & 32 & 34 & 34 & 34 & 36 & 38 & 39 \\ 29 & 32 & 34 & 35 & 36 & 38 & 39 & 41 \\ 29 & 32 & 34 & 36 & 37 & 39 & 40 & 42 \\ 31 & 33 & 35 & 37 & 39 & 42 & 44 & 46 \end{bmatrix}$$

$$G(n_1,n_2)_{LH}=\begin{bmatrix} 22 & 24 & 26 & 27 & 27 & 28 & 29 & 31 \\ 24 & 25 & 27 & 28 & 28 & 30 & 31 & 33 \\ 26 & 26 & 27 & 28 & 29 & 31 & 32 & 34 \\ 26 & 27 & 27 & 28 & 29 & 32 & 33 & 35 \\ 26 & 27 & 27 & 28 & 29 & 32 & 34 & 36 \\ 27 & 28 & 28 & 31 & 32 & 35 & 36 & 39 \\ 27 & 28 & 29 & 32 & 35 & 37 & 38 & 42 \\ 28 & 29 & 30 & 33 & 37 & 39 & 40 & 44 \end{bmatrix} \quad G(n_1,n_2)_{HH}=\begin{bmatrix} 32 & 34 & 35 & 38 & 40 & 44 & 48 & 52 \\ 34 & 36 & 38 & 41 & 44 & 49 & 53 & 58 \\ 35 & 38 & 40 & 44 & 48 & 53 & 58 & 64 \\ 37 & 40 & 43 & 48 & 52 & 58 & 64 & 70 \\ 38 & 42 & 46 & 51 & 56 & 63 & 69 & 75 \\ 42 & 47 & 51 & 57 & 63 & 70 & 76 & 83 \\ 46 & 51 & 56 & 63 & 69 & 76 & 83 & 90 \\ 51 & 57 & 65 & 74 & 80 & 87 & 95 & 104 \end{bmatrix}$$

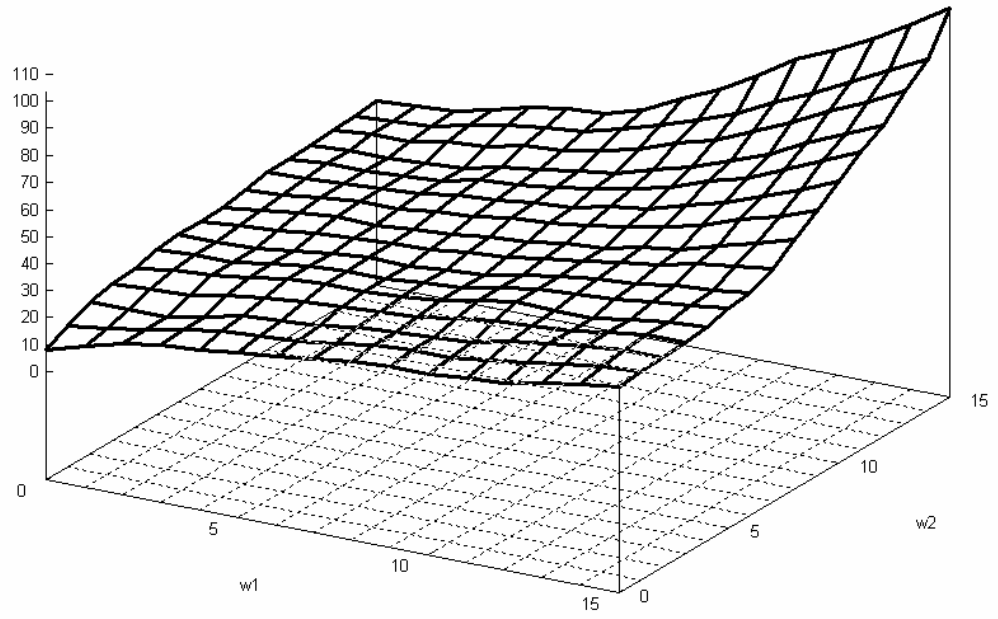


Fig. 4.9. Linearly interpolated quantization matrix for subbands (intra).

$$G(n_1,n_2)_{LL}=\begin{bmatrix} 6 & 6 & 9 & 10 & 11 & 10 & 10 & 9 \\ 7 & 7 & 9 & 10 & 11 & 11 & 10 & 9 \\ 9 & 9 & 10 & 11 & 13 & 12 & 12 & 10 \\ 9 & 10 & 11 & 12 & 14 & 13 & 12 & 11 \\ 11 & 11 & 12 & 14 & 15 & 14 & 13 & 11 \\ 10 & 11 & 12 & 13 & 14 & 13 & 12 & 10 \\ 10 & 10 & 11 & 12 & 12 & 12 & 10 & 9 \\ 8 & 9 & 10 & 10 & 11 & 10 & 9 & 8 \end{bmatrix} \quad G(n_1,n_2)_{HL}=\begin{bmatrix} 28 & 26 & 22 & 20 & 20 & 21 & 22 & 19 \\ 30 & 26 & 22 & 21 & 20 & 21 & 21 & 28 \\ 31 & 29 & 25 & 22 & 22 & 25 & 22 & 25 \\ 33 & 30 & 25 & 23 & 22 & 24 & 23 & 27 \\ 33 & 31 & 26 & 24 & 22 & 24 & 22 & 39 \\ 32 & 28 & 24 & 22 & 19 & 24 & 20 & 52 \\ 27 & 24 & 20 & 19 & 17 & 24 & 17 & 51 \\ 26 & 21 & 18 & 16 & 30 & 58 & 57 & 85 \end{bmatrix}$$

$$G(n_1,n_2)_{LH}=\begin{bmatrix} 28 & 27 & 30 & 32 & 32 & 31 & 28 & 24 \\ 24 & 25 & 28 & 28 & 30 & 28 & 25 & 21 \\ 21 & 22 & 24 & 24 & 25 & 24 & 22 & 18 \\ 20 & 19 & 21 & 22 & 23 & 22 & 20 & 18 \\ 18 & 18 & 19 & 20 & 21 & 20 & 18 & 15 \\ 18 & 19 & 20 & 21 & 23 & 22 & 19 & 17 \\ 17 & 18 & 19 & 21 & 21 & 20 & 18 & 16 \\ 18 & 19 & 20 & 22 & 23 & 22 & 35 & 47 \end{bmatrix} \quad G(n_1,n_2)_{HH}=\begin{bmatrix} 73 & 65 & 54 & 49 & 44 & 72 & 56 & 96 \\ 64 & 53 & 48 & 44 & 38 & 67 & 51 & 85 \\ 57 & 49 & 41 & 38 & 34 & 65 & 34 & 92 \\ 52 & 45 & 40 & 37 & 33 & 64 & 50 & 96 \\ 48 & 42 & 37 & 33 & 33 & 77 & 50 & 93 \\ 62 & 57 & 54 & 63 & 62 & 85 & 82 & 96 \\ 57 & 46 & 41 & 53 & 65 & 80 & 66 & 96 \\ 82 & 76 & 75 & 79 & 86 & 85 & 88 & 96 \end{bmatrix}$$

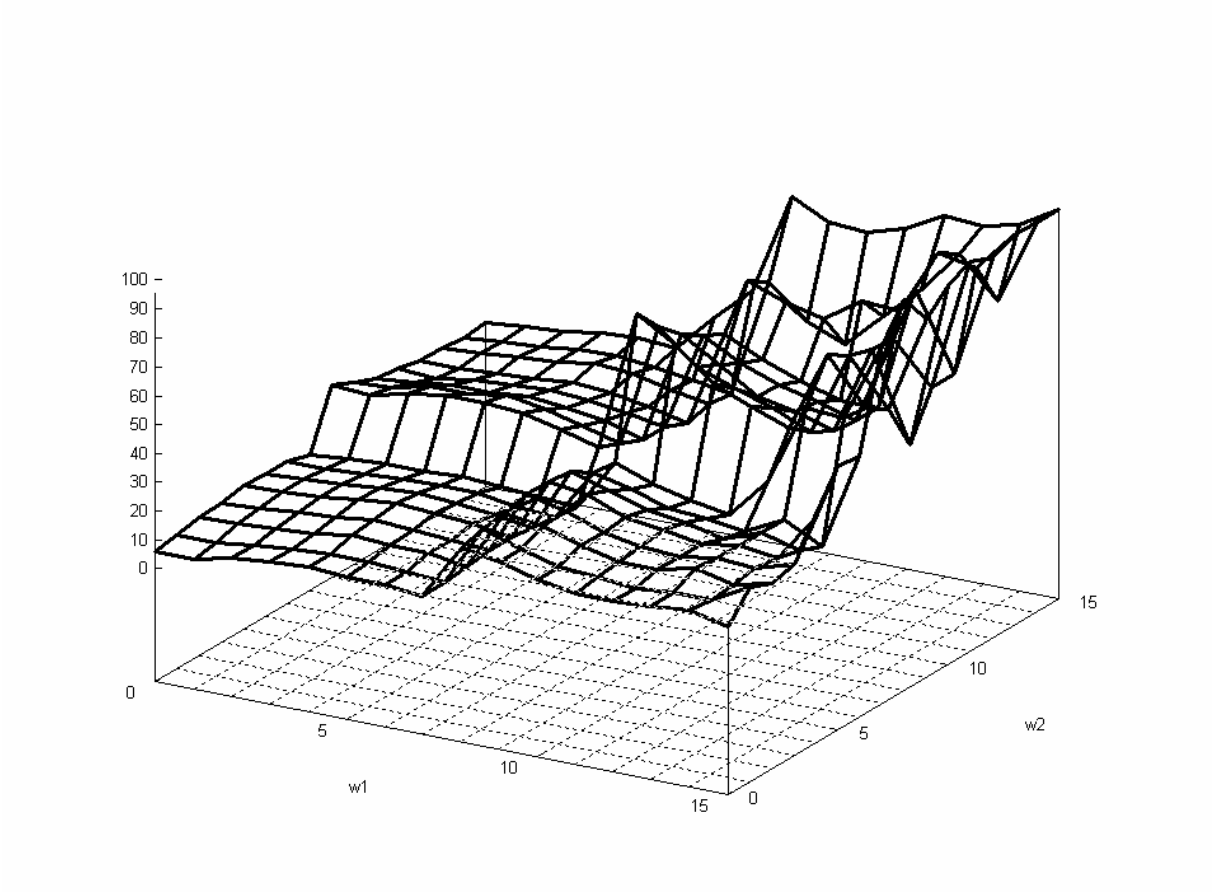


Fig. 4.10. Modified quantization matrix for subbands (intra).

The described DCT quantization tables have been tested experimentally, using a scalable encoder operating in the intra-frame mode. Simulations using several standard test sequences have been performed. The simulations presented in Tab. 4.3 are for illustration purposes and do not represent the results obtained from an optimized video

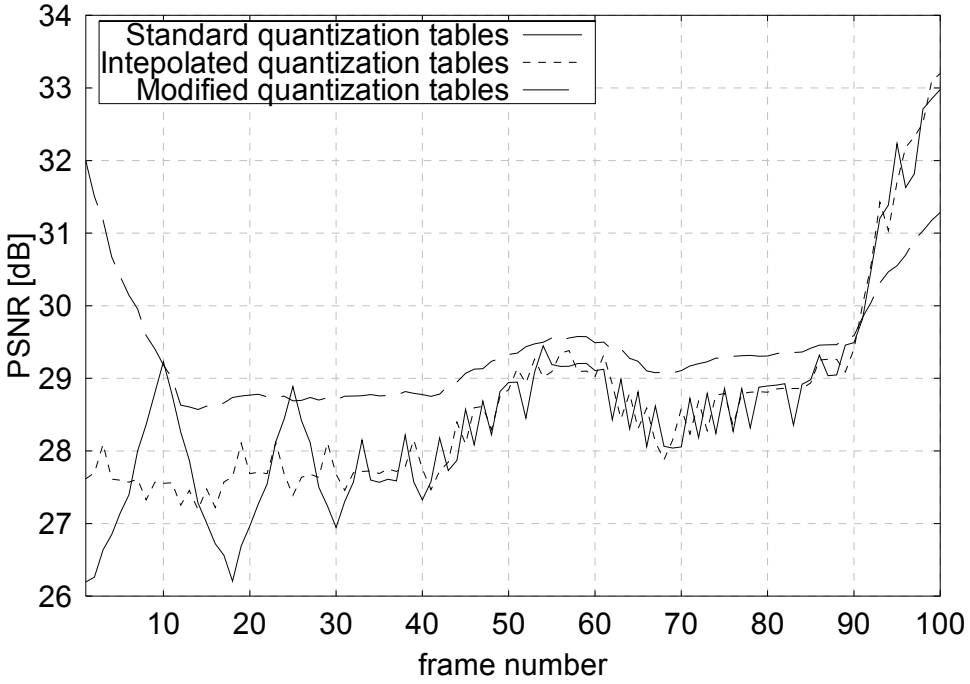
encoder. 4:2:0 progressive sequences of 720 x 576 pixels at the frame rate of 50 Hz have been used in the simulations.

Table 4.3 provides simulation results for the interpolated and modified quantization matrices. In these experiments the author has used different types of quantization matrices for subbands and has evaluated the efficiency of scalable encoding. The same scalable encoder was used in all three cases, with standard MPEG-2 quantization matrices, interpolated quantization matrices and modified quantization matrices. Table 4.3 summarizes the performance for the *Basket* and *Cheer* sequences. In this table, the PSNR are averaged for 100 pictures. In each case, PSNR is higher, i.e. better quality is obtained. Figures 4.11 and 4.12 show the PSNR of the consecutive pictures in time.

Table 4.3. Experimental results for intra frames (Scalable encoder with only intraframe coding, the results are averaged for 100 pictures).

		Test sequence	
		Basket	Cheer
Standard quantization tables	Bitstream [Mbps]	7,44	7,66
	Average PSNR [dB] for luminance – base layer	28,54	29,58
	Average PSNR [dB] for luminance – enhancement layer	28,18	30,47
	Base layer bitstream [Mbps]	2,50	2,50
	Base layer bitstream [%]	33,6	32,6
Interpolated quantization tables	Bitstream [Mbps]	7,43	7,67
	Average PSNR [dB] for luminance – base layer	28,60	29,62
	Average PSNR [dB] for luminance – enhancement layer	29,0	30,57
	Base layer bitstream [Mbps]	2,50	2,50
	Base layer bitstream [%]	33,6	32,6
Modified quantization tables	Bitstream [Mbps]	8,01	7,87
	Average PSNR [dB] for luminance – base layer	29,37	30,12
	Average PSNR [dB] for luminance – enhancement layer	29,54	30,95
	Base layer bitstream [Mbps]	2,85	2,71
	Base layer bitstream [%]	35,6	34,4

a)



b)

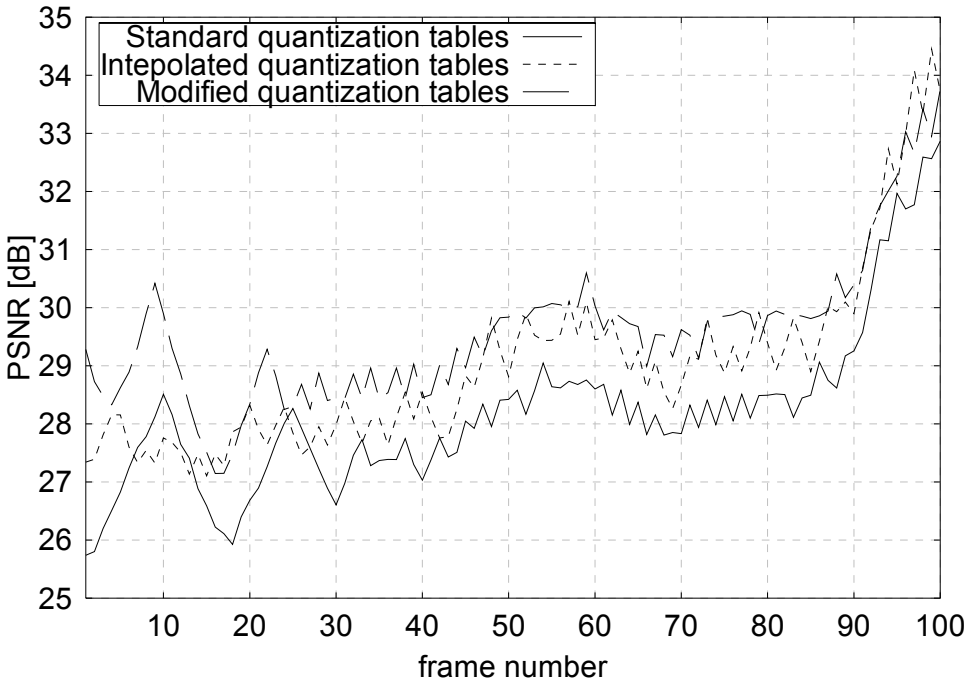
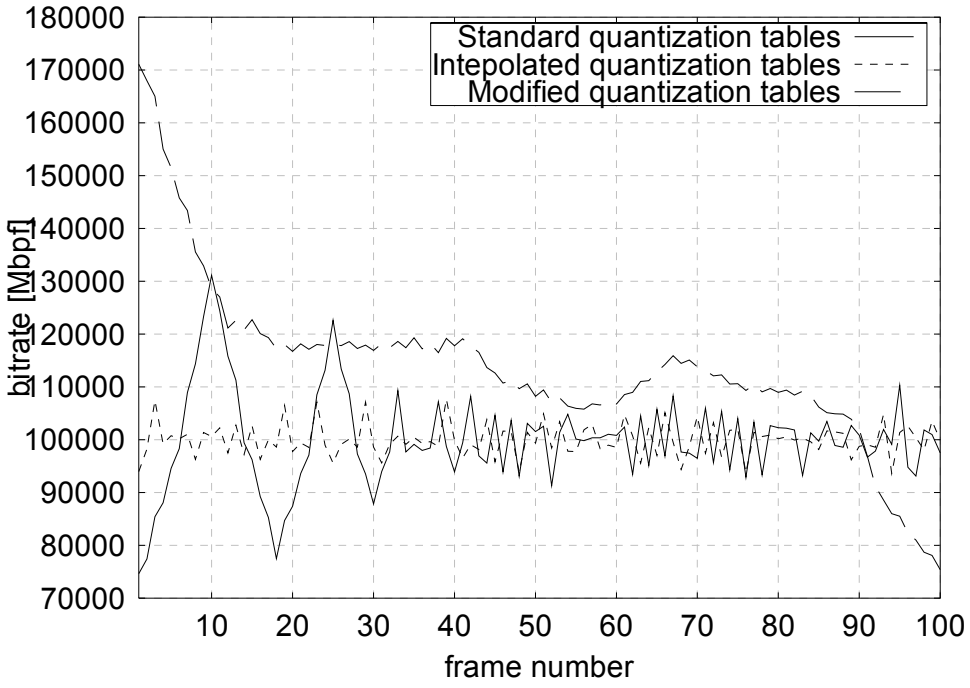


Fig. 4.11. PSNR versus frame numbers. The experimental results for scalable intraframe coding of test sequence *Basket*, (a) the base layer, (b) the enhancement layer.

a)



b)

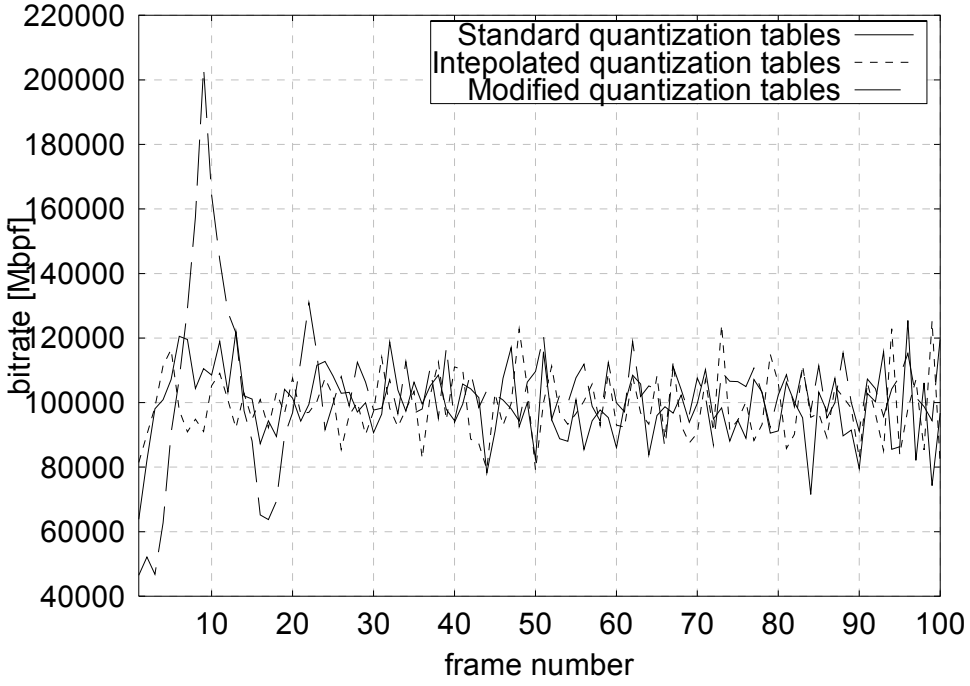


Fig. 4.12. Bitrate versus frame numbers. The experimental results for scalable intraframe coding of test sequence *Basket*, (a) the base layer, (b) the enhancement layer.

The author has compared two sets of quantizatation matrices known from the literature. The results (Table 4.3) show that improvement of compression efficiency of the scalable encoder is possible. The interpolation of the MPEG-2 standard quantization matrix values allows to achieve the same results as the use of the original MPEG-2 standard quantization matrix in individual subbands. The values of quantization matrices should be adjusted to the nature of the subbands. The results show that the MPEG-2 standard quantization matrix cannot be efficiently used to encode the subbands.

4.3.4. Modified zig-zag scan of DCT coefficients of subbands

The application of the two-dimensional DCT results in an 8x8 block of DCT coefficients, in which most energy in the block is typically concentrated in a few low-frequency coefficients. The coefficients of each block are scanned and transmitted in a zig-zag order. Two scanning orders are available in MPEG-2. The zig-zag order (Fig. 4.13a) is used in JPEG, H.261 and MPEG-1 standards [Doma98d]. The alternate scan often gives better compression for interlaced video when there is significant motion.

The appropriate order of scanning the DCT coefficients ensures better compression. The coefficients are ordered from maximal to minimal values and the differences between them are run-level encoded. Since the DCT coefficients of the subbands LH, HL and HH have different concentrations of energy in the 8 x 8 blocks, a different zig-zag order is expected. In 1996, Kronacher proposed to use the zig-zag order for the subbands (Fig. 4.13) [Kron96]. The present author uses Kronacker's scan matrices in the scalable encoder to scan the DCT coefficients in subbands LH, HL and HH.

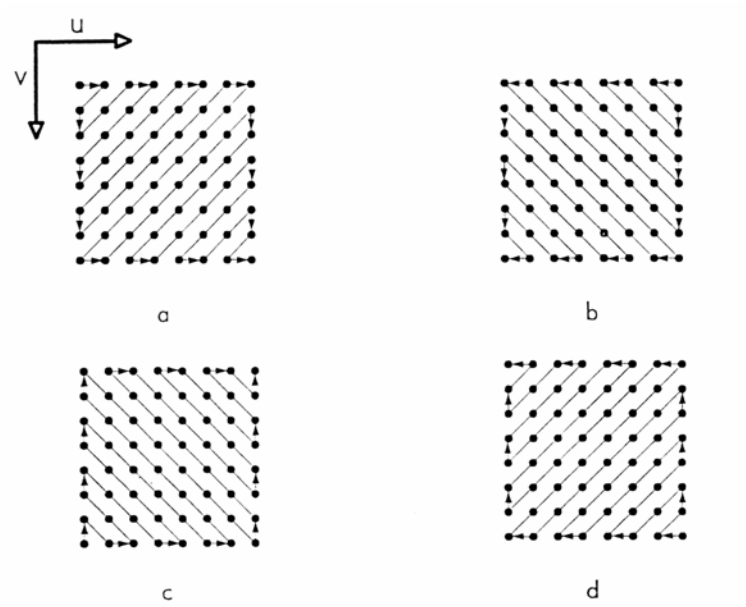


Fig. 4.13. Zig-zag scan matrix for subband a) LL, b) LH, c) HL and d) HH.

The author has compared the performance of the standard zig-zag scanning order with the perfomance of modified scanning order for subbands. The experiments have been made with 50Hz 720 x 576 progressive sequences which correspond to a 25Hz SIF progressive sequence in the base layer. The results have been presented in Table 4.4.

The enhancement layer average PSNR [dB] for luminance, which determines the image quality, has been separately presented in Figs 4.14a and 4.14b for two test sequences, for various types of quantization matrices.

Table 4.4. Experimental results for intra (Scalable encoder with intraframe coding only, the results are averaged for 100 pictures).

		Test sequence			
		Basket	Cheer	Stefan	Mobile & Calendar
Standard quantization tables	Bitstream [Mbps]	7,65	7,66	7,65	7,66
	Average PSNR [dB] for luminance – base layer	28,54	29,58	31,75	25,25
	Average PSNR [dB] for luminance – enhancement layer	29,01	30,47	32,44	24,77
	Base layer bitstream [Mbps]	2,50	2,50	2,50	2,50
	Base layer bitstream [%]	32,3	32,6	32,68	32,64
Standard quantization tables + Modified scanning tables	Bitstream [Mbps]	7,66	7,69	7,67	7,67
	Average PSNR [dB] for luminance – base layer	28,54	29,58	31,75	25,25
	Average PSNR [dB] for luminance – enhancement layer	29,36	30,70	32,94	25,12
	Base layer bitstream [Mbps]	2,50	2,50	2,50	2,50
	Base layer bitstream [%]	32,6	32,5	32,6	32,6
Interpolated quantization tables	Bitstream [Mbps]	7,67	7,67	7,67	7,67
	Average PSNR [dB] for luminance – base layer	28,60	29,62	31,84	25,31
	Average PSNR [dB] for luminance – enhancement layer	29,11	30,57	32,57	24,82
	Base layer bitstream [Mbps]	2,50	2,50	2,50	2,50
	Base layer bitstream [%]	33,6	32,6	32,6	32,6
Interpolated quantization tables + Modified scanning tables	Bitstream [Mbps]	7,67	7,69	7,67	7,68
	Average PSNR [dB] for luminance – base layer	28,60	29,62	31,84	25,31
	Average PSNR [dB] for luminance – enhancement layer	29,37	30,70	32,95	25,10
	Base layer bitstream [Mbps]	2,50	2,50	2,50	2,50
	Base layer bitstream [%]	32,6	32,5	32,6	32,6

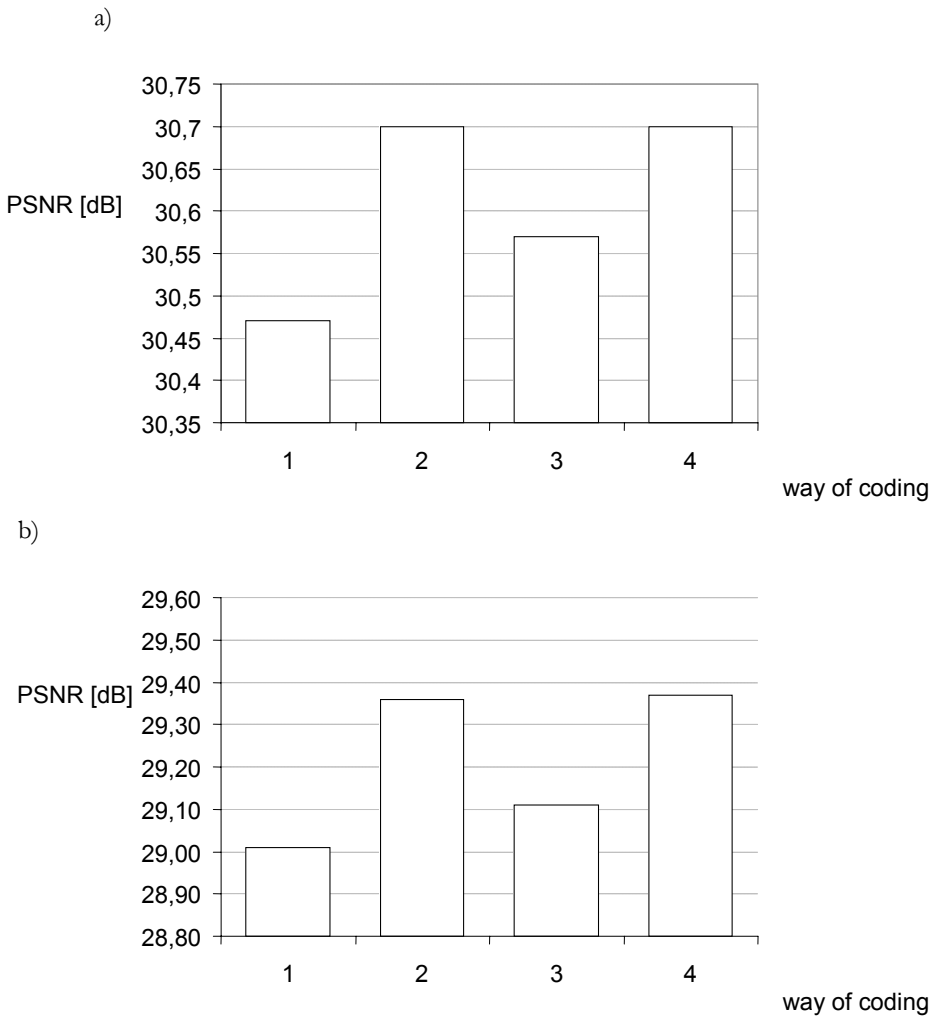


Fig. 4.14. Enhancement layer average PSNR [dB] for luminance, 7.6Mbps – for test sequences a) *Basket* and b) *Cheer*. The bars denote coding with: 1) standard quantization matrices, 2) standard quantization tables and modified scanning matrices, 3) interpolated quantization matrices, 4) interpolated quantization matrices and modified scanning matrices

The results of this comparison, also included in Table 4.4, show that improvement of compression efficiency of a scalable encoder is possible. Using modified scanning matrices gives an improvement of up to 0.5dB. The scanning matrices should be adjusted to the concentration of energy of DCT coefficients in the subbands. The results show that the MPEG-2 standard zig-zag scan cannot be used efficiently to encode the subbands.

4.3.5. Coding additional information *ΔLL*

As it was said earlier, the base layer represents the low-quality video. On the other hand, quality of the subband LL strongly influences the quality of a full-sized picture. Low quality of the LL subband restricts the full-sized picture quality to a relatively low level despite the amount of information in the remaining subbands. Therefore it is essential to transmit additional information *ΔLL* in the enhancement layer. This information is used to improve the quality of the subband LL when the latter is used to synthetize a full-sized image in the enhancement layer.

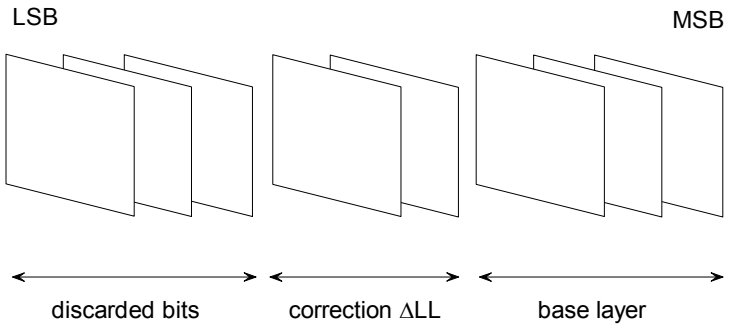


Fig. 4.15. Partitioning bitplanes between layers.

The bitstream *ΔLL* contains bitplanes correcting the transform coefficients transmitted in the base layer (Fig. 4.15). *ΔLL* is a quantized difference between the DCT coefficients from the base layer image and the dequantized values of these coefficients (see Eq. 4.4). The difference is quantized with a quantization parameter Q_e.

$$\Delta LL = Q\left(F_b(n_1,n_2) - \widetilde{F}'_b(n_1,n_2), Q_e\right), \tag{4.4}$$

where:

$F_b(n_1,n_2)$ - DCT coefficient from the base layer image,

$\widetilde{F}'_b(n_1,n_2)$ - dequantized DCT coefficient from the base layer image,

$Q(\ )$ - quantization function,

Q_e - quantization parameter in the enhancement layer.

Since ΔLL is quantized in a two-step process, the second quantization is finer. Therefore the relationship between the quantization parameter Q_e in the enhancement layer and the quantization parameter Q_b in the base layer is defined as follows (Eq. 4.5).

$$Q_e = Q_b - 2 \tag{4.5}$$

where:

Q_b - quantization parameter in the base layer.

2	-4	4	3	2	-1	2	2
-1	3	3	3	1	4	-2	5
1	-2	2	-6	-6	2	-6	-2
-1	1	-6	1	-7	9	-4	-3
1	0	-2	4	-7	0	-6	5
-1	-2	0	1	-1	1	0	-1
1	0	-1	0	0	1	1	-1

Fig. 4.16. Block of 8 x 8 values of the ΔLL signal of one base layer picture from the sequence *Mobile & Calendar*.

Since the ΔLL signal is similar to a noise signal (Fig. 4.16), it is difficult to encode. After several experiments, the author has proposed a modified bitplane technique to encode this signal. In the proposed technique, for each 8x8 ΔLL block, 64 absolute values (decimal numbers) are zig-zag ordered into an array. The decimal numbers are then transformed to binary numbers. A bitplane of each block is now defined as an array of 64 bits taken from the same bit position. For every bitplane of each block, RUN and End of Plane symbols are formed and variable-length coded to produce the output bitstream.

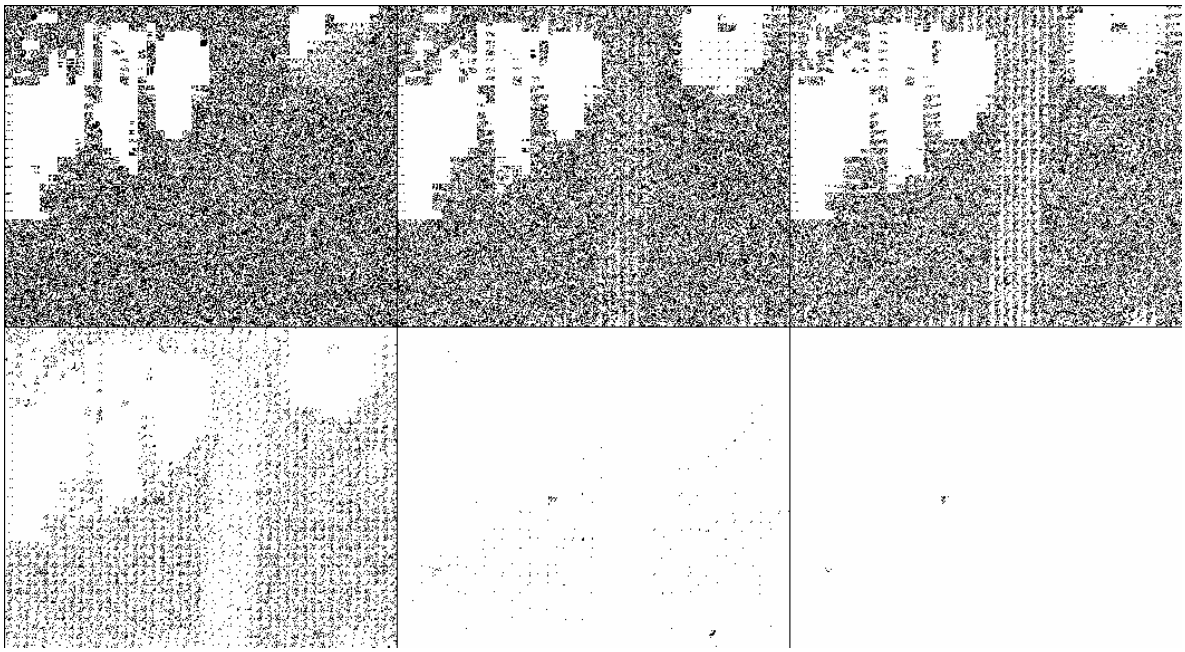


Fig. 4.17. *ΔLL* bitplanes of one base layer picture from the sequence *Flower Garden*.

4.3.6. Coding exploiting inter-subband correlation

Pictures showing natural scenes usually exhibit very high cross-correlation between sub-images corresponding to different subbands of the spatial spectrum. Therefore a technique originally successfully applied to still images [Doma94, Doma94a] can be adopted for efficient encoding of the subbands HL, LH and HH (Fig. 4.18). This technique exploits the observation that images of natural scenes exhibit power spectra which decrease more or less uniformly for increasing frequencies. Therefore, the portions of high frequency sub-images where signal values are significant usually correspond to those portions of low-frequency sub-images where some changes occur. Most portions of the high frequency sub-images exhibit only small signal values. Therefore a quantizer with a dead-zone produces a lot of zeros. It is not necessary to transmit all of them and the low-frequency sub-image can be used to identify the positions of the "active" portions of high-frequency sub-images.

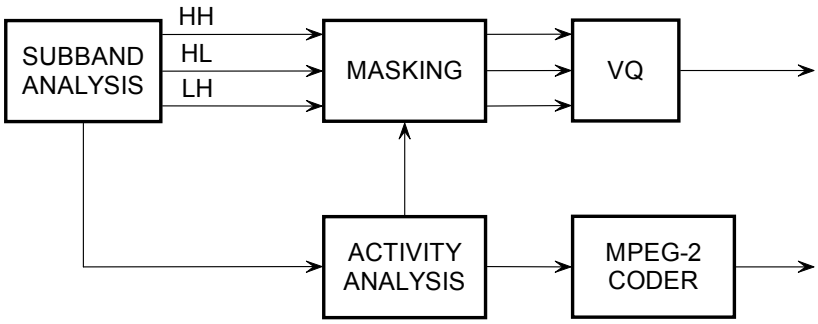


Fig. 4.18. Coding exploiting inter-subband correlation.

An 'active' portion is characterized by horizontal activity A_H and vertical activity A_V, which are defined as follows:

$$A_H = \max\left(\left|f_{LL}(n_1, n_2 - 1) - f_{LL}(n_1, n_2)\right|, \left|f_{LL}(n_1, n_2 + 1) - f_{LL}(n_1, n_2)\right|\right), \tag{4.6}$$

$$A_V = \max\left(\left|f_{LL}(n_1 - 1, n_2) - f_{LL}(n_1, n_2)\right|, \left|f_{LL}(n_1 + 1, n_2) - f_{LL}(n_1, n_2)\right|\right), \tag{4.7}$$

where $f_{LL}(n_1, n_2)$ is a pixel value from subband LL.

The image has to be retrieved uniquely, so all decisions made in the encoder must use only that information which is later available in the decoder. Thus the encoder needs a reconstructed version of the low-frequency subband.

The pixel in subband LL is considered as "active" if $A_H > T$ (or $A_V > T$), where T is a predefined threshold. The technique leads to ignoring those pixels from high-frequency sub-images whose corresponding pixels in the low-frequency sub-image are not active. The subbands HL, LH and HH are encoded using Huffman codes.

4.3.7. Control unit

MPEG-2 does not standardize any algorithm to set the encoder parameters in order to obtain a required bitrate and/or distortion level. The video coding part of MPEG-2 describes a buffer control algorithm, which is only one part of bitrate and distortion control mechanisms. Therefore it is necessary to establish the encoder

parameters individually. The basic parameter that can be used to control an encoder is the quantization factor Q, which influences the quantization of the DCT coefficients. The value of Q must be adjusted to ensure the same quality in the variable bitrate mode (VBR) or to ensure constant bitrate in the constant bitrate mode (CBR). In the communication systems, the video encoders often operate in the latter mode because of the need to match the bitstream and the communication channel throughput. The quality of the control algorithm deeply influences the overall coding performance of the video coding system. Unfortunately, despite many MPEG-2 encoders and decoders working worldwide, the construction of an efficient control algorithm is still an open problem.

In particular, the question is how to control the value of Q in order to maximize the video quality at a given bitrate. The answer is related to the two following control problems:

- global control, i.e. adjusting the value of Q in a frame header,
- local control, i.e. adjusting the value of Q in individual macroblock headers.

Łuczak and Maćkowiak have established a simple empirical model of bitrate as a function of Q, calculated for individual frames. The experimental curves (Figs. 4.19 and 4.20) have been obtained by calculating the number of bits required for encoding and transmitting the image as a function of the quantization parameter Q. The author has used this model for global control in the constant bitrate mode of operation.

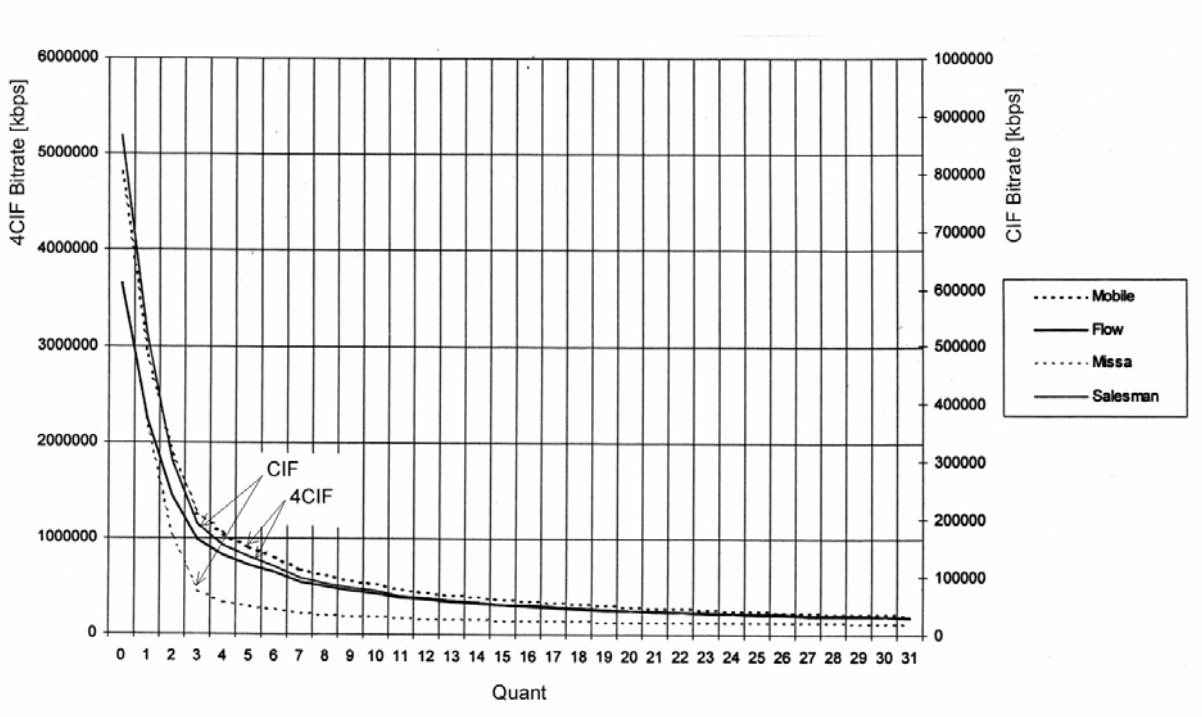


Fig. 4.19. Experimental curves of bitrate versus quantization value (quant) for I-frames.

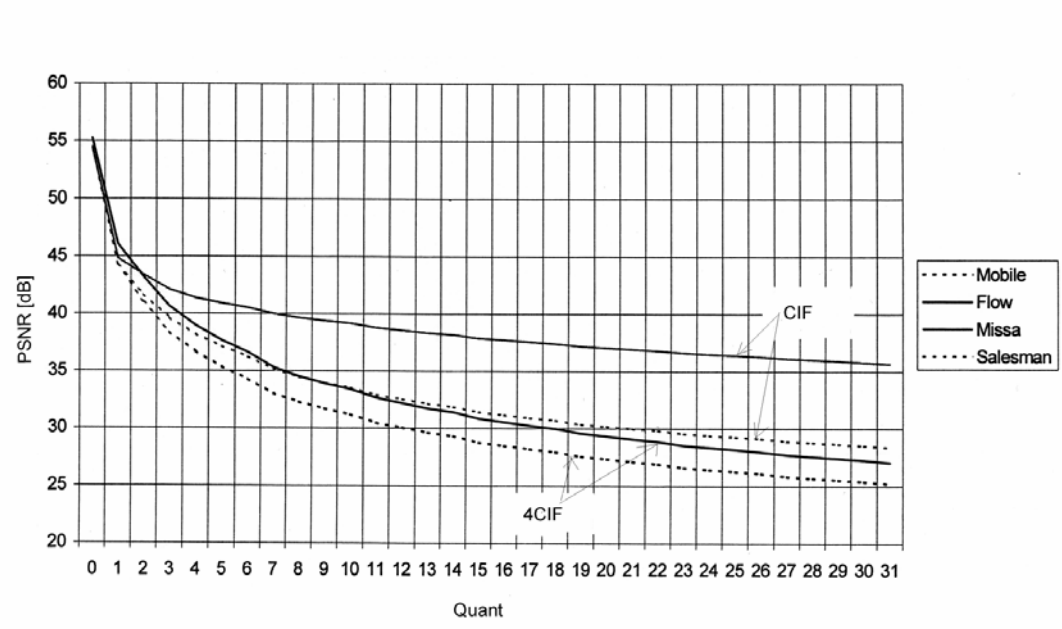


Fig. 4.20. Experimental curves of PSNR versus quant value for I-frames.

The algorithm works in three steps. In the first step, the algorithm estimates the parameter *j*, which modifies the deflection of the curve (Eqs. 4.9 and 4.11). The parameter *j* depends on the number of bits used for the encoding of the previous frame and on the quantization parameter Q_{prev} value used in the previous frame. It is performed

before coding the current picture. Next, the algorithm computes the quantization parameter Q_{cur} (Eqs. 4.10 and 4.12). Since Q_{cur} changes in the interval <0,31>, it influences the current bitrate in a wide range. In order to limit the fluctuations of the quality of consecutive images, the value of $(Q_{prev} - Q_{cur})$ is clipped (Fig. 4.21, Eq. 4.8) to the range [-2,…,2] and is added to Q_{prev}. Finally, the calculated value of $(Q_{prev} + \Delta Q)$ is used to quantize the DCT coefficients of the current image.

$$\Delta Q = NINT\left(arctg\left(\left(Q_{prev} - Q_{cur}\right)\cdot 2\right)\right) \tag{4.8}$$

where:

NINT – the nearest integer value

Q_{prev} - value of Q used in previous frame,

Q_{cur} - calculated quantization parameter for the current frame

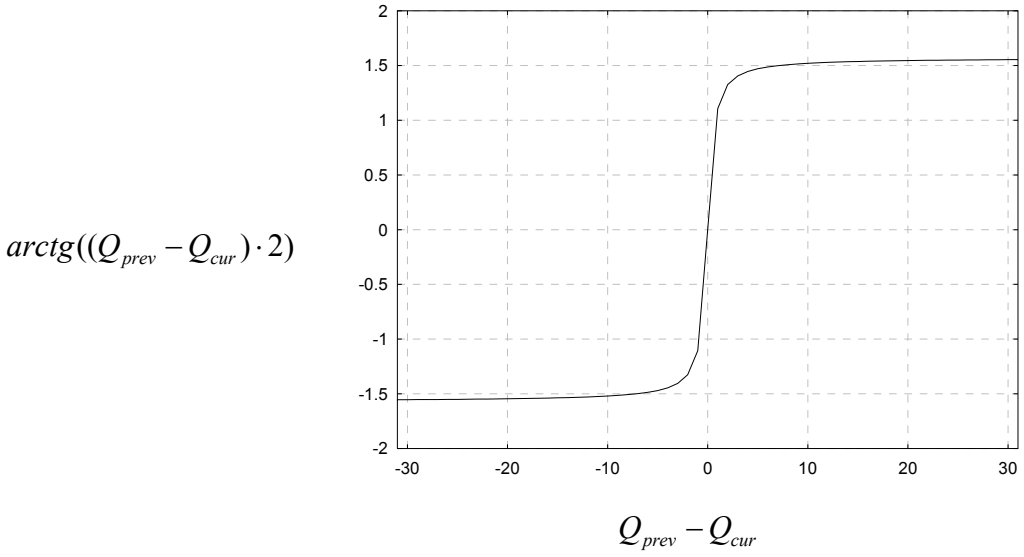


Fig. 4.21. Clipping function.

For an I-frame:

$$j = \frac{\frac{5 \cdot 10^6}{bitrate_{prev}} - 0.4}{Q_{prev}^{0.8}} \tag{4.9}$$

where:

$bitrate_{prev}$ - the number of bits in the previous I-frame,

Q_{prev} - value of Q used in the previous I-frame,

and

$$Q_{cur} = e^{\frac{5}{4} \cdot \log_{10}\left[\left(\frac{5 \cdot 10^6}{bitrate} - 0.4\right) \cdot \frac{1}{j}\right]} \tag{4.10}$$

where:

bitrate - required bitrate,

Q_{cur} - calculated quantization parameter of the current frame

Similarly, in a P-frame and B-frame:

$$j = \frac{\frac{5 \cdot 10^6}{bitrate_{prev}} - 0.4}{Q_{prev}^{1.25}} \tag{4.11}$$

where:

$bitrate_{prev}$ - the number of bits for the previous frame,

Q_{prev} - value of Q used in previous frame,

and

$$Q_{cur} = e^{\frac{4}{5} \cdot \log_{10}\left[\left(\frac{5 \cdot 10^6}{bitrate} - 0.4\right) \cdot \frac{1}{j}\right]} \tag{4.12}$$

where:

bitrate - the required bitrate,

Q_{cur} - the calculated quantization parameter of the current frame

4.3.8. Results of experiments

Several experiments have been made with 50Hz 720 x 576 progressive sequences, which correspond to 25Hz SIF progressive sequences in the base layer. The bitstream produced in the base layer encoder is an MPEG-2 bitstream that also includes motion vectors estimated with half-pel accuracy for SIF images.

In the first series of the experiments, compression efficiency in subband DCT coding was compared to compression efficiency in subband coding exploiting inter-subband correlation.

In order to evaluate compression efficiency, a verification model of the scalable encoder and an implementation of the MPEG-2 encoder have been used (see Annex B).

The experiments fulfilled the following conditions:

- motion estimation with half-pixel accuracy,
- full search of motion vectors in range [-31,32],
- independent control of the bitrate in the base and enahancement layers,
- only I-frames.

The scalable encoder uses the control algorithm presented in Section 4.3.7.

The experimental results are presented in Table 4.5. Note that DCT coding of the subbands is more efficient than coding exploiting inter-subband correlation.

Table. 4.5. Comparison of compression efficiency of subband coding (only intraframe coding).

	Test sequences			
	Flower Garden		Funfair	
	Average luminance PSNR [dB]	Scalable bitstream [Mbps]	Average luminance PSNR [dB]	Scalable bitstream [Mbps]
DCT coding	29.8	5.35	31.0	5.04
coding exploiting inter-subband correlation	28.8	5.27	29.5	5.11

In the second series of the experiments, the banks of analysis and synthesis filters have been compared. Johnston 12th order and Daubechies (9,3) filters were selected for the tests. The results are given in Table 4.6. The application of both filters results in the same compression efficiency.

For further experiments, DCT coding of the subbands and Johnston 12th analysis and synthesis filter banks were selected.

Table 4.6. Comparison of compression efficiency of intraframe coding for different banks of analysis and synthesis filters.

	Test sequences			
	Mobile & Calendar		Basket	
	Average luminance PSNR [dB]	Scalable bitstream [Mbps]	Average luminance PSNR [dB]	Scalable bitstream [Mbps]
Johnston 12th order filter	27.5	5.03	29.9	5.20
Daubechies (9,3) filter	27.6	5.20	29.8	5.13

Further experiments show the efficiency of the proposed scalable encoder. The conditions of the experiments are as follows:

- motion estimation with half-pixel accuracy,
- full search of motion vectors in the range [-31,32],
- independent control of the bitrate in base and enahancement layers,
- the following GOP structure of the enhancement layer:
  I–B–B–B–P–B–B–B– P–B–B–B,
- 12 frames in a GOP.

The scalable encoder uses the control algorithm presented in Section 4.3.7.

The first goal, i.e. to ensure that the base layer bitstream does not exceed the bitstream of the enhancement layer, was reached for all bitrates and video test sequences. Nevertheless, the coding efficiency measured by bitrate overhead for scalable coding as compared to MPEG-2 nonscalable encoder was not satisfactory for all bitrate/sequence pairs. The experiments with lower bitrates mostly resulted in a 6–15% overhead. In such

cases, the proposed encoder strongly outperforms the spatially scalable MPEG-2 encoder.

The problem remains how to control efficiently the parameters of this quite sophisticated encoder. It must be noted that the results were obtained for standard (not optimized) quantizers and standard encoding of the ΔLL stream.

The experimental results of scalable coding with I-, P- and B-frames of progressive 720 x 576, 50 Hz test sequences (”Mobile & calendar”, ”Flower garden”, ”Basketball” and ”Funfair”) are given in Table 4.5.

Table 4.5. Experimental results of scalable coding with
of progressive test sequences, I- P- and B-frames.

	Test sequence						
	Flower Garden		Mobile & Calendar	Basketball		Funfair	
Single layer (MPEG-2) bitstream [Mbps]	4.90	7.57	7.89	4.67	6.99	4.67	6.15
Average luminance PSNR [dB]	29.6	32.6	30.8	29.9	32.1	31.3	32.9
Scalable bitstream [Mbps]	5.22	8.66	10.70	5.02	9.26	4.97	8.45
Average luminance PSNR [dB]	29.6	32.4	29.9	29.6	31.1	31.0	32.3
Base layer bitstream [Mbps]	2.13	3.01	3.35	2.07	3.12	2.02	3.17
Base layer bitstream [%]	40.8	34.8	31.3	40.3	33.7	40.6	37.5
Bitstream overhead [%]	6.5	14.4	35.6	7.5	32.5	6.4	37.4

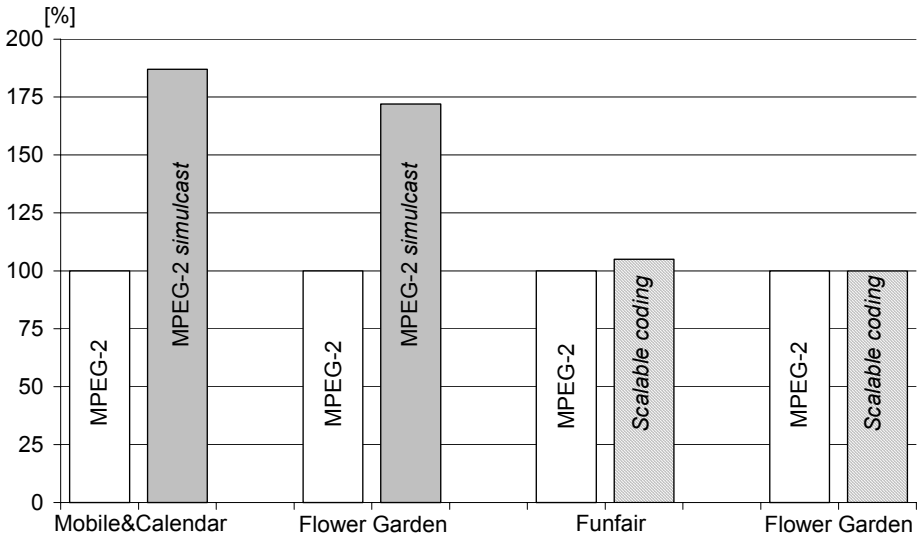


Fig. 4.22. Comparison of compression efficiency between simulcast coding and scalable coding (I-, P- and B-frames). The figure was prepared according to results from Table 4.5.

4.3.9. Discussion

Coding efficiency measured by bitrate overhead for scalable coding as compared to the MPEG-2 non-scalable encoder is satisfactory for bitrate about 5Mbps. The overhead is 6-15 %; however, the proposed scalable system is inefficient for bitrates above 5Mbps. For such bitrates, the bitrate overhead is too large.

A scalable video technique should serve a broad range of bitrates (e.g. from a few kbps to several Mbps) in networks or a wide selection of terminals with different characteristics, using only one universal scalable encoding. Therefore other proposals of scalable video coding have been developed. One such proposal, i.e. the hybrid DCT-based with spatio-temporal decomposition solution is presented in detail in the next two chapters.

Chapter 5

Multi-loop spatio-temporal encoders

5.1. Introduction

This chapter presents an original proposal for spatio-temporal scalability in hybrid DCT-based video encoders. The low resolution base layer bitstream is fully compatible with the MPEG-2 standard. The whole structure exhibits high level of compatibility with individual building blocks of MPEG-2 encoders and the enhancement bitstream exploits the MPEG-2 bitstream semantics and syntax, with some modifications. The video sequence structure is that of MPEG-2. The sequence consists of Group of Pictures (GOPs) that are defined as content access units.

5.2. General structure of encoder

The proposed scalable encoder consists of two or three motion-compensated hybrid encoders (Fig. 5.1), which encode a video sequence and produce two or three bitstreams corresponding to two or three different levels of spatial and temporal resolution. In this structure, it is also possible to encode a video sequence without temporal decomposition (as denoted in Fig. 5.1. with dotted lines).

For the experiments, the author has chosen a simplified structure of the encoder with two loops of motion estimation and compensation (Fig. 5.2).

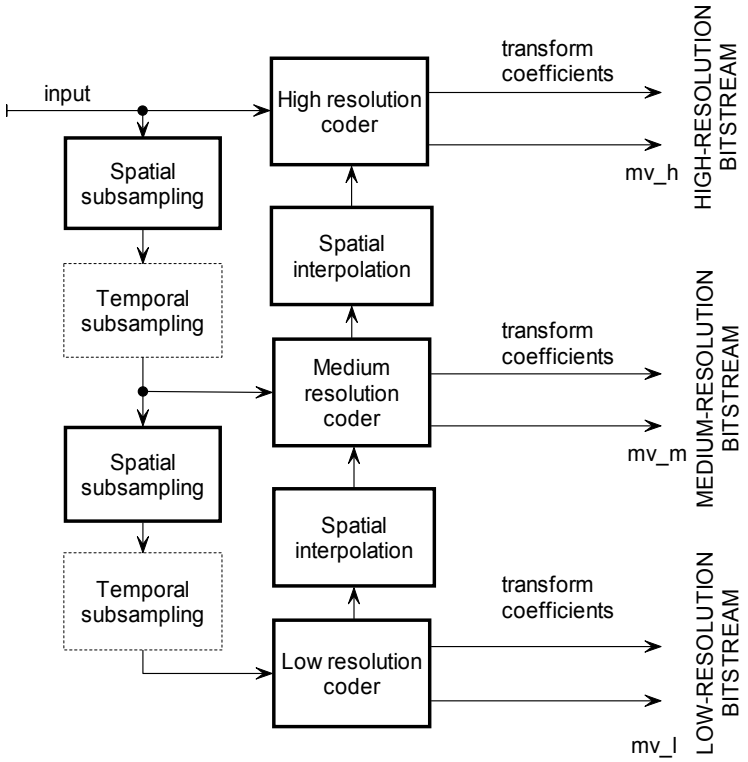


Fig.5.1. Structure of the proposed scalable encoder.

The basic structure of a group of pictures (GOP) consists of I- P- and B-frames (Fig. 5.3). The variant with 3 B-frames between two consecutive I- or P-frames has been chosen because of simple temporal decimation with factor 2.

In the proposed encoder, the temporal resolution reduction is achieved by partitioning the stream of B-frames: each second frame is skipped in the low resolution encoder. As mentioned in Section 3.3, there are two types of B-frames, i.e. BE-frames that exist in the high resolution sequence only and BR-frames that exist in both sequences (Fig. 5.5). In the experiments, in the high resolution video sequence the number of B-frames between two consecutive I- or P-frames is even.

In the enhancement layer there also exist PI-frames, i.e. frames which are encoded without motion vectors.

For the case of temporal decimation with factor 3, another GOP structure is appropriate (Fig. 5.4).

The GOP structure version without B-frames is presented in Fig. 5.5.

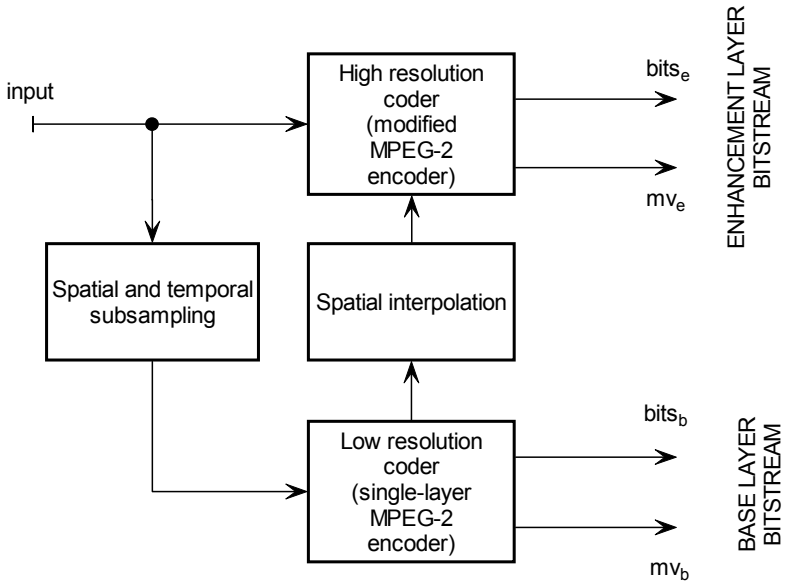


Fig.5.2. Structure of the proposed scalable encoder selected for experiments.

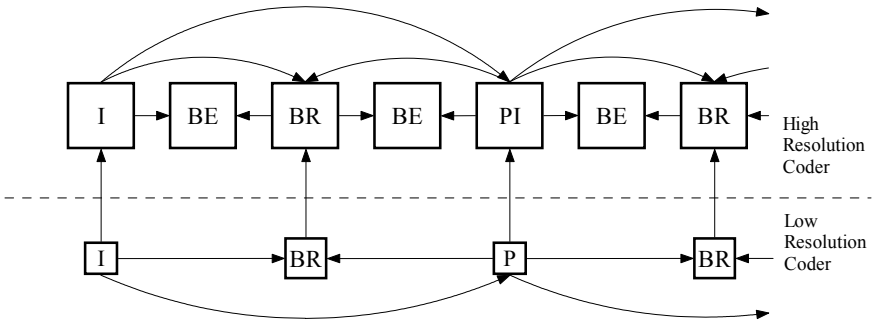


Fig. 5.3. Selected GOP structure with P- and B-frames in low and high resolution bitstreams.

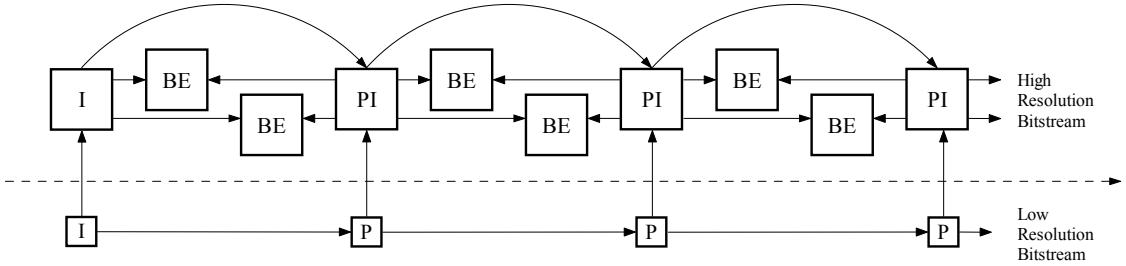


Fig. 5.4. GOP structure with temporal resolution decimation by factor 3.

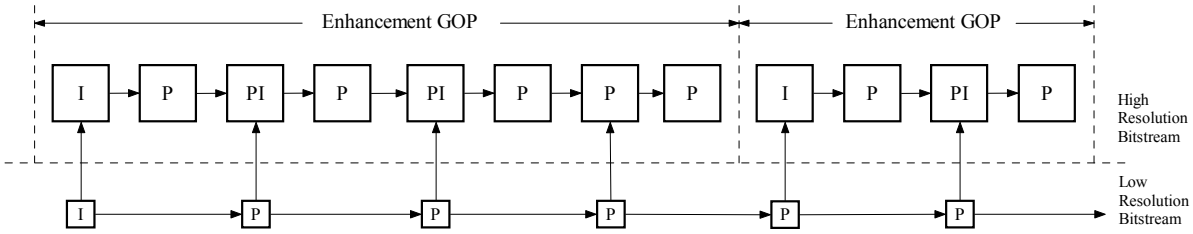


Fig. 5.5. GOP structure with only P-frames in low and high resolution layers.

5.3. Reference structure of encoder

For in-depth analysis, a two-loop encoder has been chosen. The respective GOP structure is as in Fig. 5.3.

The proposed encoder consists of a low resolution encoder and a high resolution encoder (Fig. 5.2). The low resolution encoder is implemented as a motion-compensated hybrid MPEG-2 encoder of the Main Profile@Main Level (MPEG-2 MP@ML) [ISO94, Hask96].

The low resolution sequence exhibits reduced spatial resolution. Thus, two basic video sequences are processed in the proposed encoder (Fig. 5.6):

- The low resolution sequence with reduced picture frequency and reduced horizontal and vertical resolution.
- The high resolution sequence with original resolution in time and space.

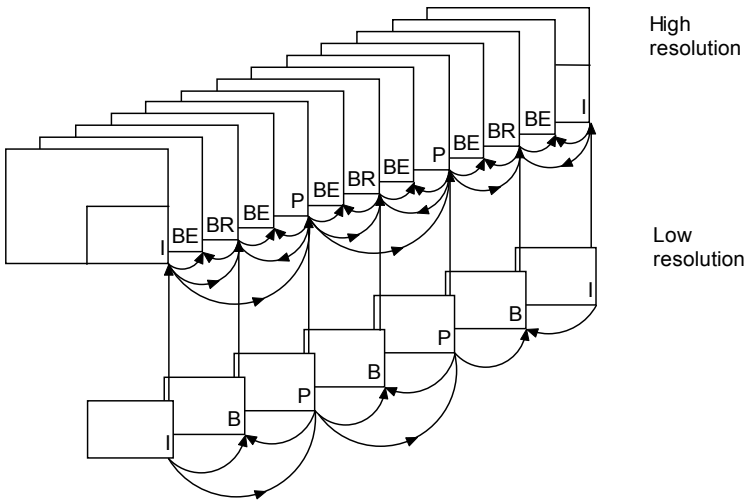


Fig. 5.6. Structure of a Group of Pictures: Arrows point to frames predicted on the basis of reference frames.

The high resolution encoder (Fig. 5.7) is a modification of the MPEG-2 encoder. In this encoder, the I- and P-frames are encoded as in the standard spatially scalable MPEG encoder, i.e. the interpolated image from the base layer is used as an additional reference frame [Doma00c, Doma01]. A high resolution P-frame is predicted using the previous reference frame, as well as the interpolated current low resolution frame encoded in the base layer. For each macroblock, the best prediction is selected from three blocks, i.e. two motion-compensated blocks and their average.

The motion-compensated predictor employed in the high resolution layer uses a new prediction for B-frames. The proposed new prediction is described in detail in the next section. As an extension to the MPEG-2 compression technique, in the new prediction those B-frames which correspond to B-frames from the base layer can be used as reference frames for predicting other B-frames in the enhancement layer (Fig. 5.6).

The low resolution encoder produces a base layer bitstream with reduced spatial and temporal resolution. Temporal resolution reduction is achieved by partitioning the stream of B-frames: each second frame is not included into the base layer. The bitstream produced in the base layer is described by MPEG-2 standard syntax.

The proposed encoder applies independent motion compensation loops in all layers. The motion vectors mv_b for the low resolution frames are estimated independently from the mv_e, which are estimated for the high resolution images unlike other recent proposals [Rose01, HeYa01, Reib01, Peng01]. This solution was chosen after some experiments. The experiments, and the reasons for making this choice, will be presented in Section 5.4.

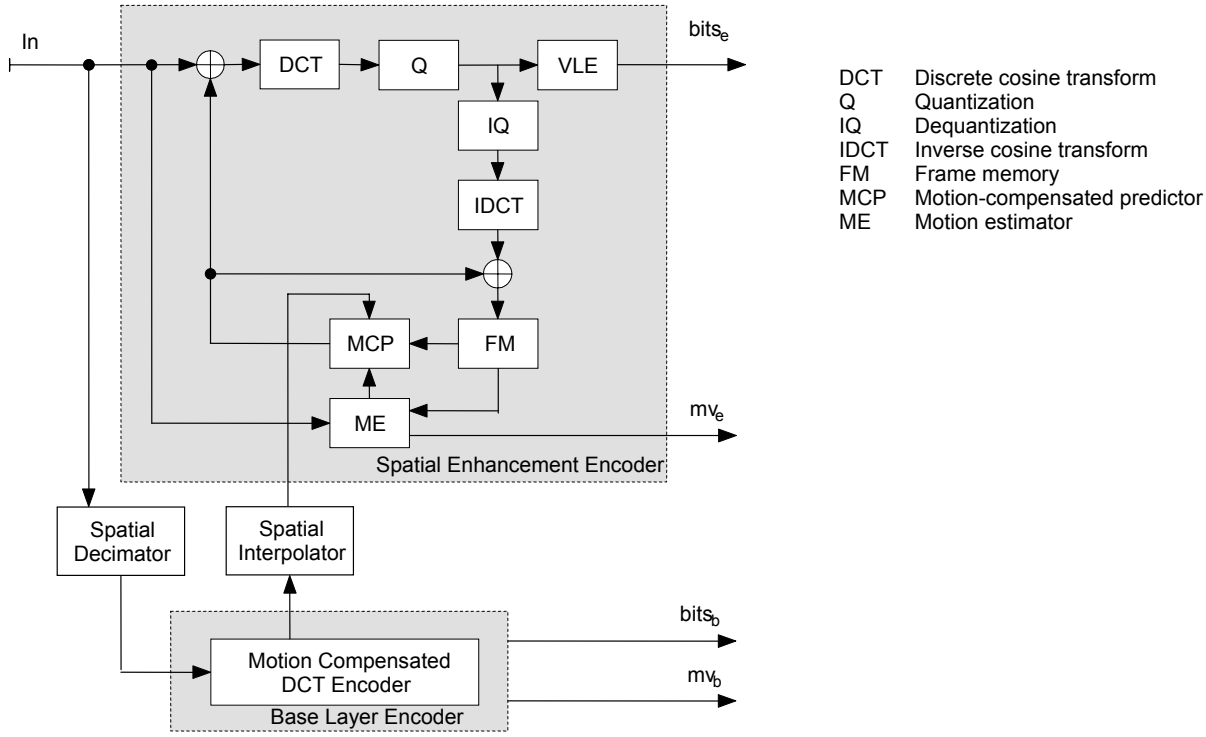


Fig. 5.7. Basic block diagram of the encoder.

5.4. Motion estimation and compensation

5.4.1 Introduction

Motion estimation is an important component of video coding systems because it enables us to exploit temporal redundancy in a video sequence. The scalable encoder proposed by the present author consists of two encoders. The low and high resolution layers produced by the low and high resolution encoders are characterized by different spatial and temporal resolution of the coded frames. Additionally, every second frame is skipped in the low resolution encoder. This means that motion estimation and compensation processes in the low and high resolution layers are performed for the frames from different time moments.

In the proposed encoder, it is possible to use independent motion estimation and compensation in the low and high resolution layers. Motion compensation processes in layers results in higher compression ratio. Therefore, the problem of efficient estimation of motion vectors in layers needs to be solved.

5.4.2 Independent motion compensation processes

High efficiency of scalable encoding requires using some information from the base layer in the high resolution encoder. The high resolution layer encoder uses the interpolated decoded frame from the base layer in the prediction of the full resolution frame. It is assumed that the base layer represents a video signal with half the spatial resolution. Therefore one macroblock in the base layer corresponds to four macroblocks in the enhancement layer (See Figs. 5.8 and 5.9).

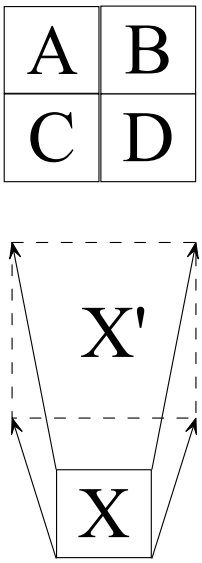


Fig. 5.8. Base layer macroblock X interpolated to X' corresponding to four macroblocks A, B, C and D in the high resolution layer.

It is assumed that there are independent motion compensation loops in the low and high resolution layers and that motion estimation and compensation is performed with half-pixel accuracy.

In the scalable encoder, motion estimation may be performed either once or twice. Estimation performed only once is characterized by low complexity, which is an advantage. Two estimations are more complex and require more calculation time. However, it may be expected that they will yield a more precisely predicted frame. The author has conducted experiments to compare the coding efficiency in both cases.

The author has compared six different cases of motion estimation. In the cases denoted A, B, C and D (Fig. 5.8), estimation is performed only once in the high resolution layer and the appropriate motion vectors from the high resolution layer are used in the low resolution layer. One of the four motion vectors from the high resolution layer is used in motion compensation in the low resolution layer. Since the interpolated base layer frame corresponds to a wider area in the high resolution frame, the motion vectors components MVx and MVy from the high resolution layer must be divided by 2.

Two additional cases were analysed. In the first case, two independent compensations were performed for estimated motion vectors. In the second case, two estimations were performed but motion vectors in the low resolution layer were further reset to zero.

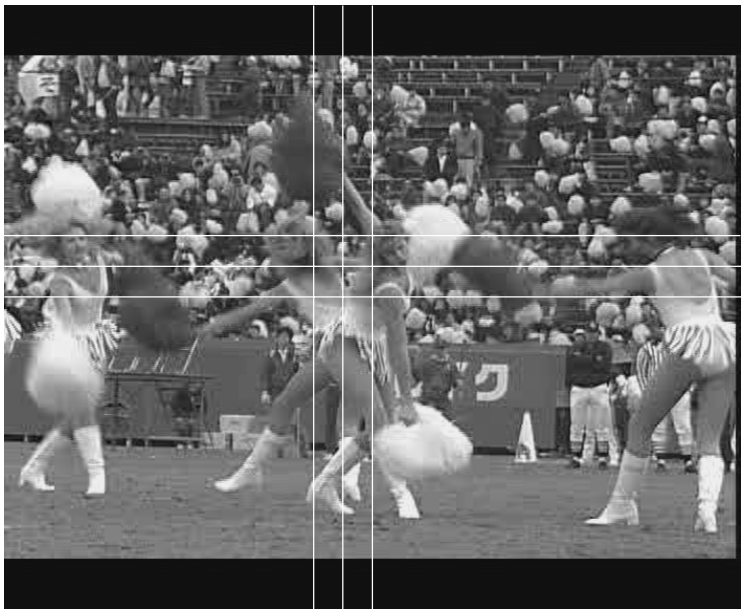
In order to evaluate the coding efficiency, a verification model of the scalable encoder and the implementation of an MPEG-2 encoder have been used (see Annex B).

The experiments fulfilled the following conditions:

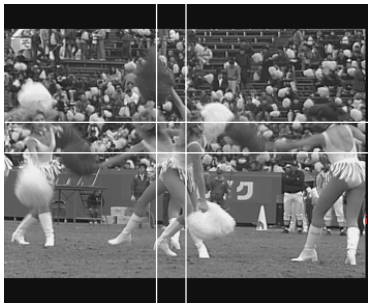
- motion estimation with half-pixel accuracy,
- full search of motion vectors in the range [-31,32],
- independent control of the bitrate in the base and enahancement layers,
- the GOP structure of the enhancement layer:
  I–BE–BR–BE–P–BE–BR–BE– P–BE–BR–BE,
- 12 frames in a GOP.

The scalable encoder uses the control algorithm presented in Section 4.3.7.

a)



b)



c)

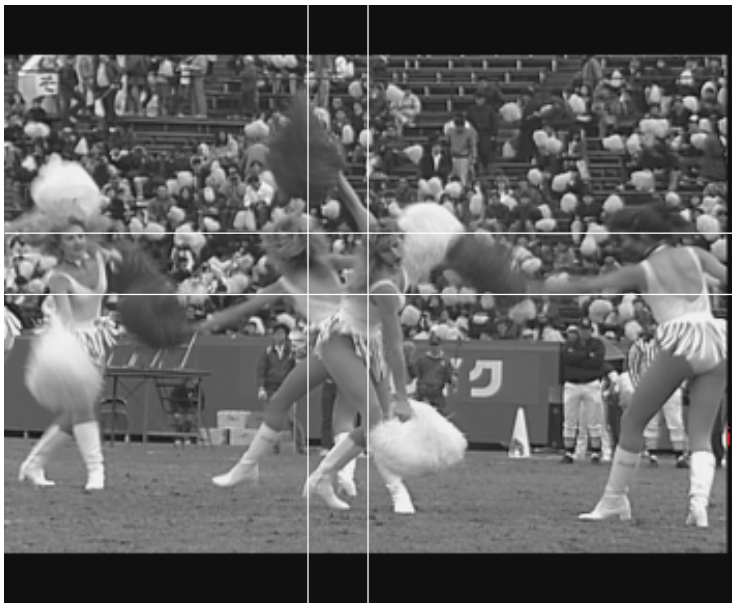


Fig.5.9. (a) The four macroblocks in a high resolution frame, (b) a macroblock in a low resolution frame, (c) a macroblock in an interpolated low resolution frame which corresponds to four macroblocks in a high resolution frame.

Table 5.1. Comparison of different motion estimation cases.

	Base layer		Enhancement layer	
	PSNR [dB]	Bitstream [Mbps]	PSNR [dB]	Bitstream [Mbps]
Case A	32.21	3.71	30.08	5.26
Case B	32.22	3.70	30.07	5.26
Case C	32.20	3.73	30.07	5.27
Case D	32.20	3.73	30.07	5.27
Two independent motion estimation and compensation processes	32.62	3.60	30.15	5.10
The base layer motion vectors reset to zero	31.78	3.97	30.06	5.42

The results of this experiment (Table 5.1) show that the highest efficiency is achieved for two independent motion estimation and compensation processes. The efficiency is lower for the cases A, B, C and D. Fig. 5.10(a) presents base layer motion vectors and Fig. 5.10(b) shows a picture with enhancement layer motion vectors. A conclusion may be drawn that motion vectors of the high resolution layer are different from the motion vectors of the estimation in the low resolution layer. Histograms of rounded differences between the base and enhancement layer motion vectors show that the motion vectors are minimally different (see Fig. 5.11). Therefore it is possible to transmit the motion vectors in the base layer and the difference between the enhancement layer motion vectors and base layer vectors in the enhancement layer.

The efficiency of the scalable encoder in which motion vectors in the low resolution layer were reset to zero is the worst.

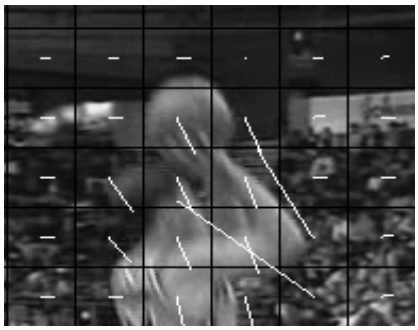
Two independent motion estimation and compensation processes yield the best results because due to the estimation and compensation in the low resolution layer, there is coarse motion compensation for slowly moving objects in the high resolution layer. The second motion estimation and compensation give more precise prediction.

Since every second frame is skipped in the low resolution encoder, motion estimation and compensation processes in the low and high resolution layers are performed for the frames from different time moments.

This explains why one estimation and compensation process is not adequate in scalable encoding.

The problem of optimal motion vectors encoding was put aside in this dissertation. The author hopes to address this problem in his further research.

a)



b)

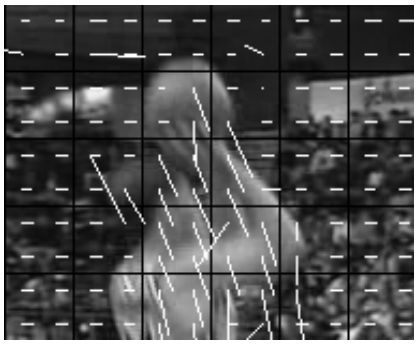


Fig. 5.10. a) Base layer motion vectors. b) Enhancement layer motion vectors.

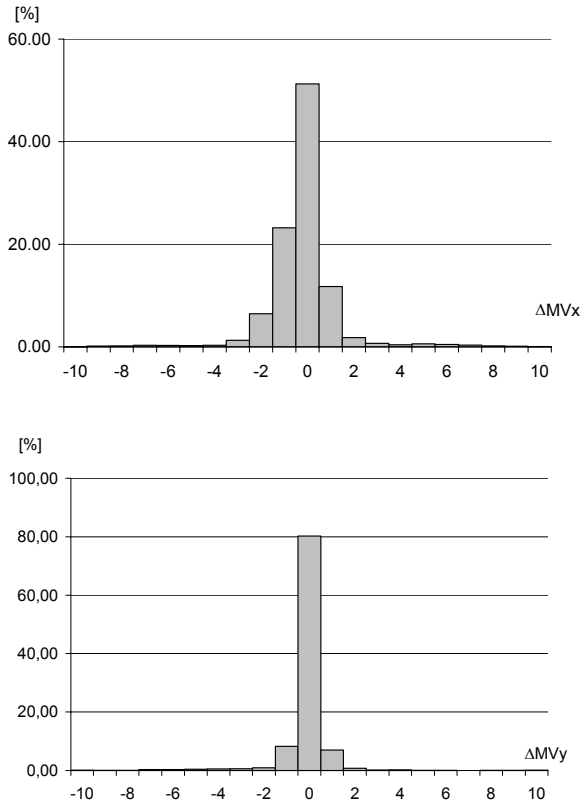


Fig. 5.11. Histograms of rounded differences between the base and enhancement layer motion vectors. The two plots correspond to two vector components.

5.4.3 Complexity

Let us assume that the high resolution encoder produces a bitstream which represents a 704 x 576 50Hz progressive test sequence and that the base layer encoder produces a bitstream which represents a 352 x 288 25Hz progressive test sequence. Therefore there are 1584 macroblocks in the high resolution layer and 396 macroblocks in the low resolution layer. In the enhancement layer the motion is estimated for high resolution images and full-frame motion compensation is performed. It means that there is a second, more precise motion compensation. Therefore the number of motion vectors mv_e estimated in the high resolution encoder is about four times bigger than the number of motion vectors mv_b estimated in the low resolution encoder. In fact, some enhancement macroblocks are interpolated from the low resolution layer and the respective motion vectors mv_e need not be sent in the high resolution layer. Therefore the total number of motion vectors is less than 25% bigger than the respective number of motion vectors of non-scalable bitstream.

Table 5.2. Comparison of motion estimation complexity of scalable and non-scalable encoders (This calculation is performed for full search estimation with half-pixel accuracy and 704 x 576, 50Hz input progressive test sequences, GOP=12, the search range D=±31).

		The number of operations for one macroblock	The number of operations per second
Non-scalable encoder (704 x 576 50Hz)		$(2D+1)^2 N_1 N_2$	128.755 x 10^9
Scalable encoder	Base layer (352 x 288 25Hz)	$(2D+1)^2 N_1 N_2$	12.875 x 10^9
	Enhancement layer (704 x 576 50Hz)	$(2D+1)^2 N_1 N_2$	128.755 x 10^9
	Overall		141.630 x 10^9

Complexity estimation constitutes another problem. Table 5.2 presents the comparison of motion estimation complexity of scalable and non-scalable encoders. This

comparison was performed under the following conditions: the calculation was performed for half-pixel accuracy. In the low and high resolution layers the full search estimation was performed. The GOP length was 12. It means that the encoded progressive 50 Hz sequence in 1 second consisted of 4 I-, 8 P- and 36 B-frames in the enhancement layer and 4 I-, 8 P- and 12 B-frames in the base layer. In bi-directional prediction two motion estimations were performed. Therefore, there were 32 motion estimations in the base layer and 80 motion estimations in the enhancement layer. The non-scalable and scalable encoders had the same GOP structure.

The scalable encoder with double independent estimation requires 12.875 x 10^9 calculations per second more than the non-scalable encoder.

5.4.4 Conclusions

Several experiments performed by the author prove that two independent motion estimation and compensation processes result in more precise prediction in the low and high resolution layers. This improvement in quality is exploited in the predicition of the next pictures. The experimental results in Table 5.1 show that one motion compensation process is insufficient for scalable coding.

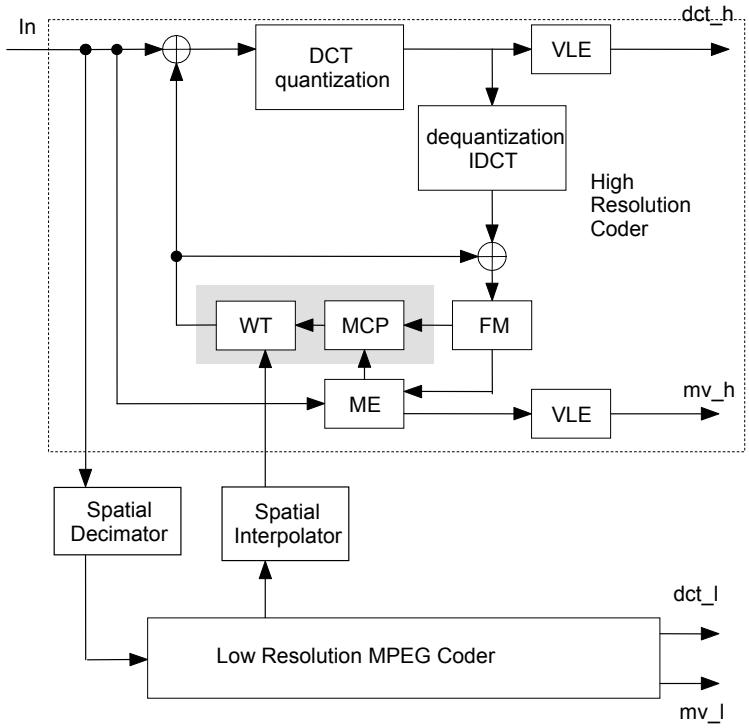
After prediction with motion compensation in the low resolution layer, the interpolated image is also used in prediction in the high resolution layer. In particular, in the enhancement layer motion is estimated for high resolution images and full-frame motion compensation is performed.

5.5. Spatial interpolation and temporal prediction of B-frames

5.5.1 Introduction

The MPEG-2 spatial scalable encoder (Fig. 5.12) uses the spatio-temporal weighted prediction [Hask96]. In spatio-temporal weighted prediction, a spatially

interpolated, locally decoded low resolution layer image is adaptively combined with motion-compensated prediction, using the predefined weights. The weight selection criterion is to minimize the prediction error. For every macroblock for which a certain weight was selected, a spatio-temporal weighted prediction is created. This prediction is used to generate a prediction error. The prediction process is presented in Fig. 5.13. The information about weights is transmitted in the bitstream for the decoder to use.



DCT	Discrete cosine transform
IDCT	Inverse cosine transform
FM	Frame memory
MCP	Motion-compensated predictor
ME	Motion estimator
WT	Weighter

Fig. 5.12. Structure of MPEG-2 encoder with spatial scalability.

Spatio-temporal prediction is obtained in two steps. In the first step, the temporal prediction with motion compensation for macroblock is performed in the enhancement layer. For every macroblock in the enhancement layer, the corresponding macroblock from the base layer is spatially up-sampled to 16 x 16 pixels. In the second step, the up-

sampled macroblock is combined with the temporally predicted macroblock (obtained in the motion compensation process in the enhancement layer) using selected weights *w =0, ½* or *1*, which ensure minimizing the prediction error.

In 2000, Łuczak proposed improved prediction of B-frames [Doma00, Doma00b, Doma00c]. In this proposal, prediction is obtained only in one step. The best temporal prediction/interpolation is chosen from the reference frame and the average of two or three reference frames, according to the criterion of smallest prediction error. This prediction is used in the proposed scalable encoder.

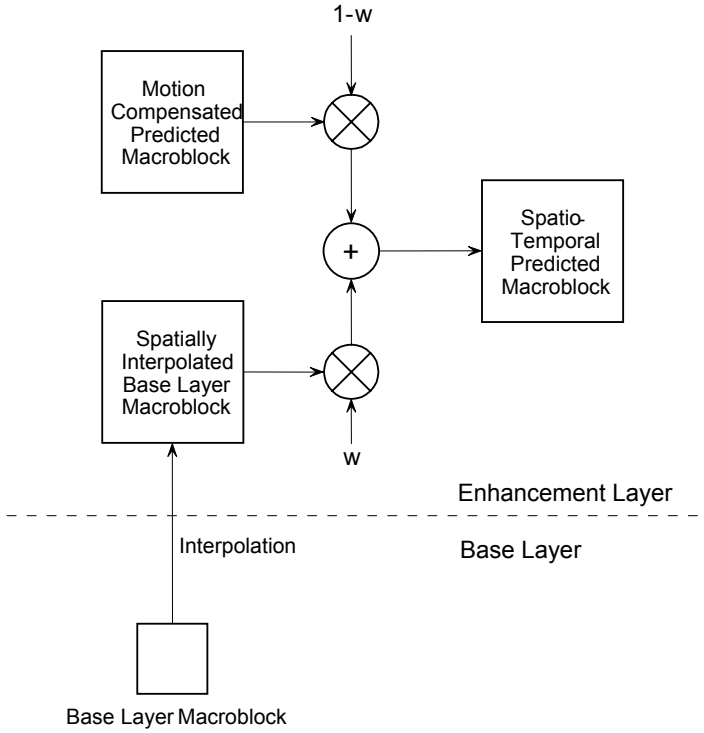


Fig. 5.13. Spatio-temporal weighted prediction in spatial scalability in MPEG-2 standard, w=0, 1/2 or 1 (Based on [Hask96]).

5.5.2 Improved prediction of BR-frames

Improved prediction is also proposed for BR-frames, which are B-frames represented in both layers (Fig. 5.14). Each macroblock in a high resolution BR-frame can be predicted from the following reference frames (Fig. 5.14):

- previous reference frame RP (I- or P-frame),
- next reference frame RN (I- or P-frame),
- current reference frame RC (BR-frame).

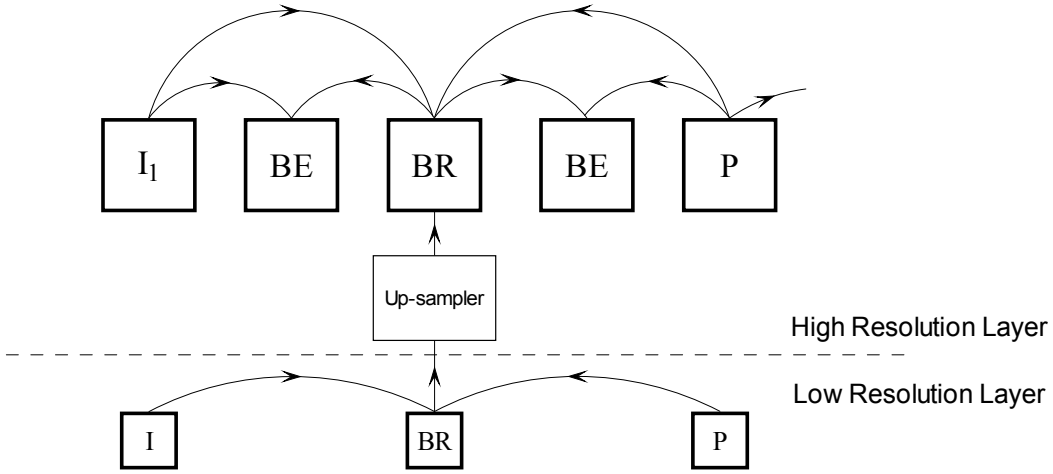


Fig. 5.14. Partitioning of a stream of B-frames.

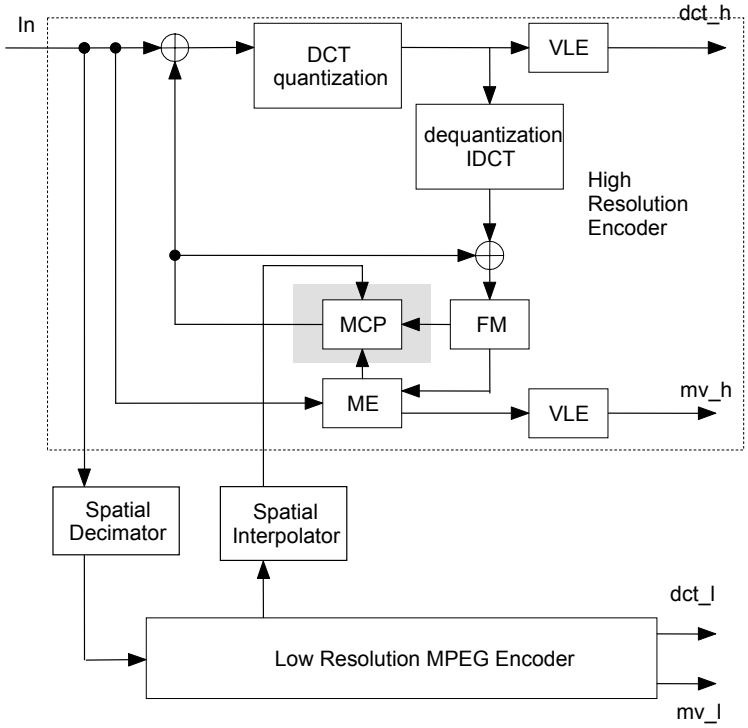
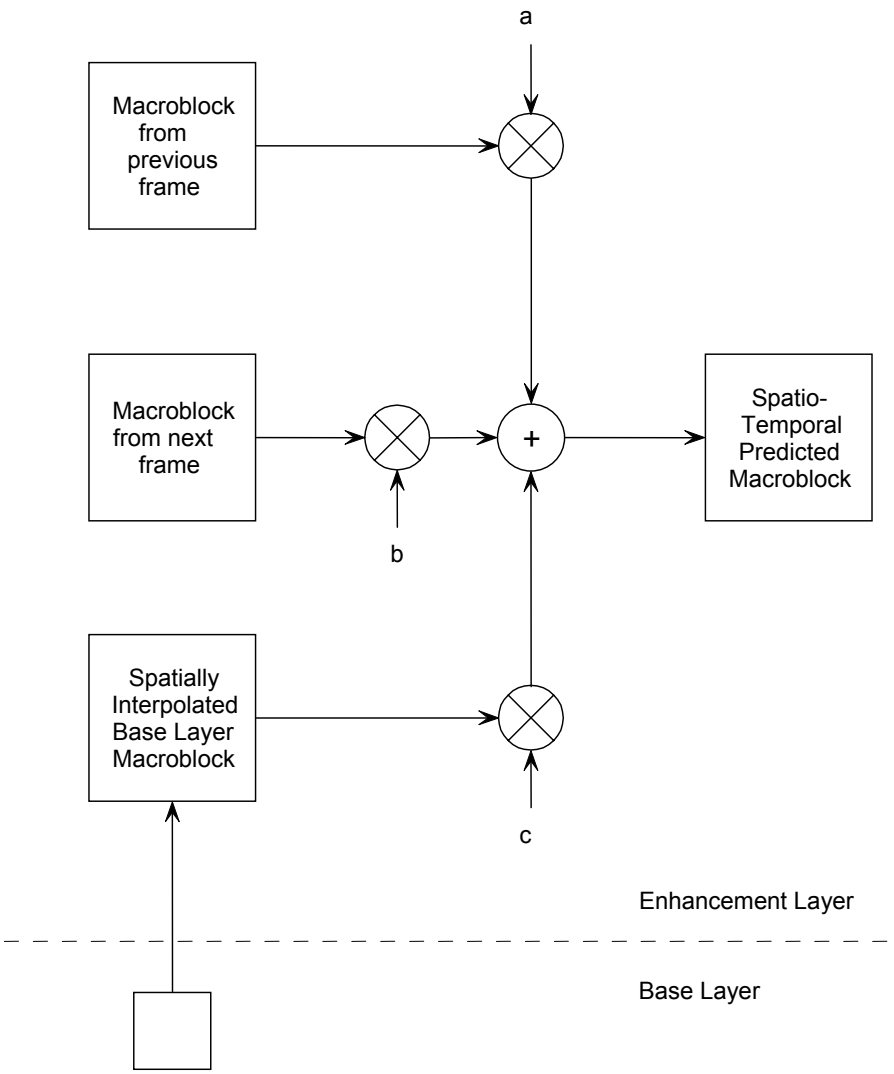


Fig.5.15. The encoder structure with improvements (The gray field indicates the difference between standard and improved prediction blocks).

The improvement of the standard MPEG-2 prediction within a single layer (Fig. 5.12) consists in a different decision strategy. In the improvement, the best prediction/interpolation is chosen from all three possible reference frames: previous, future and interpolated frames (Fig. 5.14). The data from the previous and next reference frames RP and RN are motion-compensated, and data from the current reference frame are up-sampled in the 2-D space domain. The best-suited reference frame or average of two or three reference frames is chosen according to the criterion of smallest prediction error. The structure of MPEG-2 encoder with improved prediction is presented in Fig. 5.15. This kind of prediction requires transmitting an additional bit per macroblock to identify the selected mode of prediction. This problem is presented in Section 5.6.2.



The weights:

a	b	c	
1	0	0	forward prediction from the previous reference frame
0	0	1	interpolation from the current base layer frame
0	1	0	backward prediction from the next reference frame
1/2	1/2	0	the average of the forward prediction and interpolation
1/2	0	1/2	bi-directional prediction from the previous and next reference frames
0	1/2	1/2	the average of the backward prediction and interpolation
1/3	1/3	1/3	the average of the bi-directional prediction and interpolation

Fig. 5.16. Proposed spatio-temporal weighted prediction in spatial scalability

Let m_1 and m_2 define the position of a macroblock, where m_1 =1…K, K is the number of macroblocks in the horizontal direction, and m_2 =1…L, where L is the number of macroblocks in the vertical direction, i is the frame number.

The following notation is introduced:

$f_e^i(m_1,m_2)$ - original value of macroblock (m_1,m_2) in i^{th} high resolution frame (the enhancement layer)

$f_b^i(m_1,m_2)$ - original value of macroblock (m_1,m_2) in i^{th} low resolution frame (the base layer).

$\hat{f}_e^i(m_1,m_2)$ - predicted macroblock (m_1,m_2) in i^{th} high resolution frame (the enhancement layer).

$\hat{f}_b^i(m_1,m_2)$ - macroblock (m_1,m_2) obtained from spatial interpolation of low resolution macroblock from i^{th} frame.

$g_e^i(m_1,m_2)$ - weighted predicted macroblock.

On the macroblock basis, in improved prediction there are seven possibilities to choose from:

1) forward prediction from the previous reference frame RP (I- or P-frame),

$$g_{e_1}^i(m_1,m_2)=\hat{f}_e^{i-1}(m_1,m_2), \tag{5.1}$$

2) backward prediction from the next reference frame RN (I- or P-frame),

$$g_{e_2}^i(m_1,m_2)=\hat{f}_e^{i+1}(m_1,m_2), \tag{5.2}$$

3) bi-directional prediction from the previous and next reference frame,

$$g_{e_3}^i(m_1,m_2)=\frac{\left(\hat{f}_e^{i+1}(m_1,m_2)+\hat{f}_e^{i-1}(m_1,m_2)\right)}{2}, \tag{5.3}$$

4) interpolation from the current reference frame RC (spatially up-sampled frame from the base layer),

$$g_{e_4}^i(m_1,m_2)=\hat{f}_b^i(m_1,m_2), \tag{5.4}$$

5) the average of the forward prediction from the previous reference frame RP and of the interpolation from the current reference frame RC,

$$g_{e_5}^i(m_1,m_2)=\frac{\left(\hat{f}_e^{i-1}(m_1,m_2)+\hat{f}_b^i(m_1,m_2)\right)}{2}, \tag{5.5}$$

6) the average of the backward prediction from the next reference frame RN and of the interpolation from the current reference frame RC,

$$g_{e_6}^i(m_1,m_2)=\frac{\left(\hat{f}_e^{i+1}(m_1,m_2)+\hat{f}_b^i(m_1,m_2)\right)}{2}, \tag{5.6}$$

7) the average of the forward prediction from the previous frame, backward prediction from the next frame and interpolation from the current reference frame RC,

$$g_{e_7}^i(m_1,m_2)=\frac{\left(\hat{f}_e^{i-1}(m_1,m_2)+\hat{f}_e^{i+1}(m_1,m_2)+\hat{f}_b^i(m_1,m_2)\right)}{3}, \tag{5.7}$$

Assuming mean squared error, the overall expected distortion of macroblock (m_1,m_2) in the enhancement layer is given by:

$$d_e^i(m_1,m_2)=\min_j\left(\sum_{\substack{x=0\\y=0}}^{\substack{15\\15}}\left(f_e^i(m_1+x,m_2+y)-g_{e_j}^i(m_1+x,m_2+y)\right)^2\right), \tag{5.8}$$

where j denotes the selected mode of prediction, j=1 … 7, x is the pixel index in the horizontal direction and y is the pixel index in the vertical direction.

The overall expected distortion for BR-frames is defined by:

$$d_e^i=\sum_{\substack{m_1=1\\m_2=1}}^{\substack{L\\K}} d_e^i(m_1,m_2), \tag{5.9}$$

In the interframe mode, the BR-frames prediction error e_e^i is encoded

$$e_e^i = f_e^i(m_1, m_2) - g_{e_j}^i(m_1, m_2). \tag{5.10}$$

5.5.3 Temporal prediction of BE-frames

Another improvement of standard MPEG-2 prediction consists in different temporal prediction of BE-frames in the high resolution encoder. The BE-frames are predicted from either both nearest reconstructed pictures or from one of them (Fig. 3.5). Contrary to the MPEG-2 standard, B-frames (BR-frames) are used as reference pictures for the prediction of other B-frames (BE-frames) (Fig. 5.14). Therefore higher correlation between the currently encoded BE-frame and the reference frame is achieved because of decreasing time difference. This modification reduces the number of bits allocated to BE-frames as compared to the standard scalable MPEG encoder.

5.5.4 Results of experiments

In order to evaluate the efficiency of the proposed improved prediction, the verification model of the scalable encoder and animplementation of the MPEG-2 encoder have been used (see Annex B).

The experiments fulfilled the following conditions:

- motion estimation with half-pixel accuracy,
- full search of motion vectors in range [-31,32],
- independent motion estimation and compensation in both layers,
- independent control of the bitrate in the base and enahancement layers,
- the GOP structure of the enhancement layer:
  I–BE–BR–BE–P–BE–BR–BE– P–BE–BR–BE,
- 12 frames in a GOP.

The scalable encoder uses the control algorithm presented in Section 4.3.7.

Experimental results obtained for television test sequences show that the current reference frame, i.e. the low resolution BR-frame, is used in prediction of a significant

number of macroblocks, sometimes even more than 50% of all macroblocks in BR-frames (Fig. 5.18). In particular, the interpolation from low resolution images allow more efficient predictive coding of macroblocks, which would be intra coded otherwise (Fig. 5.17).

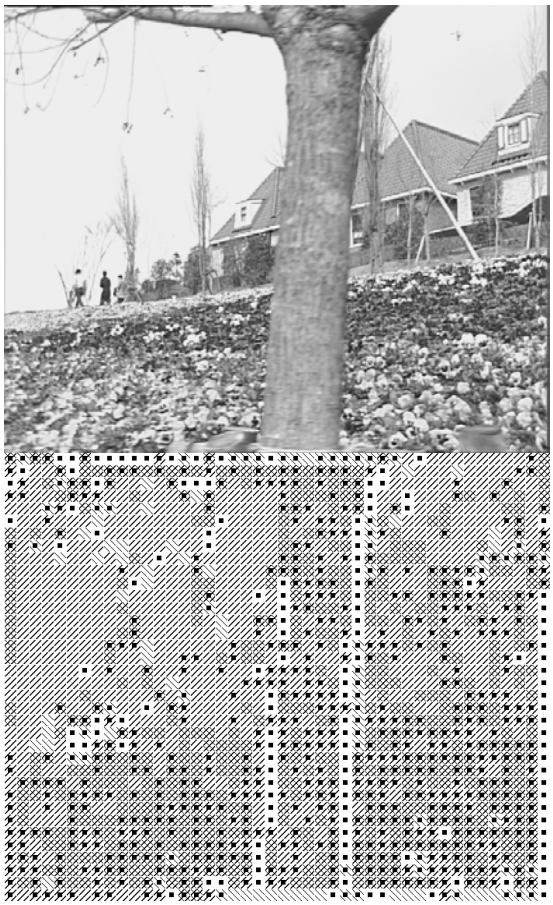


Fig. 5.17. Frame from the test sequence *Flower garden* and various types of macroblocks in the corresponding BR-frame: slash signs denote forward prediction, backslash signs denote backward prediction and square signs denote interpolation.

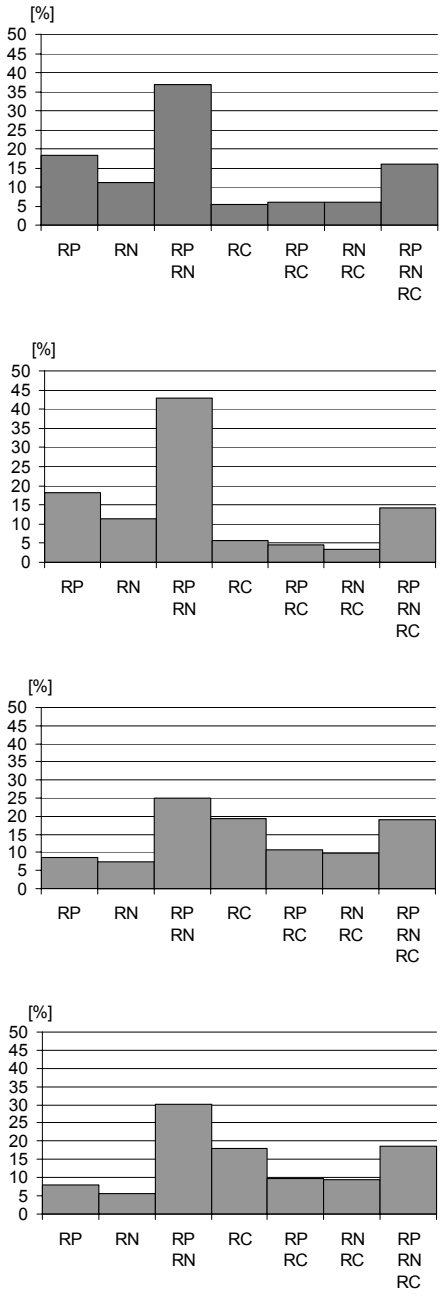


Fig. 5.18. Percentage of macroblocks predicted using individual reference frames or their averages. From top to bottom: test sequence *Flower garden* (15 frames) encoded for 5.24 and 8.76 Mbps and test sequence *Funfair* (15 frames) encoded for 5.03 and 8.48 Mbps.

The application of improved prediction leads to lower bitrates and higher PSNR, compared to standard prediction (Fig. 5.19). This improvement reduces the average size of a BR frame by about 6 - 10%, compared to spatially scalable coding defined in the MPEG-2 standard.

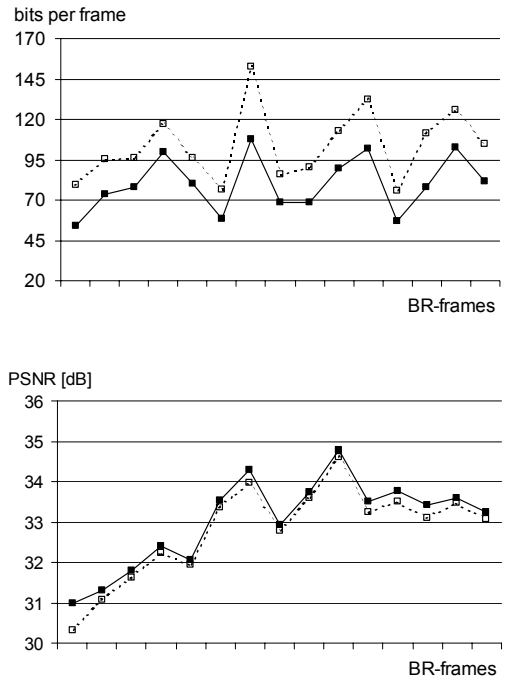


Fig. 5.19. Bitrate and PSNR for improved and standard prediction of BR-frames: solid line – improved prediction, dotted line – standard prediction used in an MPEG-2 scalable encoder.

5.6. Video bitstream syntax

5.6.1 Intraframe mode syntax

This section shows the comparison of the bitstream syntax between the base and enhancement layers for the intraframe coding mode. The analysis has been conducted to show that the scalable bitstream syntax can be more efficient than the non-scalable bitstream in the intraframe mode. The details are presented in Annex A. The analysis has been done for 704 x 576 progressive frames.

In a 704 x 576 frame there are 1584 macroblocks of 16 x 16 pixels. Non-scalable coding requires transmitting 36 bits per macroblock (see Annex A). It means that minimum 57024 bits are required to transmit a frame. In scalable coding, the frame in the base layer has 396 macroblocks of 16 x 16 pixels. Again, 36 bits are needed to transmit a macroblock, so minimum 14256 bits are needed to transmit a base layer frame. The

frame in the enhancement layer has 1584 macroblocks of 16 x 16 pixels. As 22 bits are required to send a macroblock (see Annex A), this is minimum 34848 bits per frame for the enhancement layer.

Therefore, 49104 bits are required to transmit minimum bitstream frame syntax scalable encoded in the intraframe mode. It means that a frame encoded in a scalable structure requires fewer bits to transmit. The numbers of bits required for non-scalable and scalable coding of one frame are presented in Table 5.1.

Further decrease of bitrate is possible. It was assumed that all macroblocks in the enhancement layer are encoded. Since the interframe mode of coding is used in the enhancement layer, not all macroblocks must be encoded and transmitted. The resulting decrease of bitrate allows to improve the quality of the prediction of the next pictures.

Table 5.3. The required number of bits for non-scalable and scalable coding of one frame (BT.601 progressive frame, intraframe mode).

	The number of macroblocks in frame	The number of bits per macroblock	The number of bits per frame
MPEG-2 non-scalable coding	1584	36	57024
Scalable coding:			
Base layer	396	36	14256
Enhancement layer	1582	22	34848
The total amount (base layer + enhancement layer)			49104

5.6.2 Interframe mode syntax

In the enhancement layer, improved prediction of B-frames is used. In the extreme case, this kind of prediction requires transmitting an additional bit per macroblock to identify the selected mode of prediction. To simplify the experiments, an additional bit was inserted in the macroblock header in the syntax of the enhancement layer bitstream. This problem is discussed in Annex A, section A2.

5.7 Multi-layer encoder with data partitioning

The proposed two-layer scalable encoder can be easily extended to form a three- or four-layer encoder. In the data produced by both encoders, there exist some portions of information, called granules, that can be arbitrarily assigned to consecutive layers (Figs. 5.20 and 5.21).

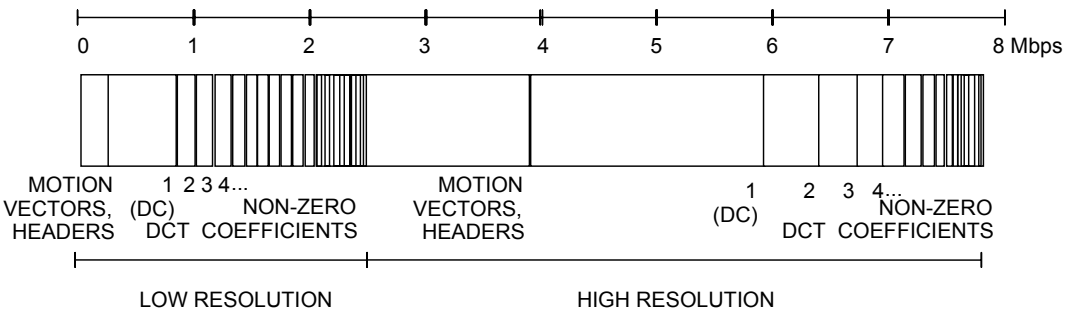


Fig. 5.20. Bitstream granules for the progressive test sequence *Cheer* with total bitrate 7.8 Mbps.

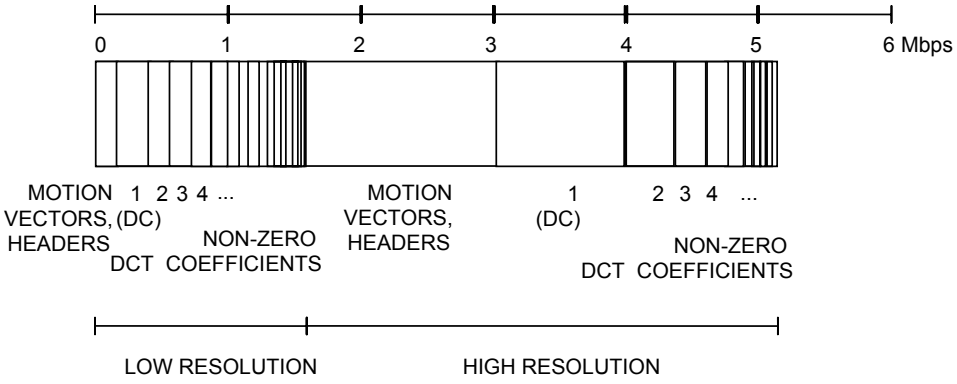


Fig. 5.21. Bitstream granules for the progressive test sequence *Fun fair* with total bitrate 5.2 Mbps.

In the three- and four-layer systemd the low resolution bitstream is not modified. The bistream produced by the high resolution encoder can be divided between some further layers by use of data partitioning.

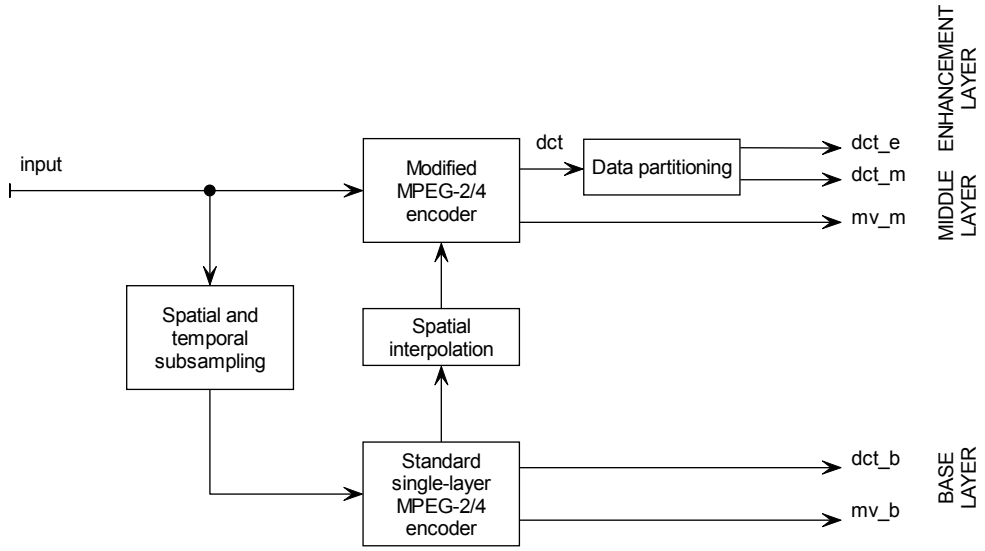


Fig. 5.22. General structure of a three- or four-layer encoder

(*dct_b* and *mv_b* – base layer motion vectors and DCT coefficients, *dct_m* and *mv_m* – middle layer motion vectors and DCT coefficients, *dct_e* – enhancement layer DCT coefficients).

In a three-layer encoder, the high resolution encoder produces a bitstream which is partitioned between the middle layer and the enhancement layer. The middle layer represents the video with full spatial and temporal resolution but with reduced quality (i.e. reduced PSNR). This layer comprises all headers and motion vectors. The enhancement layer comprises all DCT coefficients.

In a four-layer encoder, the second layer bitstream comprises all headers and motion vectors. The third layer consists of the DC coefficients from each 8 x 8 block. The fourth layer comprises the rest of DCT coefficients.

The general structure of a three-/four-layer encoder is presented in Fig. 5.22. The advantage of this solution is its simple structure and high efficiency. The efficiency will be discussed in next sections. The disadvantage is the lack of precise adaptation to network conditions. This problem is solved in Chapter 6.

5.8 Performance tests

Experiments have been conducted to check the efficiency of the proposed scalable encoder with two, three and four layers, and to verify the effectiveness of the proposed partitioning of data between the middle layer and the enhancement layers.

The first series of experiments tested the efficiency of the encoder.

In order to evaluate compression efficiency, a verification model of the scalable encoder and an implementation of the MPEG-2 encoder have been used (see Annex B). The base layer encoder is the standard MPEG 2 encoder that processes video in the SIF format. The enhancement layer is characterized by full television resolution. The bitrate of the base layer encoder is controlled using standard procedures developed for single-layer encoders.

The results are given for five video sequences only. Table 5.4 summarizes the results for 4 – 7 Mbps bitrates. The results for a single layer MPEG 2 Main Profile @ Main Level encoder are given for comparison.

The conditions of the experiments are following:

- motion estimation with half-pixel accuracy,
- full search of motion vectors in the range [-31,32],
- independent motion estimation and compensation in both layers,
- independent control of the bitrate in the base and enahancement layers,
- the GOP structure of the enhancement layer:
  I–BE–BR–BE–P–BE–BR–BE– P–BE–BR–BE,
- 12 frames in a GOP.

The scalable encoder uses the control algorithm presented in Section 4.3.7.

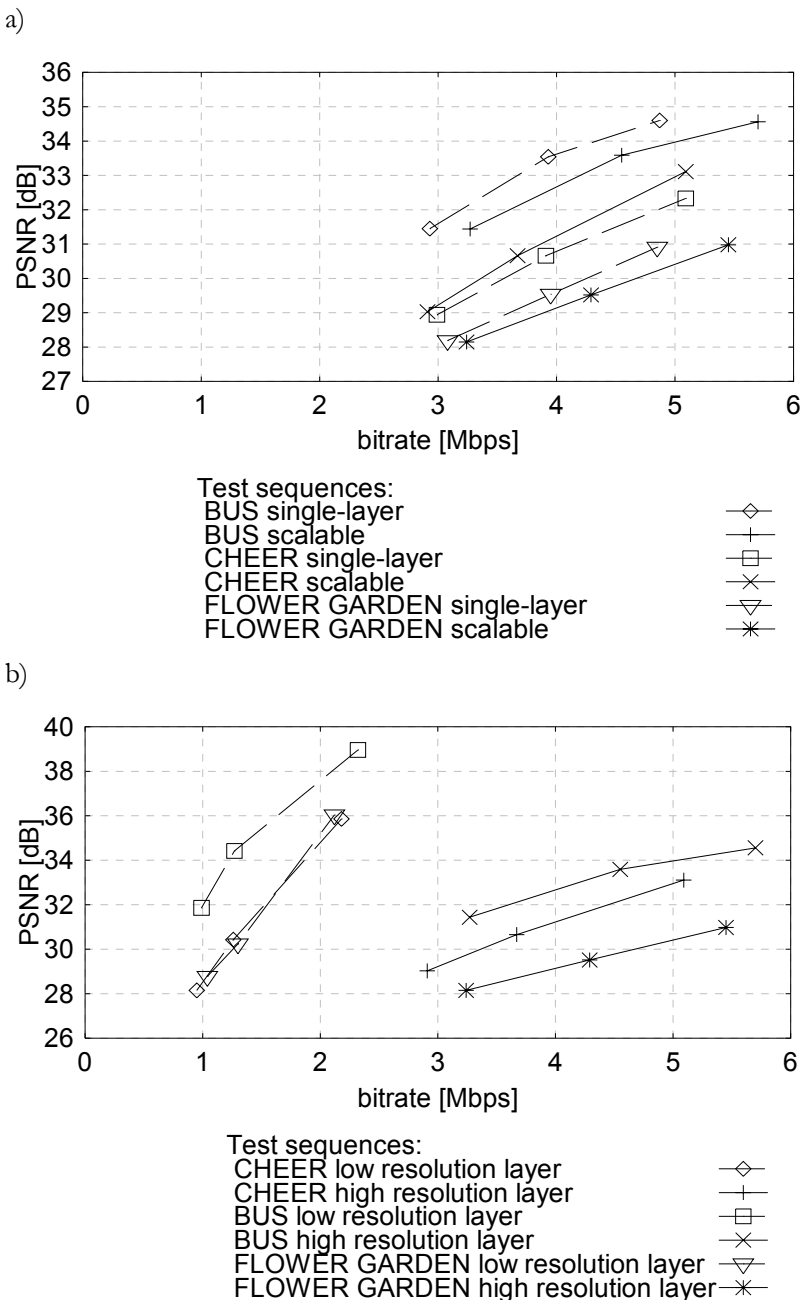
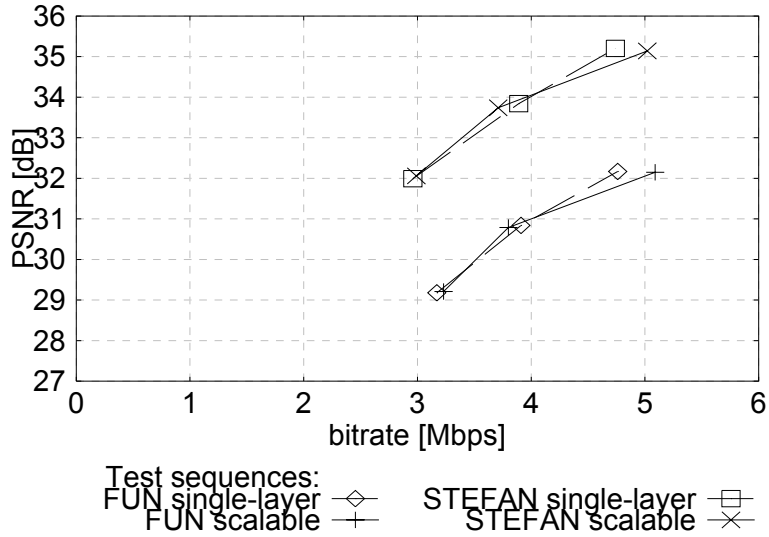


Fig. 5.23. Coding efficiency (luminance PSNR versus bitrate) for a two-loop structure (temporal subsampling factor = 2, image format = 4CIF, 50 fps), compared with the MPEG 2 single-layer encoder. a) scalable encoder and MPEG-2 single layer encoder, total bitrates. b) scalable encoder, low and high resolution layer bitrates.

a)



b)

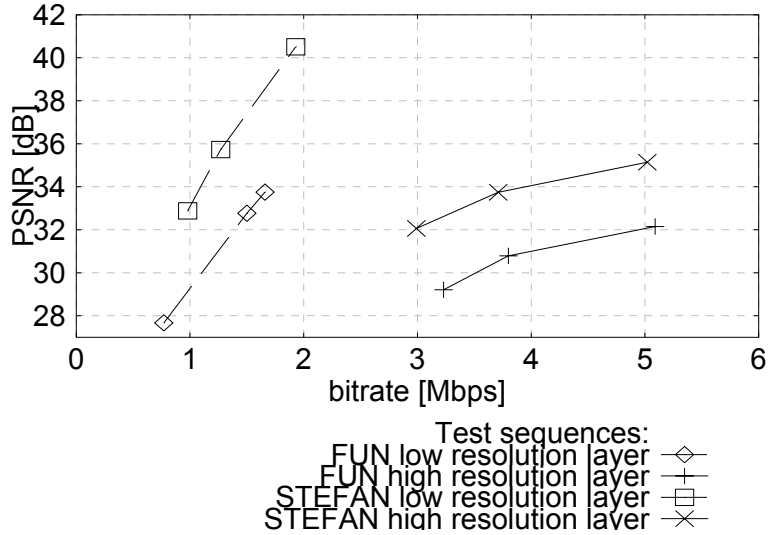


Fig. 5.24. Coding efficiency (luminance PSNR versus bitrate) for a two-loop structure (temporal subsampling factor = 2, image format = 4CIF, 50 fps), compared with the MPEG 2 single-layer encoder. a) scalable encoder and MPEG-2 single layer encoder, total bitrates. b) scalable encoder, low and high resolution layer bitrates.

Table 5.4. Experimental evaluation of the proposed two-layer scalable encoder for BT.601 progressive test sequences.

MPEG-2 - based encoder for 4CIF (704 × 576) sequences		*Cheer*	*Flower Garden*	*FunFair*	*Stefan*	*Bus*
Single-layer encoder (MPEG-2)	Bitstream [Mbps]	2.99	3.08	3.17	2.96	2.93
	Average luminance PSNR [dB]	28.94	28.19	29.18	31.99	31.45
Proposed scalable encoder	Low resolution layer bitstream [Mbps]	0.95	1.04	0.77	0.98	0.99
	Low resolution layer average PSNR [dB] for luminance	28.15	28.78	27.67	32.88	31.86
	Total bitstream [Mbps]	2.91	3.24	3.23	2.99	3.27
	Average PSNR [dB] for luminance recovered from both layers	29.03	28.15	29.21	32.06	31.44
	Bitrate overhead [%]	-2.67	5.19	1.89	1.01	11.60
Single-layer encoder (MPEG-2)	Bitstream [Mbps]	3.91	3.95	3.91	3.89	3.93
	Average luminance PSNR [dB]	30.66	29.54	30.84	33.84	33.54
Proposed scalable encoder	Low resolution layer bitstream [Mbps]	1.26	1.30	1.50	1.27	1.27
	Low resolution layer average PSNR [dB] for luminance	30.43	30.24	32.77	35.73	34.42
	Total bitstream [Mbps]	3.67	4.29	3.80	3.71	4.55
	Average PSNR [dB] for luminance recovered from both layers	30.66	29.52	30.79	33.74	33.59
	Bitrate overhead [%]	-6.14	8.61	-2.81	-4.63	15.78
Single-layer encoder (MPEG-2)	Bitstream [Mbps]	5.09	4.85	4.76	4.74	4.87
	Average luminance PSNR [dB]	32.33	30.91	32.17	35.20	34.60
Proposed scalable encoder	Low resolution layer bitstream [Mbps]	2.18	2.12	1.66	1.93	2.32
	Low resolution layer average PSNR [dB] for luminance	35.86	36.04	33.75	40.51	38.96
	Total bitstream [Mbps]	5.09	5.45	5.09	5.02	5.7
	Average PSNR [dB] for luminance recovered from both layers	33.11	30.98	32.15	35.14	34.56
	Bitrate overhead [%]	0	12.37	6.93	5.46	17.04

The results from Table 5.4 prove high efficiency of the two-layer encoder. With the same bitrate as in the MPEG-2 non-scalable profile, the proposed scalable encoder ensures almost the same quality. Bitrate overhead due to scalability is about 5% - 15%. Figures 5.23 and 5.24 show coding efficiency (luminance PSNR as a function of bitrate) for a two-loop structure.

The goal of the next experiments has been to show that the proposed data partitioning between the middle and the enhancement layers is effective.

Simulations have been carried out for constant quality coding, for two bitrates, i.e. 5 Mbps and 7 Mbps, for non–scalable MPEG–2 coding of SDTV signals. Table 5.5 contains experimental results for a three-layer scalable encoder. In this case, middle layer

comprises all headers, motion vectors and one DCT coefficient per block. The enhancement layer consists of the rest of DCT coefficients necessary to restore full quality video. As a result of such partitioning, the encoder produces three bitstreams with similar bitrates. Experiments show that each bitstream constitutes about 30-40% of the overall bitstream.

Figures 5.25, 5.26, 5.27 and 5.28 show the number of bits and PSNR ratings for particular frames for two GOPs of test sequence. The results for test sequence *FunFair* are presented in Figs. 5.25 and 5.26. Compared with non-scalable MPEG-2 coding at 5 Mbps bitrate, the increase of bitrate is between 9% and 16%. The results for test sequence *Cheer* are presented in Figures 5.27 and 5.28. Compared with non-scalable MPEG-2 coding at 7Mbps bitrate, the increase of bitrate is between 3% and 12%. Bitrate overhead between 3% and 16%, compared with the single layer MPEG-2 bitrate is acceptable for generic applications.

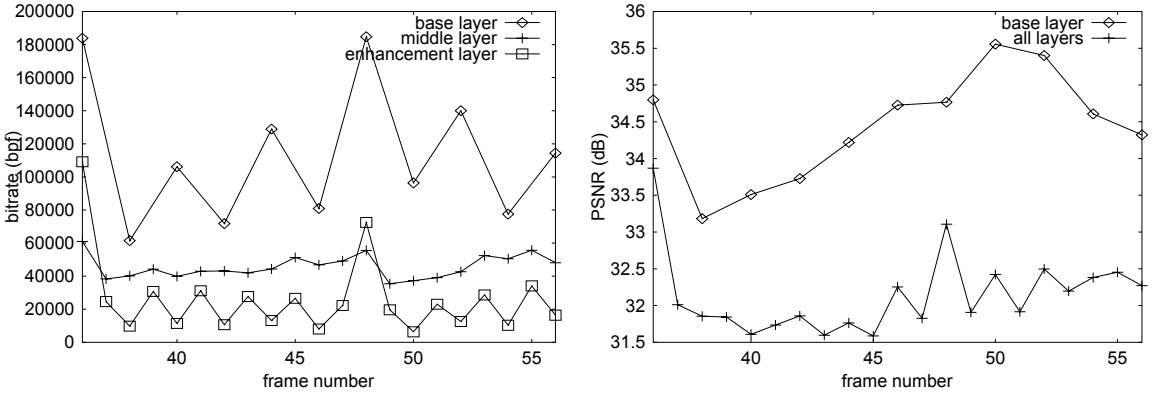


Fig. 5.25. Bits per frame and PSNR for two GOPs of test sequence FunFair, 5.7Mbps.

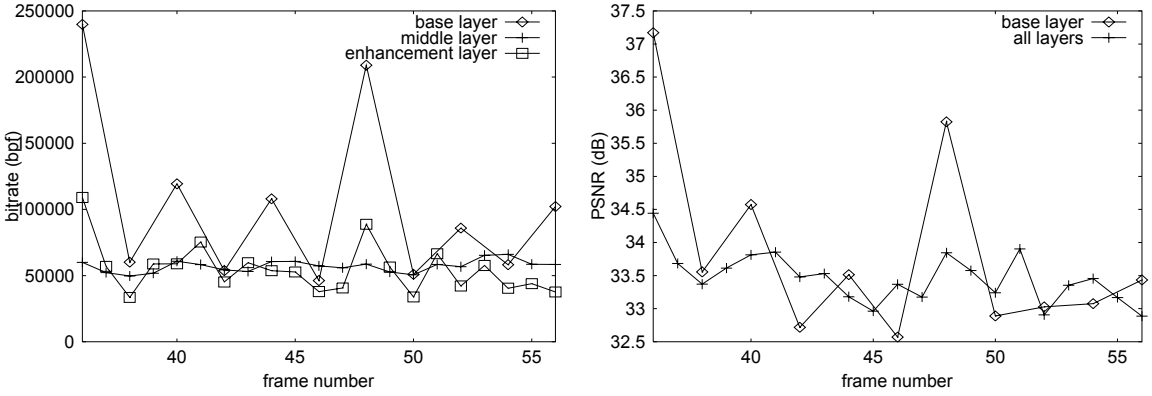


Fig. 5.26. Bits per frame and PSNR for two GOPs of test sequence FunFair, 8.1Mbps.

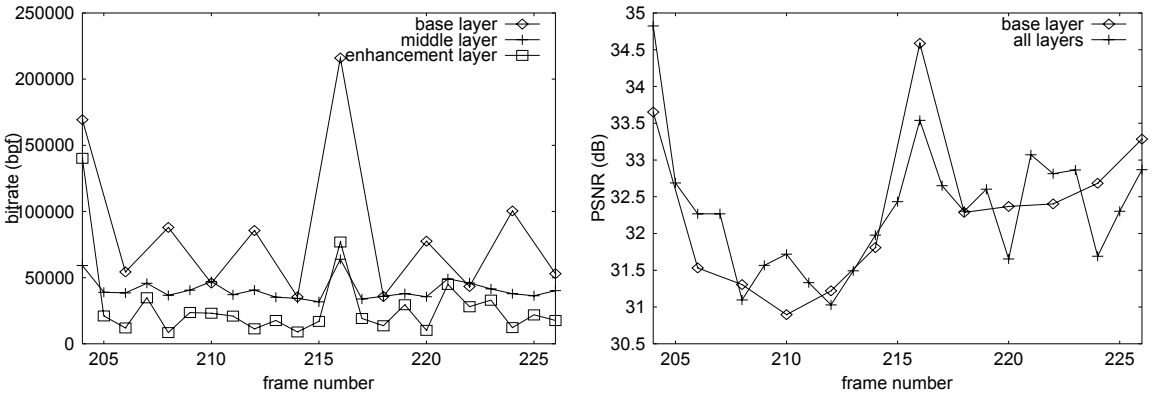


Fig. 5.27. Bits per frame and PSNR for two GOPs of test sequence Cheer, 5.4Mbps.

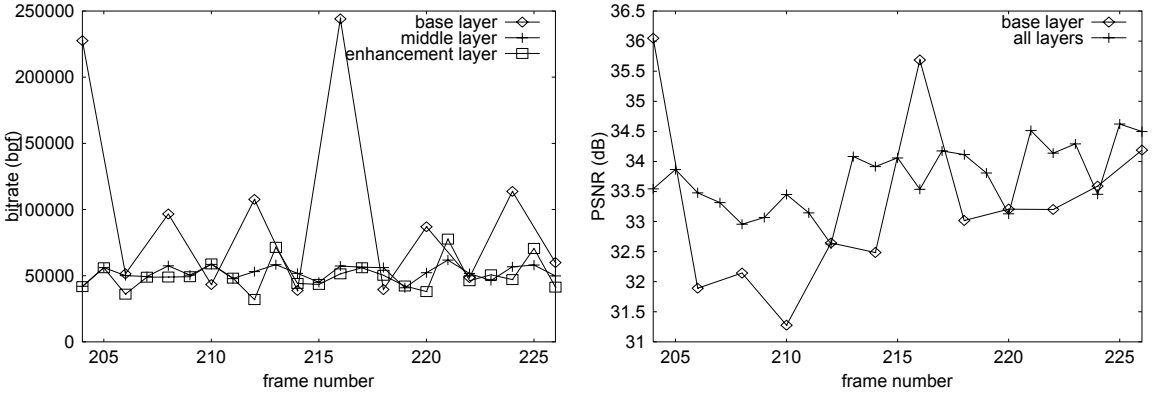


Fig. 5.28. Bits per frame and PSNR for two GOPs of test sequence Cheer, 7.8Mbps.

Table 5.5 Experimental results obtained for the proposed three-layer scalable encoder for BT.601 progressive sequences *Funfair* and *Cheer*.

		Funfair		*Cheer*	
Single layer encoder (MPEG-2)	Bitstream [Mb]	5.2	7.0	5.2	7.0
	Average PSNR [dB] for luminance	32.2	33.8	31.9	33.7
Proposed three-layer encoder	Base layer average PSNR [dB] for luminance	33.1	34.1	31.7	33.1
	Average PSNR [dB] for luminance recovered from all three layers	32.2	33.8	32.0	33.7
	Base layer bitstream [Mb]	2.2	2.5	2.1	2.5
	Middle layer bitstream [Mb]	2.2	2.8	2.0	2.7
	Enhancement layer bitstream [Mb]	1.3	2.8	1.3	2.6
	Total bitstream [Mb]	5.7	8.1	5.4	7.8
	Bitrate overhead [%]	9.6	15.7	3.8	11.1

Table 5.6 contains results for a four-layer system. The whole low resolution bitstream constitutes the base layer. Layer 2 corresponds to the headers and motion vectors of the high resolution bitstream, while Layer 3 comprises DC coefficients of all macroblocks of the high resolution bitstream. All non-zero-frequency DCT coefficients are placed the highest enhancement layer, i.e. Layer 4.

Table 5.6 comprises the values of bitrate and PSNR averaged over respective test sequences. The average bitrate overhead is about 6% for three sequences and over 24% for one sequence. These results are much better than those usually obtained for standard scalable profiles of MPEG-2, and competitive with other recent results [Benz00]. Note that the test sequences are progressive 50Hz ones. Therefore the bitrates are higher than those for standard television formats. Furthermore, the control parameters have been chosen in such a way that the base layer quality is relatively higher than the quality of the full resolution output sequence. Therefore the base layer is about 30-40% of the total bitrate. The bitrates in the other layers are more or less the same. Similar relations hold for other total bitrates selected for the experiments.

Table 5.6. Experimental results obtained for the proposed four-layer scalable encoder for various sequences, compared to non-scalable coding.

Results for 720 x 576, 50 Hz, 4:2:0 progressive test sequences.

		Funfair	*Flower Garden*	*Stefan*	*Cheer*	*Bus*
Single-layer encoder (MPEG-2)	Bitstream [Mbps]	4.88	4.97	4.85	5.21	4.69
	Average luminance PSNR [dB]	32.17	30.91	35.20	32.33	34.60
Proposed scalable encoder	Layer 1 - average luminance PSNR [dB] .	33.75	36.04	40.51	35.86	38.96
	Average PSNR [dB] for the luminance recovered from Layers 1, 2 and 3.	29.82	30.07	33.35	31.44	31.88
	Average PSNR [dB] for the luminance recovered from all layers.	32.15	30.98	35.14	33.11	34.56
	Layer 1 bitstream [Mbps]	1.70	2.17	1.98	2.23	2.38
	Layer 2 bitstream [Mbps]	1.45	1.63	1.27	1.22	1.32
	Layer 3 bitstream [Mbps]	0.89	0.89	0.90	0.86	0.94
	Layer 4 bitstream [Mbps]	1.17	0.59	1.00	0.90	1.20
	Total bitstream [Mbps]	5.21	5.28	5.15	5.21	5.84
	Layer 1 bitrate as per cent of total bitrate [%]	32.6	41.1	38.4	42.8	40.7
	Layer 2 bitrate as per cent of total bitrate [%]	27.8	30.9	24.6	23.4	22.7
	Layer 3 bitrate as per cent of total bitrate [%]	17.1	16.9	17.7	16.5	16.1
	Layer 4 bitrate as per cent of total bitrate [%]	22.5	11.1	19.3	17.3	20.5
	Bitrate overhead [%]	6.7	6.2	6.2	0	24.5

The drawback of partitioning data between layers is related to the accumulation of drift. The decoder which decodes only a part of bitstream (e.g. base and middle layer bitstream only) does not have the same data in the prediction loop as the encoder. Therefore drift between the encoder and the decoder occurs. The problem of drift is presented in Chapter 6.

The number of layers in an encoder may be increased. This aspect is further discussed in Chapter 6.

5.9 Conclusions

In this chapter the author proposed a two-layer scalable encoder based on spatio-temporal decomposition. The experiments show high efficiency of the proposed encoder. For the same bitrate as in the MPEG-2 non-scalable profile, the proposed encoder ensures almost the same quality. The bitrate overhead due to scalability is 5% - 15%. These results are much better than those usually obtained for standard scalable profiles of MPEG-2. Three- and four-layer scalable encoder were also discussed.

Chapter 6

Fine granularity scalability with motion compensation

6.1. Introduction

Efficient and flexible multilayer scalable video coding is one of the challenging problems of contemporary research on video coding. Multilayer scalable coding is a very important technique. Its applications include video streaming [LiLi00, Radh99, Scha01, Scha01b] through error-prone channels [Ghar01, Stuh99, TanZa99, TanZa99b], especially wireless networks of new generations [Giro96, Ghar97, Ghar99, WuHo01]. The existing MPEG-2 video standard provides two- and three-layer scalability (see Section 2.2.5). Nevertheless, the classic solutions for two- and three-layer scalability are rarely efficient or flexible enough to support streaming applications. The author's four-layer systems have been presented in the previous chapter. These solutions can be extended to a multilayer system.

The MPEG-4 standard [ISO98] provides fine granularity scalability [Li01b, LiLi00] as a tool for precise tuning of a layer bitstream to the channel payload (see Section 2.3). Unfortunately, MPEG-4 FGS encoders exhibit significant bitrate overhead as compared to respective single-layer (non-scalable) encoders [HeYa01].

In this chapter the author intends to show that it is possible to obtain fine granularity scalability in a scheme based on a modified version of the classic MPEG-2 scalable encoder. Fine granularity may be obtained by use of splitting the data produced on any resolution level. For example, in the case of bandwidth decrease, motion vectors and the most significant bitplanes may be received while the other bitplanes are lost. Another option is to transmit the first nonzero DCT coefficients from each block. In this way, the bitstream fed into a decoder may be well matched to the throughput available.

This chapter deals with an original proposal for an encoder with fine granularity scalability based on DCT coefficients partitioning between layers. The author assumes a high level of compatibility of the proposed solution with the MPEG-2 video standard.

The goals are:

- to ensure that the total bitrate of all layers of a scalable bitstream is possibly close to the bitrate obtained for single-layer coding,
- to guarantee flexible bit allocation in a multilayer video encoder.

6.2. Multi-layer encoder

The construction of the multi-layer encoder with fine granularity is based on the author's solution for three- and four-layer systems, presented in Chapter 5. The general structure of the multi-layer encoder is presented in Fig. 6.1. The encoder consists of a low resolution part and a high resolution part. The low resolution encoder produces a base layer bitstream which represents video with reduced spatial and temporal resolution. The high resolution encoder produces a bitstream representing video with full spatial and temporal resolution. Fine granularity is obtained by partitioning transform coefficients between layers. It means that the decoding process exploits only a part of each bitstream. The fine granularity technique is applicable for both bitstreams. For experiments, the author has chosen a simplified structure of the multi-layer encoder, in which only DCT coefficients in the high resolution bitstream are partitioned (Fig. 6.2). The problems of drift and bitstream syntax are addressed in further sections.

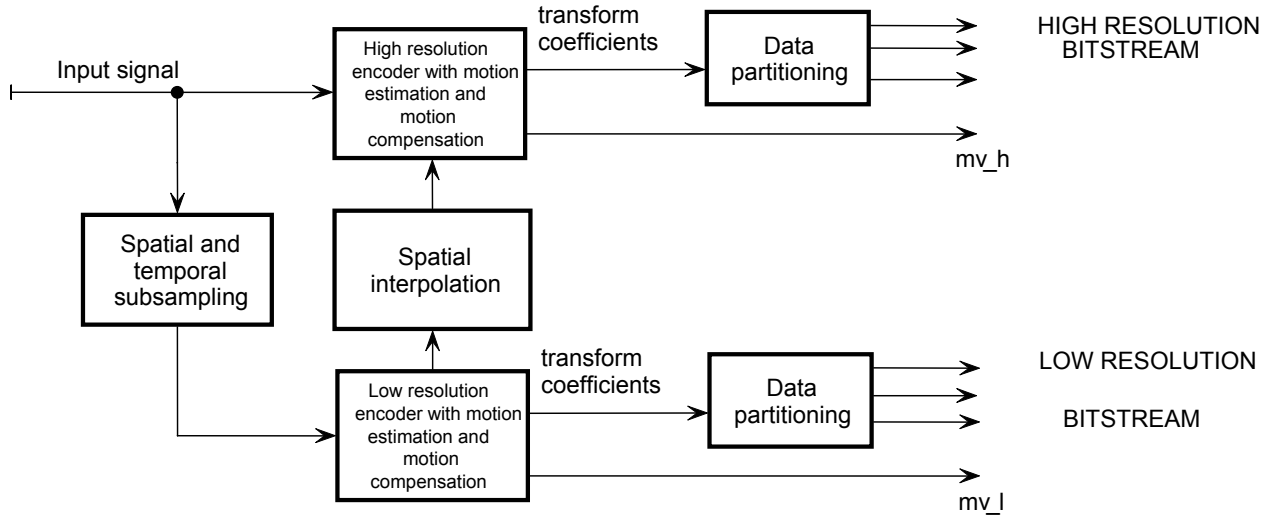


Fig. 6.1. General structure of the multi-layer encoder.

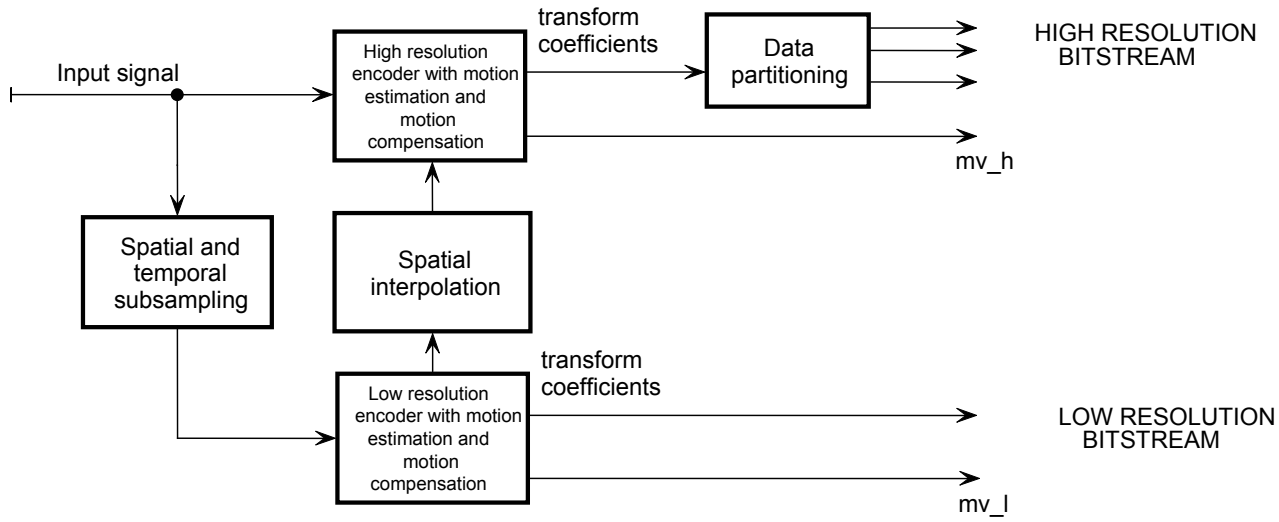


Fig. 6.2. Structure of the multi-layer encoder used in experiments.

The encoder consists of a low resolution encoder and a high resolution encoder (Fig. 6.3). Motion estimation is made independently in both encoders. The low resolution encoder is implemented as a motion-compensated hybrid MPEG-2 encoder. The bitstream produced is described by a fully standard syntax [ISO94, Hask96]. Motion vectors *mv_l* for the low resolution images are estimated independently from those

estimated for a full resolution images. The high resolution encoder is a modification of the MPEG-2 encoder, with added functionalities of spatial and temporal scalability. In particular, motion is estimated for full resolution images and full-frame motion compensation is performed. The reason for using two independent motion compensation loops is discussed in Section 5.2.2.

As mentioned earlier, there are two types of B-frames: BE-frames that exist in the full resolution sequence only and BR-frames that exist in both sequences. In the low resolution encoder, the latter are encoded in a standard way while improved predictive coding is employed in the high resolution encoder, as described in Section 5.2.3. For the prediction of BE-frames, high resolution BR-frames are used as reference in the high resolution encoder.

Further data partitioning is used in order to split into some layers the data produced by the high resolution encoder.

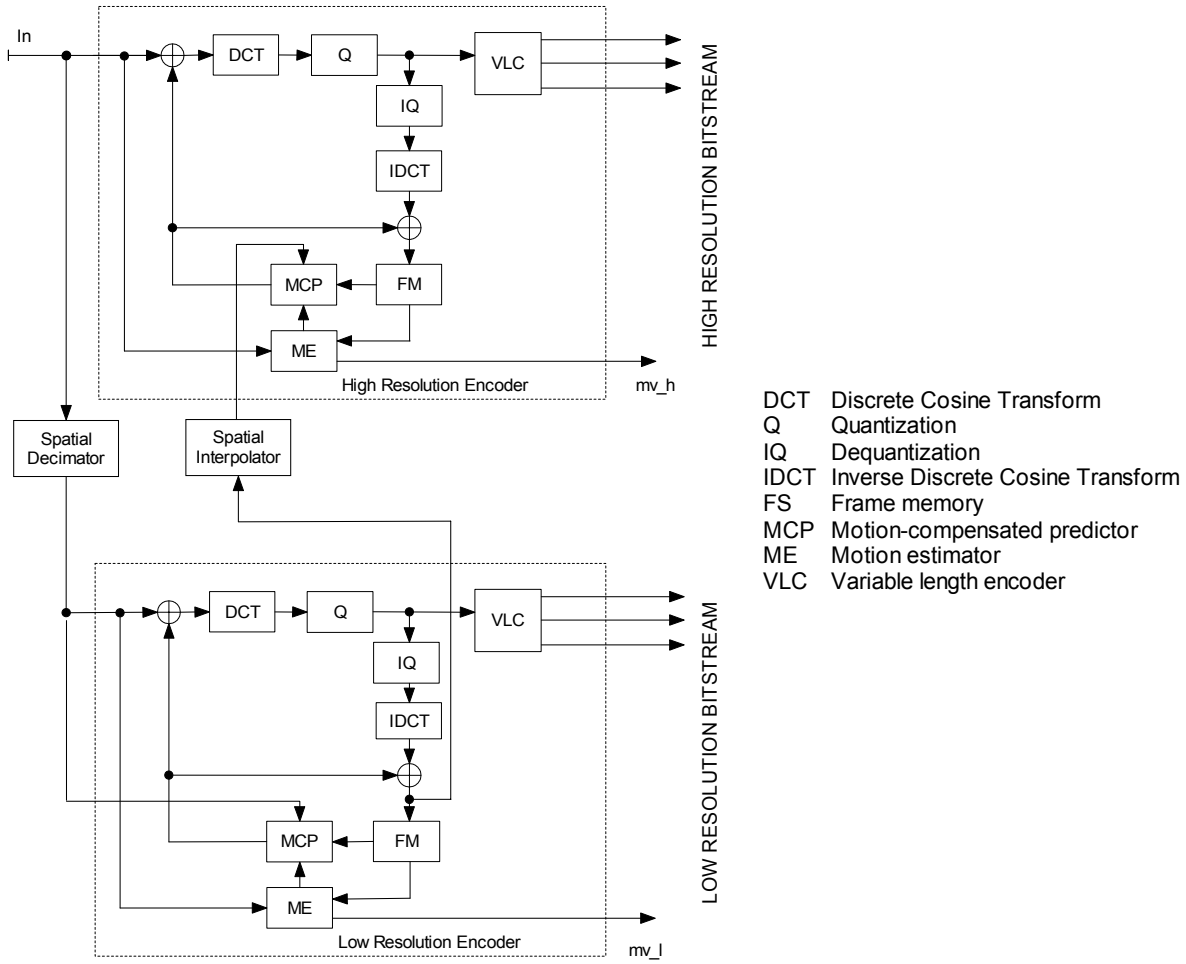


Fig. 6.3. Structure of encoder with fine granularity.

6.3. VLC-based partitioning

In the data produced by both encoders, there exist some granules that can be arbitrarily assigned to successive layers. The base layer bitrate may be reduced in the following way. For the whole low resolution bitstream, this bitrate is about 30 - 40% of the total bitrate. The base layer bitrate can be reduced if only part of non-zero DCT coefficients from the low resolution bitstream are allocated to the base layer, or more precisely, if the base layer comprises only a part of Huffman codes representing (run, level) pairs corresponding to the non-zero DCT coefficients from the low resolution bitstream. The process runs as follows. The encoder allocates headers and low resolution motion vectors *mv_l* to the base layer first. Then the codes of nonzero DCT coefficients are allocated, starting from the DC coefficients from all blocks, followed by the coefficients related to increasing frequencies in the zig-zag order [Iso94, Hask96, Ghan99] from all blocks at once. The process ends when the bit budget is exhausted, thus leaving the remaining coefficients for the enhancement layers.

In this way, the base layer bitrate may be reduced below 15% of the total bitrate of the order of a few megabits per second. A similar strategy may be used in order to create a larger number of layers from both low and high resolution bitstreams. Except from headers and motion vectors, the bitstreams can be arbitrarily split into layers and multi-layer fine granularity can be achieved. All header data and the enhancement motion vectors *mv_h* may be treated as basic granules (Figs. 5.20 and 5.21). The next granules are constituted by DCT coefficients that are encoded as (run, level) pairs, as described in the MPEG-2 standard. The lower layer contains N_m first (run, level) pairs for individual blocks. The control parameter N_m influences bit allocation to layers (see Fig. 6.4). The bitrates in subsequent layers can be controlled individually. To some extent, nevertheless, each additional layer increases the bitrate overhead because at least slice headers should be transmitted in all layers in order to guarantee resynchronization after an uncorrected transmission error. The total bitstream increases by about 3% per each layer obtained using data partitioning. The calculation will be presented in the next section.

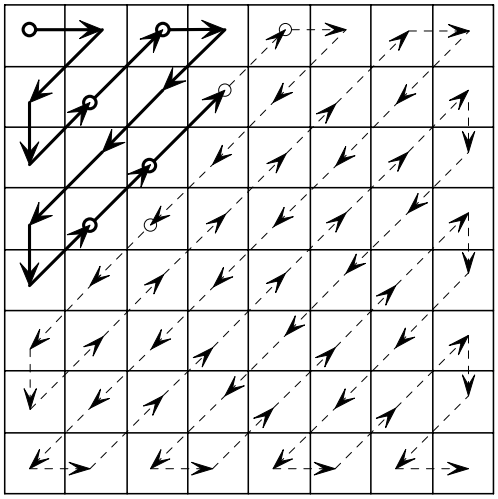


Fig. 6.4. Partitioning of DCT coefficients in a zig-zag order. The lower layer consists of N_m=5 coefficients, the remaining 2 coefficients are transmmited in the second layer (solid line – the lower layer, dash line – the second layer).

The drawback of this strategy is accumulation of drift, since the base layer decoder has no access to full low resolution reconstructions. Drift is also generated by partitioning the high resolution bitstream. Moreover, when the enhancement layer bitstream is corrupted by errors during transmission, the enhancement layer DCT coefficients cannot be properly reconstructed due to the loss of DCT information. This causes drift between the local decoder and remote decoder. It means that the decoding process exploits only the base layer bitstream.

In some applications, drift is not a significant problem [Arav96]. In particular, the MPEG-2-related encoders mostly use relatively short independently coded Groups of Pictures (GOPs), thus preventing drift from significant accumulation. In the author's solution, drift accumulation is also reduced because the total bitstream is divided into two drift-free parts, i.e. the low and the high resolution bitstream. Drift propagates within one part only when fine granularity is applied to a given bitstream. Furthermore, drift in the high resolution part may be reduced by more extensive use of the low resolution images as reference.

There are several other methods of reducing drift [Benz96b, Wern96, Wern96b, Math97]. However, they will not be discussed here.

6.4. Bitstream syntax

In this section, the author shows the syntax of a partitioned multi-layer bitstream. For simplicity, only the two-layer case will be discussed.

The control parameter N_m influences allocation of bits to the base layer and enhancement layer. Let us assume that N_m=3. All the headers (GOP, slice, macroblock) are transmitted in the base layer. For example, if block 1 in a macroblock consists of 5 (run, level) pairs of DCT coefficients, the lower layer contains only the first 3 pairs. The base layer does not contain the indicator EOB (End of Block). The remaining 2 pairs and the EOB symbol are transmitted in the next layer. If the next block consists only of 2 pairs (run, level) then both pairs and the EOB symbol are transmitted in the base layer. In this case, there are no more coefficients in the block. The absence of the EOB symbol indicates that the enhancement layer contains the remaining coefficients. The parameter N_m can be changed from block to block but EOB indicator must be additionally transmitted in the enhancement layer bitstream. If N_m is constant, no extra bits will be added to the bitstream. The situation with the parameter N_m changing from block to block is presented in Fig. 6.5.

Let us now discuss the manner in which the two layers are formed. Let us assume that a block in a non-partitioned bitstream contains 5 (run, level) pairs and N_m=3. The decoder decodes all DCT coefficients from the base layer (i.e. 3 pairs). The remaining DCT coefficients are subsequently transmitted in the enhancement layer. It should be noted that the enhancement layer does not carry any DCT coefficients from the blocks whose last DCT coefficients are included in the base layer or identified as zero blocks by the macroblock header. At the receiver, the two bitstreams are joined together to form the original scalable bitstream.

Slice headers are transmitted in each additional layer, in order to guarantee resynchronization after an uncorrected transmission error.

The slice header consists of 38 bits. The restricted slice structure [ISO94] requires minimum 36 slices per frame. It results in 1368 bits per frame in every additional layer. Therefore the total bitstream increases by about 3% per each layer obtained using data partitioning.

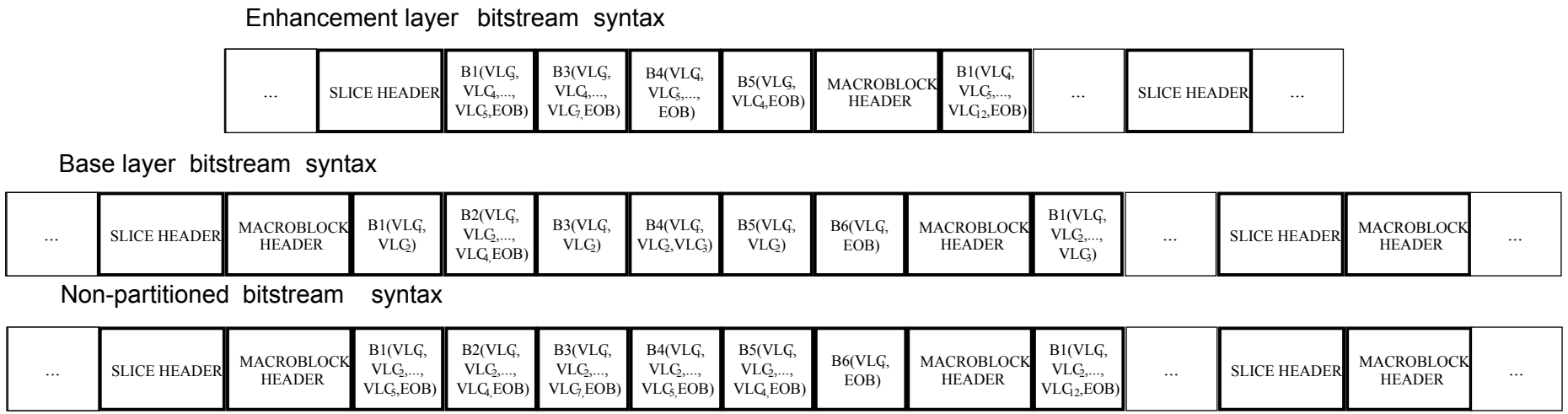


Fig. 6.5. Two-layer partitioning of bitstream (the parameter N_m is changed from block to block).

6.5. Results of experiments

In order to evaluate the proposed partitioning of DCT coefficients, a verification model has been written in the C++ language and is currently available for progressive 4CIF (704 x 576), 50 Hz, 4:2:0 video test sequences (see Annex B). This software also provides an implementation of the MPEG-2 encoder, which has been cross-checked with the MPEG-2 verification model [Mpeg96]. The experiments were conducted to test the efficiency of allocation of DCT coefficients between layers. The experiments fulfilled the following conditions:

- motion estimation with half-pixel accuracy,
- full search of motion vectors in range [-31,32],
- independent motion estimation and compensation in both layers,
- independent control of the bitrate in the base and enahancement layers,
- the GOP structure of the enhancement layer:
  I–BE–BR–BE–P–BE–BR–BE– P–BE–BR–BE,
- 12 frames in a GOP.

The scalable encoder uses the control algorithm presented in Section 4.3.7.

Beside the overall performance of the scalable encoder, the performance (Figs. 6.6 and 6.7) for 5Mbps bitrate has also been measured in order to estimate the efficiency of fine granularity scalability. For the sake of simplicity, only the non-zero coefficient allocation scheme for FGS has been implemented. The application of bitplane coding improves the efficiency of the scalable encoder. A number of non-zero DCT coefficients allocated to a given layer controls smoothly the bitrate of this layer (Figs. 6.6 and 6.7). The respective plots are quite similar for various test sequences.

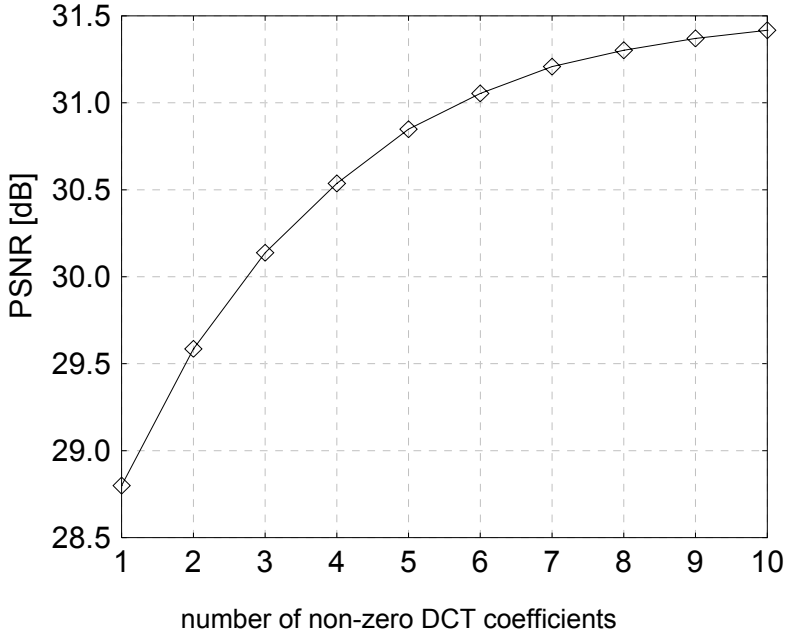


Fig. 6.6. Average values of PSNR [dB] plotted versus the number of non-zero DCT coefficients allocated to the high resolution bitstream. Test sequence *Funfair*, total bitrate 5 Mbps, base layer bitrate about 1.66 Mbps, length of GOP=12.

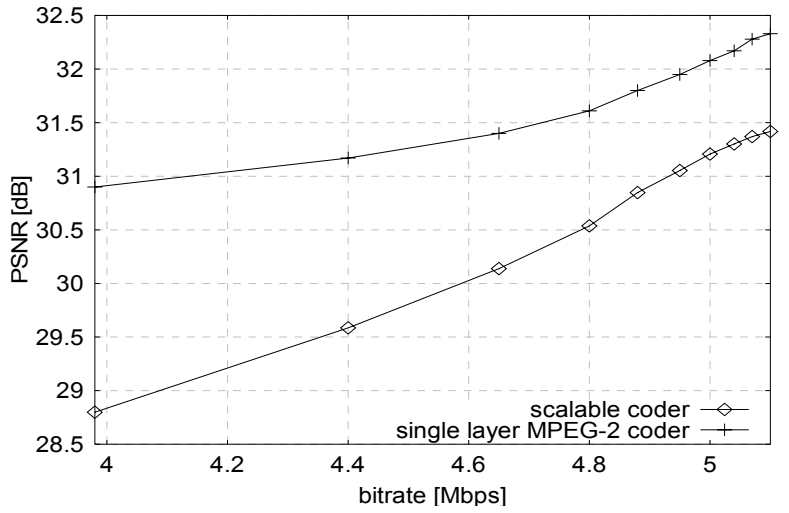


Fig. 6.7. Fine granularity scalability in a two-loop encoder (lower curve) compared to a single layer MPEG-2 encoder (upper curve). Test sequence Funfair, total bitrate 5 Mbps, base layer bitrate about 1.66 Mbps, length of GOP=12.

Apart from suffering from standard coding artifacts, the reconstructed video quality is additionally degraded by drift. The drift artifacts are stronger in the last picture of the Group of Pictures or if the number of allocated coefficients is small. For one or

two allocated coefficients, when no additional measures against drift accumulation are used, the estimated average influence of drift is about 2 dB in the high resolution bitstream. This figure decreases as the number of coefficients grows (Fig. 6.8). The influence of drift in the low resolution layer is less critical as the enhancement layer usually exhibits higher overall quality.

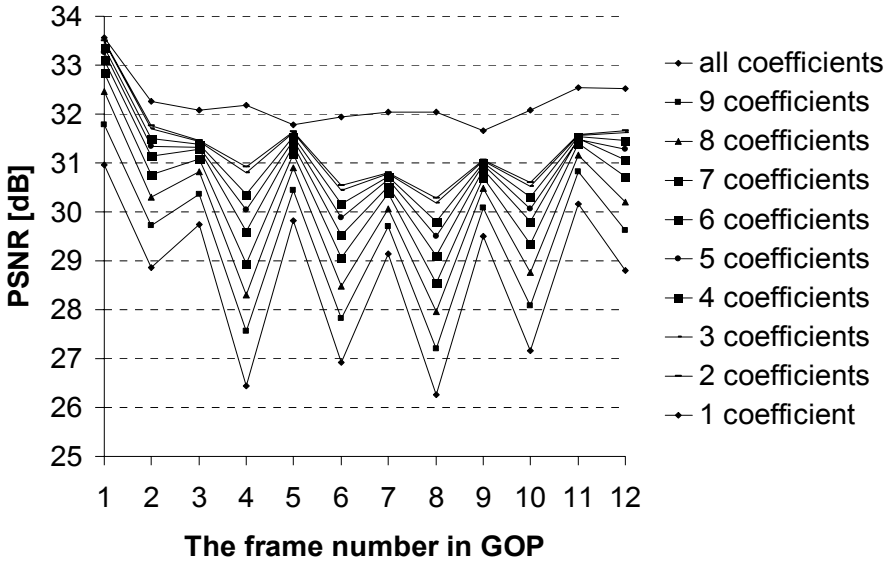


Fig. 6.8. Decrease of signal-to-noise ratio, resulting from drift, for various number of DCT coefficients per block transmitted in the enhancement layer to the decoder.
Test sequence *Funfair* with an average bitrate 5Mbps, length of GOP=12.

6.6. Conclusions

In this chapter the author has proposed and discussed a multi-layer system with fine granularity, based on a slightly modified structure of the classic MPEG-2 scalable encoder. This solution is characterized by flexible bit allocation and low bitstream overhead. The total bitstream increases by about 3% per each layer when data partitioning is used.

In 2001, Gharavi and Alamouti [Ghar01] proposed a similar technique of video partitioning, based on the separation of the variable-length coded DCT coefficients within each block. In the splitting process, the number of bits assigned to each of the two

partitions is adjusted according to the requirements of the unequal error protection scheme employed. Partitioning is applied according to the ITU-T H.263 [ITUT96] coding standard.

Chapter 7

Results and conclusions

7.1. Main results of the dissertation

The author's research has focused on introducing the functionality of scalability in video coding, especially for progressive SDTV signals. This work is expected to have wide prospective applications. Efficient, flexible scalable video coding is one of the challenging problems of contemporary research on video coding. The proposed methods are more efficient than the solutions defined in the MPEG-2 standard.

The main goal of the thesis was to modify the existing video compression algorithms in order to obtain scalable video coding systems with possibly low scalability overhead. The thesis of this dissertation was that it is possible to improve the efficiency of scalable video coding using spatio-temporal scalability. This thesis was proved in several experiments, in which the proposed coding techniques were used.

The main original results of the dissertation are:

- The development of a scalable encoder structure with individual loops with independent motion estimation and compensation (Chapter 5).
- The proposal of an encoder with fine granularity scalability (Chapter 6).

- Detailed experimental examination of the features of the proposed encoders' structures.
- The comparison of the features of different proposed modifications of scalable encoder structures.

Other original achievements of the dissertation are:

- A scalable video encoder with three-dimensional filter banks for spatio-temporal analysis (Section 4.2).
- A 2-D subband scalable system, where some B-frames are allocated to the enhancement layer (Section 4.3).
- The results of experiments of scalable coding with two different types of quantization matrices and modified zig-zag scan tables (Sections 4.3.3 and 4.3.4).
- A modified bitplane technique to encode the ΔLL signal (Section 4.3.5).
- Encoding of the subbands, exploiting inter-subband correlation (Section 4.3.6).
- A simple empirical model for global control of the encoder parameters (Section 4.3.7).
- The implementation of improved predictions of B-frames in the scalable encoder (Section 5.5).
- The analysis of the bitstream for non-scalable and scalable solutions in the interframe mode (Section 5.6.1).
- The modification of bitstream syntax in the case of improved prediction (Section 5.6.2).
- The verification models of proposed scalable encoders (Annex B).

As a result of using spatio-temporal decomposition in the proposed encoder schemes, satisfactory coding results are achieved, which is proved by a thorough comparison with the MPEG-2 standard coding. The presented solutions show that it is possible to achieve an efficient scalable video coding system based on spatio-temporal scalability while preserving an acceptable bitstream overhead.

7.2. Conclusions and future developments

The dissertation provides new solutions for scalable video coding and transmission over a variety of heterogeneous networks. The development of scalable encoding methods and transmission is just starting and a rapid growth can be expected in the coming years. Therefore there will be a growing demand for scalable video encoders and other coding tools.

The author has obtained very promising results for mixed spatio-temporal scalability. The solutions presented in the dissertation achieve higher efficiency than MPEG-2 scalable techniques. For the same bitrate as in MPEG-2 non-scalable encoding, the proposed scalable encoder reaches almost the same quality. Bitrate overhead due to scalability is mostly below 15%. It is important to stress that the elements of the proposed encoders are highly optimized. Therefore, in order to further decrease bitrate overhead, other ways must be sought for.

One approach to increasing coding efficiency and extending scalability features is using metadata during the encoding and decoding steps. Metadata describe audio-visual material. They are used to index its content.

Currently, MPEG is working on standard description of the content of audio-visual material [MPEG01c, IEEE00, IEEE01, Manj02]. SMPTE has recently released the specifications of a standard for metadata [SMPT01a, SMPT01b, SMPT01c]. So far, metadata generated by the indexing process have been used in search and retrieval scenarios. However, some kinds of metadata can be used to improve the efficiency of compression. Note that in the future a very large amount of audio-visual material will be indexed. Therefore, if future encoders have access to such metadata before the encoding process, it may increase efficiency considerably.

Some new metadata-based coding and prediction tools will be integrated into the proposed encoders. Some of the metadata-based coding tools will only influence the encoder, while others will also influence the decoder and therefore some metadata will need to be transmitted. This will result in a trade-off between bitrate overhead and the coding efficiency. The bitstream syntax will provide hooks to attach metadata to the video bitstream. Some of the new coding tools will leave the bitstream syntax unchanged;

others will require modifications of the syntax. Either way, the result will be improved codecs.

An attempt to improve scalable coding efficiency with the use of metadata is realized in the MASCOT project [MASC] as a joint effort of eight research groups in Europe. The project is supported by the European Commision under the FET part of the Information Society Technologies Programme of the Fifth Framework, as project number IST-2000-26467. The goal of the MASCOT project is to develop new video coding schemes and tools that provide increased coding efficiency and extend scalability features compared to the technology available today. This shall be achieved by new metadata-based coding tools, new spatio-temporal decompositions and new prediction schemes [MASC01, MASC02, Smol01].

References

Author's contributions

[Doma98] Domański M., Łuczak A., Maćkowiak S., Świerczyński R., "Hybrid coding of video with spatio-temporal scalability using subband decomposition", Signal Processing IX: Theories and Applications, pp. 53-56, Typorama, 1998.

[Doma98b] Domański M., Maćkowiak S., "Badania kodeka MPEG-2 standardu telewizji cyfrowej", Krajowe Sympozjum Telekomunikacji, pp. 190-198, Bydgoszcz 1998.

[Doma98c] Domański M., Łuczak A., Maćkowiak S., Świerczyński R., "MPEG-Based Scalable Video Codec with 3-D Subband Decomposition", Proceedings of the XXIst National Conference on Circuit Theory and Electronic Networks, pp. 487-492, Poznań, 1998.

[Doma99] Domański M., Łuczak A., Maćkowiak S., Świerczyński R., "MPEG-Based Scalable Video Codec with 3-D Subband Decomposition", w Bulletin of the Polish Academy of Sciences Technical Sciences, Vol. 47, No. 3, pp. 247-253, Warszawa, 1999.

[Doma99b] Domański M., Łuczak A., Maćkowiak S., Świerczyński R., "Hybrid coding of video with spatio-temporal scalability using subband decomposition", Proc. Of SPIE, vol. 3653, Visual Communication and Image Processing, pp. 1018-1025, 1999.

[Doma99c] Domański M., Łuczak A., Maćkowiak S., Świerczyński R., Benzler U., "Spatio-Temporal Scalable Video Codecs with MPEG-Compatible Base Layer", Proceedings of Picture Coding Symposium, pp. 45-48, Portland, 1999.

[Doma99d] Domański M., Łuczak A., Maćkowiak S., "MPEG-2 Compatible Hierarchical Video Coders with B-frame Data Partitioning" Proceedings IWSSIP'99 6th. Int. Workshop on Systems, Signals and Image Processing, pp. 34-37, Bratislava, 1999.

[Doma00] Domański M., Łuczak A., Maćkowiak S., „Spatio-Temporal Scalability for MPEG" IEEE Transactions on Circuits and Systems for Video Technology, Vol. 10, No. 7, pp. 1088-1093, October 2000.

[Doma00b] Domański M., Łuczak A., Maćkowiak S., "Spatio-Temporal Scalability using modified MPEG-2 predictive video coding", Proceedings of European Signal Processing X, pp. 961-964, Tampere 2000.

[Doma00c] Domański M., Łuczak A., Maćkowiak S., "On improving MPEG Spatial Scalability", Proceedings of the IEEE International Conference on Image Processing ICIP'2000, Vancouver, pp. II-848 - II-851, 2000.

[Doma00d] Domański M., Łuczak A., Maćkowiak S., "Scalable MPEG Video Coding with improved B-frame prediction", Proceedings of IEEE International Symposium Circuits and Systems ISCAS 2000, pp. II-273 II-276, Geneva, 2000.

[Doma00e] Domański M., Łuczak A., Maćkowiak S., "Video Coding with Spatio-Temporal Scalability for applications in heterogeneous communication networks", Proceedings of IWSSIP'2000 7th. Int. Workshop on Systems, Signals and Image Processing, pp.127-130, Maribor, 2000.

[Doma00f] Domański M., Łuczak A., Maćkowiak S., "Ulepszony koder skalowalny dla systemu MPEG-2", Systemy i technologie telekomunikacji multimedialnej STM 2000, pp. 153-158, Łódź, 2000.

[Doma01] Domański M., Maćkowiak S., "Modified MPEG-2 Video Coders with Efficient Multi-Layer Scalability", Proceedings of the IEEE International Conference on Image Processing ICIP'2001, vol. II, pp. 1033-1036, Thessaloniki, 2001.

[Doma01b] Domański M., Maćkowiak S., "Three-Layer MPEG-2 Based Video Codecs with Spatio-Temporal Scalability", EURASIP Conference on Digital Signal Processing for Multimedia Communications and Services, pp. 189-192, Budapest, 2001.

[Doma01c] Domański M., Maćkowiak S., "An MPEG based Multilayer Video Coder with Spatio-Temporal Scalability and DCT Data Partitioning", International Conference on Signals and Electronic Systems ICSES 2001, pp. 203-208, Łódź, 2001.

[Doma01d] Domański M., Maćkowiak S., "MPEG 2 Based video coding with three-layer mixed scalability", 9th International Conference on Computer Analysis of Images and Patterns CAIP'2001, Warszawa 2001, Springer Verlag Lecture Notes in Computer Science Volume 2124, Issue, pp. 110-117, 2001.

[Doma02] Domański M., Maćkowiak S., Błaszak Ł., „Efficient hybrid video coders with spatial and temporal scalability", Proceedings IEEE International Conference on Multimedia and Expo ICME'2002 Lausanne, Switzerland, August 26-29, 2002 (accepted).

[Doma02b] Domański M., Maćkowiak S., Błaszak Ł., "Efficient Structure of Video Coders with Motion-Compensated Fine-Granularity Scalability", Proceedings of XI European Signal Processing Conference EUSIPCO'2002, Toulouse, France, September 3-6, 2002 (accepted).

[Doma02c] Domański M., Maćkowiak S., Błaszak Ł., "Fine Granularity In Multi-Loop Hybrid Coders With Multi-Layer Scalability", Proceedings of IEEE Conference on Image Processing 2002, Rochester, NY, USA, 2002 (accepted).

[Doma02d] Domański M., Maćkowiak S., Błaszak Ł., "Scalable video compression for wireless systems", URSI X Nat. Symposium of Radio Sciences, pp. 336-340, Poznań, March 2002.

[Doma02e] Domański M., Maćkowiak S., Błaszak Ł., „Scalable hybrid video coders with double motion compensation", Proceedings of ISCE'2002, September 23-26, 2002 Erfurt, Germany, 2002 (accepted).

[Lucz98] Łuczak A., Maćkowiak S., "Uniwersalne oprogramowanie do realizacji koderów wizyjnych", Poznańskie Warsztaty Telekomunikacyjne, pp. 3.5-1 3.5-4, Poznań, 1998.

[Lucz99] Łuczak A., Maćkowiak S., Rakowski K., Szymański Z., „System multimedialny dla telemedycyny", Poznańskie Warsztaty Telekomunikacyjne, pp. 3.4-1 3.4-6, Poznań, 1999.

[Lucz00] Łuczak A., Maćkowiak S., „Multimedia a medycyna - programowa realizacja zdalnego dostępu do danych patologicznych", V Konferencja Internetu Medycznego, Poznań, 2000.

[Lucz00b] Łuczak A., Maćkowiak S., Dedykowane oprogramowanie do zastosowań telepatologicznych", Telemedycyna 2000, pp. 63-68, Łódź, 2000.

[Lucz00c] Łuczak A., Maćkowiak S., „Medyczny system multimedialny ze zdalnym dostępem", Systemy i technologie telekomunikacji multimedialnej STM 2000, pp. 231-236, Łódź, 2000.

[Mack00] Maćkowiak S., "Hierarchiczna kompresja cyfrowych sygnałów wizyjnych", Poznańskie Warsztaty Telekomunikacyjne, pp. 3.6.1-3.6.4, Poznań, 2000.

[Mack01] Maćkowiak S., "Model weryfikacyjny trójwarstwowego skalowalnego kodera wizyjnego wykorzystującego struktury koderów MPEG-2", Krajowa Konferencja Radiotelekomunikacji Radiofonii i Telewizji KKRITT 2001, pp. 7.3-1 7.3-4, Poznań, 2001.

References used in mention in the dissertation

[Akan96] Akansu A. N., Smith M .J. T., "Subband and Wavelet Transforms Design and Applications", Kluwer Academic Publishers, 1996.

[AlSh99] Al-Shaykh O., Miloslavsky E., Nomura T., Neff R., Zakhor A., "Video Compression Using Matching Pursuits", IEEE Transactions on Circuits and Systems for Video Technology, 17(2), pp.123-143, February 1999.

[Andr02a] Andreopoulos Y., Munteanu A., Van der Auwera G., Barbarien J., Schelkens P., Cornelis J., „Wavelet-based fine granularity scalable video coding with in-band prediction", ISO/IEC JTC1/SC29/WG11MPEG2002/m7906, Jeju Island, March 2002.

[Andr02b] Andreopoulos Y., Munteanu A., Van Der Auwera G., Schelkens P. and Cornelis J. "Wavelet-Based Fully-Scalable Video Coding with In-Band Prediction," Proc. Signal Processing Symposium, SPS'02, Leuven, BE, pp. 217-220, March 2002.

[Andr02c] Andreopoulos Y., Munteanu A., van der Auwera G., Schelkens P., Cornelis J., "A new method for complete-to-overcomplete discrete wavelet transforms," Proc. Digital Signal Processing, DSP'02, Santorini, GR, accepted.

[Arav96] Aravind R., Reha Civanlar M., Reibman A., "Packet loss resilience of MPEG-2 scalable video coding algorithms," IEEE Transactions on Circuits and Systems for Video Technology, Vol. 6, pp. 426 435, October 1996.

[Asbu99] Asbun E., Salama P., Delp E.J. Delp "Encoding of predictive error frames in rate scalable video codecs using wavelet shrinkage", Proceedings of the IEEE International Conference on Image Processing, Kobe, Japan, October 1999.

[Bani01] Banister B.A., Fischer R., "Quadtree Classification and TCQ Image Coding", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 11, No. 1, pp. 3-8, January 2001.

[Benz96] Benzler U., "Scalable Multiresolution Video Coding using Subband Decomposition", First International Workshop on Wireless Image/Video Communication, Loughborough, U.K., pp. 109-114, September 1996.

[Benz96b] Benzler U., Werner O., "Improving multiresolution motion compensating hybrid coding by drift reduction", Picture Coding Symposium (PCS '96), Melbourne, Australia, March 1996.

[Benz99] Benzler U., "Scalable multi-resolution video coding using a combined subband-DCT approach", Picture Coding Symposium (PCS'99), Portland, USA, pp. 21-23, April 1999,.

[Benz99b] Benzler U., Pestel-Schiller U., „Result of Core Experiment on Fine Granularity Scalability for Video (part 2: MC with drift)", ISO/IEC JTC1/SC29/WG11, MPEG99/M4847, July 1999.

[Benz00] Benzler U.: Spatial scalable video coding using a combined subband-DCT approach, IEEE Transactions on Circuits and Systems for Video Technology, vol. 10, pp. 1080-1087, October 2000.

[Benz00b] Benzler U., "Hierarchische DCT-Teilband-Codierung" DFG Schwerpunktprogramm "Mobilkommunikation", Abschlusskolloquium, München, 2000.

[Bert98] Berts J. and Persson A., "Objective and subjective quality assessment of compressed digital video sequences", Master Thesis, Department of Signals and Systems, Chalmers University of Technology, Göteborg, Sweden, 1998.

[Blät00] Blättermann G., Heising G., Marpe D., "A Quality Scalable Mode for H.26L", ITU-T SG16/Q.15, Q15-J24, Osaka, Japan, October 2000.

[Bosv93] Bosveld F., Lagendijk R.L., Biemond J., "Compatible video compression using subband and motion compensation techniques", Proc. Of International Workshop on HDTV, vol.II, 1993.

[Bosv96] Bosveld F., "Hierarchical video compression using SBC," Ph.D. dissertation, Delft University of Technology, Delft 1996.

[Bott01] Bottreau V., Benetiere M., Felts B. and Pesquet-Popescu B., "A Fully Scalable 3D Subband Video Codec", Proceedings of IEEE International Conference on Image Processing, ICIP'2001, vol. II, pp. 1017-1020, Thessaloniki, Greece, October 7-10, 2001.

[Bovi00] Bovik Al., „Handbook of Image & Video Processing", Academic Press, 2000.

[Brün00] Brünig M. and Wien M., "Scalable wavelet video coding using long-term memory motion-compensated prediction", in Proc. SPIE Image and Video Communications and Processing 2000, Vol. 3974, pp. 617-624, San Jose, California, USA, January 2000.

[Camp95] Campbell, A., Eleftheriadis, A. and Aurrecoechea, C., "Meeting End-to-End QoS Challenges for Scalable Flows in Heterogeneous Multimedia Environments", Proc. 6th IFIP International Conference on High Performance Networking (HPN95), Palma de Mallorca, Spain, September 1995.

[Chen99a] Chen Y., van der Schaar M., Radha H. Schuster B., "Residual Computation for Fine Granularity Scalability", ISO/IEC JTC1/SC29/WG11 MPEG99/M4937, July 1999.

[Chen99b] Chen Y., van der Schaar M., and Radha H., "Residual computation for fine granularity scalability: results, analysis and complexity study", ISO/IEC JTC1/SC29/WG11 MPEG 99/M5090, September 1999.

[Chia99] Chiang T., Anastassiou D., "Hierarchical HDTV/SDTV Compatible Coding Using Kalman Statistical Filtering", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 9, No. 3, pp. 424-437, April 1999.

[Chen96] Chen Y., Pearlman W.A., "Three-Dimensional Subband Coding of Video Using the Zero-Tree Method", Proc. Of Symposium on Visual Communications and Image Processing SPIE, Vol. 2727, pp. 1302-1309, 1996.

[Cho01] Cho S., Pearlman W. A., "A Full-Featured, Error Resilient, Scalable, Wavelet Video Codec Based on the Set Partitioning in Hierarchical Trees (SPIHT) Algorithm", CNGV TR 01-01, Center for Next Generation Video, Rensselaer Polytechnic Institute, Troy, New York 12180-3590, http://www.rpi.edu/web/NGV, March 2001.

[Choi97] Choi S.-J., Woods J.W., "Three-Dimensional Subband/Wavelet Coding of Video with Motion Compensation", Proc. Of Visual Communications and Image Processing '97, Vol. 3024, San Jose 1997.

[Choi99] Choi S.-J., Woods J.W., "Motion-Compensated 3-D Subband Coding of Video", IEEE Transactions on Circuits and Systems for VideoTechnology, Vol. 8, no. 2, pp. 155-167, February 1999.

[Conk99] Conklin G.J., Hemami S.S., "A Comparison of Temporal Scalability Techniques", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 9, No. 6, pp. 909-919, September 1999.

[Daub92] Daubechies I., "Ten lectures on wavelets", SIAM, 1992.

[Demo95] Demos G., "Temporal and Resolution Layering in Advanced Television", DemoGraFX, Santa Monica, CA, http://www.demografx.com, 1995.

[Doma94] Domański M., Świerczyński R., "Subband coding of images using hierarchical quantization", Signal Processing VII: Theories and Applications, Edinbourgh, pp. 1218-1221, 1994.

[Doma94a] Domański M., Świerczyński R., „Subband coding of colour images in the YUV space", Of XVII-th National Conference Circuit Theory and Electronic Networks, Wrocław – Polanica Zdrój, Poland, pp. 221-226, 1994.

[Doma95] Domański M., Świerczyński R., „Video 3-D Subband Coding for Low Bit Rate Channels", Proc. Of XVIII-th National Conference Circuit Theory and Electronic Networks, Polana Zgorzelisko, Poland, Vol. 2, pp. 467-472, 1995

[Doma96] Domański M., Świerczyński R., "3-D subband coding of video using recursive filter banks", Signal Processing VIII: Theories and Applications, Trieste, pp. 1359-1362, 1996.

[Doma97] Domański M., Łuczak A., Świerczyński R., „Skalowalne kodowanie sygnałów wizyjnych" Poznańskie Warsztaty Telekomunikacyjne PWT'97, pp. III 5-1 – III 5-4, Poznań, 1997.

[Doma98d] Domański M., Bartkowiak M., „Compression", The Colour Image Processing Handbook, Sengwine S. J., Horne R. E. N. (ed), Chapman & Hall, London, 1998.

[Doma98e] Domański M. „Zaawansowane techniki kompresji obrazów i sekwencji wizyjnych", Wydawnictwo Politechniki Poznańskiej, Poznań, 1998.

[Dovs00] Dovstam K., „Video coding in H.26l", Master of science thesis, Royal Institute of Technology, Kungl Tekniska Högskolan, 2000.

[Felt00] Felts B. and Pesquet-Popescu B., "Efficient Context Modeling in Scalable 3D Wavelet-Based Video Compression", Proc. "IEEE International Conference on Image Processing", Vancouver, Canada, September 2000.

[Fuld96] Fuldseth A., Ramstad T.A., "Subband video coding with smooth motion compensation", In IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP'1996, Vol. 4, pp. 2333, 1996.

[Gall01] Gallant M., Kossentini F., "Rate-distortion optimized layered coding with unequal error protection for robust internet video," IEEE Transactions on Circuits and Systems for Video Technology, Vol. 11, pp. 357-372, March 2001.

[Gemm95] Gemmell J., Vin H. M., Kandlur D. D., Rangan P. V., Rowe L. A., "Multimedia Storage Servers: A Tutorial", IEEE Computer, Vol. 28, No. 5, pp. 40-49, 1995.

[Ghan99] Ghanbari M., Video coding, London, IEE 1999.

[Ghar97] Gharavi H., Richards C.I., "Partitioning of MPEG coded video bitstreams for wireless transmission," IEEE Signal Processing Letters, Vol. 4, pp. 153 155, June 1997.

[Ghar99] Gharavi H. and Alamouti S. M., "Multipriority Video Transmission for Third-Generation Wireless Communication Systems", Proceedings Of the IEEE, Vol. 87, No. 10, pp. 1751-1763, October 1999.

[Ghar01] Gharavi H., Alamouti S. M., "Video Transmission for Third Generation Wireless Communication Systems", Journal of Research of the National Institute of Standards and Technology, Vol. 106, No. 2, pp. 455–469, March–April 2001.

[Giro96] Girod B., Horn U., Belzer B., "Scalable Video Coding for Multimedia Applications and Robust Transmission over Wireless Channels", Proceedings 7th International Workshop on Packet Video, Brisbane, Australia, pp. 43-48, March 1996.

[Giro97] Girod B., Färber N., Horn U., "Scalable Codec Architectures for Internet Video-on-Demand" Proc. 1997 Asilomar Conference on Signals and Systems, Pacific Grove, CA, USA, November 1997.

[Giro98] Girod B., Färber N. and Stuhlmüller K., "Internet Video Streaming", IEEE 10th Tyrrhenian Workshop on Digital Communications (Multimedia Communications), Ischia, September, 1998.

[Giro99] Girod B., Färber N. and Stuhlmüller K., "Internet Video-Streaming", In Multimedia Communications, F. De Natale and S. Pupolin, Eds., Springer, ISBN 1-85233-135-6, pp. 547-556, 1999.

[Giro00] B. Girod, N. Färber, "Wireless Video," in A. Reibman, M.-T. Sun (eds.), Compressed Video over Networks, Marcel Dekker, 2000.

[Hanz01] Hanzo L., Cherriman P., Streit J., Wireless video communications: second to third generation systems and beyond, New York, IEEE Press, 2001.

[Hask96] Haskell B.G., Puri A., Netravali A.N., Digital video: an introduction to MPEG-2, New York, Chapman & Hall, September 1996.

[HeWu02] He Y., Wu F., Li S., Zhong Y., Yang S., "H.26L-based fine granularity scalable video coding", Proc. of ISCAS 2002, accepted, 2002.

[Held96] Held G., Marshall T. R., "Data and image compression: tools and techniques", John Willey and Sons Ltd., NY, 1996.

[HeYa01] He Y., Yan R., Wu F., Li S., "H.26L-based fine granularity scalable video coding", ISO/IEC JTC1/SC29/WG11, MPEG2001/M7788, December 2001.

[Hema99] Hemami S. S., "Distortion Analyses for Temporal Scalability Coding Techniques," Proc. IEEE Int. Conf. on Image Processing, Kobe, Japan, October 1999.

[Horn97] Horn U. and Girod B., "Scalable video transmission for the Internet", Journal on Computer Networks and ISDN Systems, Vol. 29, No. 15, pp. 1833-1842, 1997.

[Hsu99] Hsu Y.-F., Hsieh C.-H., Chen Y.-C., „Embedded SNR scalable MPEG-2 video encoder and its associated error resilience decoding procedures" Signal Processing: Image Communication 14 (1999) pp. 397-412, 1999.

[IEEE99] Special issue on video transmission for mobile multimedia applications, Proc. of the IEEE, vol. 87, October 1999.

[IEEE00] Special Issue on MPEG-7 Technology, Image Communication, September 2000.

[IEEE01] Special issue on MPEG-7, IEEE Transactions on Circuits and Systems for Video Technology , Vol. 11, Issue: 6 , June 2001.

[IETF97] IETF Internet Draft, Shenker S., Partridge C., Guerin R., "Specification of Guaranteed Quality of Service", RFC 2212, September 1997.

[IETF98] IETF Internet Draft, Deering S., Hinden R., "IPV6 Internet Protocol, Version 6 (IPv6) Specification 2460", December 1998.

[ISO94] ISO/IEC IS 13818-2 / ITU-T Rec. H.262, "Generic coding of moving pictures and associated audio, part 2: video", November 1994.

[ISO98] ISO/IEC IS 14496-2, "Generic coding of audio-visual objects, Part 2: Visual", December 1998.

[ISO00] ISO/IEC 14496-2/FPDAM4, "Coding of Audio-Visual Objects, Part-2 Visual, Amendment 4: Streaming Video Profile, July 2000.

[ISO01] ISO/IEC JTC1/SC29/WG11 N3908, "MPEG-4 Video Verification Model, ver. 18.0," Pisa, January 2001.

[ISO01b] "AHG on Exploration of Interframe Wavelet Technology in Video", ISO/IEC JTC1/SC29/WG11/N4474, Pattaya, December 2001.

[ISO01c] ISO/IEC JTC1/SC29/WG11/N4473, "AHG on FGS", Pattaya, December 2001.

[ISO01d] ISO/IEC JTC 1/SC 29/WG 11N4341, "Proposed Work Plan for the Joint Video Project between ITU-T SG16 and ISO/IEC JTC 1/SC 29/WG 11", July 2001.

[ISO01e] ISO/IEC JTC 1/SC 29/WG 11N4355, "Anticipatory Joint Model (JM) of Enhanced-Compression Video Coding", July 2001.

[ISO02] ISO/IEC JTC 1/SC 29/WG 11/N4751, "AHG on Exploration of Interframe Wavelet Technology in Video", Fairfax, VA, USA, May 2002.

[ISO02b] ISO/IEC JTC1/SC29/WG11 MPEG2002/N4747, "Description of Exploration Experiments in Interframe Wavelet Coding", Fairfax, May 2002.

[ISO02c] ISO/IEC JTC1/SC29/WG11 MPEG2002/N4584, "Experimental conditions for wavelet coding exploration", March 2002.

[ITUR] Recommendation ITU-R BT.601-4, Encoding parameters of digital television for studios, Geneva, 1994.

[ITUR94] Recommendation ITU-R BT.500-6, Methodology for the subjective assessment of the quality of television pictures, 1994.

[ITUT96] ITU-T, "Video coding for narrow telecommunication channels at < 64kbit/s", Recommendation H.263, 1996.

[ITUT98] "Video Codec Test Model Near -Term Number 10 (TMN -10)", ITU-T/SG 16/VCEG (formerly Q.15, now Q.6), Tampere, April 1998.

[ITUT98b] ITU-T Rec. H.245, "Infrastructure of audiovisual services – Communication procedures - Control protocol for multimedia communication", September 1998.

[ITUT01] "H.26L Test Model Long-Term Number 7 (TML-7)", ITU-T/SG 16/VCEG (formerly Q.15 now Q.6), Doc. VCEGM81, April 2001.

[John80] Johnston J., "A filter family designed for use in quadrature mirror filter banks", Proc. IEEE Int. Conf. Acoustics Speech Signal Proc., pp. 291-294, 1980.

[Kall01] Kalluri R., van der Schaar M., "Single-Loop Motion-Compensated based Fine-Granular Scalability (MC-FGS), with cross-checked results", ISO/IEC JTC1/SC29/WG11, MPEG2001/6831, January 2001.

[Kara01] Karande S., „Motion Compensated FGS (MC-FGS): Cross Checking Results", ISO/IEC JTC1/SC29/WG11, MPEG2001/6830, January 2001.

[Kim00] Kim B.-J., Xiong Z., Pearlman W.A., "Low Bit-Rate Scalable Video Coding with 3-D Set Partitioning in Hierarchical Trees (3-D SPIHT)", IEEE Transactions on Circuits and Systems for Video Technology, Volume: 10, No. 8, December 2000.

[Kim97] Kim B.-J., Pearlman W., "An embedded wavelet video coder using three-dimensional set partitioning in hierarchical trees (SPIHT)," in Proceedings of Data Compression Conference, pp. 251--260, 1997.

[Koen97] Koenen R., Pereira F., Chiariglione L., "MPEG-4: Context and objectives", Signal Processing: Image Communication, Vol. 9, Issue: 4, pp. 295-304, May 1997.

[Kond98] Kondi L. P., Ishtiaq F., and Katsaggelos A. K., "On Video SNR Scalability", Proc. 1998 IEEE International Conf. on Image Processing, vol. 3, pp. 934-936, Chicago, IL, October 1998.

[Kron96] Kronacher F., "Anpassung der Wichtungsmatrizen fur DCT – Koeffizienten an eine hierarchische Quellencodierung mit Teilbandzerlegung", Institut für Theoretische

Nachrichtentechnik und Informationsverarbeitung, Universität Hannover, Studienarbeit, July 1996.

[Kurc01] Kurceren R., Karczewicz M., "Improved SP-frame encoding", ITU-T SG16/Q.15, VCEG-M73, Austin, USA, April 2001.

[Lage96] Lagendijk R. L., Bosveld F., Biemond J., Chapter 8.: "Subband Video Coding", in "Subband and Wavelet Transforms Design and Applications" edited by Akamsu A. N., Smith M. J.T., Kluwer Academic Publishers 1996.

[LeGa92] Le Gall D.J. "MPEG: A Video Compression Standard for Multimedia Applications", Communications of the ACM 34 (4) 1991, pp. 46-58, 1991.

[Li99] Li W., Chen Y., „Experiment Result on Fine Granularity Scalability", ISO/IEC JTC1/SC29/WG11, MPEG99/M4473, March 1999.

[Li99b] Li W., "Syntax of Fine Granularity Scalability for Video Coding", ISO/IEC JTC1/SC29/WG11 M4792, July 1999.

[Li00] Li W., "Fine Granularity Scalability in MPEG-4 for Streaming Video" Proc. of ISCAS'2000, Vol. I, pp. 299-302, 2000.

[Li01] Li W., "Fine Granularity Scalability Using Bit-Plane Coding of DCT Coefficients", ISO/IEC JTC1/SC29/WG11, MPEG98/M4204, December 1998.

[Li01b] Li W. „Overview of Fine Granularity Scalability in MPEG-4 Video Standard", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 11 Issue: 3, pp. 301 - 317, March 2001.

[Li02] Li W., "Fine Granularity Scalability for MPEG-4 Part 10", ISO/IEC JTC1/SC29/WG11 M8003 JeJu, March 2002.

[LiJi99] Li W., Jiang H., "Complexity Analysis of Two Residue Computation Methods for FGS", ISO/IEC JTC1/SC29/WG11 MPEG98/M5082, October 1999.

[LiLi00] Li W., Ling F., Chen X., "Fine Granularity Scalability in MPEG-4 for Streaming Video", Proc. of ISCAS 2000 - IEEE International Symposium on Circuits and Systems, Geneva, pp. I-299 – I.302, 2000.

[LiLi97] Li W., Ling F., Sun H., "Bit-Plane Coding od DCT Coefficients", ISO/IEC JTC1/SC 29/WG 11 MPEG97/M2691, October 1997.

[Ling99] Ling F., Li W., Sun H., "Bitplane Coding of DCT Coefficients for Image and Video Compression", Proc. Of Visual Communications and Image Processing VCIP'99, pp. 500-508, San Jose, January 1999.

[Ling01] Ling F., Li W., "Verification of Macroblock Based PFGS Results", ISO/IEC JTC1/SC29/WG11 M6767, Pisa, January 2001.

[LiWu99] Li S., Wu F., Zhang Y.-Q., „Study of a new approach to improve FGS video coding efficiency", ISO/IEC JTC1/SC29/WG11, MPEG99/M5583, December 1999.

[LiWu00] Li S., Wu F., Zhang Y.-Q., „Experimental Results with Progressive Fine Granularity Scalable (PFGS) Coding", ISO/IEC JTC1/SC29/WG11, MPEG99/M5742, March 2000.

[Macn99] Macnicol J., Frater M., Arnold J., "Results on Fine Granularity Scalability", ISO/IEC JTC1/SC29/WG11, MPEG99/M5122, October 1999.

[Mall93] Mallat S., Zhang Z., "Matching pursuits with time-frequency dictionaries", IEEE Transactions on Signal Processing, Vol. 41, No.12, pp. 3397--3415, 1993.

[Manj02] Manjunath B.S., Salembier P., Sikora T., "Introduction to MPEG-7: Multimedia Content Description Interface", John Wiley & Sons, Ltd., 2002.

[Marc00] Marcellin M. W., Gormish M. J., Bilgin A., Boliek M. P.,"An Overview of JPEG-2000", Proc. of IEEE Data Compression Conference, pp. 523-541, 2000.

[MASC] http:// www.cwi.nl/projects/mascot

[MASC01] Dissemination and Use Plan, Report Version: 0.2, IST–2000-26467 Metadata for Advanced Scalable Video Coding Tools, 2001.

[MASC02] Deliverable D1.1: "Specification of Architecture of the MASCOT codec", Report Version: 4.0, IST–2000-26467 Metadata for Advanced Scalable Video Coding Tools, 2002.

[Math97] R. Mathew, J.F. Arnold, "Layered coding using bitstream decomposition with drift correction," IEEE Trans. Circ. Syst. Video Techn., vol. 7, pp. 882-891, December 1997.

[Moul00] Moulin P., "Multiscale Image Decompositions and Wavelets", in Handbook of Image and Video Processing, A. C. Bovik, Ed., Academic Press, 2000.

[MOMUS] http://www.cordis.lu/infowin/acts/analysys/products/thematic/mpeg4/momusys/momusys.htm

[MPEG96] MPEG Software Simulation Group, MPEG-2 Encoder Decoder version 1.2, July 19, 1996, ftp://ftp.mpeg.org/pub/mpeg/mssg.

[MPEG01] "Proposed Work Plan for the Joint Video Project between ITU-T SG16 and ISO/IEC JTC 1/SC 29/WG 11", ISO/IEC JTC 1/SC 29/WG 11N4341, July 2001.

[MPEG01b] "Anticipatory joint mode (JM) of enhanced-compression video coding", ISO/IEC JTC1/ SC29/WG11, N4355, July 2001.

[MPEG01c] "Overview of the MPEG-7 Standard (version 6.0)", ISO/IEC JTC1/SC29/WG11 N4509, December 2001.

[MPEG01d] "JNB Comment on JVT activity", ISO/IEC JTC 1/SC 29/WG 11, MPEG2001/M7782, Pattaya, December 2001.

[MPEG01e] "Report of the Ad-hoc Group on Requirements for Joint Video Project", ISO/IEC JTC1/SC29/WG11 M7667, Pattaya, December 2001.

[Neff97] Neff R., Zakhor A., "Very low bit rate video coding based on matching pursuits", IEEE Trans. Circuits and Systems for Video Technology, Vol.7, No.1, pp. 158-171, February 1997.

[Ohm93] Ohm J.-R., "Three-Dimensional Motion-Compensated Subband Coding," Proceedings International Symposium on Video Communications and Fiber Optic Services, SPIE, vol. 1977, Berlin, pp. 188-197, April 1993.

[Ohm93b] Ohm J.-R., "3-D SBC-VQ with Motion Compensation" Proceedings Picture Coding Symposium (PCS'93), Lausanne, pp. 11.5-1 - 11.5-2, March 1993.

[Ohm94] Ohm J.-R., "Three-Dimensional Subband Coding with Motion Compensation," IEEE Transactions on Image Processing, vol. IP-3, no. 5, pp. 559-571, September 1994.

[Ohm95] Ohm J.-R., "Three-Dimensional Subband Coding with Motion Compensation", ISO/IEC/JTC1/SC29/WG11 M0333, MPEG95, November 1995.

[Ohm01] Ohm J.-R., Li W., Ling F., Chun K.-W., Li S., Wu F., Zhang Y.-Q., van der Schaar M., Jiang H., Chen X., Vetro A., Sun H., Wollborn M., "Summary of Discussions on Advanced Scalable Video Coding", ISO/IEC JTC1/SC29/WG11 M7016, Singapore, March 2001.

[Ohm01b] Ohm J.-R., van der Schar M., "Scalable Video Coding", Tutorial Material SM.T3, IEEE International Conference in Image Processing ICIP'2001, Thessaloniki, Greece, 2001.

[Ohm02] Ohm J.-R., Hanke K., Rusert T., „Improved capabilities of motion-compensated wavelet filters", ISO/IEC JTC1/SC29/WG11 MPEG02/M8360, Fairfax, May 2002.

[Ohm02b] Ohm J.-R., Ebrahimi T., „Report of Ad hoc Group on Exploration of Interframe Wavelet Technology in Video", ISO/IEC JTC1/SC29/WG11 MPEG02/M8359, Fairfax, May 2002.

[Okub95] Okubo S., McCann K., Lippmann A., "MPEG-2 requirements, profiles and performance verification – Framework for developing a generic video coding standard", Signal Processing: Image Communication 7 (1995) pp. 201-209, 1995.

[Pere97] Pereira, F.; O'Connell, K.; Koenen, R.; Etoh, M., "MPEG-4, Part 1: Invited papers", Signal Processing: Image Communication, Vol. 9, Issue: 4, pp. 291-294, May 1997.

[Pesq00] Pesquet-Popescu B., Benetiere M., Bottreau V., Felts B., "Improved color coding for scalable 3D wavelet video compression", Proc. of EUSIPCO2000, pp. 1481-1484, 2000.

[Pesq01] Pesquet-Popescu B. and Bottreau V., "Three-Dimensional lifting schemes for Motion Compensated Video Compression", Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing 2001, ICASSP'2001, IMDSP-L3.6, Salt Lake City, Utah, May 7-11, 2001.

[Pesq02] Pesquet-Popescu B., Piella G., Heijmans H., "Adaptive update lifting with gradient criteria modeling high-order differences" , accepted for publication in Proc. Int. Conf. Acoustic, Speech and Signal Proc. 2002, Orlando, FL, May 2002.

[Peng01] Peng W., Chen Y.-K., "Mode-adaptive fine granularity scalability", Proc. of IEEE International Conference on Image Processing, ICIP'2001, vol.II, pp. 993-996, Thessalonike 2001.

[Pita00] Pitas I., "Digital Image Processing Algorithms and Applications", John Wiley & Sons Inc., NY, 2000.

[Podl95] Podlichuk Ch., Jayant N., Farvardin N., "Three-dimensional subband coding of video", IEEE Transactions on Image Processing, Vol. 4, pp.125-284, 1995.

[Puri93] Puri A., Wong A., "Spatial Domain Resolution Scalable Video Coding", Proc. of SPIE Visual Communications and Image Processing '93 8-11, Cambridge Massachussets, pp. 718-729, November 1993.

[Puri94] Puri A., Yan L., Haskell B.G.," Temporal resolution scalable video coding", Proc. of IEEE International Conference on Image Processing, ICIP'1994, pp. 947-951, 1994.

[Radh99] Radha H., Chen Y., Parthasarathy K., Cohen R., "Scalable internet video using MPEG-4" Signal Processing: Image Communication 15 (1999) pp. 95-126, 1999.

[Reib01] Reibman A., Bottou L., Basso A., "DCT-based scalable video coding with drift", Proc. of IEEE International Conference on Image Processing, ICIP'2001, Thessaloniki 2001, vol.II, pp. 989-992, 2001.

[Robe97] Robers M. A., "SNR Scalable Video Coder Using Progressive Transmission of DCT Coefficients", Master's thesis, Northwestern University, Department of Electrical and Computer Engineering, May 1997.

[Robe98] Robers M. A., Kondi L. P., Katsaggelos A. K., "SNR Scalable Video Coder Using Progressive Transmission of DCT Coefficients," Proc. SPIE Conf. on Visual Communications and Image Processing, pp. 201-212, SPIE vol. 3309, San Jose, CA, Jan. 28-30, 1998.

[Rose01] Rose K. and Regunathan S., "Toward optimality in scalable predictive coding," IEEE Transactions on Circuits and Systems for Video Technology, vol. 11, pp. 965-976, July 2001.

[Said96] Said A. and Pearlman W. A., "A new fast and efficient image codec based on set partitioning in hierarchical trees," IEEE Transactions on Circuits and Systems for Video Technology, vol. 6, pp. 243--250, June 1996.

[Scha99] van der Schaar M., Radha H., Chen Y., "An all FGS solution for hybrid Temporal-SNR scalability", ISO/IEC JTC1/SC29/WG11, MPEG 99/M5552, November 1999.

[Scha00] van der Schaar M., Radha H., "Results for all FGS hybrid temporal-SNR scalability", ISO/IEC JTC1/SC29/WG11, MPEG00/M5782, March 2000.

[Scha00b] van der Schaar M., Radha H., " Motion-Compensation based Fine-Granular Scalability (MC-FGS)", ISO/IEC JTC1/SC29/WG11, MPEG 2000/M6475, October 2000.

[Scha00c] van der Schaar M., "All FGS temporal-SNR-spatial scalability", ISO/IEC JTC1/SC29/WG11 MPEG 2000/M6490 La Baule, October 2000.

[Scha01] van der Schaar M., Radha H., "A Hybrid Temporal-SNR Fine-Granular Scalability for Internet Video", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 11, No. 3, pp. 318-331, March 2001.

[Scha01b] van der Schaar M., Radha H., "The MPEG-4 Fine-Frained Scalable Video Coding Method for Multimedia Streaming Over IP", IEEE Transactions on Multimedia, Vol. 3, No. 1, March 2001.

[Seeg98] A. Seeger, U. Benzler, "Hierarchische Quellen- und Kanalcodierung für digitales Video", ITG-Fachtagung "Codierung für Quelle, Kanal und Übertragung", Aachen, März 1998.

[Shen99] Shen K., Delp E.J., "Wavelet Based Rate Scalable Video Compression", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 9, No. 1, pp. 109-122, February 1999.

[Skar93] Skarbek W., "Metody reprezentacji obrazów cyfrowych", Akademicka Oficyna Wydawnicza PLJ, Warszawa, 1993.

[Skar98] Skarbek W. (ed), "Multimedia. Algorytmy i standardy kompresji", Akademicka Oficyna Wydawnicza PLJ, Warszawa, 1998.

[Smol01] Smolić A. and Wiegand T., "High-Resolution Video Mosaicing", in Proc. IEEE International Conference on Image Processing, Vol. III pp. 872-875, Thessaloniki, Greece, September 2001.

[SMPT01a] SMPTE 336M-2001, Television - Metadata Dictionary Structure, 2001.

[SMPT01b] SMPTE 335M-2001, Data Encoding Protocol Using Key-Length-Value, 2001.

[SMPT01c] SMPTE RP210.2-2001, Metadata Dictionary, 2001.

[Sted93] R. Stedman, H. Gharavi, L. Hanzo, and R. Steele, "Transmission of subband coded images via mobile channels," IEEE Transactions on Circuits and Systems for Video Technology, vol. 3, pp. 15 26, February 1993.

[Stuh99] Stuhlmüller K.W., Link M., Girod B., Horn U., "Scalable Internet Video Streaming With Unequal Error Protection", Packet Video Workshop, New York, April 1999.

[Sull98] Sullivan G. J. and Wiegand T., Rate-Distortion Optimization for Video Compression, IEEE Signal Processing Magazine, pp. 74-90, November 1998.

[Sull01] Sullivan G., Wiegand T., Stockhammer T., "Using the Draft H.26L Video Coding Standard for Mobile Applications", Proc. Of IEEE Iinternational Conference on Image Processing 2001, pp. 573-576, Thessaloniki, Greece, October 2001.

[Sull02a] Sullivan G., Luthra A., and Wiegand T., "Report of the First Meeting of the Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG", ISO/IEC JTC1/SC29/WG11MPEG2002/M8149, Jeju Island, March 2002.

[Sull02b] Sullivan G., Wiegand T. and Luthra A.„JVT (of ISO/IEC MPEG and ITU-T Q.6/16 VCEG) 2nd Meeting Report,Jan 29 – Feb 1 2002", ISO/IEC JTC1/SC29/WG11 MPEG2002/M8150, Jeju Island, March 2002.

[Sun02] Sun X., Wu F., Li S., „The Description and Proposed Syntax of JVT-based FGS", ISO/IEC JTC1/SC29/WG11, MPEG2002/8204, March 2002.

[Sun01] Sun X., Wu F., Li S., Gao W., Zhang Y.-Q., " Macroblock-based Progressive Fine Granularity Scalable-Temporal (PFGST) video coding", ISO/IEC JTC1/SC29/WG11, MPEG2001/6780, January 2001.

[Sun01b] Sun X., Wu F., Li S., Gao W., Zhang Y.-Q., "Macroblock-based temporal-SNR progressive fine granularity scalable video coding", IEEE International Conference on Image Processing ICIP'2001, 1025-1028, Thessaloniki, Greece, October, 2001.

[TanZa99] Tan W., Zakhor A., „Real-time Internet Video Using Error Resilient Scalable Compression and TCP-friendly Transport Protocol", IEEE Trans. Multimedia, Vol. 1, No. 2, pp 172-186, June 1999.

[TanZa99b] Tan W., Zakhor A., "Multicast Transmission of Scalable Video using Receiver-driven Hierarchical FEC", Packet Video Workshop 99, April 1999.

[Taub94] Taubman D. and Zakhor A., "Multirate 3-D subband coding of video," IEEE Trans. Circ. Syst. Video Techn., vol. 3, pp. 572-588, September 1994.

[Taub02] Taubman D. S., Marcellin M. W., JPEG2000 Image Compression Fundamentals Standards and Practice, Kluwer Academic Publishers, 2002.

[Topi98] Topiwala P.N., "Wavelet image and video compression", Kluwer Academic Publishers, 1998.

[Tsun94] Tsunashima K., Stampleman J.B., Bove V.M., "A Scalable Motion-Compensated Subband Image Coder", IEEE Transactions on Communications, vol. 42, pp. 1894-1901, January 1994.

[Tura00] Turaga D. S. and Chen T., Book Chapter "Fundamentals of Video Coding: H.263 as an example", Compressed Video Over Networks, The Signal Processing Series, Marcel Dekker, 2000.

[Tura02] Turaga D. S. and van der Schaar M., "Unconstrained motion compensated temporal filtering", ISO/IEC JTC1/SC29/WG11 M8388, Fairfax, May 2002.

[Uz91] Uz B., Vetterli M., LeGall D. J., ”Interpolative multiresolution coding of advanced television with compatible subchannels”, IEEE Transactions on Circuits and Systems for Video Technology, Vol.1, No.1, pp. 86-99, 1991.

[Wern96] Werner O., “Drift analysis and drift reduction for multiresolution hybrid video coding,” Signal Processing: Image Communication, vol. 8, pp. 387-409, July 1996.

[Wern96b] Werner O., “Driftanalyse und Driftreduktion für hierarchische Bewegtbildcodierung”, Universität Hannover, Doktor-Ingenieur Dissertation, 1996.

[Wieg96] Wiegand T., Lightstone M., Mukherjee D., Campbell T. G., Mitra S. K., “Rate-Distortion Optimized Mode Selection for Very Low Bit Rate Video Coding and the Emerging H.263 Standard”, IEEE Transactions on Circuits and Systems for Video Technology, vol. 6, no. 2, pp. 182-190, April 1996.

[Wils96] Wilson D., Ghanbari M., „Transmission of SNR Scalable Two-Layer MPEG-2 Coded Video Through ATM Networks”, Proc. Of 7th International Workshop on Packet Video, Brisbane, Australia, pp. 185-189, 1996.

[Wood91] Woods J. (ed.), “Subband image coding”, Kluwer, Boston , 1991.

[Wood02] Woods J. W. and Chen P., “Improved MC-EZBC with Quarter-pixel Motion Vectors”, ISO/IEC JTC1/SC29/WG11 MPEG2002/M8366, Fairfax, VA, May 2002.

[WuHo01] Wu D., Hou Y., Zhang Y., “Scalable video coding and transport over broadband wireless networks,” Proc. of the IEEE, vol. 89, pp. 6-20, January 2001.

[WuLi01] Wu F., Li S., Sun X., Yan R., Zhang Y.-Q., “Macroblock-based progressive fine granularity scalable coding”, ISO/IEC JTC1/SC29/WG11 N6779, MPEG 2001, January 2001.

[Vlee00] De Vleeschouwer C., Macq B., "SNR Scalability using Matching Pursuits", IEEE transactions on Multimedia, 2 (4), pp. 198-208, December 2000.

[Vlee00b] De Vleeschouwer C., Macq B., "SNR Scalability with Matching Pursuits", Packet Video Workshop, Cagliary, May 2000.

[Yan01] Yan R., Wu F., Li S., Zhang Y.-Q., “Macroblock-based Progressive Fine Granularity Spatial Scalability (mb-PFGSS)”, ISO/IEC JTC1/SC29/WG11, MPEG2001/M7112, March 2001.

Other related papers

Akramullah S. M., Ahmad I. and Liou M. L., “A Portable and Scalable MPEG-2 Video Encoder on Parallel and Distributed Computing Systems", Proceedings of SPIE, Vol. 2727, Pt. 2, Symposium on Visual Communications and Image Processing '96, pp. 973-984, Orlando, Florida, USA, March 1996.

Alshaykh O., Banham M., Dejaco A., Neff R., Ramanzin Y., Sethuraman S., Sodagar I., Suzuki T., Tsai C.-J., "Request for a New Level of Simple Scalable for Wireless Devices", ISO/IEC JTC 1/SC 29/WG 11 M7003, March 2001.

Alshaykh O., Banham M., Sodagar I., Tsai C.-J., Dejaco A., Sethuraman S., "Request for a New Level of Simple Scalable for Wireless Devices" ISO/IEC JTC 1/SC 29/WG 11 N6651, October 2000.

Asbun E., Salama P., Delp E.J., "Encoding of predictive error frames in rate scalable video codecs using wavelet shrinkage", Proceedings of the IEEE International Conference on Image Processing, ICIP'1999, 1999.

Asbun E., Salama P., and Delp E. J., "Preprocessing and Postprocessing Techniques for Encoding Predictive Error Frames in Rate Scalable Video Codecs," Proceedings of the 1999 International Workshop on Very Low Bitrate Video Coding, Kyoto, Japan, pp. 148-151, October 1999.

Asbun E., Salama P., Shen K., and Delp E. J., "Very Low Bit Rate Wavelet-Based Scalable Video Compression," Proceedings of the IEEE International Conference on Image Processing, Chicago, Illinois, October 1998.

Baras J. S., "3D Wavelet-Based Video Codec with Human Perceptual Model", Master's Thesis, The Center for Satellite and Hybrid Communication Networks University of Maryland, 1999.

Bayarakeri S., Merserau R. M., „MPEG-2 Nonlinear Temporally Scalable Coding and Hybrid Quantization", In IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP'1997 t.4, 1997.

Boucherok F., Vial J.F., "Compatible multi-resolution coding scheme", Proc. Of International Workshop on HDTV'92, 1992.

Brigger P., Illgner K., Unser M., „Centered Pyramids", IEEE Transactions On Image Processing, Vol. 8, No. 9, September 1999.

Campisi P., Gentile M., Neri A., "Three Dimensional Wavelet Based Approach for a Scalable Video Conference System", Proceedings of the IEEE International Conference on Image Processing, ICIP'1999, 1999.

Chang E. and Zakhor A., "Disk-based Storage for Scalable Video," IS&T/SPIE International Symposium on Electronic Imaging: Science and Technology, Vol. 3020: Multimedia Computing and Networking 1997, San Jose, February 1997.

Chang E. and Zakhor A., "Disk-based Storage for Scalable Video," IEEE Transactions on Circuits and Systems for Video Technology, Vol. 7, No. 5, pp. 758-770, October 1997.

Chang E. and Zakhor A., "Scalable Video Coding Using 3-D Subband Velocity Coding and Multi-Rate Quantization," IEEE International Conference on Acoustics, Speech, and Signal Processing, Minneapolis, pp. V:574-577, April 1993.

Chang E. and Zakhor A., “Scalable Video Data Placement on Parallel Disk Arrays," IS&T/SPIE International Symposium on Electronic Imaging: Science and Technology, Vol. 2185: Image and Video Databases II, San Jose, pp. 208-221, February 1994.

Chang E. and Zakhor A., “Velocity-based Architectures for 3-D Subband Video Coding," IEEE Visual Signal Processing Workshop, Raleigh, pp. 245-251, September 1992.

Chen J.-R., Huang S.-Y., Hsieh H.-Y., Wang J.-S., “A New Two-Layer Coding Scheme Utilizing Background Information”, Proc. Of SPIE Visual Communications and Image Processing ’95, 24-26, Taipei Taiwan, Vol. 2501, part 3/3 pp. 1416-1425, May 1995.

Chen Y. and Thebault B., „Results on Selective Enhancement for Fine Granularity Scalability”, ISO/IEC JTC1/SC29/WG11 MPEG 99/M5091, September 1999.

Chen Y. and Thebault B., “Selective Enhancement for Fine Granular Scalable Video”, ISO/IEC JTC1/SC29/WG11 MPEG 99/M5551, December 1999.

Chung-How J. T. and Bull D. R., “Joint Audio and Video Internet Coding (JAVIC)”, Final report, University of Bristol, Department of Electrical and Electronic Engineering, November 1999.

Comer M. L., Shen K. and Delp E. J., "Rate-scalable Video Coding Using a Zerotree Wavelet Approach," Ninth Workshop on Image and Multidimensional Signal Processing, pp. 162-163, Belize City, Belize, March 1996.

Delp E. J., Salama P., Asbun E., Saenz M., and Shen K., "Rate Scalable Image and Video Compression Techniques," Proceedings of the 42nd Midwest Symposium on Circuits and Systems, Las Cruces, New Mexico, August 1999.

Diepold K., “Report on the AHG on the Call for Proposals for new tools to further improve video coding efficiency”, ISO/IEC JTC1/SC29/WG11 M 6991 Singapore, March 2001.

Dubois E., Baaziz N., and Matta M., "Impact of scan conversion methods on the performance of scalable video coding," in Digital Video Compression: Algorithms and Technologies 1995, vol. 2419, pp. 119-128, February 1995.

Eleftheriadis A. and Anastassiou D., “Optimal Data Partitioning Of MPEG-2 Coded Video”, Proceedings 1st International Conference on Image Processing ICIP’1994, pp. 273-277, Austin, Texas, November 1994.

Gharavi H., “Multilayer Subband-Based Video Coding”, IEEE Transactions on Communications, vol. 39 No. 9, September 1991.

Girod B., Hartung F., Horn U., „Multiresolution Coding of Image and Video Signals“ Proc. Of EUSIPCO 98, pp. 1957-1960, Rhodes, Greece.

Girod B., Horn U., Belzer B., "Scalable Video Coding with Multiscale Motion Compensation and Unequal Error Protection", in: Wang Y., Panwar S., Kim S.-P. and

Bertoni H. L. (eds), Multimedia Communications and Video Coding, pp. 475-482, Plenum Press, New York, October 1995.

Glenn, Marcinka, and Dhein, Simple Scalable Video Compression Using 3-D Subband Coding, SMPTE Journal: Compression pp. 105-140, March 1996.

Gonzales C., Viscito E., "Flexibly scalable digital video coding", Signal Processing: Image Communication 5(1993) pp. 5-20 Vol. 5 No. 1-2, 1993.

He Y., Yang S., Zhong Y., "Block-based FGS Coding with Optimized Truncation for MPEG-4 Streaming Video", ISO/IEC JTC1/SC29/WG11 MPEG02/M7991, Jeju, Korea, March 2002.

Hee C., Kim J.-K., "An Adaptive Error Concealment in SNR Scalable System", Proc. Of SPIE Visual Communications and Image Processing '95, vol. 259, Part 3/3, pp. 1380-1387, Taipei Taiwan, May 1995

Hidalgo J.R. and Salembier P., Robust segmentation and representation of foreground key-regions in video sequences, In IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP'2001, Vol. 3: IMDSP_L5, Salt Lake City, Utah, USA, May 2001.

Horn U., Girod B., "Performance Analysis of Multiscale Motion Compensation Techniques in Pyramid Coders", Proceedings of the IEEE International Conference on Image Processing ICIP'1996, Vol. III, pp. 255-258, Lausanne, September 1996.

Horn U., Girod B., "Pyramid Coding using Lattice Vector Quantization for Scalable Video Applications", Proc. PCS 1994, Sacramento, CA, 1994.

Horn U., Stuhlmüller K., Link M., Girod B., "Robust Internet video transmission based on scalable coding and unequal error protection", Signal Processing: Image Communication 15 Special Issue on Real-time Video over the Internet, (1999) pp. 77-94, 1999.

Houlding D., "Pyramid Coding for Image and Video Compression", Master thesis M.A.Sc., Engineering Science, Simon Fraser University, 1994.

Hsiang S.T., Woods J. W., "Embedded video coding using invertible motion compensated 3-D subband/wavelet filter bank", Signal Processing: Image Communication 16 (2001) 705–724, 2001.

Hsiang S.-T. and Woods J. W., "Embedded video coding using motion compensated 3-D subband/wavelet filter bank," Packet Video Workshop , Sardinia, Italy, May 2000.

Hsiang S.-T. and Woods J. W., "Invertible three-dimensional analysis/synthesis system for video coding with half-pixel-accurate motion compensation," Proc. SPIE 3653, Visual Communications and Image Processing'99, pp. 537-546, January 1999.

Huang H.-C., Wang C.-N., Chiang T. and Hang H.-M., "A Robust Fine Granularity Scalability (RFGS) Using Predictive Leak", ISO/IEC JTC1/SC29/WG11 MPEG 2002/M8409, May 2002.

Illgner K., Menser B., and Müller F., “Bewegungskompensation zur skalierbaren videocodierung”, in Proc. 9, Aachener Kolloquium "Signaltheorie": Bild- und Sprachsignale, pp. 63-68, Aachen, Germany, March 1997.

Illgner K. and Müller F., “Scalable video coding at very low bit rates employing resolution pyramids”, in VIIIth European Signal Processing Conference EUSIPCO'96, Vol. 2, pp 1343-1346, Trieste, Italy, September 1996.

Ishtiaq F. and Katsaggelos A.K., "A Rate Control Method for H.263 Temporal Scalability," Proc. 1999 IEEE International Conf. on Image Processing, Kobe, Japan, October 1999.

ISO/IEC JTC1/SC29/WG11 MPEG99/M5056, Melbourne, “Report of Ad hoc Group on Fine Granularity Scalability in MPEG-4 Video”, October 1999.

ISO-IEC/JTC1/SC29/WG11 N2605, “Report Of The Formal Verification Tests On MPEG-4 Temporal Scalability in Simple scalable Profile”, December 1998.

ISO-IEC/JTC1/SC29/WG11 N2823, “Report Of The Formal Verification Tests On MPEG-4 Temporal Scalability in Core Profile”, July 1999.

ISO/IEC JTC1/SC29/WG11 N2925, “Information Technology – Generic Coding Of Audio-Visual Objects: Visual ISO/IEC 14496-2 / Amd X Working Draft 2.0”, Melbourne, October 1999.

ISO/IEC JTC1/SC29/WG11 N2831, “AHG on Fine Granularity Scalability in MPEG-4 video”, July 1999.

ISO/IEC JTC1/SC29/WG11 N2926, “FGS Verification Model Version 2.0 Draft of 7 October, 1999” , Melbourne, October 1999.

ISO/IEC JTC1/SC29/WG11 N2937, “AHG on Fine Granularity Scalability in MPEG-4 video”, October 1999.

ISO/IEC JTC1/SC29/WG11 N3095, “Information Technology – Generic Coding Of Audio-Visual Objects: Visual ISO/IEC 14496-2 / Amd X Working Draft 3.0”, Maui, December 1999.

ISO/IEC JTC1/SC29/WG11 N3097, “FGS Verification Model Version 3.0 Draft of 9 December, 1999”, Maui, December 1999.

ISO/IEC JTC1/SC29/WG11 N3317 “FGS Verification Model Version 4.0 Draft of 11 April, 2000”, Noordwijkerhout, March, 2000.

ISO/IEC JTC1/SC29/WG11 N3528, “AHG on Fine Granularity Scalability in MPEG-4 video”, Beijing, July 2000.

ISO/IEC JTC1/SC 29/WG 11 N3670, “Study of ISO/IEC 14496-2 :1999/FPDAM4”, La Baule, October, 2000.

ISO/IEC JTC1/SC29/WG11 N3678, “AHG on Fine Granularity Scalability in MPEG-4 video”, La Baule, October 2000.

ISO/IEC JTC1/SC29/WG11 M8002, “Report on MPEG-4 Visual Fine Granularity Scalability Tools Verification Test”, JeJu, March 2002.

ISO/IEC JTC 1/SC 29/WG 11/N4750, “AHG on Advanced Fine Granularity Scalability”, Fairfax, VA, USA, May 2002.

Ito N., Baroncini V., “Revised test conditions and test plan for video verification test on Temporal Scalability in Core Profile”, ISO/IEC JTC1/SC29/WG11 MPEG98/N2601, December 1998.

Kawada R., Matsumoto S., „Flat multi-scalable coding for failure-free video transmission”, Proceedings of the IEEE International Conference on Image Processing ICIP’99, 1999.

Khansari M., Zakauddin A., Chan W., Dubois E., and Mermelstein P., "Approaches to layered coding for dual-rate wireless video transmission," in Proc. IEEE Int. Conf. Image Processing, pp. I-258-I-262, November 1994.

Kim K., Cheong W.-S., Park G. H. and Kim J., “Study on Advanced Fine Granularity Scalability in MPEG-4 video”, ISO/IEC JTC1/SC29/WG11 MPEG02/M8014, Jeju, March 2002.

Kodama M., Kasai H., Tominaga H., “The Hierarchical Image Coding and its Video Data Sescription Methods in Multimedia Scalability Packages” Proc. Of Picture Coding Symposium ‘99, Portland Oregon, pp.37-40, April 1999.

Kondi L. P. and Katsaggelos A. K., “An Operational Rate-Distortion Optimal Single-Pass SNR Scalable Video Coder”, IEEE Transactions On Image Processing, Vol. 10, No. 11, pp. 1613-1620, November 2001.

Kondi L. P. and Katsaggelos A.K., "An Optimal Single Pass SNR Scalable Video Coder," Proc. 1999 IEEE International Conf. on Image Processing, Kobe, Japan, October 1999.

Krishnamurthy R., Moulin P., and Woods J. W., "`Multiscale motion models for scalable video coding," Proc. IEEE Int. Conf. Image Processing, pp. I. 965-968, Lausanne, Switzerland, 1996.

Kunkelmann T., Horn U., “Partial Video Encryption Based on Scalable Coding“, Proc. of IWSSIP’1998, pp. 215-218, 1998.

Lee J.-Y., “A Novel Approach to Multi-Layer Coding of Video for Playback Scalability”, Ph.D. Dissertation, Dept. of Electronics Engineering, Graduate School, Korea University, June 1999.

Lee J.-Y., Kim T.-H., Ko S.-J., “Motion Prediction Based on Temporal Layering for Layered Video Coding” Proc. Of ITC-CSCC’98, Vol.1, July 1998.

Li W., “Result on Drift Prediction for FGS”, ISO/IEC JTC1/SC29/WG11 MPEG98/M5084, October 1999.

Li W., Ling F., Chen X., „Comparison of FGS with Simulcast”, ISO/IEC JTC1/SC29/WG11 MPEG98/M5083, October 1999.

Li W., Ling F., Li S., Wu F., Zhang Y.-Q., van der Schaar M., Jiang H., Chen X., Vetro A., Sun H., “Advanced Fine Granularity Scalability for High Quality Video Distribution”, ISO/IEC JTC1/SC29/WG11 M6766 Pisa, January 2001.

Lim C. S. and Tan T. K., “FGS Residue Presentation and Prediction”, ISO/IEC JTC1/SC29/WG11 MPEG99/M5324, December 1999.

Marquant G., Pateux S. and Labit C., “Mesh-Based Scalable Video Coding with Rate-Distortion Optimization”, VCIP 2000 (Visual Communications and Image Processing 2000), Perth, Australia, June 2000.

Mathew R., Arnold J.F., “Efficient layered video coding using data partitioning”, Signal Processing: Image Communication 14 (1999) pp. 761-782, 1999.

Moyano E., Quiles F.J., Garrido A., Orozco-Barbosa L., Duato J., “Efficient 3D Wavelet Transform Decomposition For Video Compression”, Proceedings of the Second International Workshop on Digital and Computational Video (DCV'01), February 2001.

Nakamura M., Sawada K., „Scalable Coding Schemes based on DCT and MC Predicition”, Proceedings of the IEEE International Conference on Image Processing ICIP’1995, Washington DC, vol. II, pp. 575-578, October 1995.

Nguyen A. and Dubois E., "Spatio-temporal adaptive interlaced-to-progressive conversion," in Signal Processing of HDTV, IV. Proc. International Workshop on HDTV'92, Kawasaki, Japan, November 1992 (Dubois E. and Chiariglione L., eds.), (Amsterdam), pp. 749-756, Elsevier, 1993.

Nicholls, J.A. and Monro, D.M., "Scalable Video By Software", In IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP’1996, Vol. 4, pp. 2005-2008, 1996.

Nicoulin A., Mattavelli M., Li W., Basso A., Popat A., Kunt M., Chapter 8. Image Sequence Coding Using Motion-Compensated Subband Decompsition” in “Motion analysis and image sequence processing” Edited by Ibrahim M., Kluwer Academic Publishers 1993.

Ohm J.-R., "3-D SBC-VQ with Motion Compensation" Proceedings Picture Coding Symposium (PCS'93), Lausanne, March 1993, pp. 11.5-1 - 11.5-2, 1993.

Ohm J.-R., "Layered VQ and SBC Techniques for Packet Video Applications," Proceedings 5th International Workshop on Packet Video, Berlin, March 1993, pp. B.4-1 -B.4-6, 1993.

Ohm J.-R., “Report of Ad hoc Group on Fine Granularity Scalability in MPEG-4 Video”, ISO/IEC JTC1/SC29/WG11 MPEG00/M6451, La Baule, October 2000.

Ohm J.-R., “Report of Ad hoc Group on Fine Granularity Scalability in MPEG-4 Video”, ISO/IEC JTC1/SC29/WG11 MPEG01/M6981, Singapore, March 2001.

Ohm J.-R., "Skalierbare Videocodierung für asynchrone und heterogene Datennetze," FREQUENZ 50 (1996), Nr. 9-10, Sept./Okt. 1996.

Ohm J.R., Benzler U., Wollborn M., “On the Requirements for Advanced Scalable Video Coding”, ISO/IEC JTC1/SC29/WG11 MPEG2001/7054, March 2001.

Ohm J.-R., Frater M., “Report of Ad hoc Group on Fine Granularity Scalability in MPEG-4 Video”, ISO/IEC JTC1/SC29/WG11 MPEG00/M6284, Beijing, July 2000.

Ohm J.-R., Li W., Kim K., „Report of Ad hoc Group on Fine Granularity Scalability in MPEG-4 Video”, ISO/IEC JTC1/SC29/WG11 MPEG02/M8206, Jeju, March 2002.

Ohm J.-R., Li W., Kim K., „Report of Ad hoc Group on Fine Granularity Scalability in MPEG-4 Video”, ISO/IEC JTC1/SC29/WG11 MPEG02/M8358, Fairfax, May 2002.

Park G. H., Cheong W.-S., Kim K., Lee Y. J., Lim Y. K., and Kim J., “Water Ring Scan Method for MPEG-4 and H.26L based FGS Methodologies”, ISO/IEC JTC1/SC29/WG11 MPEG2002/M8023, Jeju, March 2002.

Park G. H., Lim Y. K., Lee Y. J., Kim S. T., Yoon J. H., Lee M. H., and Ahn C., “Water ring scan order for improving subjective quality of MPEG-4 FGS”ISO/IEC JTC1/SC29/WG11 MPEG2000/m6183, Beijing, July 2000.

Piella G. and Heijmans H., "An adaptive update lifting scheme with perfect reconstruction ", Proceedings of the IEEE International Conference on Image Processing ICIP’2001, Thessaloniki, Greece, 2001.

Piella G. and Heijmans H., "Adaptive lifting schemes with perfect reconstruction", Report PNA-R0104, CWI, Amsterdam, February 2001.

Piella G., Heijmans H., Pesquet-Popescu B., "Adaptive wavelet decompositions driven by a weighted norm of the gradient" , accepted for publication in Proc. of the 3rd IEEE Benelux Signal Proc. Symposium (SPS-2002), Leuven, Belgium, March 2002.

Radha H., Chen Y., van der Schaar M., Thebault B., “Proposal for streaming profiles based on Fine-Granular-Scalability”, ISO/IEC JTC1/SC29/WG11 M5258, Melbourne, September 1999.

René J. van der Vleuten and Mihaela van der Schaar, “Quality Field for MPEG-4 Fine Granularity Scalability”, ISO/IEC JTC1/SC29/WG11 MPEG2000/M6483 La Baule, October 2000.

Robers M. A., “SNR Scalable Video Coder Using Progressive Transmission of DCT Coefficients”, Thesis, Nortwestern University 1997.

Robers M. A., Kondi L. P., and Katsaggelos A. K., "SNR Scalable Video Coder Using Progressive Transmission of DCT Coefficients," Proc. SPIE Conf. on Visual Communications and Image Processing, pp. 201-212, SPIE vol. 3309, San Jose, CA, January 1998.

Rose K. and Regunathan S. L., "Towards Optimal Scalability in Predictive Video Coding," Proceedings of the IEEE International Conference on Image Processing ICIP'1998, pp. 929-933, Chicago, October 98.

Rose K., Wu P. and Regunathan S. L., "Efficient SNR Scalability in Predictive Video Coding," Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing, May 1998.

Sawada K., Kinoshita T., „Subband-based scalable coding schemes with motion compensated prediction", Proc. Of SPIE Visual Communications and Image Processing '95, Taipei Taiwan vol. 259 Part 3/3 pp.1470-1477, May 1995.

Sawada K., Yoshida T., „Scalable Coding Based on Adaptive Subband for Interlaced Video Sequences", Proceedings of the IEEE International Conference on Image Processing ICIP'1997, vol. II, 1997.

Schafer R., Heising G., Smolic A., „Improving Image Compression Is It Worth the Effort?", Proc. of Eusipco, pp. 677-680, Tampere, 2000.

van der Schaar M., „Using S-Frames for fast switching between FGS streams and switching between MC-FGS structures to limit prediction-drift", ISO/IEC JTC1/SC29/WG11 M8140, Jeju, March 2002.

van der Schaar M., Chen Y. and Radha H., "Adaptive Quantization Modes for Fine-Granular Scalability", ISO/IEC JTC1/SC29/WG11 MPEG 99/M4938, July 1999.

van der Schaar M., Rajendran R. K., Chang S.-F., „FGS+: A framework for improved Joint Spatio-Temporal Video Quality of Fine Grained Scalable Coding", ISO/IEC JTC1/SC29/WG11 M8128, Jeju, March 2002.

Senbel S., Abdel-Wahab H., „Scalable and robust image compression using quadtrees", Signal Processing: Image Communication 14 (1999) pp. 425–442, 1999.

Shanableh T., "Pursing support for incorporating error resilience tools in the Simple Scalable Profile - Current status", ISO/IEC JTC1/SC29/WG11 MPEG2002/m8317, May 2002.

Shanableh T. and Ghanbari M., "The Importance of the Bi-Directionally Predicted Pictures in Video Streaming", IEEE Transactions On Circuits And Systems For Video Technology, Vol. 11, No. 3, pp. 402- 414, March 2001.

Shanableh T. and Hobson P., "A proposal for syntax amendment to use the error resilience tools in the MPEG-4 Simple Scalable Profile", ISO/IEC JTC1/SC29/WG11 MPEG2002/m7919, March 2002.

Shen K. and Delp E. J., "Wavelet Based Rate Scalable Video Compression," IEEE Transactions on Circuits and Systems for Video Technology, Vol. 9, No. 1, pp. 109-122, February 1999.

Sikora T., "Digital Video Coding Standards and Their Role in Video Communications", in Signal Processing for Multimedia J.S. Byrnes (Ed.), IOS Press, pp. 225-252, 1999.

Simon S., Bruyland I., "Layered and packetized subband coding for compatible digital HDTV", Image Processing: Theory and Applications, pp. 115-118, 1993.

Simon S. F. and Hürtgen B., "Optimale Quellencodierung für hierarchische Übertragung", in ITG-Fachbericht 130, Codierung für Quelle, Kanal und Übertragung, pp. 243-250, München, Germany, October 1994.

Stockhammer T., Buchner C., Marpe D., Blättermann G. and Heising G., "Efficient Fine Granular Scalable Video Coding " IEEE International Conference On Image Processing 2001, Vol. II, pp. 997-1000, Thessaloniki, Greece, October 2001.

Stockhammer T. and Weiss C., "Channel and Complexity Scalable Image Transmission", IEEE International Conference On Image Processing 2001, Vol. I, pp.102-105, Thessaloniki, Greece, October 2001.

Sun X., Wu F., Li S., Gao W., „The framework for seamless switching of scalable bitstreams", ISO/IEC JTC1/SC29/WG11, MPEG2002/8214, Jeju Island, March 2002.

Sun X., Wu F., Li S., Gao W. and Zhang Y.-Q., "Macroblock-based progressive fine granularity scalable video coding", IEEE International Conference on Multimedia and Expo (ICME), 461-464, Tokyo, August, 2001.

Sun X., Wu F., Li S., Gao W. and Zhang Y.-Q., "Seamless switching of scalable video bitstreams for efficient streaming", Proc. of ISCAS2002, (accepted), 2002.

Tan W., Chang E., and Zakhor A., "Real Time Software Implementation of Scalable Video Codec", Proceedings of the IEEE International Conference on Image Processing ICIP'1996, Vol. 1, pp. 17-20, September 1996.

Tan T. K., Ngan K. N., "A Frequency Scalable Coding Scheme Employing Pyramid and Subband Techniques", IEEE Transactions on Circuits and Systems for Video Technology, vol. 4, No. 2, pp. 203-207, April 1994.

Tan W. and Zakhor A., "Internet Video using Error Resilient Scalable Compression and Cooperative Transport Protocol", Proceedings of the IEEE International Conference on Image Processing ICIP'1998, Vol. 3, pp. 458-462, October 1998.

Tan W. and Zakhor A., "Resilient Compression of Video for Transmission over the Internet", Proc. 32nd Asilomar Conf. Signal, Sys. and Computers, November 1998.

Taubman D., Ordentlich E., Weinberger M., Seroussi G., "Embedded block coding in JPEG 2000", Signal Processing: Image Communication 17 (2002) 49–72, 2002.

Thayer G. and Jiang H., “Results for Hybrid FGS-Temporal Scalability”, ISO/IEC JTC1/SC29/WG11 MPEG2000/M5939, March 2000.

Vass J., Chai B.-B., and Zhuang X., "3DSLCCA - A highly scalable very low bit rate software-only wavelet video codec," in Proceedings of IEEE Workshop on Multimedia Signal Processing, Los Angeles, CA, December 1998.

Vetro A., Sun H., DaGraca P., and Poon T., “Minimum drift architectures for three-layer scalable DTV decoding," IEEE Trans. Consumer Electronics, vol. 44, no. 3, August 1998.

Vetro A., Sun H. and Wang Y., “Object-based transcoding for scalable quality of service," Proc. IEEE International Symposium on Circuits and System, Geneva, Switzerland, pp. IV-17 IV-20, May 2000.

Wang Q., Wu F., Li S., Xiong Z., Zhang Y.-Q., Zhong Y., “A new rate allocation scheme for progressive fine granular scalable coding”, IEEE International Symposium on Circuits and Systems (ISCAS), vol 2, pp. 397-400, Sydney, May, 2001.

Wang Q., Wu F., Li S., Zhong Y., Zhang Y.-Q., "Fine-granularity spatially scalable video coding", 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Volume: 3, pp. 1801 -1804, 2001.

Wang Q., Xiong Z., Wu F., and Li S., “Optimal rate allocation for progressive fine granularity scalable video coding," IEEE Signal Processing Letters, vol. 9, pp. 33-39, February 2002.

Wang Q., Xiong Z., Wu F., and Li S., ``Optimal rate allocation for progressive fine granularity scalable video coding," Proc. of International Conference on Coding and Computing ITCC'01, Las Vegas, NV, April 2001.

Wang Z., Lu L., and Bovik A. C., "Rate scalable video coding using a foveation-based human visual system model," Proc. of IEEE Inter. Conference on Acoustics, Speech, & Signal Processing, May 2001.

Wee S.J., Polley M.O., and Schreiber W.F., “A Generalized Framework for Scalable Video Coding”, Proceedings of the International Symposium on Multimedia Communications and Video Coding, Brooklyn, NY, October 1995.

Wee S.J., Polley M.O., and Schreiber W.F., “A Scalable Source Coder for a Hybrid HDTV Terrestrial Broadcasting System”, IEEE International Conference on Image Processing, Volume 1, pp. 238-242, Austin, TX, November 1994.

Wiegand T., Zhang X., and Girod B., "Block-Based Hybrid Video Coding Using Motion-Compensated Long-Term Memory Prediction", in Proc. Picture Coding Symposium, Berlin, Germany, September 1997.

Wien M. and Brünig M., “Optimized bit allocation for scalable wavelet video coding”, in Proc. SPIE Image and Video Communications and Processing 2000, Vol. 3974, pp. 153-160, San Jose, California, USA, January 2000.

Wien M. and Menser B., “Adaptive scalable video coding using wavelet packets”, in Proc. SPIE Visual Communications and Image Processing 2000, Vol. 4067, pp. 1281-1287, Perth, Australia, June 2000.

Wu F., Li S., Sun X., Yan R., Zhang Y.-Q., “Comparisions between the one-loop and two-loop solutions for improving the coding efficiency of FGS”, The second workshop and exhibition on MPEG-4, 79-82, San Jose, June, 2001.

Wu F., Li S., Zeng B., Zhang Y.-Q., “Drifting Reduction in Progressive Fine Granular Scalable Video Coding”, Picture Coding Symposium (PCS), Seoul, April 2001.

Wu F., Li S. and Zhang Y.-Q., "DCT-Prediction Based Progressive Fine Granularity Scalable Coding", Proceedings of the IEEE International Conference on Image Processing ICIP’2000, Vancouver, September 2000.

Wu F., Li S., Zhang Y.-Q., “Progressive Fine Granularity Scalable (PFGS) video using Advance-Predicted Bitplane Coding (APBIC)”, IEEE International Symposium on Circuits and Systems (ISCAS), vol 5, 97-100, Sydney, May, 2001.

Xu J., Li S. and Zhang Y.-Q., "A Wavelet Codec Using 3-D ESCOT" IEEE-PCM2000, December 2000.

Xu J., Wu F., „Cross-check on the Robust Fine Granularity Scalable coding scheme”, ISO/IEC JTC1/SC29/WG11 MPEG2002/M8427, Fairfax, VA, May 2002.

Yan S., Wu F., Li S., „Comparisons between PFGS and JVT SP”, ISO/IEC JTC1/SC29/WG11 MPEG2002/M8426, Fairfax, VA, May 2002.

Yan R., Wu F., Li S. and Tao R., "Error Resilience Methods for FGS Video Enhancement Bitstream", The First IEEE Pacific-Rim Conference on Multimedia (IEEE-PCM 2000), Sydney, Australia, December 2000.

Yan R., Wu F., Li S., Tao R., Wang Y., “Efficient Video Coding with Hybrid Spatial and Fine-Grain SNR Scalabilities”, SPIE Visual Communications and Image Processing, 2002.

Yan R., Wu F., Li S., Tao R., Wang Y. and Zhang Y.-Q., "Error robust coding for the FGS enhancement bitstream", Third International Conference on Information, Communications, Singapore, October, 2001.

Yoon S. H., Chellapilla K., and Rao S., "Wavelet Based Very Low Bit Scalable Video Coding using Bitplane Encoding," 1997 Proc. Of IEEE International Symposium on Circuits and Systems, Hong Kong, June 1997.

Zafar S., Zhang Y.-Q., Jabbari B., “Multiscale Video Representation Using Multiresolution Motion Compensation and Wavelet Decomposition”, IEEE Journal on Selected Areas in Communications, Vol. 11, No. 1, January 1993.

Zeng W., Li J. and Lei S., "Adaptive wavelet transforms with spatially varying filters for scalable image coding," Proceedings of the IEEE International Conference on Image Processing ICIP'1998, Chicago, 1998.

Zhang R., Regunathan S. L. and Rose K., "Optimal Estimation for Error Concealment in Scalable Video Coding," Thirty-Fourth Asilomar Conference on Signals, Systems and Computers, October 2000.

Zhang R., Regunathan S. L. and Rose K., "Scalable Video Coding with Robust Mode Selection, " Packet Video Workshop 2000, May 2000.

Zhang R., Regunathan S. L. and Rose K., "Switched Error Concealment and Robust Coding Decisions in Scalable Video Coding," Proceedings of the IEEE International Conference on Image Processing ICIP'2000, Vancouver, Sept. 2000.

Zhang T. and Xu Y., "General modelling of layered packet video for multicast systems", in Proceedings of SCI'99 and ISAS'99, pp.261-268, Orlando, Florida, USA, July - August 1999.

Zhang T. and Xu Y., "Unequal Packet Loss Protection for Layered Video Transmission", in IEEE Transaction on Broadcasting, Vol. 45, No. 2, June 1999.

Zhang, Ya-Qin "Scalable Wavelet Coding for Synthetic/Natural Hybrid Images" IEEE Transactions on Circuits and Systems for Video Technology, Vol.9, No.2, pp. 244-254, March 1999.

Annex A

Video bitstream syntax

A.1. Intraframe mode

This section presents the comparison of bitstream syntax between the non-scalable and scalable bitstream for the intraframe mode of coding. The analysis shows that the scalable bitstream syntax is more efficient than the non-scalable bitstream syntax in the intraframe mode.

The terminology is used in accordance with the MPEG-2 standard bitstream syntax [ISO94, Hask96].

Non-scalable coding:

a) The minimal requirement in bits for bitstream syntax description per macroblock for non-scalable coding of a full resolution BT.601 progressive sequence is as follows:

- 1 bit for the *macroblock_address_increment*. This variable-length coded integer indicates the difference in the position of consecutive macroblocks. Since in the intraframe mode all macroblocks have to be transmitted, the minimum value of *macroblock_address_increment* is 1.

- 2 bits for *macroblock_type*. This variable-length integer indicates the method of coding and the content of the macroblock according to Tables B.2 through B.8 in the MPEG-2 standard [ISO94, Hask96]. For intraframe coding with quantization, 2 bits are transmitted.
- 5 bits for *quantiser_scale_code*. It is a constant length integer.
- 3 bits for *dct_dc_size_luminance* for 4 blocks of the luminance component. DC coefficients in blocks in intra macroblocks are encoded as a variable length code denoted *dct_dc_size*, as defined in Tables B.12 and B.13 in the MPEG-2 standard [ISO94, Hask96]. 3 bits per block are the minimum required number of bits to be transmitted. It means that 12 bits per macroblock are transmitted.
- 2 bits for *dct_dc_size_chrominance* for 2 blocks of the chrominance component in the 4:2:0 chroma format. 2 bits per block are the required minimum number of bits to be transmitted. It means that 4 bits per macroblock are transmitted.
- All AC coefficients may be variable-length encoded, using Tables B.14, B.15 and B.16. Let us assume, however, that AC coefficients are not encoded. Only the *end_of_block* indicator is sent. It means that 2 bits per block are required to be transmitted. In total, 12 bits per macroblock are required to transmit *end_of_block*.

In a 704 x 576 resolution frame there are 1584 macroblocks of 16 x 16 pixels. Since 36 bits per macroblock are required for transmission, this is minimum 57024 bits per frame for intraframe non-scalable coding.

Scalable coding:

b) The scalable coding of a full resolution BT.601 sequence requires the minimal following syntax description per macroblock, in bits:

- For the base layer:
  - 1 bit for the *macroblock_address_increment*.
  - 2 bits for *macroblock_type*.
  - 5 bits for *quantiser_scale_code*.

- 3 bits for *dct_dc_size_luminance* for 4 blocks of the luminance component. 3 bits per block are the required minimum number of bits to be transmitted. It means that 12 bits per macroblock are transmitted.
- 2 bits for *dct_dc_size_chrominance* for 2 blocks of the chrominance component in the 4:2:0 chroma format. 2 bits per block are the required minimum number of bits to be transmitted. It means that 4 bits per macroblock are transmitted.
- Let us assume that we do not encode AC coefficients are not encoded. Only the *end_of_block* indicator is sent. It means that 2 bits per block have to be transmitted. In total, 12 bits per macroblock are required to transmit *end_of_block*.

In the base layer a 352 x 288 resolution frame is encoded. There are 396 macroblocks of 16 x 16 pixels. Since 36 bits per macroblock must be transmitted in the bitstream, this is minimum 14256 bits per frame for the base layer encoding.

- For the enhancement layer:
  - 1 bit for the *macroblock_address_increment*. It is the extreme case because in the enhancement layer frames are encoded in the interframe mode. It means that not all macroblocks must be encoded. For simplicity of calculation however, let us assume that all the macroblocks are encoded.
  - 1 bit for *macroblock_type*.
  - 5 bits for *quantiser_scale_code*.
  - 3 bits for *pattern_code*. This integer indicates which blocks in the current macroblock are encoded. 3 bits per macroblock are required to indicate all encoded blocks.
  - In the interframe mode it is not possible to transmit *dct_dc_size_luminance* and *dct_dc_size_chrominance*. All coefficients are encoded differentially. Let us assume that coefficients are not encoded. Only the *end_of_block* indicator is sent. It means that 2 bits

per block must be transmitted. In total, 12 bits per macroblock are required to transmit *end_of_block*.

In the enhancement layer a 704 x 576 resolution frame is encoded. There are 1584 macroblocks of 16 x 16 pixels. Since 22 bits per macroblock must be transmitted in the bitstream, this is minimum 34848 bits per frame for the enhancement layer encoding.

In total, 49104 bits are required to transmit minimum syntax in the scalable encoding in the intraframe mode.

A.2. Interframe mode

On a macroblock basis, in the improved prediction there are seven possibilities to choose from. In the extreme case, this kind of prediction requires transmitting an additional bit per macroblock to identify the selected mode of prediction.

In the MPEG-2 standard, the mode of prediction is indicated by the *macroblock_type* which is variable length encoded. The value of this indicator is constructed from several flags: *macroblock_quant, macroblock_motion_forward, macroblock_motion_backward, macroblock_pattern, macroblock_intra, spatial_temporal_weight_code_flag, permitted_ spatial_temporal_weight_classes*. The *macroblock_motion_forward* and *macroblock_motion_backward* flags are used to indicate the type of prediction in the MPEG-2 standard. Since the enhancement layer is not fully compatible with the MPEG-2 standard, an additional bit was inserted in the macroblock header in the syntax of the enhancement layer bitstream.

In order to decrease the number of bits of the *macroblock_type* indicator, a new variable length code should be calculated.

Below, the prediction modes and appropriate flags are presented.

1) forward prediction from the previous reference frame RP (I- or P-frame) is indicated by setting the *macroblock_motion_forward* flag,
2) backward prediction from the next reference frame RN (I- or P-frame) is indicated by setting the *backward_motion_forward* flag,
3) bi-directional prediction from the previous and next reference frame is indicated by setting the *macroblock_motion_forward* and *backward_motion_forward* flags,

4) interpolation from the current reference frame RC (spatially up-sampled frame from the base layer) is indicated by setting the *macroblock_interpolation* flag,
5) the average of forward prediction from the previous reference frame RP and interpolation from the current reference frame RC is indicated by setting the *macroblock_motion_forward* and *macroblock_interpolation* flags,
6) the average of the backward prediction from the next reference frame RN and interpolation from the current reference frame RC is indicated by setting the *macroblock_motion_backward* and *macroblock_interpolation* flags,
7) the average of the backward prediction from the next reference frame RN, forward prediction from the previous reference frame RP, and interpolation from the current reference frame RC is indicated by setting the *macroblock_motion_forward, macroblock_motion_backward* and *macroblock_interpolation* flags.

Annex B

Verification model

B.1. Introduction

The creation of a very efficient tool to research the author's concept of a scalable video coding was one of the main goals of the work.

Since the author wanted to achieve flexibility of the software, clarity of the source code and possibility of fast modification of the parameters of the encoder, the procedural writing of the source code was rejected and the implementation was created in an object oriented language instead.

The construction of the verification model of the scalable encoder is based on universal software [Lucz98] which was created in the Laboratory of the Division of Multimedia Telecommunications and Radioelectronics of Poznań University of Technology. The software is used to implement the video coders such as H.263 [ITUT96, Tura00] and MPEG-2. This software was chosen because in all solutions proposed by the author the base layer is compatible with the MPEG-2 standard and the whole structure shows a high level of compatibility with the particular blocks of the MPEG-2 coder.

B.2. Language

In the earlier phase of research, the implementation was written in the CC variant of C++ language. GNU CC is the most popular object oriented C language compiler in UNIX. This compiler was chosen because of easy transfer of the source code into different environments and computer platforms. The compilation of the same source code is possible on the SUN and PC platforms, and under Solaris and Windows NT system. However, high technology offered by the Microsoft Visual C++ let us construct a more efficient implementation of the scalable video coder. Consequently, the verification model has been prepared as software written in Microsoft Visual C++. At present, the software consists of about 14000 lines of object code.

B.3. Features of object oriented programming languages

For the sake of brevity, only the basic features of the encoder are described. The actual coding algorithm is very sophisticated. The software prepared is enormous. In order to achieve flexibility in changing the parameters, standard MPEG-2 software [Mpeg96] was not used. The MPEG-2 coder was replaced by new implementation.

One of the main advantages of using an object oriented programming language to write a source code of a scalable encoder is the possibility of arranging objects in a structure which is directly taken from the general diagram of this coder. The diagram of the data flow in the encoder is translated into the main part of the software. This part of software consists only of the main operations of the algorithm. Since it is necessary to ensure synchronous work of both encoders (base and enhancement layer encoders) which jointly utilize the same input data, successive call of object methods is implemented. A fragment of the source code listing is presented in Fig. B.1.

The purpose of the experiments is to compare various encoder structures, establish their properties and find the best structure. Therefore the software is not optimized for run time. The most important feature is its flexibility, allowing tests of different variants of the coding algorithm.

```
.
.
.

// Taking an image from the input sequence.


B0.Get(pn,input);


// Spatial decimation of the input image.


qmfy.Split(B0.Y,LL.Y,LH.Y,HL.Y,HH.Y);
qmfc.Split(B0.U,LL.U,LH.U,HL.U,HH.U);
qmfc.Split(B0.V,LL.V,LH.V,HL.V,HH.V);

// Coding the image in the base layer.


BaseLayer.EncodePFrame(LL,31);
RLL = BaseLayer.Reconstructed;

// Spatial interpolation of the output image obtained from the base layer encoder.


LH  = 0;
HL  = 0;
HH  = 0;

qmfy.Merge(RLL.Y,LH.Y,HL.Y,HH.Y,B1.Y);
qmfc.Merge(RLL.U,LH.U,HL.U,HH.U,B1.U);
qmfc.Merge(RLL.V,LH.V,HL.V,HH.V,B1.V);

// Modified prediction and motion compensation in the enhancement layer.


mep.MPEG2PCompensation(B2,B0,T0,31);
mep.MPEG2PCompensation(T2,B0,T1,31);
interpolacja(B0,T0,B1,T1,T2);

// Coding the image in the enhancement layer.


EnhP.MakeBFrame(pn,B0,T2,T0,T1,PB);


// Calculation of PSNR of the images decoded from the base and enhancement layers.


psnrBY = LL.Y.PSNR(RLL.Y);
psnrBU = LL.U.PSNR(RLL.U);
psnrBV = LL.V.PSNR(RLL.V);

psnrY = B0.Y.PSNR(PB.Y);
psnrU = B0.U.PSNR(PB.U);
psnrV = B0.V.PSNR(PB.V);
.
.
.
```

Fig. B.1. Fragment of the listing of the scalable video encoder.

B.4. Verification

Correct operation of the encoder was verified during several tests, consisting in comparing selected blocks of the scalable encoder with the MPEG-2 verification model [Mpeg96]. The values of PSNR of selected images and their bitrate were tested. The software library also provides an implementation of the MPEG-2 encoder, which has been cross-checked with the MPEG-2 verification model [Mpeg96].

The progress of the encoding is reported by several indicators which are stored. The information, warnings and errors are reported on the screen and in the report file.

The described schemes of the encoders proposed by the author were examined with standard test sequences. Their luminance format is progressive BT.601 [ITUR], i.e. 720 x 576 pixels and 50 frames per second. The chrominance subsampling scheme is 4:2:0. It means that the base layer provides SIF frames.

B.5. Specification of input and output data

B.5.1. Verification model input data

The input data are:

- progressive 4:2:0 video sequences CIF(352x288), 4CIF(704x576),
- configuration file.

The verification model input configuration file describes the parameters and functionalities of the encoder. The file has a hierarchical structure and its content can be extended if required. The configuration file is divided into sections. The beginning of each section is indicating by the following header:

```
[NAME:NAME]
{
…
// body of section
…
};
```

The body of a sections consist of variables whose names are preceded by '$', Variables can be strings of chars, integers, floats, binary numbers and hexadecimal numbers.

Example:

```
[SECTION:EXAMPLE]
{
 $IntraQuant      12;              // integer
 $XParam          7.121;           // float
 $FileName        "file.cif";      // string
 $ESCAPECode      0b000011;        // binary number
 $Size            352,288;         // table of integers

$Matrix           10,10,20,30;     //
                  20,30,40,50;     // matrix of integers [4,2]
};
```

Sections can include sub-sections, which constitute the hierarchical structure of the configuration file.

```
[SECTION:ONE]
{
 [SECTION:ONE_A]
 {
  …
 };

 [SECTION:ONE_B]
 {
        …
 };

};

SECTION:TWO]
{
…
};
```

Example of a fragment of verification model input configuration file of the scalable encoder:

```
[MPEG2STREAM:BASE_ENHP]
{
  $BitRate                    3700000;
  $FrameRate                  50;
  $VBVBufferSize              128000;

  $Profile                    MAIN;
  $Level                      MAIN;
  $VideoFormat                "PAL";
  $ColourPrimaries            "ITUR_BT470_2";
```

```
$ProgressiveSequence         1;
$DCPrecision                 8;

$IQName                      "INTRA_QUANT_TABLE";
$NIQName                     "NON_INTRA_QUANT_TABLE";

$TimeCodeHours               0;
$TimeCodeMinutes             0;
$TimeCodeSeconds             0;
$TimeCodePictures            0;

$TransferCharacteristics     5;
$MatrixCoefficients          5;
$AspectRatio                 2;
$RepeatFirst                 0;

$StartIQuant                 4;
$StartPQuant                 11;
$StartBQuant                 10;

$FactorI                     120;
$FactorP                     80;
$FactorB                     120;

$PFrameNumber                3;
$BFrameNumber                3;

  [VLC:MACROBLOCK_ADRESS_INCREMENT]
    {
   $COUNT       33;
   $DIMENSION   1;
   $VECTORS
                  1 , 0b1;
                  2 , 0b011;
                  3 , 0b010;
                  4 , 0b0011;
                  5 , 0b0010;
                  6 , 0b00011;
                  7 , 0b00010;
                  8 , 0b0000111;
                  9 , 0b0000110;
                 10 , 0b00001011;
                 11 , 0b00001010;
                 12 , 0b00001001;
                 13 , 0b00001000;
                 14 , 0b00000111;
                 15 , 0b00000110;
                 16 , 0b0000010111;
                 17 , 0b0000010110;
                 18 , 0b0000010101;
                 19 , 0b0000010100;
                 20 , 0b0000010011;
                 21 , 0b0000010010;
                 22 , 0b00000100011;
                 23 , 0b00000100010;
                 24 , 0b00000100001;
                 25 , 0b00000100000;
                 26 , 0b00000011111;
                 27 , 0b00000011110;
                 28 , 0b00000011101;
                 29 , 0b00000011100;
```

```
                    30 , 0b00000011011;
                    31 , 0b00000011010;
                    32 , 0b00000011001;
                    33 , 0b00000011000;
    }
.
.
.
}
```

B.5.2. Verification model output data

The output signals are:

- base and enhancement layer bitstreams,
- the report files (See Fig. B.1).

The enhancement layer bitstream is almost compatible with the MPEG-2 scalable profile. The base layer bitstream is 100% compatible with the MPEG-2 Main Profile @ Main Level.

```
>> I   0 : ..   5 Base( Y-36.004[dB] ,U-39.7[dB] ,V-39.3[dB] ,STR-201920[bpf])  5| 4| 7| 4| 31160| 38280| 24648|  4224| Enh( Y-33.464[dB] ,U-38.4[dB] ,V-38.3[dB] ,STR- 98312[bpf],  48841,  54639)
>> P   4 : ..   7 Base( Y-34.666[dB] ,U-38.1[dB] ,V-37.7[dB] ,STR-107952[bpf]) 14|  |  |  |      |      |      |      | Enh( Y-32.688[dB] ,U-37.7[dB] ,V-37.4[dB] ,STR- 54688[bpf],  35520,  20460)
>>  B   2 : .  10 Base( Y-33.172[dB] ,U-38.2[dB] ,V-37.8[dB] ,STR- 48056[bpf]) 14|  |  |  |      |      |      |      | Enh( Y-32.721[dB] ,U-37.8[dB] ,V-37.7[dB] ,STR- 52216[bpf],  30888,  22620)
>>  B   1 : .                                                                  13|  |  |  |      |      |      |      | Enh( Y-32.942[dB] ,U-37.8[dB] ,V-37.8[dB] ,STR-103152[bpf],  63309,  41135)
>>  B   3 : .                                                                  14|  |  |  |      |      |      |      | Enh( Y-32.666[dB] ,U-37.5[dB] ,V-37.4[dB] ,STR- 91520[bpf],  58201,  34611)
>> P   8 : ..   8 Base( Y-33.672[dB] ,U-37.4[dB] ,V-36.9[dB] ,STR- 98048[bpf]) 13|  |  |  |      |      |      |      | Enh( Y-32.670[dB] ,U-37.3[dB] ,V-37.2[dB] ,STR- 78208[bpf],  50053,  29447)
>>  B   6 : .  11 Base( Y-32.367[dB] ,U-37.4[dB] ,V-36.9[dB] ,STR- 45936[bpf]) 15|  |  |  |      |      |      |      | Enh( Y-32.342[dB] ,U-37.3[dB] ,V-37.2[dB] ,STR- 64928[bpf],  45815,  20405)
>>  B   5 : .                                                                  14|  |  |  |      |      |      |      | Enh( Y-32.514[dB] ,U-37.3[dB] ,V-37.2[dB] ,STR- 92848[bpf],  57087,  37053)
>>  B   7 : .                                                                  15|  |  |  |      |      |      |      | Enh( Y-32.408[dB] ,U-37.2[dB] ,V-37.1[dB] ,STR- 82504[bpf],  53877,  29919)
>> I  12 : ..   5 Base( Y-36.084[dB] ,U-40.0[dB] ,V-39.7[dB] ,STR-199296[bpf])  4| 5| 6| 5| 67072| 26160| 30632|  2168| Enh( Y-34.134[dB] ,U-39.1[dB] ,V-39.1[dB] ,STR-126032[bpf],  51332,  79868)
>>  B  10 : .  10 Base( Y-33.085[dB] ,U-38.1[dB] ,V-37.6[dB] ,STR- 48488[bpf]) 16|  |  |  |      |      |      |      | Enh( Y-32.535[dB] ,U-38.0[dB] ,V-37.9[dB] ,STR- 53200[bpf],  40329,  14163)
>>  B   9 : .                                                                  15|  |  |  |      |      |      |      | Enh( Y-32.312[dB] ,U-37.3[dB] ,V-37.3[dB] ,STR- 80592[bpf],  53259,  28625)
>>  B  11 : .                                                                  14|  |  |  |      |      |      |      | Enh( Y-33.097[dB] ,U-38.2[dB] ,V-38.2[dB] ,STR- 84936[bpf],  55236,  30992)
>> P  16 : ..   9 Base( Y-33.181[dB] ,U-37.6[dB] ,V-37.1[dB] ,STR- 82552[bpf]) 12|  |  |  |      |      |      |      | Enh( Y-33.061[dB] ,U-37.9[dB] ,V-37.8[dB] ,STR- 94824[bpf],  55622,  40494)
>>  B  14 : .   9 Base( Y-33.560[dB] ,U-38.5[dB] ,V-37.9[dB] ,STR- 55552[bpf]) 15|  |  |  |      |      |      |      | Enh( Y-32.785[dB] ,U-38.1[dB] ,V-38.0[dB] ,STR- 56736[bpf],  42316,  15712)
>>  B  13 : .                                                                  14|  |  |  |      |      |      |      | Enh( Y-32.930[dB] ,U-38.2[dB] ,V-38.2[dB] ,STR- 84400[bpf],  54363,  31329)
>>  B  15 : .                                                                  13|  |  |  |      |      |      |      | Enh( Y-33.106[dB] ,U-37.8[dB] ,V-37.9[dB] ,STR- 81976[bpf],  46396,  36872)
>> P  20 : ..   8 Base( Y-33.567[dB] ,U-37.4[dB] ,V-36.9[dB] ,STR- 97592[bpf]) 13|  |  |  |      |      |      |      | Enh( Y-32.759[dB] ,U-37.4[dB] ,V-37.4[dB] ,STR- 76576[bpf],  49856,  28012)
>>  B  18 : .   8 Base( Y-33.838[dB] ,U-37.8[dB] ,V-37.4[dB] ,STR- 65528[bpf]) 12|  |  |  |      |      |      |      | Enh( Y-33.306[dB] ,U-37.9[dB] ,V-37.9[dB] ,STR- 80088[bpf],  52538,  28842)
>>  B  17 : .                                                                  11|  |  |  |      |      |      |      | Enh( Y-33.575[dB] ,U-37.9[dB] ,V-38.1[dB] ,STR-114056[bpf],  63279,  52069)
>>  B  19 : .                                                                  12|  |  |  |      |      |      |      | Enh( Y-33.329[dB] ,U-37.7[dB] ,V-37.8[dB] ,STR- 99600[bpf],  58693,  42199)
>> I  24 : ..   4 Base( Y-37.249[dB] ,U-40.7[dB] ,V-40.5[dB] ,STR-233448[bpf])  4| 6| 6| 4| 44912| 19176| 32056|  4160| Enh( Y-34.040[dB] ,U-39.3[dB] ,V-39.4[dB] ,STR-100304[bpf],  50251,  55221)
>>  B  22 : .   9 Base( Y-33.846[dB] ,U-38.8[dB] ,V-38.4[dB] ,STR- 50360[bpf]) 13|  |  |  |      |      |      |      | Enh( Y-33.363[dB] ,U-38.4[dB] ,V-38.4[dB] ,STR- 68640[bpf],  47410,  22522)
>>  B  21 : .                                                                  14|  |  |  |      |      |      |      | Enh( Y-32.727[dB] ,U-37.6[dB] ,V-37.7[dB] ,STR- 80616[bpf],  52050,  29858)
>>  B  23 : .                                                                  15|  |  |  |      |      |      |      | Enh( Y-32.962[dB] ,U-38.4[dB] ,V-38.4[dB] ,STR- 74496[bpf],  49341,  26447)
>> P  28 : ..   7 Base( Y-34.787[dB] ,U-38.4[dB] ,V-38.1[dB] ,STR-106240[bpf]) 12|  |  |  |      |      |      |      | Enh( Y-33.248[dB] ,U-38.0[dB] ,V-38.1[dB] ,STR- 80976[bpf],  51048,  31220)
>>  B  26 : .   8 Base( Y-34.390[dB] ,U-39.0[dB] ,V-38.7[dB] ,STR- 58184[bpf]) 16|  |  |  |      |      |      |      | Enh( Y-32.751[dB] ,U-38.2[dB] ,V-38.3[dB] ,STR- 47712[bpf],  37937,  11067)
>>  B  25 : .                                                                  15|  |  |  |      |      |      |      | Enh( Y-32.826[dB] ,U-38.4[dB] ,V-38.5[dB] ,STR- 70600[bpf],  47261,  24631)
>>  B  27 : .                                                                  14|  |  |  |      |      |      |      | Enh( Y-32.860[dB] ,U-37.9[dB] ,V-38.0[dB] ,STR- 79720[bpf],  51364,  29648)
>> P  32 : ..   6 Base( Y-35.517[dB] ,U-38.3[dB] ,V-38.0[dB] ,STR-138304[bpf]) 12|  |  |  |      |      |      |      | Enh( Y-33.290[dB] ,U-37.8[dB] ,V-37.9[dB] ,STR- 81344[bpf],  54081,  28555)
>>  B  30 : .   8 Base( Y-34.177[dB] ,U-38.3[dB] ,V-37.9[dB] ,STR- 61472[bpf]) 13|  |  |  |      |      |      |      | Enh( Y-33.113[dB] ,U-37.9[dB] ,V-38.0[dB] ,STR- 73640[bpf],  53817,  21115)
>>  B  29 : .                                                                  12|  |  |  |      |      |      |      | Enh( Y-33.206[dB] ,U-37.9[dB] ,V-38.0[dB] ,STR- 97824[bpf],  56636,  42480)
>>  B  31 : .                                                                  13|  |  |  |      |      |      |      | Enh( Y-32.973[dB] ,U-37.6[dB] ,V-37.8[dB] ,STR- 80608[bpf],  44381,  37519)
>> I  36 : ..   4 Base( Y-37.170[dB] ,U-40.5[dB] ,V-40.4[dB] ,STR-239880[bpf])  3| 5| 7| 3| 99648| 26008| 28296|  9920| Enh( Y-34.443[dB] ,U-39.7[dB] ,V-40.1[dB] ,STR-163872[bpf],  59999, 109041)
>>  B  34 : .   8 Base( Y-34.456[dB] ,U-38.9[dB] ,V-38.6[dB] ,STR- 63928[bpf]) 14|  |  |  |      |      |      |      | Enh( Y-33.173[dB] ,U-38.5[dB] ,V-38.7[dB] ,STR- 66888[bpf],  51310,  16870)
>>  B  33 : .                                                                  13|  |  |  |      |      |      |      | Enh( Y-32.936[dB] ,U-37.7[dB] ,V-37.9[dB] ,STR- 90128[bpf],  53360,  38060)
>>  B  35 : .                                                                  14|  |  |  |      |      |      |      | Enh( Y-33.126[dB] ,U-38.6[dB] ,V-38.9[dB] ,STR- 85944[bpf],  52922,  34314)
>> P  40 : ..   7 Base( Y-34.574[dB] ,U-38.2[dB] ,V-37.9[dB] ,STR-119344[bpf]) 12|  |  |  |      |      |      |      | Enh( Y-33.098[dB] ,U-38.0[dB] ,V-38.1[dB] ,STR- 88224[bpf],  53904,  35612)
>>  B  38 : .   9 Base( Y-33.556[dB] ,U-38.4[dB] ,V-38.2[dB] ,STR- 60096[bpf]) 15|  |  |  |      |      |      |      | Enh( Y-32.569[dB] ,U-38.1[dB] ,V-38.5[dB] ,STR- 59824[bpf],  44024,  17092)
>>  B  37 : .                                                                  14|  |  |  |      |      |      |      | Enh( Y-32.823[dB] ,U-38.4[dB] ,V-38.7[dB] ,STR- 79584[bpf],  45582,  35294)
>>  B  39 : .                                                                  14|  |  |  |      |      |      |      | Enh( Y-32.617[dB] ,U-37.7[dB] ,V-38.1[dB] ,STR- 93328[bpf],  55847,  38773)
>> P  44 : ..   8 Base( Y-33.514[dB] ,U-37.3[dB] ,V-36.9[dB] ,STR-107984[bpf]) 13|  |  |  |      |      |      |      | Enh( Y-32.456[dB] ,U-37.3[dB] ,V-37.4[dB] ,STR- 87048[bpf],  55025,  33315)
>>  B  42 : .  10 Base( Y-32.718[dB] ,U-37.4[dB] ,V-37.2[dB] ,STR- 53768[bpf]) 15|  |  |  |      |      |      |      | Enh( Y-32.252[dB] ,U-37.4[dB] ,V-37.6[dB] ,STR- 63448[bpf],  45040,  19700)
>>  B  41 : .                                                                  14|  |  |  |      |      |      |      | Enh( Y-32.412[dB] ,U-37.4[dB] ,V-37.6[dB] ,STR- 84984[bpf],  47184,  39092)
>>  B  43 : .                                                                  15|  |  |  |      |      |      |      | Enh( Y-32.216[dB] ,U-37.1[dB] ,V-37.4[dB] ,STR- 87912[bpf],  54442,  34762)
>> I  48 : ..   5 Base( Y-35.824[dB] ,U-39.6[dB] ,V-39.6[dB] ,STR-208968[bpf])  4| 4| 7| 4| 70888| 38600| 28368|  4384| Enh( Y-33.845[dB] ,U-38.8[dB] ,V-38.9[dB] ,STR-142240[bpf],  58611,  88797)
>>  B  46 : .  11 Base( Y-32.571[dB] ,U-38.0[dB] ,V-37.8[dB] ,STR- 46272[bpf]) 16|  |  |  |      |      |      |      | Enh( Y-32.319[dB] ,U-37.8[dB] ,V-38.1[dB] ,STR- 67088[bpf],  50273,  18107)
>>  B  45 : .                                                                  15|  |  |  |      |      |      |      | Enh( Y-32.048[dB] ,U-37.2[dB] ,V-37.5[dB] ,STR- 78456[bpf],  44952,  34796)
>>  B  47 : .                                                                  15|  |  |  |      |      |      |      | Enh( Y-32.610[dB] ,U-38.0[dB] ,V-38.2[dB] ,STR- 82880[bpf],  52792,  31380)
>> P  52 : ..   9 Base( Y-33.027[dB] ,U-37.5[dB] ,V-37.3[dB] ,STR- 85856[bpf]) 14|  |  |  |      |      |      |      | Enh( Y-32.325[dB] ,U-37.4[dB] ,V-37.6[dB] ,STR- 76240[bpf],  50966,  26566)
```

```
>>  B  50 : .  10 Base( Y-32.889[dB] ,U-37.9[dB] ,V-37.9[dB] ,STR- 50904[bpf]) 16|  |  |  |     |     |     |     | Enh( Y-32.270[dB] ,U-37.7[dB] ,V-37.9[dB] ,STR- 56176[bpf],  42690,  14778)
>>  B  49 : .                                                                 15|  |  |  |     |     |     |     | Enh( Y-32.478[dB] ,U-37.8[dB] ,V-38.0[dB] ,STR- 75176[bpf],  46508,  29960)
>>  B  51 : .                                                                 14|  |  |  |     |     |     |     | Enh( Y-32.526[dB] ,U-37.4[dB] ,V-37.7[dB] ,STR- 93208[bpf],  59758,  34742)
>> P  56 : ..   8 Base( Y-33.434[dB] ,U-37.2[dB] ,V-37.1[dB] ,STR-102160[bpf]) 15|  |  |  |     |     |     |     | Enh( Y-32.080[dB] ,U-36.9[dB] ,V-37.1[dB] ,STR- 65904[bpf],  48730,  18466)
>>  B  54 : .   9 Base( Y-33.076[dB] ,U-37.4[dB] ,V-37.3[dB] ,STR- 58248[bpf]) 15|  |  |  |     |     |     |     | Enh( Y-32.293[dB] ,U-37.3[dB] ,V-37.4[dB] ,STR- 70192[bpf],  54971,  16513)
>>  B  53 : .                                                                 14|  |  |  |     |     |     |     | Enh( Y-32.375[dB] ,U-37.2[dB] ,V-37.4[dB] ,STR- 89704[bpf],  56640,  34356)
>>  B  55 : .                                                                 15|  |  |  |     |     |     |     | Enh( Y-32.181[dB] ,U-37.0[dB] ,V-37.2[dB] ,STR- 79992[bpf],  51992,  29292)
 >>CALKOWITY STRUMIEN BITOW 7612680 [bits/s] (2896368,4716312)

 >>Computation time : 174.969000 sec. (29 Frames)
 >>STRUMIEN BITOW 2496868 [bits/s]
 >>PSNRY 34.1369 [dB] PSNRU 38.3501 [dB] PSNRV 38.0246 [dB]
 >>Sekwencja: v:\4cif\fun.4cif
 >>Zalozona przeplywnosc: 2500000 [bits/s]
 >>Zalozona liczba obrazow na sekunde: 25.000000
 >>Data: Wednesday 28 February 2001   10:11:26  Central European Standard Time

<END>

 >>(57 Frames)
 >>STRUMIEN BITOW 4137115 [bits/s]
 >>PSNRY 32.8451 [dB] PSNRU 37.8515 [dB] PSNRV 37.9515 [dB]
 >>Sekwencja: v:\4cif\fun.4cif
 >>Zalozona przeplywnosc: 4100000 [bits/s]
 >>Zalozona liczba obrazow na sekunde: 50.000000
 >>Data: Wednesday 28 February 2001   10:11:26  Central European Standard Time

<END>

 >>Warstwa posrednia: 2543779
```

```
>>Warstwa rozszerzajaca: 1674935
>>PSNR Warstwa posrednia: 29.5982
>>Udzial warstwy posredniej: 60
```

Fig. B.1. The result report of scalable coding.