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

Adam Łuczak

Sterowanie prędkością transmisji cyfrowych sygnałów wizyjnych z wykorzystaniem modeli koderów

Rozprawa Doktorska
Przedłożona Radzie Wydziału Elektrycznego
Politechniki Poznańskiej

Promotor: Prof. dr hab. inż. Marek Domański

Poznań, 2003

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

Adam Łuczak

Digital Video Bitrate Control Using Coder Models

Doctoral Dissertation

Advisor: Prof. Marek Domański

Poznań, 2003

Contents

Acknowledgements

I am deeply indebted to many people who shared with me their knowledge, expertise and time, offered advice, encouragement and friendship.

First and foremost, I would like to thank my advisor Prof. Marek Domański for his continuous encouragement, optimism and confidence in me to make it possible to write this dissertation. Prof. Marek Domański not only put enormous amount of time and effort into directing this research but he also helped in editing the documents and confirming every aspect of research results to the tiniest level of detail. I thank him for giving invaluable advices and great help

I would like to express my gratitude to my colleagues and to those whom I was working with in our lab. I owe very much to Sławomir Maćkowiak for efficient and fruitful cooperation in creating the library for the software simulation of video codecs. I also would like to thank Maciej Bartkowiak and Krzysztof Rakowski for discussing the English language phraseology.

Adam Łuczak

Poznań, 20.06.2003.

Abstract

Dissertation is dealing with video coder models and control algorithms for hybrid video coders.. Microscopic and global video bitstream models are presented in this work and they both are the author's original contribution. Accuracy analysis of the proposed models for H.263 and MPEG-2 systems has been made. Experimental analysis of the video coder control algorithms with those models incorporated into MPEG-2 system are presented. Experimental results prove high efficiency and low computational cost of the proposed algorithms.

Streszczenie

Rozprawa poświęcona jest modelowaniu hybrydowych koderów wizyjnych oraz algorytmom sterowania. W pracy przedstawiono oryginalne zaproponowane przez autora modele strumienia wizyjnego: model globalny oraz model mikroskopowy. Praca zawiera analizę dokładności proponowanych modeli dla koderów standardu H.263 i MPEG-2 oraz wyniki badań eksperymentalnych algorytmów sterowania opartych o nowe modele strumienia dla kodera MPEG-2. Wyniki te pokazują dużą efektywność działania koderów wykorzystujących zaproponowane algorytmy sterowania oraz niską złożonść obliczeniową tych algorytmów.

List of symbols and abbreviations

16CIF	- *16CIF* video format: progressive video sequence with 1408×1142 pixels and 4:2:0 chrominance sampling,
2-D	- two-dimensional,
4CIF	- *4CIF* video format: progressive video sequence with 704×576 pixels and 4:2:0 chrominance sampling,
AC	- DCT coefficient for which the frequency in one or both dimensions is non-zero,
AVC	- Advanced Video Coding,
B_{YDC}	- bitstream of encoded DC coefficients of DCT of luminance,
B_{CDC}	- bitstream of encoded DC coefficients of DCT of chrominance,
B_{CTRL}	- bitstream of coder control data,
B_{MV}	- bitstream of encoded motion vectors,
BMA	- block marching algorithm,
B_{YAC}	- encoded bitrate of AC coefficients of DCT of luminance,
B_{CAC}	- encoded bitrate of AC coefficients of DCT of luminance,
CBR	- constant bitrate mode,
CIF	- *CIF* video format: progressive video sequence with 352×288 pixels and 4:2:0 chrominance sampling,
D	- distortion function,
DC	- the DCT coefficient that corresponds to zero frequency in both dimensions,
DCT	- Discrete Cosine Transform,
DVD	- Digital Versatile Disk,
EOB	- End of Block,
F_{ij}	- DCT coefficient *i,j*,
F'_{ij}	- quantized DCT coefficient *i,j*,
FIR	- Finite Impulse Response,
FM	- frame memory,
fps	- number of frames per second,
GOP	- Group of Pictures,
H_{ij}	- Histogram if *(i,j)*-th DCT coefficient,
HDTV	- High Definition Television,
HSV	- Human Visual System,
IDCT	- Inverse Discrete Cosine Transform,
Intra	- intra-frame,

ITU	- International Telecommunication Union,
JM	- Joint Model,
JVT	- Joint Video Team,
Kbps	- kilobits per second,
Level	- the absolute value of this non-zero coefficient.
MADF	- mean absolute difference function
Mbps	- megabits per second,
ME	- motion estimator,
ML	- Main Level (level of MPEG-2),
MPEG	- Motion Pictures Expert Group,
MP	- Main Profile (profile of MPEG-2),
MOS	- mean opinion score,
MVx	- motion vector, component *x*,
MVy	- motion vector, component *y*,
NMSE	- normalized mean squared error
N_1	- size of a frame in horizontal direction,
N_2	- size of a frame in vertical direction,
PSNR	- Peak Signal to Noise Ratio,
Q	- quantization parameter,
Q( )	- quantization function,
QCIF	- *QCIF* video format: progressive video sequence with 176×144 pixels and 4:2:0 chrominance sampling,
qc	- maximum allowed change of quantization parameter,
R	- rate value for some bitstream component,
Run	- number of zero coefficients fallowed by a non-zero coefficient in the scan order. The absolute value of this non-zero coefficient is called "*Level*",
SDTV	- Standard Definition Television,
SIF	- *SIF* video format: progressive video sequence with 360×288 pixels and 4:2:0 chrominance sampling,
SNR	- Signal to Noise Ratio,
T_{ij}	- threshold for *i,j*-th DCT coefficient,
T(Q)	- threshold function
TM	- Test Model,
VLC	- Variable Length Coding,
VBR	- Variable Bitrate Mode,
VBV	- Virtual Buffer Verifier,

Chapter 1

Introduction

1.1 The problem of video coder control

This dissertation deals with one of the key problems in lossy video compression algorithms [Siko99, Skar98, Bovi00, Doma98] – the problem of coder control.

Over the last years video compression techniques were developed very rapidly and nowadays they became very efficient providing high compression ratio with small degradation of quality. Lossy coding algorithms provide efficient data representation at the cost of quality degradation. The contemporary techniques also support new functionalities, for example scalability [Doma00, Horn97]. Hence, dissemination of video and communication services via available communication networks becomes possible. Diversity of products has been developed for a wide range of emerging applications, such as video on demand, digital TV/HDTV broadcasting, and multimedia image/video database services.

Video compression algorithms are complex and tunable by many parameters that should be set automatically by a coder control procedure. The main goal of a control algorithm is to maximize the quality of an encoded video sequence under constraints of limited bandwidth available for transmission or available capacity of storage medium.

Other requirements, like maximum acceptable video signal delay have to be taken into account as well. Hence, a control algorithm is supposed to find the best vector of control parameters in multi-dimensional space of parameters. The optimal set of control parameters can be found but with high computational cost. Some simplifications enable to find a sub-optimal solution. Each video compression technique and each application of video coding may require a different control algorithm.

The two most important types of video coders are hybrid and wavelet coders. Hybrid coders exploit three key techniques of bitrate reduction of encoded video signal: motion-compensated inter-frame prediction, block-based transform coding, and entropy coding. This scheme came into its own as a very efficient technique for moving images. Since first standardizations of hybrid techniques like CCIRe.723 (now J.81) [ITUT93], H.261 [ITUT93b] and MPEG-1 [ISO93] many features were added and a lot of improvements were made. Nowadays, hybrid coders constitute the most widely used group of video coders. Most current products either follow standards like H.263 [ITUT96], MPEG-2 [ISO94], MPEG-4 [ISO00b] and new forthcoming AVC [ITUT01] or are proprietary variants of them, e.g. DivX.

Wavelet coders [Soda99, Ohm00, Kim97a] instead of block-based transform coding exploit 3-D or 2-D multilevel sub-band decomposition (2-D spatial and 1-D temporal or only 2-D spatial, respectively). These coders use new wavelet compression algorithms which are very complex with high computation cost. Compression is obtained by quantizing and encoding coefficients in subbands. In fact, the wavelet-based coding algorithms are still being developed and are not commonly used except for still images area where they have been standardized in well-known standard JPEG2000 [ISO00, ISO02].

Nowadays, hybrid video coding is the most widespread scheme and only these methods will be considered further in the dissertation. Hybrid video coding provides high encoding efficiency with acceptable complexity of encoder and decoder and is recently used in most of video services. Target applications include digital television and video on demand (MPEG-2, MPEG-4) [Giro97, ISO00a], videoconferencing (MPEG-1, H.261, H.263), telemedicine [Lucz99, Lucz00, Lucz00b Lucz00c], digital cameras, DVDs and media storage (MPEG-2, MPEG-4), video on IP (H.263, MPEG-4)[Giro99a] etc.

Because a heterogeneous communication network (Fig 1.1) has various bandwidths, especially the wireless networks, it is necessary to adapt hybrid coders to work in such environment [Giro99, Horn97, Doma00e, ITUT98b]. On the other hand, digital television coder should perform in multichannel mode that requires encoding several channels by cooperating coders. It enables transmission of several TV channels through a constant bitrate broadband channel.

Various types of media have to be taken into account. There are transmission media with constant or variable bitrate and storage media. Storage media are characterized by the limited capacity as well as limited data transfer rate and determined access time to data.

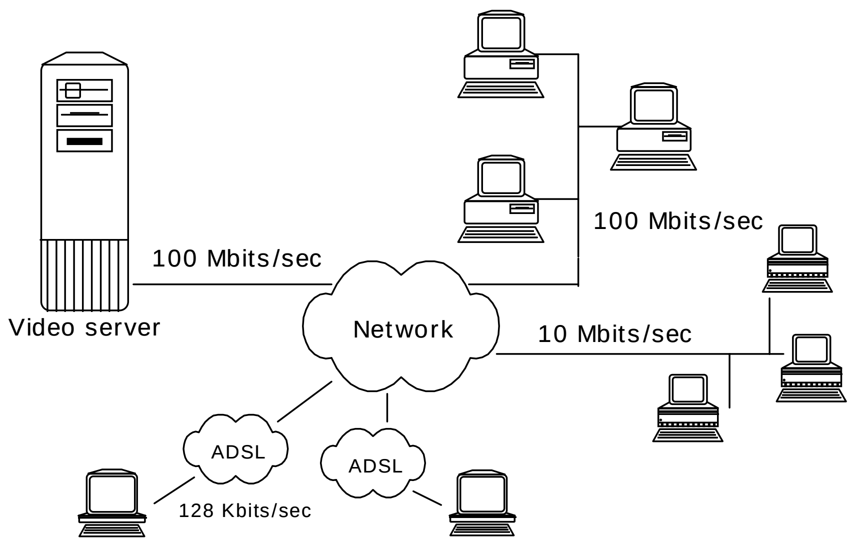


Fig. 1.1 Heterogeneous communication network with video services like video on demand or videoconferencing.

Today, most of video network services use standard coders like H.263 (video-conferencing) or MPEG-2 and MPEG-4 with H.263 as the core coder (streaming media [Giro97]). All those coders need control algorithms which would be able to deal with such various working environments. However, all the standards (such as MPEG-1, MPEG-2, MPEG-4, H.261, H.263 and H.264) do not explicitly define coders. Standards rather define the syntax of an encoded video bitstream together with the method of decoding that bitstream. They do not standardize any algorithm to set the coder parameters in order to obtain required bitrate and/or distortion level. Some video coding standards (like MPEG-2 or MPEG-4) include description of buffer control algorithm but it is only a part of the bitrate and distortion control mechanisms.

Although many coders and decoders are working worldwide, defining efficient control algorithms is still an open problem that gains a lot of attention recently, e.g. [Wang00, Kim00, Wang02a, Lee03, Grec03]

1.2. Goals and thesis of the work

The control algorithm has to determine the optimal set of coding parameters in order to minimize image distortion for a given bitrate in a given working environment. The video hybrid coding algorithm is complex and the set of coding parameters is large. Most parameters are set in order to obtain the best quality for a given bitrate as well as to adapt the coding algorithm to the content and application requirements. The quantization scale factor Q is the only parameter (see chapter 2) that allows controlling the bitstream in a wide range. That is why this parameter is used for video bitstream modeling.

The main goal of the work is to propose bitstream models as a function of quantization scale factor Q. These models should to be simple and easy to use. The objective is also to prove that the proposed bitstream models can be very useful in control algorithms of hybrid video coders.

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

- There exist simple models $B=f(Q)$ that can be identified either by a coding experiment or by transform coefficients analysis.
- Such models can be efficiently applied for bitrate control.

Detailed analysis of the references proved that such models are not described hitherto (chapter 3). The models will be proposed (chapter 4 and 5) and the thesis will be verified by extensive experiments with video test sequences. In order to limit the many time-consuming experiments to reasonable number, only the standard MPEG-2

MP@ML and H.263 coders have been considered for progressive and interlaced sequences with 704×576 pixels and 4:2:0 chrominance sampling.

The experimental conditions mentioned above were chosen because of their compliance with widespread application like satellite digital television (SDTV), video on demand (VoD), videoconferencing etc. Nevertheless, the author has made some experiments that indicate applicability of the models for other resolutions. These results are not included in the dissertation.

1.3 Overview of the dissertation

The dissertation is organized in seven chapters. Chapter 2 describes compression algorithm of hybrid coders as well as methods of quality measurement of coded video sequences. The main features of standards MPEG-2 and H.263 are briefly presented.

Chapter 3 describes the problem of video coding control. The literature review is presented. In particular, the default control mechanism of MPEG-2 video coder is described.

The new original bitstream model is introduced in chapter 4. The new global model of the video coder bitstream is shown, and the way such model can be used in control algorithm of hybrid coder is explained. Experimental results and performance comparisons between original TM5 (Test Model 5) control algorithm and the proposed algorithm based on bitstream modeling are included.

Chapter 5 describes new microscopic bitstream model based on DCT histogram analysis. Control algorithm based on this model is presented. The bitstream model accuracy analysis and results of numerous experiments with new control algorithm are given.

In chapter 6 the scalable coder control is considered. The two-layer scalable coder and coder control algorithm based on bitstream model are presented. The improved prediction of Inter-frames is described and its application for scalable coders. This chapter also includes experimental results.

In chapter 7 the original achievements of the author are listed and final conclusion is provided.

This dissertation is supported by the Polish National Committee for Scientific Research under Grant 4 T11D 007 22 ("New efficient techniques of video coder control").

Chapter 2

Hybrid Video Coders

2.1 Digital video compression

The goal of video compression is bitrate reduction for video storage and transmission by exploiting both statistical and subjective redundancies. The performance of video compression techniques depends on the amount of redundancy contained in the image data as well as on the actual compression techniques used. For video data reduction, quite many various techniques are described in the references [Bovi00, Salo00]. Efficient video compression needs adoption of many complementary techniques that may be combined into sophisticated compression algorithms. Recently, wavelet video codecs have gained a lot of attention, but the hybrid coding algorithm is most often used in video compression. Such algorithm includes three key compression techniques: motion compensated prediction, transform coding, and entropy coding [Held96, Hask97].

Due to transmission medium limitations, the low bitrate on video coder output is expected, therefore most often lossy compression (irreversible) mode is applied. Lossy compression in hybrid coders is obtained by quantization of transform coefficients. Such operation decreases image quality and introduces distortions which manifest themselves as blocking effects or ringing effects on edges [Wink99]. For high compression ratios those artifacts may be strongly visible and quite annoying.

Low bitrate yields poor quality of video sequence and for better quality higher bitstream is needed. Therefore, a trade-off between image distortion and bitrate value (high compression with sufficient quality) is sought (Fig. 2.1). In practice also trade-off between coding performance and implementation complexity is desired [Yang98, Kimr97, Kimr00].

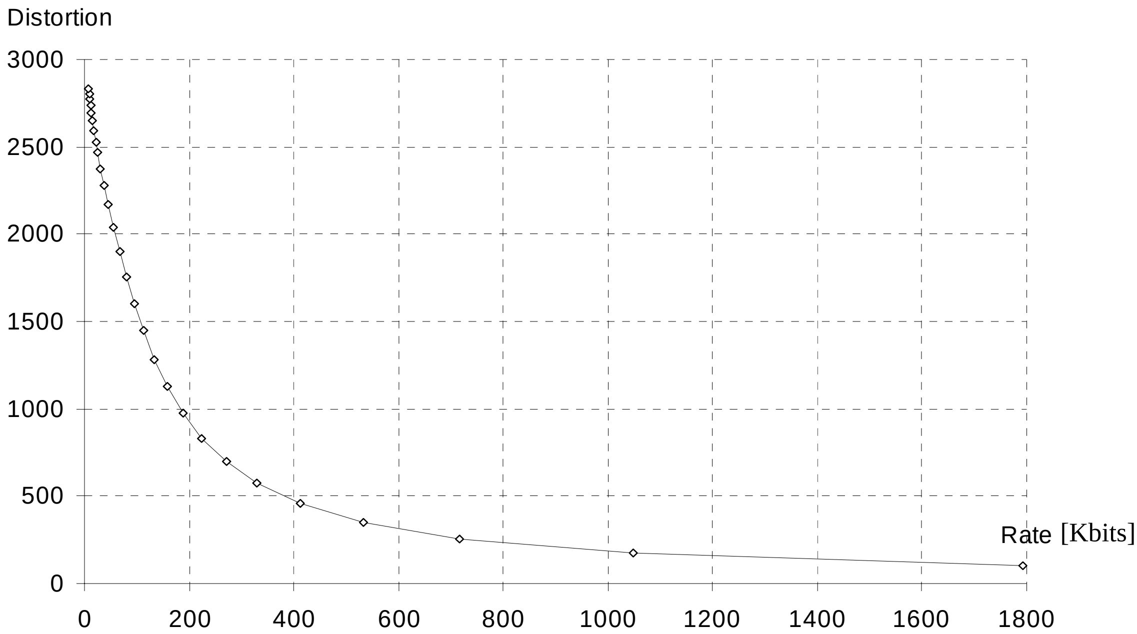


Fig 2.1 An exemplary rate-distortion curve plotted for the whole interval of quantization scale factor Q (MPEG-2 system). Distortion is measured as a sum of absolute value of squared error of each pixel of encoded image.

Therefore, one of the key components of video coder is a control unit. The main objective of control algorithm is to set all coding parameters in order to achieve the output bitrate closest to target bit budget and to achieve the best possible encoded image quality as well. The bitrate can be measured, but measuring the quality of image or video sequence remains problematic.

2.2 Video quality measurement

In order to evaluate the quality of decompressed images two kinds of measures can be used: subjective and objective.

Subjective assessment is made by a group of viewers who either evaluate video quality or evaluate video distortion by comparing the original and decompressed images. The obtained averaged score is called mean opinion score (MOS). The assessment rules of television images like lighting condition, timing, and number of viewers are normalized by International Telecommunication Union (ITU) [ITUR94a]. The outline of common procedure is as fallows: a group of viewers (at least 15) are shown multiple sequence pairs consisting of reference and encoded sequence, which should be short. The reference and encoded sequence are presented twice but in a random order. The viewers are not informed this. They rate each viewed sequence on a continuous scale from "bad" to "excellent". Such technique is called double stimulus continuous quality scale method (DSCSQ). There are also two other similar techniques like double stimulus impairment scale method (DSIS) and single stimulus continuous quality scale (SSCQE) [Wink99].

There were attempts at introducing algorithmic subjective measures [Taoh94, Wats94]. Such algorithmic measure is mostly based on a model of human visual processing. They fit empirical measurements of the response properties of neurons in the primary visual cortex and the psychophysics of spatial pattern detection. An exemplary subjective quality measure algorithm uses subband transform [Taoh94]. Algorithm squares each coefficient of transform, and finally applies the divisive normalization mechanism. Each squared and normalized coefficient represents the response of hypothetical neuron. Exploiting phenomena of contrast sensitivity, contrast masking, and spatial masking the image integrity can be predicted. This scaled integrity coefficient gives perceptual distortion measure.

For online video coding, the fast algorithm of perceptual distortion measure would be useful. Unfortunately, the proposed methods are characterized by high complexity [Masr01, Masr03, Wink99, Bass96]. In most cases this feature excludes them from use. Moreover, video coding control algorithms need fast, mathematical way of objective quality assessment. The advantages of objective measure are impartiality, reliability, repeatability, and efficiency.

Objective evaluation uses normalized mean squared error (NMSE), signal to noise ratio (SNR) or peak to noise ratio (PSNR) [Bert98, Doma98c, Skar98] but only the PSNR measure is used in practice. Definition of PSNR for 8 bits per pixel representation is as follows:

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

(2.1)

where:

N – number of pixels in image,

255 – maximum possible magnitude of pixel in the original image,

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

The PSNR measure is appropriate for analog video systems where the Gaussian noise is the most common distortion. In such systems objective measure of waveform distortion can be related to perceived quality with relatively high precision. Assessment of image quality in digital systems is a more difficult task. Lossy compression algorithms introduce artifacts like blocking effect, blurring, ringing, color blending and motion compensation mismatches. Moreover, those artifacts strongly depend on image content and differ from those in analog systems that PSNR measure sometimes does not coincide with subjective evaluation [Wink98]. However, in most cases of video coding the PSNR measure gives results similar to subjective assessment. In order to compare quality of encoded video sequences, when the coding artifacts in video sequences are similar, the PSNR is also applicable.

Despite some disadvantages of the PSNR measure it is still used by reason of simplicity, reliability and repeatability. Hence, the PSNR is widely used for comparison of different algorithms and in this dissertation will be used as well.

2.3 Coding algorithm

2.3.1 Input video

Usually, an input video sequence consists of images with samples represented in space YC_RC_B. The 4:2:0 format of chrominance is mostly used and denotes picture with two chrominance components decimated by factor 2 in each direction (Fig. 2.2). Therefore, in the dissertation, all the video sequences are assumed to be 4:2:0 sequences.

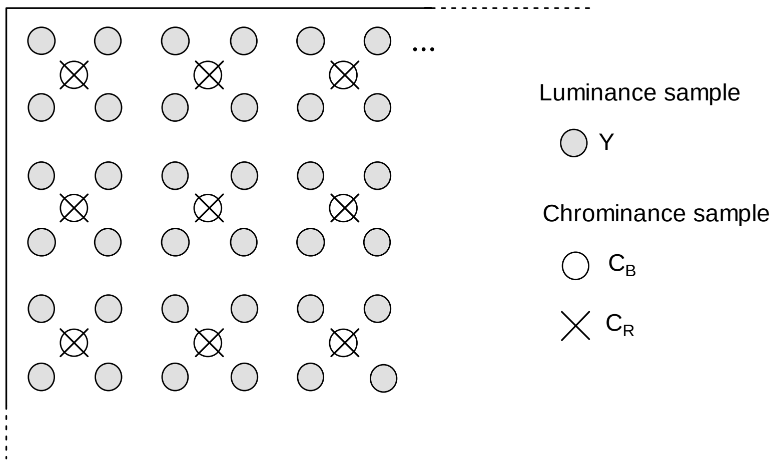


Fig. 2.2 The 4:2:0 luminance and chrominance sampling.

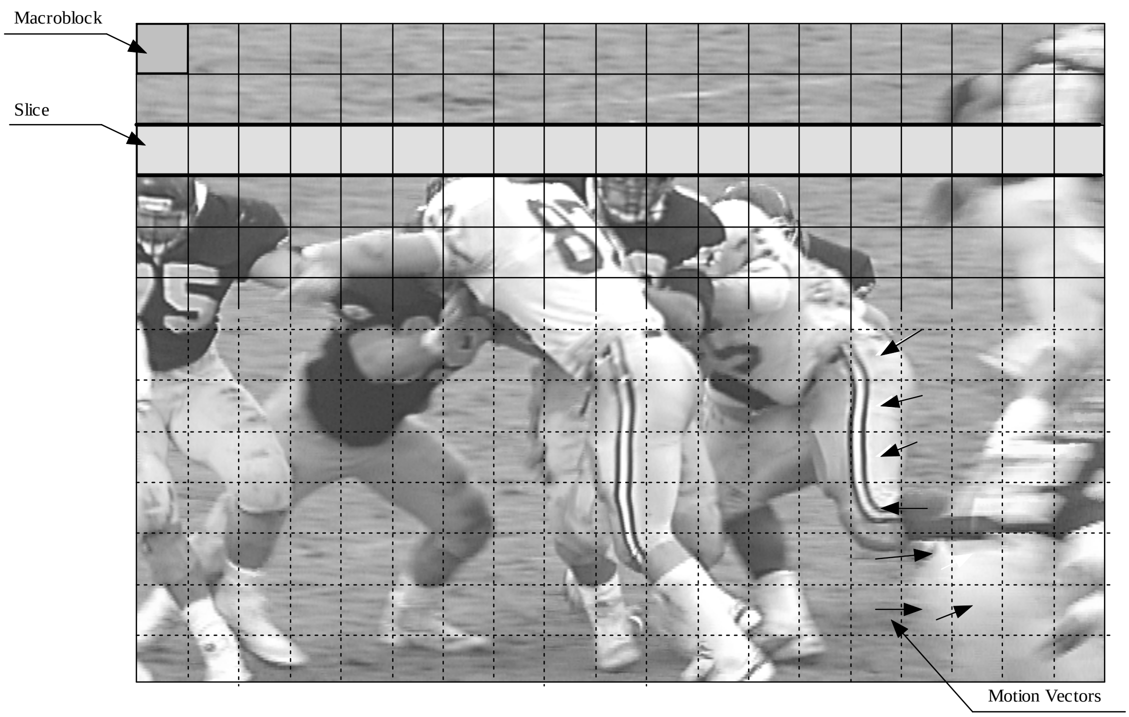


Fig. 2.3 Slices and macroblocks partitioning of example picture.

Input image for encoding is partitioned into slices and macroblocks (Fig 2.3). A slice is one row of consecutive macroblocks. A macroblock is a block 16×16 of pixels of luminance samples together with corresponding chrominance samples. Therefore, in a 4:2:0 sequence each macroblock (16×16 pixels) consists of four luminance blocks (8×8 pixels) and two chrominance blocks (Fig. 2.4). Some modifications of the macroblock structure exist for interlaced video, where each frame is transmitted in two fields. Then, the macroblock has only two blocks of luminance (two blocks in each field).

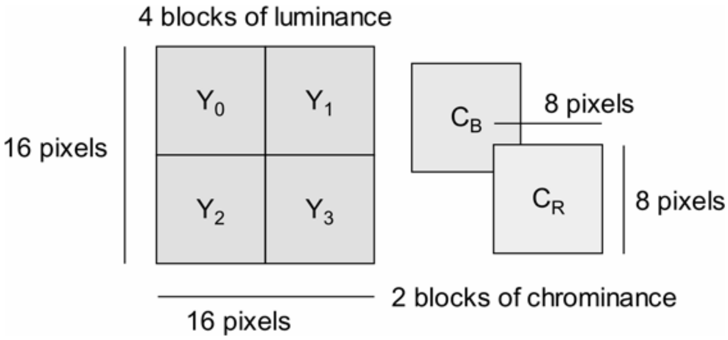
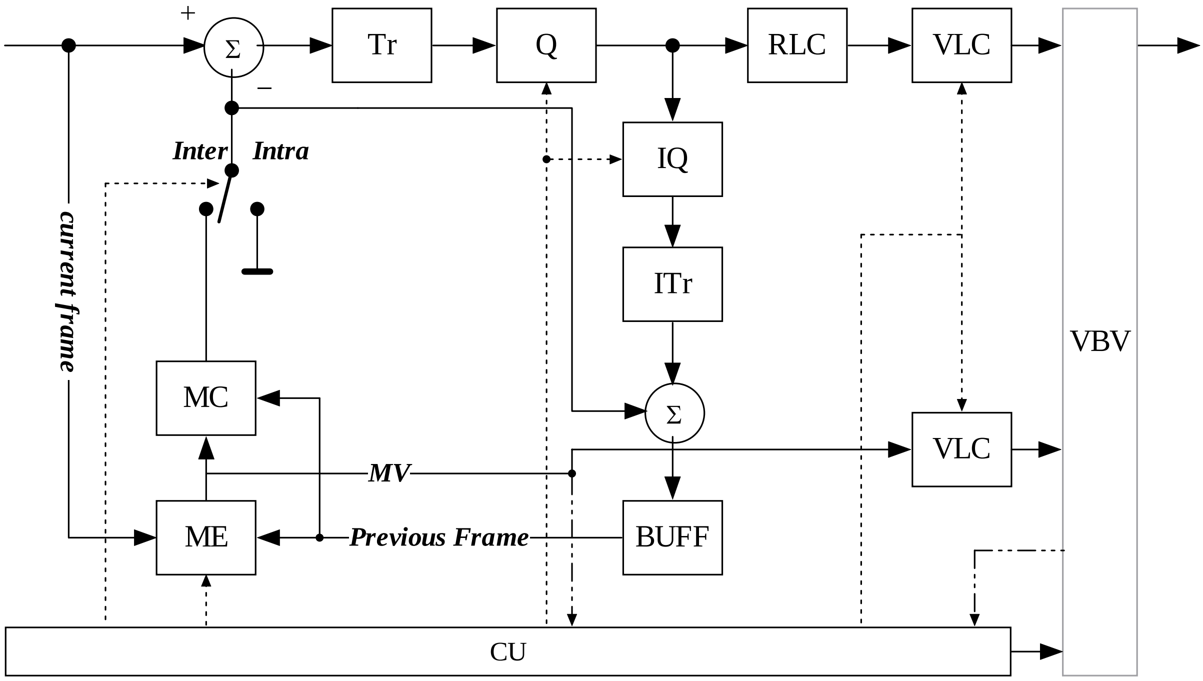


Fig. 2.4 Macroblock structure

A important advantage of sequential encoding and decoding macroblock by macroblock is reduction of required memory and other resources. It enables cheap and fast hardware and software implementations of hybrid coder (see point 2.4) and flexibility of the algorithm modified macroblock by macroblock.

2.3.2 Coder structure

The general structure of hybrid coder (Fig. 2.5) with some small modifications in functional blocks is used by most of common video coders [Tura00]. The main data processing path consist of transform coding with quantization process and entropy coding. Motion compensated prediction is performed in a coder loop. Data from VLC coders are combined with side information like synchronization headers, control data, motion vectors and special flags in one bitstream. The whole bitstream is stored in output buffer (VBV). As generated video bitstream on buffer input is always variable, buffer is indispensable in order to obtain constant bitrate on the output.

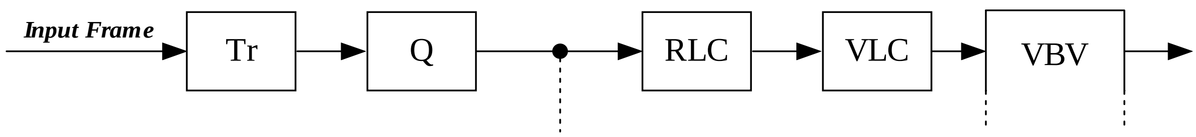


solid lines	– path of the video data;	Tr	– transform;
doted lines	– path of control information;	Q	– quantization;
VBV	– virtual buffer	MC	– motion compensation
RLC	– run length coding	ME	– motion estimation
BUFF	– frame buffer	CU	– control unit

Fig. 2.5 General structure of hybrid coder.

2.3.3 Intra-frame mode of coding

Frame encoded without reference to any past or future frames is called Intra-frame (I-frame). The technique of Intra-frame coding is similar to that of still-image coding (like JPEG) (Fig. 2.6). The first frame in a video sequence or in a Group of Pictures (GOP) is always encoded in Intra mode. In the first step, the encoder applies the transform coding to each 8×8 luminance and chrominance block. Subsequently, all coefficients are quantized with appropriate value of quantize scale factor *Q*, and then the quantized coefficients are *zig-zag* scanned and RL encoded. The obtained *RL*-pairs are encoded by means of Huffman codes. All stages of Intra-frame encoding are described in details in the next points (2.3.3.1 ÷ 2.3.3.4).



Tr	– transform;	solid lines	– path of the video data;
Q	– quantization;	RLC	– run length coding;
VBV	– virtual buffer	VLC	– variable length coding

Fig. 2.6. A piece of the coder structure with Intra encoding path for Intra mode.

2.3.3.1 Transform coding

In hybrid coders two types of transforms are used: Discrete Cosine Transform (DCT) and Integer Transform. The DCT transform for 8×8 blocks is defined as follows:

$$F(u,v)=\frac{2}{N}C(u)C(v)\sum_{x=0}^{N-1}\sum_{y=0}^{N-1}f(x,y)\cdot\cos\left(\frac{(2x+1)u\pi}{2N}\right)\cdot\cos\left(\frac{(2y+1)v\pi}{2N}\right) \tag{2.2}$$

where

$$C(u),C(v)=\begin{cases}\frac{1}{\sqrt{2}}\ for\ \ u,v=0\\ 1\ otherwise\end{cases}$$

where *x, y* denotes pixel index in the horizontal and vertical direction, *u, v* are frequency indexes in both directions, and *N* is the number of transformed pixels. For images with pixels stored on 8 bits, sufficient precision of transform coefficients for video coding is 12 bits. The DCT computed in 8×8 blocs of pixels is used in such standards like MPEG-1, MPEG-2, MPEG-4, H.261, and H.263.

Integer transform [Wien03] coding is used only in the most recent standard H.264. This transformation is computed in 4×4 blocs of pixels. The transform matrices are defined as follows:

$$T_4=\begin{bmatrix}1 & 1 & 1 & 1\\ 2 & 1 & -1 & -2\\ 1 & -1 & -1 & 1\\ 1 & -2 & 2 & -1\end{bmatrix}, \tag{2.3}$$

$$T_{4,INV} = \begin{bmatrix} 1 & 1 & 1 & 1/2 \\ 1 & 1/2 & -1 & -1 \\ 1 & -1/2 & -1 & 1 \\ 1 & -1 & 1 & -1/2 \end{bmatrix}, \tag{2.4}$$

where T_4 is matrix of transform, $T_{4,INV}$ is matrix of inverse transformation and transformation is defined as follows:

$$F = T_4 \cdot P \cdot T_4^T \tag{2.5}$$

where the P is 4×4 matrix of pixels which are transformed. Thank to this transformation implementation becomes very efficient and requires only eight additions and two bit shifts.

2.3.3.2 Quantization

After transformation each of the 64 DCT coefficients is uniformly quantized [Gray98]. Quantization eliminates small coefficients lives and leaves only coefficients with significant energy. In Intra mode of coding mostly quantizer with dead zone is used.

$$w_{ij} = \begin{bmatrix} 16 & 17 & 18 & 19 & 20 & 21 & 22 & 23 \\ 17 & 18 & 19 & 20 & 21 & 22 & 23 & 24 \\ 18 & 19 & 20 & 21 & 22 & 23 & 24 & 25 \\ 19 & 20 & 21 & 22 & 23 & 24 & 25 & 26 \\ 20 & 21 & 22 & 23 & 24 & 25 & 26 & 27 \\ 21 & 22 & 23 & 24 & 25 & 26 & 27 & 28 \\ 22 & 23 & 24 & 25 & 26 & 27 & 28 & 29 \\ 23 & 24 & 25 & 26 & 27 & 28 & 29 & 30 \end{bmatrix}$$

a)

$$w_{ij} = \begin{bmatrix} 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\ 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\ 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\ 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\ 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\ 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\ 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\ 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \end{bmatrix}$$

b)

Fig. 2.7 Default quantization weights matrices for a) intra quant mode and b) non-intra quant mode (MPEG-2 system [ISO94]).

The strength of quantization is adjusted by two parameters: quantizer scale factor Q and weight matrices. Quantization weights matrix for Intra mode favors low frequency coefficients in order to minimize the impact on visual quality by quantization process. Dependently on coding system the matrices are constant or can be changed to optimal ones [Smoo96, Lee98, Wats94].

The quantizer scale factor *Q* is transmitted to decoder on macroblock, slice or frame level. The process of quantization reduces the number of non-zero coefficients and the number of possible values of coefficients magnitudes. The non-zero quantized values of the remaining DCT coefficients and their locations are RLC encoded. It means that *Q* factor controls a number of nonzero coefficients. Hence, quantization scale factor *Q* has the greatest impact on the output bitstream and its value can be changed in wide range, therefore it is the best coding parameter for coder control. Typically, each macroblock in a frame can have assigned different quantization scale factor *Q*. These features allow video coding systems to operate at several channel rates and to provide the benefits of a good bit allocation within a given frame.

2.3.3.3 Zig-zag scan and RLC coding

After transformation and quantization the output coefficients are scanned and ordered in linear table by using a zig-zag order (Fig. 2.8).

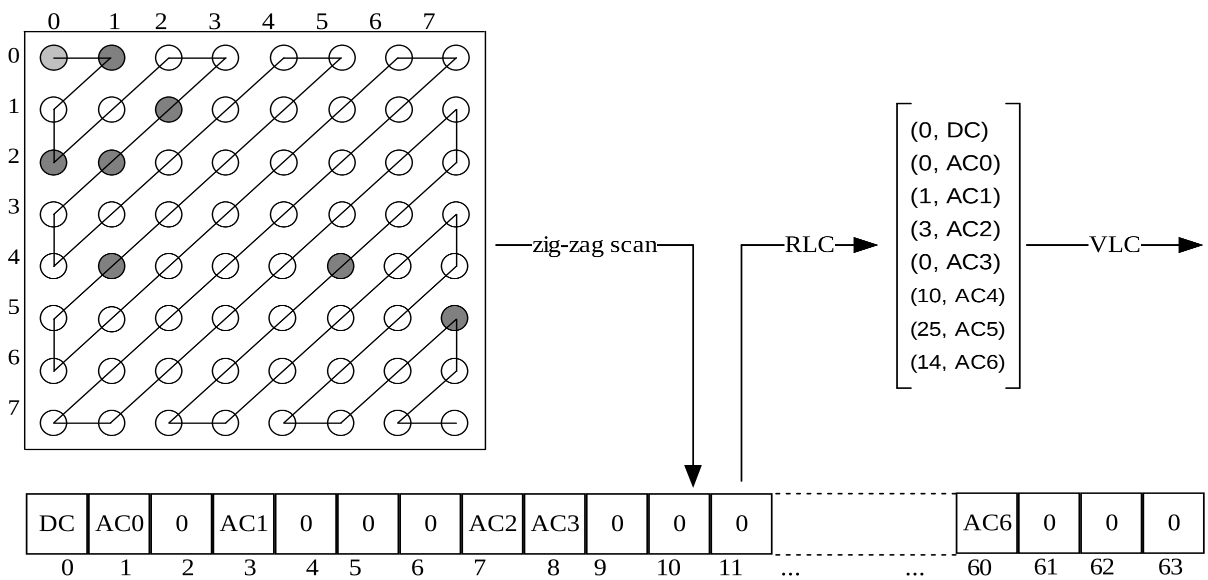


Fig 2.8 DCT coefficients and zig-zag scam pattern for 8x8 block operation.

The non-zero AC coefficient values (*Level*) are detected along the scan line as well as the distance (Run) between two consecutive non-zero coefficients. Each consecutive (*Run*, *Level*) couple is encoded by transmitting only one VLC codeword because only the non-zero quantized DCT-coefficients are encoded. The purpose of zig-zag scanning is to trace the low-frequency DCT coefficients (containing most energy) before tracing the high-frequency coefficients. It means that the zig-zag scan attempts to trace the DCT coefficients according to their significance. The lowest DC coefficient is treated differently from the remaining AC coefficients. It is not quantized and is encoded using prediction.

2.3.3.4 Variable length encoding

For all output data at the end of processing path the entropy coding is applied. All data are encoded with Variable Length Codes (VLC). Each coding system (except recent H.264 system) has defined constant Huffman tables for VLC coding of control information, motion vectors, DCT coefficients and many others. However, for *RL*-pairs obtained from zig-zag scanning only a small part of Huffman table is defined. This part corresponds to a set of most probable *RL*-pairs. The pairs outside that set are ESCAPE coded, which means that *RL*-pair is sent emploing constant length code (Fig. 5.2). For example, MPEG-2 defines only 110 most probable *RL*-pairs with Huffman codes (there are 262 144 of all the possible *RL*-pairs). Despite this limitation *RL*-pairs outside this set occur very rarely.

2.3.4 Inter-frame mode of coding

Except the first frame in sequence or the first in Group of Pictures (GOP) all remaining frames are encoded in Inter-frame mode. In this mode the temporal correlation between two or tree successive frames is exploited (Fig. 2.9) by applying motion-compensated prediction. This operation removes the temporal redundancy present in video sequences. When the frame rate is sufficiently high, there is a great amount of similarity between neighboring frames. It is more efficient to code the difference between frames rather than the frames themselves. An estimate for the frame being coded is obtained from the previous and/or next frame and the difference between the motion-compensated prediction and the current frame is sent. This concept is similar to Predictive Coding and Differential Coding techniques.

2.3.4.1 Motion-compensated prediction

Motion compensation is performed in two stages. First the motion vectors are estimated on a macroblock basis (or block basis). Motion estimation algorithm has to find the best motion vectors which minimize the error between current macroblock and the macroblock-size area indicated by this one. Motion vectors indicate most similar area in reference frame rather than real vector of movement. The most common motion estimation algorithms apply block matching methods but also other algorithms like optical flow or wavelet domain estimation are used. Motion estimation provides only the translatory motion displacement of macroblock [Lee95, Wieg00], none the less it is very complex algorithm. Hence, beside the optimal, the fast algorithms are sought as well [Lee96, Lee97a].

In the second stage image prediction is made by means of estimated motion vectors. In general, new block of current frame is created by averaging adequate – indicated by motion vectors – blocks from previous and/or next frame (Fig. 2.9). In some systems prediction is a more complex operation, for example H.263 coder includes optional Advanced Prediction Mode (APM) which exploits 6 different blocks from source frame to create one block in current (predicted) frame.

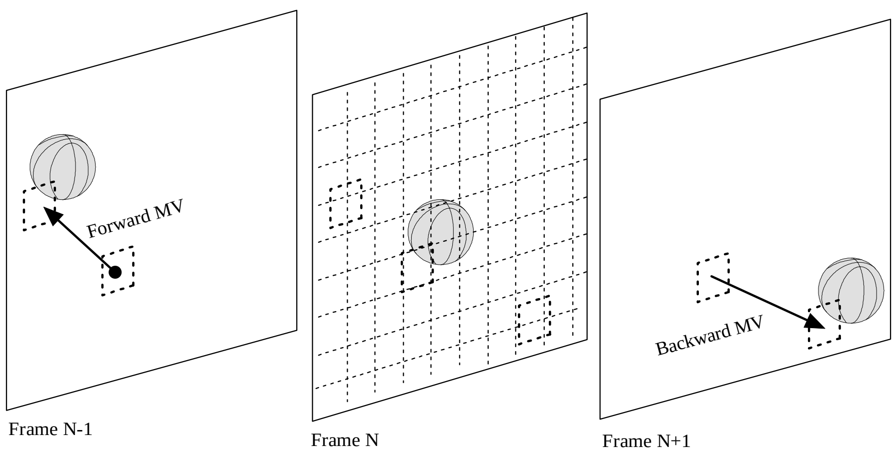


Fig. 2.9 Bidirectional (forward and backward) motion estimation example.

For Intra-frame mode of coding the two types of frame are distinguished P-frames and B-frames. In P-frames the motion compensation exploits only the previous coded frame for forward prediction, whereas in B-frame the bidirectional prediction exploiting two neighboring frames is used. Some video coding algorithms (e.g. H.264) allow using several reference frames for prediction [Wieg97, Wieg99].

The motion compensated prediction error (residual) is calculated by subtracting each pixel in a macroblock with its shifted counterpart in the previous (and next) frame. The obtained prediction error is further coded like in Intra-mode. The advantage of the motion compensated prediction is significant reduction of encoded data.

2.3.5 Bitstream structure

The bitstream produced by hybrid coders has a hierarchical structure (e.g. MPEG-2 system, Fig 2.10). Therefore, the parameters can be changed on several levels of bitstream structure: sequence, group of pictures (GOP), picture, slice, and macroblock. The control on the sequence and GOP level consists mainly in detection of scene editing like fading or cuts in order to insert intra-frames in proper places of video sequence. On the frame level and below, it is possible to choose an image or macroblock type (Inter or Intra), motion vector range, global quantization scale factor Q, VLC tables, progressive or interlaced format, and many others. In this work the picture/frame level was chosen for control video coder.

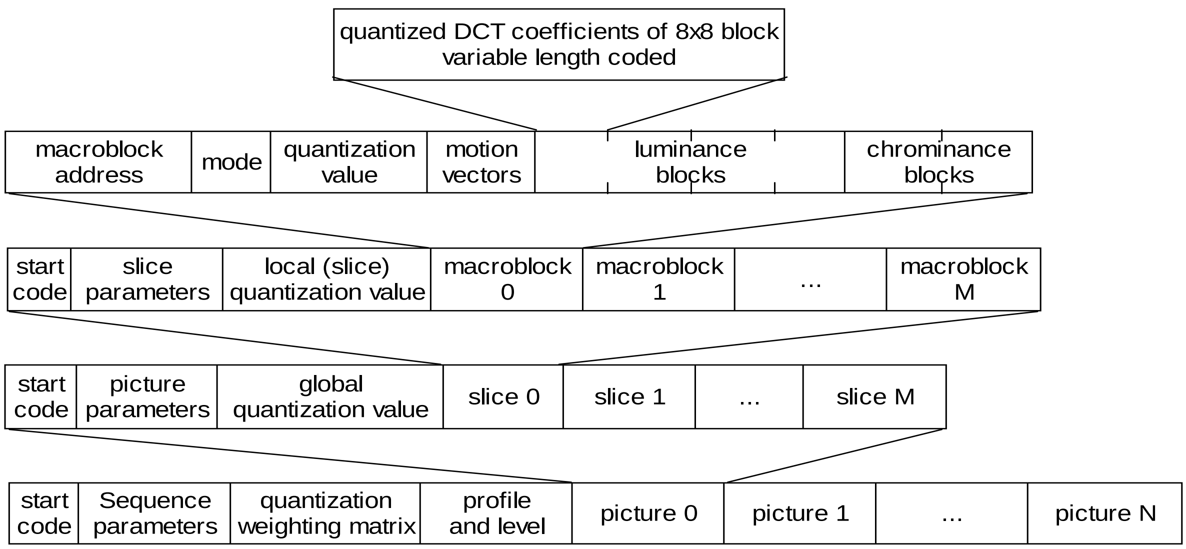


Fig 2.10 Hierarchical structure of the MPEG-2 bitstream [ISO94].

2.3.6 Control unit

The main task of the control algorithm is to set all parameters of coding process. Moreover, it has to take into account coding constraints and imposed requirements like output bitrate or level of quality. In order to meet the requirements and correctly set the value of parameters some analysis of the processed video sequence and previous encoded data should be done. Parameters can be changed on several levels of bitstream hierarchy:

- Sequence and Group of Pictures (GOP) level:

  Control consists mainly in detection of scene editing like fading or cuts in order to adapt GOP by insertion Intra frames in proper places of video sequence. Temporal and spatial resolution and type of encoded picture can be set as well.
- Frame level:

  All the most important parameters are set on frame level. There are distinguished Global and Local parameters. Global parameters are the quantization scale factor *Q*, motion vector search range, coefficients scan type (different for progressive and interlaced frames), VLC table selection, picture coding type (Intra, Inter), frame format (progressive or interlaces), VBV buffer size, quantization matrices and several others, but less important. Local parameters – that is for slice and macroblock - are macroblock coding type and also quantization scale factor *Q*.

However, most important parameter is quantization factor *Q* (see point 2.3.3.2) which impacts on generated bitstream the most, enabling easy way of bitrate control. The VBV buffer should be used for grater flexibility of coder control. Relying on these two key elements (quantization factor and output VBV buffer) the efficient video coder control algorithm can be created.

In general, control unit monitors a virtual encoder buffer (VBV), and periodically adjusts the global and local quantization scale factor *Q* and other coding parameters depending on the video content and activity to ensure that the video buffers will never overflow - while at the same time targeting to keep the buffers as full as possible to maximize image quality. In theory the overflow of buffers can always be avoided by using a large enough video buffer. However, besides the possibly undesirable costs of the implementation of large buffers, there may be additional disadvantages of applications which required low-delay between encoder and decoder, e.g. the real-time and interactive services.

The rate control algorithm used to compress video is not a part of any standard, and thus is left to the implementers to develop efficient strategies. It is worth emphasizing that the efficiency of selected rate control algorithms to compress video at a given bit rate heavily impacts on the visible quality of the video reconstructed at the decoder.

2.4. Video coding standards

The main aim of standardization is to foster implementations of image and video coding equipment and software. Therefore, international standards do not necessarily represent the best technical solutions, but rather attempt to achieve a compromise between the amount of flexibility supported by the standard and the implementation complexity required and achieved compression efficiency. Recent progress in digital technology has made the widespread use of compressed digital video signals practical. Standardization is very important in the development of common compression methods to be used in the new services and products. This allows new services to interoperate with each other.

The most popular and commonly used standards are ITU-T H.261, H.263/H.263+, ISO/IEC MPEG-1, MPEG-2 and MPEG-4. Among all those video coding standards, MPEG-2 [ISO01, Hask97] is extremely successful because of strong commitment from industries, cable and satellite operators, and broadcasters to use this standard. Services like Digital TV broadcasting, pay TV, pay-per-view, video-on-demand, interactive TV or DVD-Video use MPEG-2 coding system. Others, like H.263, are exploited for videoconferencing and for tasks required low bitrates. Now, the new MPEG-4 platform slowly replaces H.263 giving a lot of new functionalities and higher compression ratios.

These well-known international standards are related to innumerable software and hardware implementations but they actually define mostly semantics and syntax of the compressed bitstream and not coder implementations which are described in the informative parts of some standards as examples only.

It needs to be noticed that all the standards and recommendations do not standardize any algorithm to set the coder parameters in order to obtain required bitrate and/or distortion level.

2.4.1. MPEG-2

The structure of MPEG-2 coder directly corresponds to the figure 2.5. In this system the frames can be coded as I-frames, P-frames and B-frames. Progressive and interlaced video signal is allowed with wide range of horizontal and vertical resolutions. Motion compensation is performed on macroblock-bases with half-pixel accuracy. Almost all data is encoded with Huffman codes. The MPEG-2 standard has fixed tables of Huffman codes and only quantization weight matrices can be changed.

Produced bitstream has a hierarchical structure (Fig. 2.10) with a few synchronization headers. This makes the system more resistant to transmission errors.

The quantization scale factor Q can be modified on each level of this hierarchical structure and also Q has the strongest impact on output bitstream enabling easy bitrate controlling. This parameter enables coder control in wide range of bitrates and wide range of sequence quality. It means that MPEG-2 can be used in order to produce either consumer quality bitstream (PSNR<40 dB and bitrate< 8 Mbits/sec) or studio quality bitstream (PSNR>45dB and bitrate>10 Mbits/sec). All this can be achieved only by quantizer scale factor Q adjusting. MPEG-2 standard offers high efficiency of encoding and is capable of compressing of standard-definition 4:2:0 video down to about 2-15 Mbits/s. At the lower bit rates in this range, the impairments introduced by the MPEG-2 coding and decoding process become increasingly objectionable and some coding artifacts become strongly visible.

Headers of macroblocks consist only of simple flags, while slices include extended headers with synchronization markers. The frame can include optional headers with general coding parameters and synchronization information.

2.4.2. H.263/H.263+

The baseline H.263 video coding algorithm is based on techniques common to previous video coding standards [Giro95, Faer97]. This standard has been targeted for videoconferencing and for very low bitrates. Therefore it supports only progressive sequences with limited resolution and also its bitstream structure is much simpler without too many headers. Furthermore, existed headers are rather short and do not protect against transmission errors.

The new baseline of H.263 coder, known as H.263+ [Cote98], has some improvements over H.261, H.263 and MPEG-2 standards by supporting several negotiable advanced coding modes [ITUT96].

Table 2.1 Advanced modes supported by H.263++ video coder.

Additional Advanced Modes Of Coding			
1	Unrestricted Motion Vector (UMV) mode,	7	Reference Picture Selection (RPS) mode,
2	Syntax-based Arithmetic Coding (SAC) mode,	8	Independent Segment Decoding (ISD) mode,
3	Advanced Prediction (AP) mode,	9	Alternative INTER VLC (AIV) mode,
4	Advanced INTRA Coding (AIC) mode,	10	Modified Quantization (MQ) mode,
5	Deblocking Filter (DF) mode,	11	Reference Picture Resampling (RPR) mode,
6	Slice Structured (SS) mode,	12	Reduced-Resolution Update (RRU) mode.

These options increase compression ratio and allow developers to trade off between compression performance and computational complexity [Erol98, Ghar96]. Several parameters may be varied to control the rate of generation of coded video data. These include processing prior to the source coder, the quantizer, block significance criterion, and temporal sub-sampling scheme. So many additional modes of coding complicate bitstream analysis and make the control algorithm very complex. As in the previous standard the quantization scale factor Q is the best parameter for controlling coder.

Chapter 3

Video coder control

3.1 Video coder control scenarios

The goal of control algorithm is to find the optimal encoding parameters that minimize distortion for a given bitrate, for a given frame, given coding algorithm with a given set of control parameters. Distortion level for certain frame unequivocally determines bitrate. Therefore the working point (bitrate-quality) on a Rate-Distortion (*RD*-curve) curve has to be chosen (Fig. 3.1). It is done in an indirect way by choosing Q that corresponds to given R (Fig. 3.2b) or D. The quantization scale factor Q from previous frame can also influence the R-D relation for current frame.

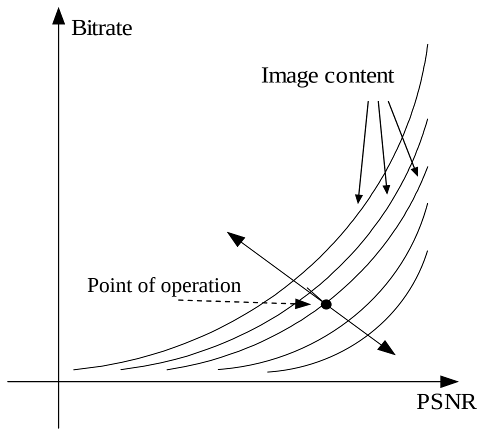
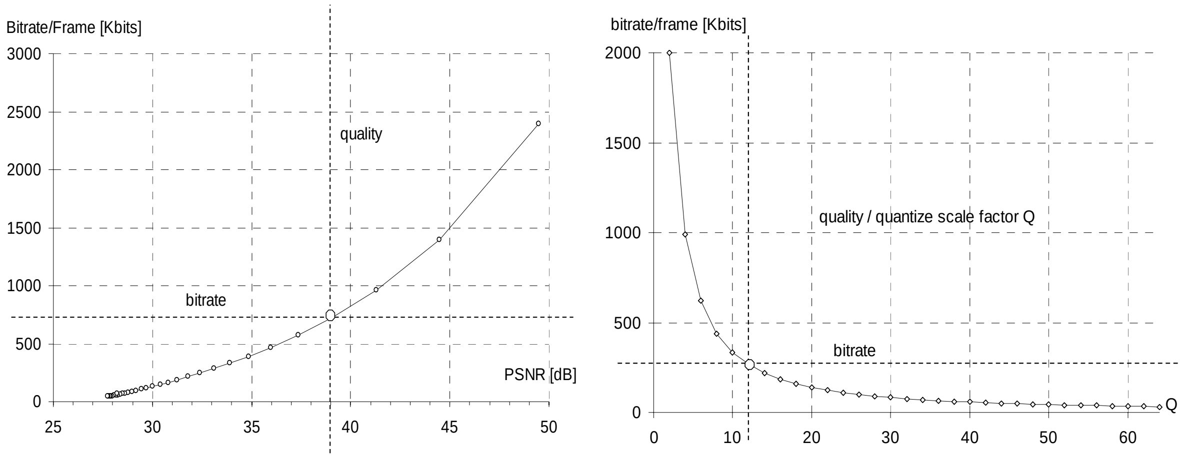


Fig. 3.1 Rate-Distortion curves for different image contents.

Unfortunately, due to varying in time of video content (Fig. 3.1) the relation R-D cannot be set once. Consequently, in order to fix the working point on *RD*-curve, the quantization scale factor Q has to be set individually for every frame.

When the achieving of constant bitrate is the most important, the *R-Q* curve has to be chosen, whereas the constant quality is the aim, then the *D-Q* relation must be determined. In any case the *R-D* relation may be used to establish one relation from the other (Fig. 3.2). In this dissertation the *R-Q* relation (Fig. 3.2) will be used because the CBR mode is used in telecommunication application the most.



a) b)

Fig. 3.2 The Rate-Distortion and Rate-Quantization curves for one frame of video sequence *Basket*: a) bitrate versus encoded image quality and b) bitrate versus quantization scale factor *Q*.

The parameters of coding algorithm depend on a kind of transmission medium as well. The limitations of medium like bandwidth or capacity and some hardware/software limitations form a set of constraints. According to these constraints and user requirements, two ways of encoding are possible: constant bitrate mode (CBR) and variable bitrate mode (VBR), the latter one enables achieving near-constant encoding quality.

3.1.1 Constant bitrate mode (CBR)

Constant bitrate (*CBR*) mode is appropriate for transmission channels with constant or almost constant bandwidth, for example satellite channel [Dalg95, Assu98]. It means that bitstream at the encoder output has to be constant (Fig. 3.3).

Because bitstream generated during encoding an image is various in nature, the output buffer is needed. Capacity of the buffer depends on a transmission bitrate (the higher bitrate the larger buffer capacity) and maximum admissible delay (the smaller delay the smaller buffer capacity). In real-time services this buffer has to be relatively small in order to limit delaying of video signal. For example, digital satellite television uses MPEG-2 system for video encoding and make available 22 Mbits/sec constant bitrate channels. Such channel is shared between 7 digital TV channels with about 3 Mbits/sec for each of them.

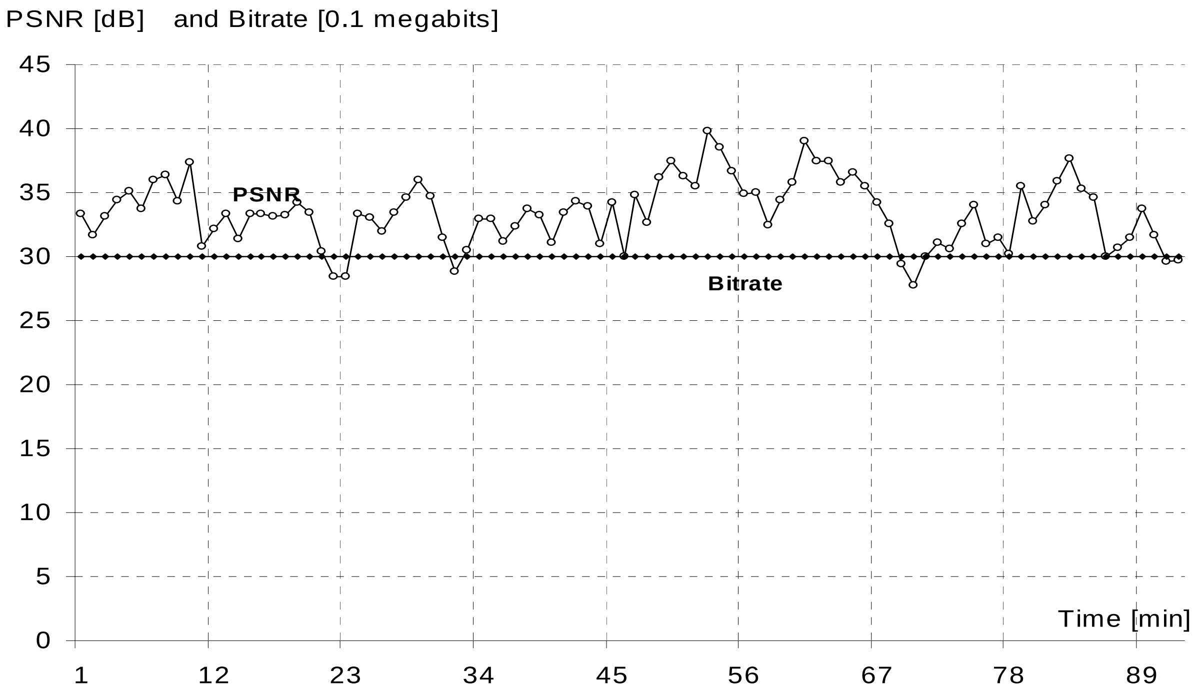


Fig 3.3 Bitrate and PSNR curves for example sequence encoded with constant bitrate (e.g. Satellite Digital TV).

For such an exemplary bitrate and for MPEG-2 coder the 1 Mbits output buffer is demanded by standard. The video coder achieves constant bitrate of coded video sequence by using this buffer (Fig. 3.3). If the buffer underflow appears the sequence of zeros will be sent, but if overflow appears data will be skipped and it will cause an error of video sequence decoding. Such buffer introduces additional delay but is acceptable in the broadcast services.

Unfortunately in this mode of coding it is hardly possible to maintain constant quality throughout the whole encoded video sequence. When CBR mode is chosen the main task of the control algorithms is to obtain the highest possible quality of video sequence being encoded, and to avoid overflow and underflow of the buffer.

3.1.2 Variable bitrate mode (VBR)

In some applications available medium like CD or DVD offers a wide bandwidth for reading and writing a stream, but has finite capacity. Therefore, another strategy of encoding has to be applied. All types of media have limited maximum value of bitstream but it is higher than, the average bitrate imposed by total capacity.

This limitation depends on medium type, for example Digital Versatile Disk for Video (DVD-Video) medium provides maximum bitrate about 10 Mbits/s. In that case, it is possible to let the MPEG-2 encoder produce a video sequence with a constant visual quality over time (fig 3.4). Encoder has a variable bitrate (VBR) on the output and no buffer is needed. Moreover off-line processing is applied enabling initial analysis of video sequence which cause the fact that the available bits are appropriately distributed over various video segments hence constant visual quality is obtained [Hamd97, West99, Yuzh01]. Of course, assumed quality level should be reasonable in order to keep within the limits of bitrate and to avoid buffer overflow if exist.

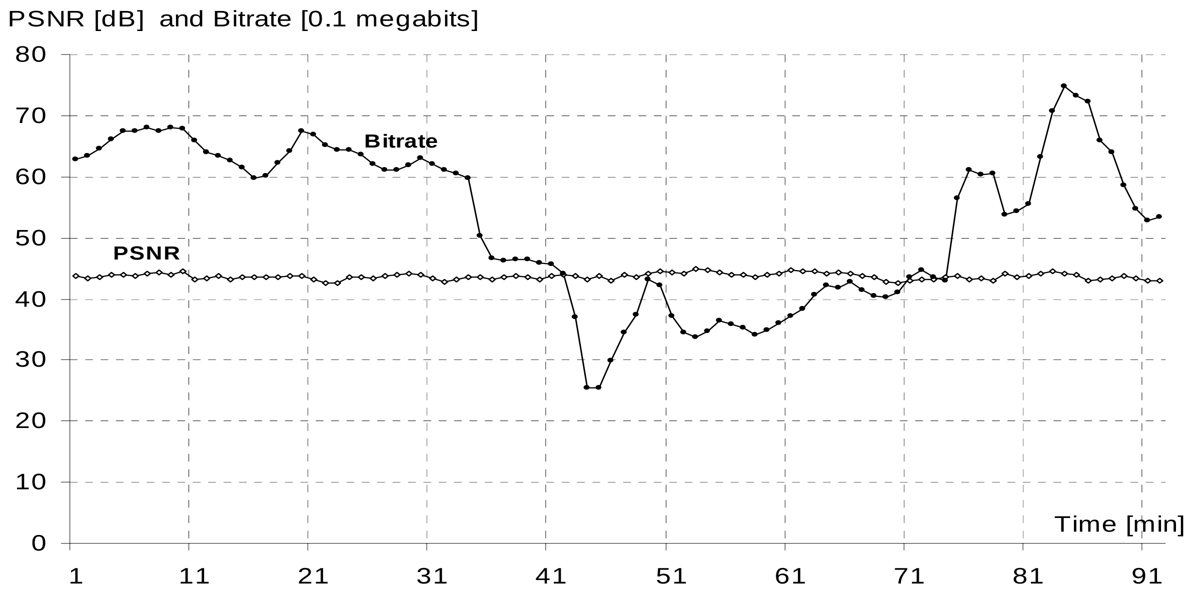


Fig 3.4 Bitrate and PSNR curves for exemplary movie encoded with very high constant quality and variable bitrate (e.g. DVD-Video).

The advantage of using a variable bit rate is mainly the gain of encoding efficiency. For fixed storage mediums the variable bit rate is ideal. By reducing the amount of space needed to store the video - retaining very high quality, it leaves more space on the medium for inclusion of other features e.g. multiple language soundtracks, extra subtitle channels, interactivity, etc. Another feature of the variable bitrate encoding mode is constant video quality for all complexities of program material content. The difference between CBR mode and VBR mode can be described as follows:

Table 3.1 Comparison of the CBR and VBR mode:
(AVR is average bitrate and the MR is the maximum bitrate available for certain channel)

	CBR mode	***VBR mode***
Objectives	AVR $\cong$ MR	AVR < MR
Primary:	To maintain constant average bitrate ***(R)*** equal maximum bitrate in short-term.	To maintain near constant quality ***(PSNR)*** of encoded sequence.
Secondary	To maximize the quality (PSNR) for given R of encoded sequence.	To minimize bitrate for given sequence quality avoiding to overflow or underflow of the buffer.

The CBR and VBR modes differ only in priority of targets, and these targets are reached by controlling the coder with the same set of parameters. The CBR mode with very large VBV buffer converges to VBR mode.

3.2 Video coding constraints

Some limitations were introduced by standardization organizations and by group of manufacturers in order to decrease ranges of coding parameters. An encoding mode with such sets of limitations is called constraints coding mode. It is guarantied that bitstream produced in that mode will be decodable by each hardware and software decoder. For example: MPEG-2 constraints assure hardware decoding of bitstream set, the maximum motion vector range, buffer size, delay limit, profile and level choosing method. Some restrictions depend on the storage media, for example there is about 10 Mbits/sec limit for bitrate for DVD standard and about 2 Mbits/sec for Video CD (VCD). Those restrictions enable the manufacturers to create hardware versions of coders.

3.3 Scalable mode of coding

Scalability means the coder produces bitstream partitioned into layers [Giro97, Doma98b, Doma99, Doma99b, Ohm01, Mack02]. The layers represent various qualities or various spatial and/or temporal resolutions of encoded video. Lower layers represent video with reduced quality or resolution. Those can be decoded independently from higher layers. Video sequence decoded from lower layers can be improved by decoding additional higher layers which are decoded with respect to the previous layers.

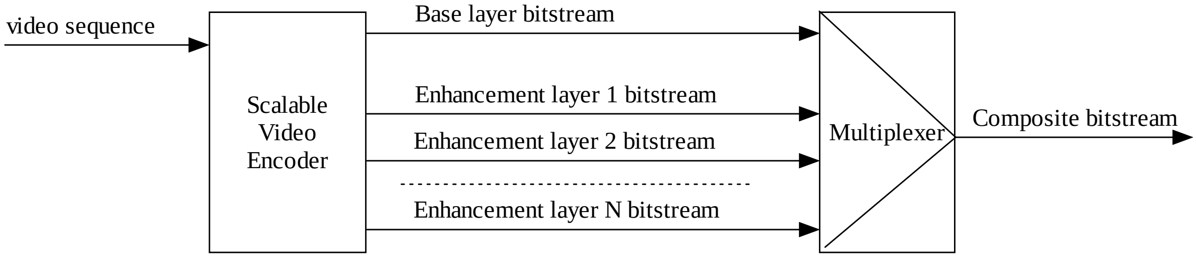


Fig. 3.5 Multi-layer scalable video coding system producing N independent bitstream layers [Mack02].

In such complex scheme of coding bits have to be allocated among two or more layers. The first way of controlling scalable encoder is to fix bitrate for each layer and apply the independent control mechanism for each. Such mechanism can, for example, employ optimization techniques based on Lagrangian minimization (see point 3.4.1) [Gall99].

The second way is to treat scalable coder as one coder allowing the bitstream of enhancement layers to vary slightly [Hask96, Mack02] (see: chapter 6). The base layer is controlled independently from enhancement layers. The Enhancement layer control unit adjusts bitrate of enhancement layers in order to obtain a constant bitrate for the sum of bitstreams of all layers. It means that variation of base layer bitstream affects the enhancement layers bitstream. Owing to these variations the best fit to required bitrate can be achieved.

3.4 Review of video coder control techniques

This point briefly reviews the control algorithms solutions. There are many proposals of control algorithms, some of them are appropriated only for one type of encoder or one certain task, and others are more versatile. In this thesis, controls techniques have been classified into three groups that are not necessarily disjoint. The other way of classification could be also possible. The classification is as follows:

- Lagrange multiplayer methods,
- Modeling methods,
- Sequence feature analysis

3.4.1 Lagrangian multiplier method

As it was shown before, the hybrid coder has many coding options like macroblock coding type (e.g. for MPEG-2 system there are: Intra, Inter P, Inter B), quantization scale factor *Q*, range of motion vectors, quantization matrices and changeable tables for variable length coding. Direct applying of some optimization method in order to find the best values for parameters is often used.

The optimization task is to choose the most efficient coding representation (best set of coding parameters) for each macroblock in the rate-distortion sense. It means that several mutually dependant parameters have to be adjusted in order to achieve the best performance/the smallest distortion for a given bitstream. Optimization task is additionally complicated by reason that various coding options reveal varying efficiency at various bitrates (or levels of quality) and with various scene contents. Such optimization is rather considered as a minimization problem.

The authors [Choi94] [Orte96] [Schu97] proposed to use the Lagrangian approach for the case of multiple control parameters and constraints. The problem is to minimize distortions for a given bitrate:

$$\min_{x_0, \ldots, x_n} D(x_0, \ldots, x_n) \quad \textit{subject to}: \quad R(x_0, \ldots, x_n) \le R_{\max}, \tag{3.1}$$

where $D(x_0,\ldots,x_n)$ is the distortion measure function, $R(x_0,\ldots,x_n)$ is the bitrate function, R_{max} is maximum allowed bitrate, and $x_0,\ldots,x_n$ are some parameters of coding. Unfortunately, the x parameters are dependant. Therefore, the Lagrange multipliers correspond to each of the constraints are introduced.

$$J_\lambda(x_0, \ldots, x_n) = D(x_0, \ldots, x_n) + \lambda \cdot R(x_0, \ldots, x_n), \tag{3.2}$$

where $J_\lambda(x_0,\ldots,x_n)$ is a cost function that is minimized. Such technique can help to find the suboptimal solution. The Authors propose to use the iterative technique to solve this set of linear equations. These equations include all basic set of constraints and parameters like bitrate limitation, buffers size, coding mode, and quantization scale factor Q. Others researchers extend these simple method including additional coding parameters and constraints [Orte98, Riba99].

A. Ortega and Y. Hsu proposed a control algorithm based on Lagrangian optimization [Orte98] but in their control algorithm approach the delay and channel bandwidth are considered as additional constraints which can be translated into rate constrains at the video encoder. They formulate the rate constraints imposed on each block of encoded video due to the real time operation of the system and the available channel capacity which altering in time. In typical video communication systems, the end-to-end delay ΔT consists of the following delay components:

$$\begin{aligned}\Delta T = \Delta T_e + \Delta T_{eb} &+ \\ \Delta T_c &+ \\ \Delta T_{db} + \Delta T_d&.\end{aligned} \tag{3.3}$$

when ΔT_e is the encoding time, the ΔT_{eb} is the time of data buffering in encoder (VBV buffer), the ΔT_c is the transmission time thought a channel, the ΔT_{db} is the time of data buffering in decoder, and the ΔT_d is the decoding time. Due to there are only two variable components ΔT_{eb}, ΔT_{db}, than ΔT_e, ΔT_c and ΔT_d are known constants:

$$\Delta T_{eb} + \Delta T_{db} = \Delta T - \Delta T_e - \Delta T_c - \Delta T_d, \tag{3.4}$$

and for given duration of a frame interval T_f, the total number of frames in either the encoder or decoder buffers ΔN will also be constant.

$$\Delta N = \frac{\Delta T_{eb} + \Delta T_{db}}{T_f} \tag{3.5}$$

The source encoding rate is constrained by the available capacity and end-to-end delay. The time to transmit one packet will be denoted as T_p, hence, the time index t will be integer. One video frame spans F packet intervals with:

$$F = \frac{T_f}{T_p} \tag{3.6}$$

and the *n-th* frame is encoded and released to the encoder buffer at time $n \cdot F$. Due to delay constraint ΔN, a frame has to be transmitted by time $(n+\Delta N)\cdot F$. The $R(i)$ is defined as the number of bits used for encoding *i-th* frame, the $R'(i)$ as the number of bits of *m-th* frame that is still in the encoder buffer, and $C(k)$ as the number of bits transmitted through the channel at time k. The condition for *i-th* frame to arrive at the decoder is that all the data of *i-th* frame have to be transmitted by the due time $(i+\Delta N)\cdot F$, hence:

$$R^{'}(m) + \sum_{j=m+1}^{i} R(j) \leq \sum_{k=t+1}^{(i+\Delta N)\times F} C(k), \quad for\ m = 0,1,..,i. \tag{3.7}$$

where m is the frame in the encoder buffer, and t is the time index of current encoded frame. It is assumed in the presented control algorithm that the bitrates of those video frames which are still in the encoder buffer can be dynamically adjusted before transmission. It is possible by storing data encoded using various values of quantization scale factor Q in separate buffers, so the transmitter can select the data source from one of the buffers according to the control algorithm.

$$Q = \arg\min_{Q} \sum_{j=m+1}^{n} D_{X_j}(j) + \sum_{i=m+1}^{n} \lambda_i \cdot \left(\sum_{j=m+1}^{i} R_{X_j} \right) \tag{3.8}$$

such equation has (*n-m*) Lagrange multipliers which replace the (*n-m*) constraints. The problem of finding out the appropriate values of each λ_i is solved by iteratively increasing the lower bounds of the multipliers, in such a way that the violation of rate constraints can be prevented, and adjusting the values of λ_i. In comparison to standard control algorithm (TM5) that solution can improve coding efficiency giving the average quality gain about 0.5 dB.

The control algorithm described above is complex and perform iteratively. This algorithm requires high computational cost and large capacity of output buffer memory in order to maintain frames which are encoded with various quantization scale factors Q.

Most of Lagrangian methods are not used, because of their complexity. Therefore one method which is practically used is selected by the author and is presented below in detail. This method is employed for controlling hybrid video coder (H.263, MPEG-4) and its limited set of constraints results in low complexity and low computational cost. T. Wiegand and J. Sullivan propose Lagrange Multiplayer Method (LMM) as a solution to the problem [Sull98] which they defined as a minimization of distortion function D, subject to a constraint R_C on the number of used bits R.

$$\min D \quad for \quad R < R_C . \tag{3.9}$$

The Lagrangian formulation of the problem (3.5) is given by:

$$\min J, \quad where \quad J = D + \lambda R, \tag{3.10}$$

where the Lagrangian rate-distortion function J is minimized for particular value of Lagrange multiplier λ and each solution of this equation for a given value of λ corresponds to an optimal solution for equation (3.5) with particular value R_C. Now we can choose which parameters will be minimized. Minimizing functions proposed by T. Wiegand, for example, comprise of motion estimation and prediction parameters (λ_{MOTION}), prediction mode with quantization step decision (λ_{MODE}). To obtain a relationship between Q factor and λ_{MODE}, the minimization of the Lagrangian cost function is extended by the macroblock mode type decision over the set of macroblock modes (for H.263):

$$\{INTRA, SKIP, INTER, INTER + 4V, INTRA + Q, INTER + Q\} \tag{3.11}$$

and followed equation is minimized :

$$\min[D_{REC}(S_K, I_K \mid Q) + \lambda_{MODE} \cdot R_{REC}(S_K, I_K \mid Q)], \tag{3.12}$$

calculating independently for each macroblock S_K, and each macroblock mode I_K. The distortions D_{REC} are measured as the sum of squared differences (SSD) between reconstructed and the original macroblock pixels. The rate R_{REC} is the rate that results after RLC and VLC coding. Rate constrained motion estimation is obtained from Lagrangian minimization of another cost function:

$$\min[D_{DFD}(S_i, m) + \lambda_{MOTION} \cdot R_{MOTION}(S_i, m)] \tag{3.13}$$

where D_{DFD} is the distortion measure and R_{MOTION} is the bit budget required for encoding motion vectors and the m is vector search range. The authors, [Sull99] have made experiments and obtained statistics (Fig 3.6) which show the conditional probability of chosen macroblock quantizer values of Q_INDEX for several values of λ_{MODE}. One can see the strong dependency between λ_{MODE} and Q_INDEX. This experiments yield:

$$\lambda_{MODE} = c \cdot Q^2, \tag{3.14}$$

and $$\lambda_{MOTION} = \sqrt{\lambda_{MODE}}, \tag{3.15}$$

where c and d are some constant values dependent on the image content. Proposed control algorithm results in better quality of sequence for a given bitstream. It means that coder makes the most of available bandwidth. Unfortunately, those methods do not take into account variations in quality. Therefore, they do not make any effort to keep them constant or even prevent from violent changes of quality of video sequence.

The described method of control is rather complex but gives very good results of coder parameters adjusting and increases encoded sequence quality about 0.2-0.4 dB in comparison to TM5 control algorithm.

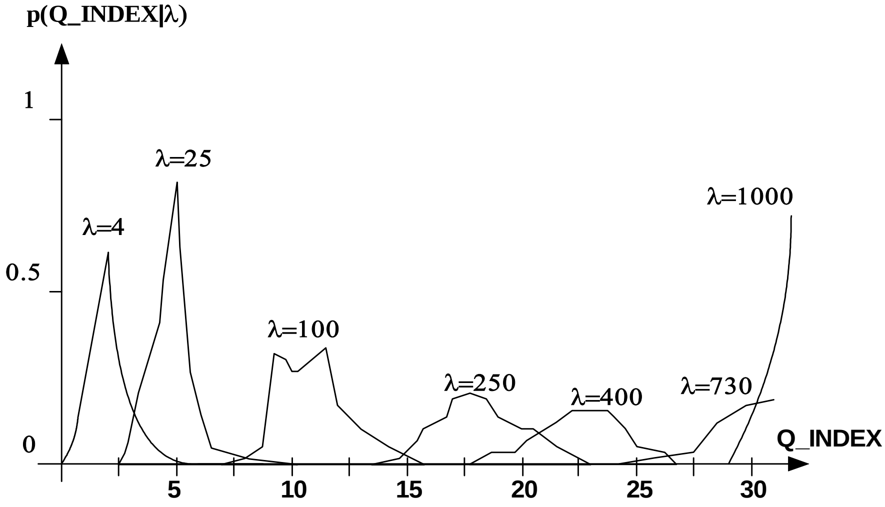


Fig. 3.6 Conditional probability versus macroblock quantization factor Q for various values of the Lagrange multiplier λ_{MODE} (on the basis of Wiegand and Sullivan paper [Sull98]).

In general, coders with control algorithms based on an optimization method are very efficient. Consequently many authors still develop this technique by changing the set of parameters used for optimization in order to find the best coding efficiency with minimal computational cost [Kees96, Wieg01, Lee01].

These optimization methods improve coding efficiency, giving constant bitrate on the video coder output and gain in sequence quality. Nevertheless, they are often too complex to be used in video coders, especially in real-time coders.

3.4.2 Modeling methods

Another way for adjusting parameters is modeling the bitstream in order to predict how many bits will be needed for encoding an image with certain coding parameters and image content.

Various aspects of video coding can be modeled, for example, video source [Laza94, Mall98] modeling, video traffic modeling [Liu01], Rate-Distortion relation modeling [Lino96, Lino98], parameters estimation [Sawg97, Sawg98] etc. Both video sources and traffic modeling rely on analyzing of video content. The first one concentrates on image content and the latter one on the differences between images (motion analysis). Both the above methods are useful for VBR mode of coding if constant sequence quality is demanded. However, the Rate-Distortion modeling [Chia97] is much more versatile, and can be used in either VBR or CBR mode of coding. Therefore, this one will be briefly presented below.

3.4.2.1 Rate-Distortion curve modeling

G. Siemek proposes control algorithm with a rate-distortion model that selects the number of significant (nonzero) coefficients to be encoded in a macroblock [Siem01]. Hence, this is model of the number of encoded coefficients for given quantization scale factor Q. The average cost of encoding the significant coefficient is modeled by the following formula:

$$\frac{R_i}{M_i} = C + \log_2 \frac{N}{M_i}, \tag{3.16}$$

where N is the block size, M_i is the number of significant coefficient to be coded, R_i is the estimated bit cost per block and C is an empirical model constant. Further, the distortion function is described by the simple formula:

$$D_i = L\frac{D_{MAX\,i}}{M_i}, \tag{3.17}$$

where $D_{MAX\,i}$ denotes maximum distortion value and L is the empirical model constant. In order to find the number of the nonzero coefficients $M_1, \ldots, M_k$ that minimizes the distortion D_i with the constraint of total number of bits, the new expression is obtained:

$$M_1,\ldots,M_k,\lambda = \underset{M_1,\ldots,M_k,\lambda}{\arg\min} \sum_{i=1}^{K} L\frac{D_{\max i}}{M_i} + \lambda\left[\sum_{i=1}^{K} M_i C + M_i \log_2 \frac{N}{M_i} - R\right], \tag{3.18}$$

This control algorithm finds rate-distortion trade-off by the Lagrange optimization of formula 3.18. As a result, the number of coefficients which should be encoded is indicated.

Another, more accurate and much simpler model is proposed by Y. Kim, Z. He and K. Mitra [Kim01]. They tried to find the best expression for the bitrate *R* in terms of the quantization scale factor *q*.

The *p* parameter is introduced and denotes percentage of zeroes among the quantized transform coefficients. The value of *p* monotonically increases with *q*. Hence, the coding bitrate *R* will be function of *p*, denoted by *R(p)*. It means that rate is expressed in *p* domain. For MPEG-2, H.263 and MPEG-4 video coding systems *R(p)* is always a linear function (Fig. 3.7). Moreover, for any type of video frames such as I-frames, P- and B-frames the relationship between *R* and *p* is always linear. However, this approach is not accurate enough. Therefore, proposed bitstream estimation contains significant approximation error (see: chapter 5 and figure 5.7).

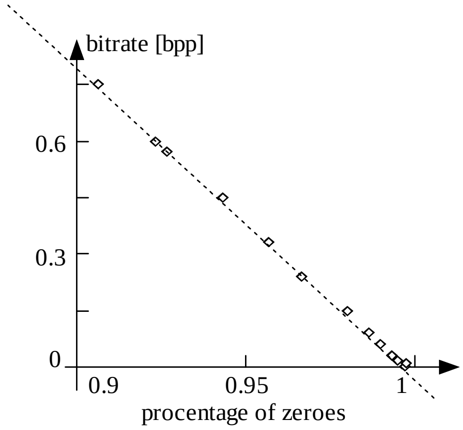


Fig. 3.7 The near-linear relation between generated bitstream and the percentage of non-zero coefficients.

The authors proposed following linear source model:

$$r(p) = \theta \cdot (1 - q), \tag{3.19}$$

where θ is a constant for each frame. Let *$D_0(x)$* and *$D_1(x)$* be the distributions of the DCT coefficients, then for any *q*, the corresponding *p* can be computed as follows:

$$p(q) = \frac{1}{M}\sum_{[x]\le q} D_0(x) + \frac{1}{M}\sum_{|x|\le 1.25\cdot q} D_1(x), \tag{3.20}$$

where M is the number of coefficients in current frame. The only parameter in 3.19 is the slope θ. Unfortunately θ has a large variation. The authors proposed to estimate this value according to the following formula:

$$\theta = \frac{R_m}{384 \cdot N_m - p_m}, \tag{3.21}$$

where R_m is a number of bits used to encode N_m macroblocks. Despite the fact that the value of θ has a large variation and previous encoded macroblocks are used for estimation of θ, the value of θ is determined with sufficient accuracy. A simple formula is used for bitrate control:

$$R = R_T - B_0 + \alpha \cdot B_T \tag{3.22}$$

where R_T is the target bitrate per frame, B_T is the encoder buffer size and B_0 is number of bits in the buffer. Parameter α is set by default to 0.2. Control algorithm determines the quantization scale factor Q in three steps:

Step 1: Initialize. Set $N_m=R_m=p_m=0$, compute the distributions $D_0(x)$ and $D_1(x)$ and set $\theta = 7.0$ which is the average value.

Step 2: Determine the quantization scale factor Q. According to formula 3.11 the number of zeroes produced by remaining macroblocks should be:

$$p = 384 \cdot (M - N_m) - \frac{R - R_m}{\theta}, \tag{3.23}$$

and from 3.18 the Q is determined.

Step 3: Update. All parameters and especially θ are updated.

The control algorithm based on R-D modeling has been implemented by the authors in the H.263 and MPEG-2 video coders. Presented results show that rate control algorithm meets the target bitrate more accurately than control algorithm in the test model, and also gives better average quality of encoded sequence. Gain is about 0.1÷0.5 dB for low bitrates modes (comparing to standard control algorithm TM5 results).

These models are very simple, however they do not estimate the average codeword length and have all the disadvantages of control algorithms with buffer feed-back. Main feature of these algorithms is an easy and fast computation/update of models parameters with low computational cost and simple control algorithm. Unfortunately they do not look-ahead in order to improve the accuracy of models parameters estimation. Moreover, none of the proposed algorithms takes into account the changes of encoded video sequence quality. They allow fluctuations of quality which can give subjectively poor quality of video sequence.

Similar approach is described in papers [Ding96, Lee96, Stre97, Stuh00]. Those papers however, concentrate on real-time application. The real-time control algorithm operation is achieved but modeling accuracy is decreased. Two pass approach is proposed by Westering [West99]. This approach is more efficient but it can be used only for off-line processing.

3.4.3 Sequence feature analysis

Some techniques based on video sequence feature analysis, like picture and motion activity measure or edge and texture activity, can help bit allocation algorithm to achieve near optimal allocation with maximized image quality for a given bitstream [Dawo98, Dawo98a]. Several methods of image analysis are presented below. Some of them are used in local control algorithm, others are much too complex to use them for real-time or even for off-line encoding.

3.4.3.1 Picture activity

Problem of activity measure is connected with transform coding. It means that the block activity measure should be correlated with quality of encoded area and with bitstream needed for encoding of this block (by transform and VLC coding) [Feri02, Kim99, Libh95].

Ferin proposed picture activity analysis method in order to introduce adaptive quantization. Such adaptive quantization applied on macroblock level aims to reduce the amount of quantization noise in areas where it is most visible to the Human Visual System (HSV). Quality of area with fine, high-contrast texture is decreased, and additional bits gained in this way are allocated to areas where HVS is more sensitive to noise or some coding artifacts. The analyzed blocks are divided into three groups: textured blocks, flat blocks and mixed blocks (Fig. 3.8).

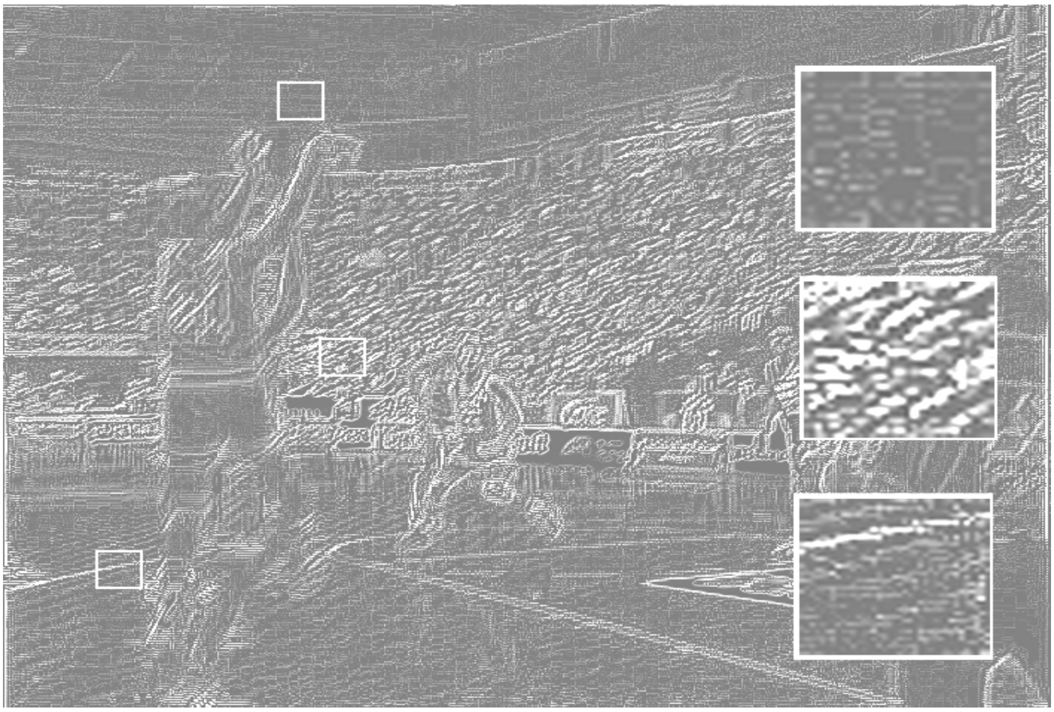


Fig 3.8 Sample picture filtered in order to detect edge and texture. Three rectangles mark flat surface, textured surface and detected edge.

Bits from textured region are moved to mixed region. For each block indicator q_{noise} is defined which serves as a measure of perceptual visibility of the quantization noise:

$$q_{noise} = \alpha \cdot rng^{\gamma} - \beta \cdot act^{\delta}, \tag{3.24}$$

where $\alpha, \beta, \gamma, \delta$ are empirically determined constants. The parameter *rng* describes amount of ringing noise. Proposed activity measure for block is defined as follows:

$$block\_act_{(x,y)} = \sum_{i,j} \left| \frac{\partial f}{\partial x}(x+i, y+i) + \frac{\partial f}{\partial y}(x+i, y+i) \right|, \tag{3.25}$$

The sub-block activity is high for textured blocks and low for flat ones. Macroblock activity is described as follows:

$$act_{(x_0,y_0)} = \sum_{i,j} block\_act_{(x_0+4i, y_0+4j)}, \tag{3.26}$$

and:

$$rng_{(x,y)} = \sum_{(x_a,y_a,x_b,y_b)\in P} \left| block\_act_{(x_a,y_a)} - block\_act_{(x_b,y_b)} \right|, \tag{3.27}$$

By adding such an algorithm to the local control mechanism, a slightly better perceptual image quality is obtained. Image quality according to objective measure (PSNR) is not always improved. Bits spent for encoding a certain frame maintain unchanged. Other similar algorithms [TMN5, Cort91, Kim99, Krun97a, Ryu00, Schu97a] use different activity measure functions and various bit allocation strategy, but the main goal is the same – to improve image quality by local modification of quantization scale factor *Q* and to match better to a given bitrate. Great advantage of those methods consists in taking into account the whole image, so the nearly optimal bit allocation is achieved.

3.4.3.2 Motion parameters estimation

Some methods of estimation and determination of motion parameters have been developed [Serv98, Guin99]. By motion parameters we mean horizontal shift, vertical shift, rotation and zoom, as well as motion activity through the whole video sequence [Giun99]. G. Giunta proposes method of measurements for several parameters of motion. First, he defines a motion model using complex variables for denoting the Cartesian coordinates in the form $z=x+jy$. The two frames are considered: reference frame $s(x,y)=s(z)$ and moved frame $r(x,y)=r(z)$. Each point in affine motion model can move during zoom, rotation or translation. Such model is described as:

$$z' = a \cdot z + d, \tag{3.28}$$

where the translation is denoted as:

$$d = d_x + jd_y, \tag{3.29}$$

and zoom with rotation:

$$a = pe^{j\theta}, \tag{3.30}$$

all this together gives a movement description:

$$r(z) = a\left[(z-d)/a\right] + e(z), \tag{3.31}$$

where $e(z)$ is an error image. Proposed motion estimation in its first step uses a standard block matching algorithm (BMA) for evaluating of displacement ζ. Standard matching criterion (mean absolute difference function MADF) have been used:

$$MADF_i(\zeta) = \frac{1}{N}\sum_{z \in B_i} \left| r(z) - s(z+\zeta) \right|, \tag{3.32}$$

In the second step the sub pixel accuracy of estimation is achieved by 2-D parabolic interpolation. A dense motion field or sparse one can be used, according to the application. Subsequently the obtained results are used in control algorithm. The obtained information is very detailed and tells us about each kind of movement.

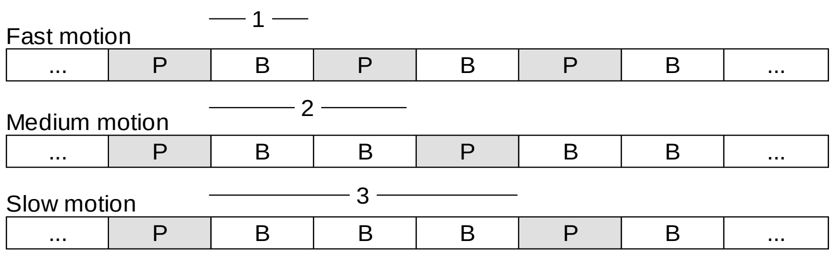


Fig. 3.9. Exemplary GOP parameters adjustment according to movement estimation.

Moreover, the value of movement parameters determines if the movement is fast or slow. Such information is helpful for GOP parameters adjustment (global control mechanism) like GOP length or time distance between two P frames. Also detailed local movement information helps to segment the image and to allocate bits for every region of movement respectively. On the figure 3.9 one can see various GOP structures. Having information about movement in video sequence the coder can switch between these structures. In case of slow motion, the number of B-frames can be increased causing coding efficiency improvement.

3.4.3.3 Scene editing detection

The previous methods aim to improve local quality of encoded image to make the most of given bitstream, but they perform on local level of control. Methods of scene editing detection improve coder performance by adjusting global control parameters like GOP structure or frame format (Inter/Intra) [Katt95, Lee97, Lou97]. Furthermore the detected feature could be used for further video sequence edition, sequence indexing and metadata generation. The obtained information about sequence also ameliorates local control.

Let us consider a scene cut case. Such situation is shown on figure 3.10 when sudden scene change occurs. Mostly constant GOP length is used under encoding video sequence, typical parameters of GOP are: GOP length = 15 and two B-frames between P-frames. In such case, a scene cut which appears between two coding reference frames causes the increase in the bitstream needed for encoding them (Fig. 3.10) because one of reference frames becomes useless.

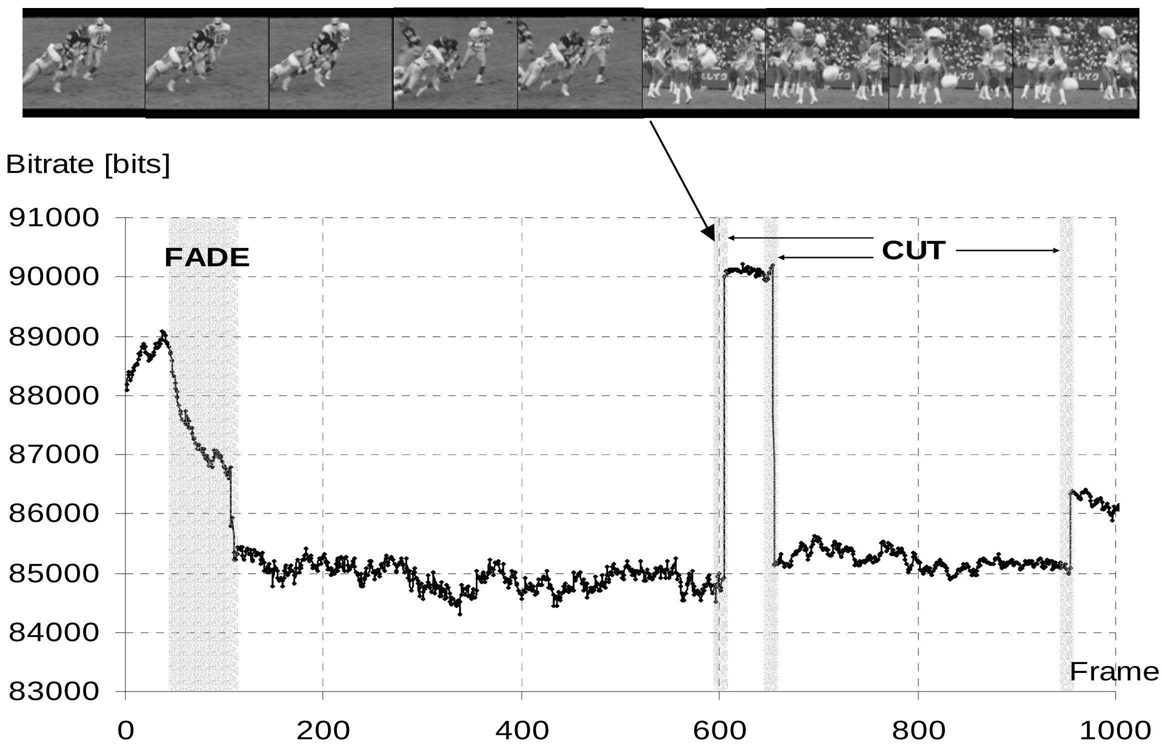


Fig. 3.10 Exemplary sequence with scene cut and the bitstream of encoded DC coefficients (MPEG-2) of this sequence.

As the standard bit-allocation algorithms do not use look-ahead, the coder is not aware of scene change and allocates bits in a wrong way. For example, D. Ferin proposes technique of GOP structure adjustment through scene change detection based on pictures difference histogram measures [Feri02]. Look-ahead of $(s_{max}+1)$ frames has to be applied. First of all, value of histogram differences $d(n,n+1)$ founded for all pairs $(n,n+1)$ of successive pictures is searched. The value which is above certain threshold t_u appoints scene cut causing new GOP to start with an I-frame (Fig. 3.11).

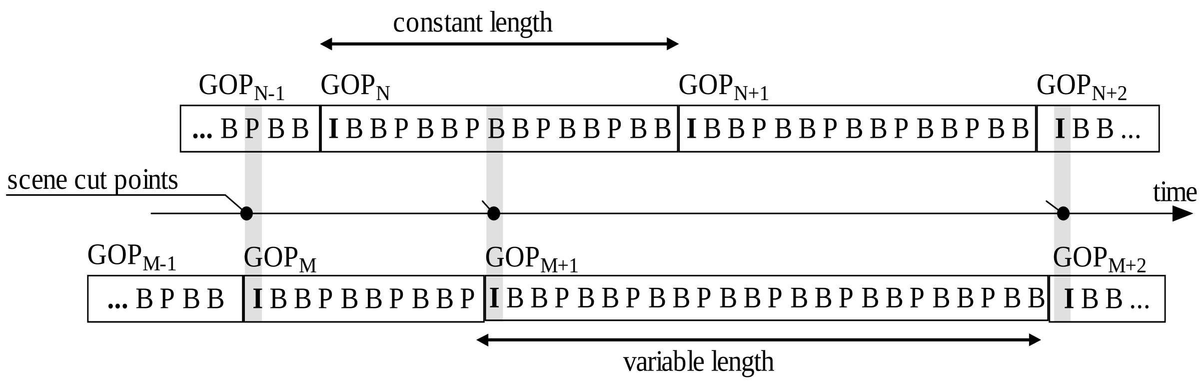


Fig. 3.11 Standard GOP structure and the GOP structures fitted well to sequence characteristic.

Ferrin introduces a scene change detection weighting factor w_i which is defined as follows:

$$w_i = \begin{cases} \frac{1}{2}\left(1+\left(\frac{i-s_{\min}}{m-s_{\min}}\right)^{\alpha}\right) & \text{for} \quad i \leq m, \\ \frac{1}{2}\left(1+\left(\frac{s_{\max}-i}{s_{\max}-m}\right)^{\alpha}\right) & \text{for} \quad i > m, \end{cases} \tag{3.33}$$

where i is the number of frame which is marked as the first frame in a new scene, s_{min} and s_{max} denote minimum and maximum available GOP length. This parameter changes the GOP length and sets detection range. By adding the second threshold t_l it is possible to find fast movement or fade pieces of a video sequence. The differences histogram for those two thresholds determines the cases mentioned above.

The authors of [Sawg96] combined feed-forward buffering, scene change detection with control unit of the quantization scale factor Q. To re-order frame types for encoding the input frame buffer in used. Therefore, the advantage of the frame delay can be taken. MPEG-2 encodes P and I frames first, if a scene change occurs in a B frame the encoder realizes this in advance of encoding the P or I frame. The authors show that this feature can effectively be used for advance adjustment of quantization scale factor. They propose a non-linear curve to estimate future VBV buffer occupancy, the SCF function (scene change function), and also the MVF function (motion vector function). These functions calculate the ratio of the variance of a difference frame to the variance of an input frame and the mean value of the motion vector function in a slice. The output values of SCF and MVF impact control block of the quantization scale factor Q. Te GOP structure in this solution is not changed only Q factor is adjusted.

Other authors [Mott00, Park96, Yuzh98] also improve control algorithm by adding sequence editing detection tools. The disadvantage of all those methods is the necessity of buffering of many forward frames. It means that large encoding delay is introduced. That is why those methods are designed rather for off-line coding than real-time. On the other hand, such GOP adjustment can significantly decrease needed bitstream of about 20% and enables to avoid sudden changes of quality level or bitstream values.

3.4.4 Default bitrate control algorithm of Test Mode 5 (TM5) of MPEG-2 video coder standard

All the described methods and also the author's methods of bitrate control have been compared with default control mechanism included in TM5 MPEG-2 verification coder model. Therefore, this default control mechanism [Test5] will be presented in this section. That control mechanism is based on buffer feedback control and picture activity measure.

The MPEG-2 standard describes only a buffer control mechanism of bitstream decoder. In consequence each encoder uses a different control algorithm in order to obtain the required quality and bitrate of video sequence. Encoding efficiency of different encoders is compared to Test Model (actually ver.5 – TM5). This encoder model has implemented standard bitrate control algorithm with adaptive quantization mechanism. That control mechanism allocates bits in bitstream GOP by GOP. Every GOP is given the same number of bits.

The implemented algorithm works in there steps. The first step is to estimate the number of bits available for encoding next I, P or B-frame. Next, sets the reference value of the quantization parameter Q for each macroblock. At last, the reference value of the quantization parameter Q according to spatial activity in the macroblock is modulated. To estimate the number of bits for one frame of certain type, the complexity factor X is calculated:

$$
\begin{aligned}
X_I &= S_I \cdot Q_I \\
X_P &= S_P \cdot Q_P \\
X_B &= S_B \cdot Q_B
\end{aligned}
\tag{3.34}
$$

where S_I, S_P, S_B are the number of bits generating by encoding actual picture, and Q_I, Q_P, Q_B are the average quantization parameter for all encoded macroblock including skipped macroblocks. Initial values in case of new GOP are:

$$
\begin{aligned}
X_I &= \frac{C_I \cdot bit\_rate}{115} \\
X_P &= \frac{C_P \cdot bit\_rate}{115} \\
X_B &= \frac{C_B \cdot bit\_rate}{115}
\end{aligned}
\tag{3.35}
$$

where variable *bit_rate* is the required bitrate per second in bits. The constants C_I, C_P, C_B are empirical factors and they give best bit allocation between different types of frames in standard applications. Default values of these constants for MPEG-2 system are as follows:

$$
\begin{aligned}
C_I &= 160, \\
C_P &= 80, \\
C_B &= 42,
\end{aligned}
\tag{3.36}
$$

Next, the target number of bits is computed:

$$T_I = \max\left[\frac{R}{1+\frac{N_P \cdot X_P}{X_I \cdot X_P}+\frac{N_B \cdot X_B}{X_I \cdot X_B}}, \frac{bit\_rate}{8 \cdot picture\_rate}\right],$$

$$T_P = \max\left[\frac{R}{N_P + N_B \cdot \frac{K_P \cdot X_B}{K_I \cdot X_B}}, \frac{bit\_rate}{8 \cdot picture\_rate}\right], \tag{3.37}$$

$$T_B = \max\left[\frac{R}{N_B + N_P \cdot \frac{K_B \cdot X_P}{K_P \cdot X_B}}, \frac{bit\_rate}{8 \cdot picture\_rate}\right],$$

where N_P, N_B is a number of P- and B- frames remaining to encoding in the current GOP (Fig 3.1), K_P and K_B are constants which depend on the quantization matrices. In case when standard matrices are used, the constants K_P = 1.0 and K_B = 1.4. The remaining number of bits for a GOP is marked as R, and its value is updated after encoding a picture as follows:

$$\begin{aligned} R &= R - S_I && \text{for encoded I-frame,} \\ R &= R - S_P && \text{for encoded P-frame} \\ R &= R - S_B && \text{for encoded B-frame} \end{aligned} \tag{3.38}$$

where is S_I, S_P, S_B is the number of bits spent to just encode picture (as I-, P- or B-frame. Before encoding the first in a GOP it is necessary to initiate variables R:

$$R = G + R$$

where G is given as

$$G = bit\_rate * N / picture\_rate \tag{3.39}$$

where N is the number of pictures in the GOP (Fig. 3.11).

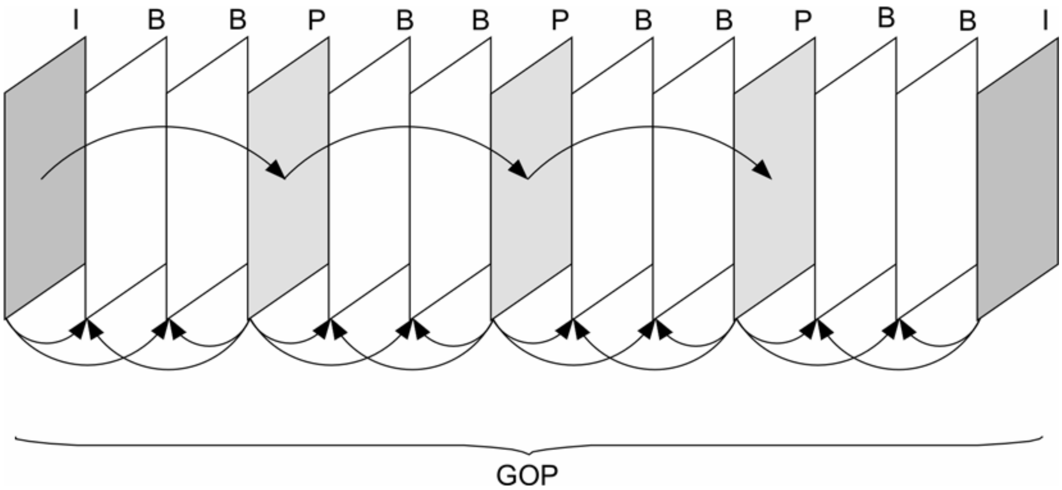


Fig. 3.12 Sample GOP structure with N=12.

The following step is to estimate *Q* factor for macroblock. Before macroblock encoding the fullness of the appropriate virtual buffer has to be computed:

$$\begin{aligned} d_j^I &= d_0^I + B_{j-1} - \left\lceil \frac{T_I \cdot (j-1)}{mbc} \right\rceil \\ d_j^P &= d_0^P + B_{j-1} - \left\lceil \frac{T_P \cdot (j-1)}{mbc} \right\rceil \\ d_j^B &= d_0^B + B_{j-1} - \left\lceil \frac{T_B \cdot (j-1)}{mbc} \right\rceil \end{aligned} \tag{3.40}$$

where d_j^I, d_j^P and d_j^B are a initial fullness of virtual buffers - one for each picture type. The *Bj* denotes the number of bits spent by encoder for all macroblocks in the picture up to the present one and including macroblock indexed as j. *T*he number of macroblocks in the picture is marked as *mbc*. The final fullness of the virtual buffer is used for encoding the next picture of the same type. Q factor can be computed as follows:

$$Q_j = \left(\frac{d_j \cdot 31}{r} \right) \tag{3.41}$$

where *r*:

$$r = 2 \cdot \frac{bit\_rate}{picture\_rate} \tag{3.42}$$

The initial value for the virtual buffer fullness is:

$$
\begin{aligned}
d_0^I &= 10 \cdot \frac{r}{31} \\
d_0^P &= K_P \cdot d_0^I \\
d_0^B &= K_B \cdot d_0^I
\end{aligned} \tag{3,43}
$$

Q factor can adapt to local picture activity. Adaptive quantization enables to match to the required bitrate. Adaptation based on picture activity factor act_j computed as follows:

$$act_j = 1 + \min(vblk_1, vblk_2, \ldots, vblk_8) \tag{3.44}$$

where $vblk_n$ is spatial activity measure for the set of eight luminance blocks in a macroblock, four organized as the frame and four organized as the field:

$$vblk_n = \frac{1}{64} \cdot \sum_{k=1}^{64} (P_k - P\_mean)^2 \tag{3.45}$$

where

$$P\_mean_n = \frac{1}{64} \cdot \sum_{k=1}^{64} P_k^n \tag{3.46}$$

where P_k is value of the sample in n-th block.
Normalised activity act_j is given as:

$$N\_act_j = \frac{(2 \cdot act_j) + avg\_act}{act_j + (2 \cdot avg\_act)} \tag{3.47}$$

Now the new value of adapted Q factor can be obtained:

$$mquant_j = Q_j \cdot N\_act_j \tag{3.48}$$

The obtained quantization scale factor Q can be changed in the macroblock, slice or picture header. Adaptive quantization helps to obtain the required bitrate but each change of Q costs 5 bits (MPEG-2 system). This method does not handle scene cut and VBV buffer compliance is not guaranteed.

3.5 Conclusions

The efficient control of video coders is still an open problem that gains a lot of attention. A lot of papers have been published to improve the rate control behavior and the various approaches have been described, but still every technique has some disadvantages. The goal of the control algorithm is to maximize the video quality by given bitrate constrains, or minimize the bitrate by assumed quality.

The approaches based on Lagrange multiplier give optimal set of encoding parameters. They adjust quantization scale factor Q (locally or globally) in an optimal way but such algorithms turn out very complex with high computational cost.

The methods of sequence and image feature analysis improve coding efficiency providing significant gain in sequence quality, but these methods control parameters of coding of the whole video sequence rather than a single frame. Hence, these methods themselves cannot constitute the control mechanism and are not suitable for the applications where small delay is required, because they are rather complex and need buffering of several future frames

The approaches based on image activity analysis improve the subjective quality and allocate the bits better. Such control algorithms operate on slice/macroblock level (locally). The most promising application of these methods would be additional modifying of the local quantization scale factor Q. One of the advantages of such approach is that the whole image is analyzed taking the incoming data into account. High computational complexity is its main disadvantage.

Other approaches to coder control which are based on feed-forward control expose difficulties with discontinues in the input signal characteristic [Choi94]. For example, for encoding flat portion of an image, quite small number of bits is required. The value of quantization scale factor Q is being slowly decreased in such flat area. After exceeding the edge of the textured region, the value of Q turns out to be too small because of unacceptably high bitrate. Therefore, the control unit tries to increase the quantization scale factor Q step by step (Fig. 3.12). During this process, some portion of the available bit budget is allocated in an improper way.

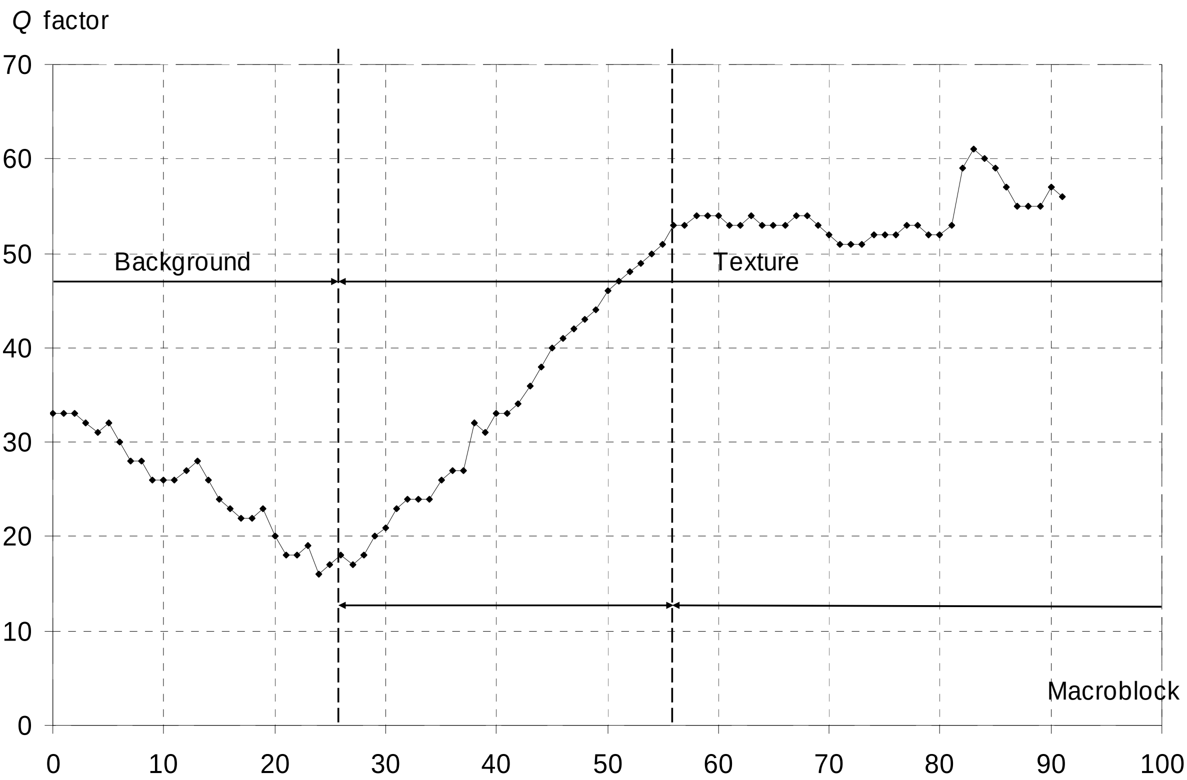


Fig. 3.12. Variations of the quantization scale factor *Q* versus the macroblock number *MN* in a picture of the test sequence *Basket.4cif*.

The control algorithm estimates the average global quantization scale factor *Q* and helps to avoid the mentioned situation (Fig. 3.12). The global value of quantization scale factor *Q* appears in picture header, whereas local value of *Q* is sent in the individual slice or macroblock headers if it is needed. The local control should modify the values of quantization scale factor *Q* in certain range around global *Q* in order to match the encoding process to local statistics within a frame. Moreover, local control algorithm helps to use output buffer in optimal way.

The goal of the dissertation is to establish a simple empirical model for a bitrate as function of *Q* set for individual frames. The model should be valid for typical conditions of operation of a coder. The model can be used for global control in the constant bitrate mode of operation CBR).

Chapter 4

Global model of video bitstream

4.1 Introduction

In this chapter the author presents his original model of video bitstream for the hybrid video coders. This global model was already published by author in [Doma02, Doma02b]. All the examples will be presented for the MPEG-2 video coder (main level with main profile) and for 4CIF sequences, both progressive and interlaced. Such configuration (MPEG-2 processing 4CIF video sequences) is mostly used in real applications for example Digital TV, DVD-Video etc

4.2 Video bitstream modeling

As the global quantization scale factor Q is sought, and the whole encoded image is considered, then the video coder will be treated as a black box. It is assumed that video coder is optimized in such a way that for particular set of parameters and for given quantizer scale factor Q gives as high quality of encoded image as possible for the required video bitstream.

The number of bits B allocated to an individual frame is a sum of the component B_{CONST} that does not depend on the quantizer scale factor Q and the component $B_{\mathrm{VAR}}(Q)$ that depends on Q,

$$B = B_{\mathrm{CONST}} + B_{\mathrm{VAR}}(Q). \tag{4.1}$$

The B_{CONST} part is defined as follows:

$$B_{CONST} = B_{CTRL} + B_{\mathrm{YDC}} + B_{\mathrm{CDC}} + B_{\mathrm{MV}} \tag{4.2}$$

The bitstream B_{CTRL} is the number of bits needed for headers, B_{YDC} and B_{CDC} for Intra DC coefficients, and B_{MV} for motion vectors. As the bitstream B_{CTRL} consist of picture headers, slices and macroblocks that little depends on Q, but this variation can be neglected. (Table 4.1). The components B_{YDC} and B_{CDC} exist only in Intra coded macroblocks, whilst B_{MV} exists in P- and B-frames. The components B_{YDC} and B_{CDC} are independent from Q factor and are constant even in time (Fig. 4.2b).

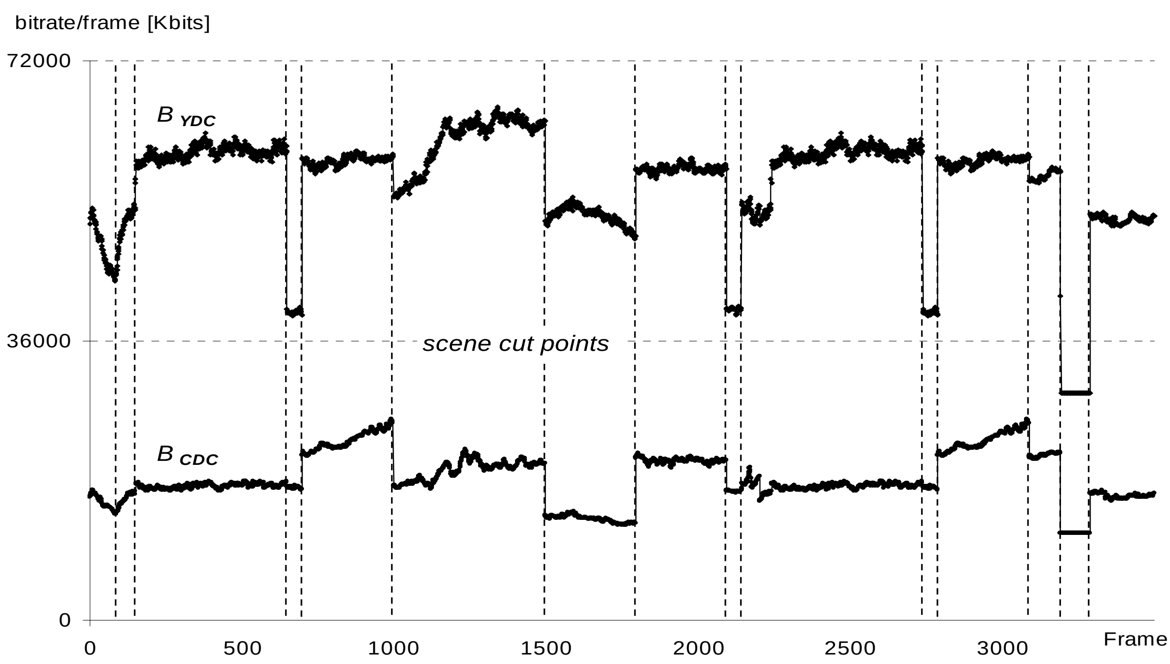
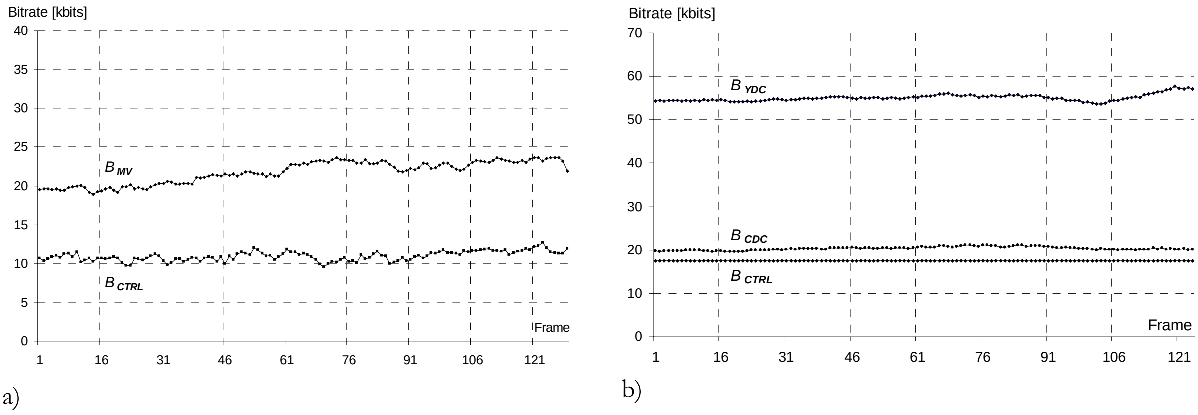


Fig.4.1. Changes of the B_{YDC} and B_{CDC} components if some scene cuts points exist. Test performed for several sequences in 4CIF format (MPEG-2 system).

As the coding mode (Intra or Inter) is chosen independently from the current value of the quantizer scale factor Q, the value B_{CONST} can be calculated during the first stage of the frame encoding process, i.e. during those coding operations that do not depend on the quantizer scale factor Q and do not vary in time, but one additional assumption should be done that only fragment of vide sequence between two scene cut points is considered.

Figure 4.1 shows an exemplary sequence containing many pieces of various sequences. The placement of scene cut points and their frequency impact on the results of coding. All bitstream components change rapidly in the sequence cut points, while between cuts within the section of sequence these bitstreams are relatively stable. Therefore, we assumed that the sequence is encoded between two scene cuts. Each scene editing point should be detected and all the control algorithm parameters should be recalculated and the new GOP with Intra frame should start.

On the other hand, the value of B_{CONST} depends on a frame type (I, P or B) and frame content, but does not depend on the quantizer scale factor Q. The analysis of changes in time of bitstream components shows that the B_{CONST} can be predicted using information about the previous encoded frame of this same type (I, P or B). However, when the previous frame is unavailable (at the GOP beginning for example) the B_{CONST} has to be predicted by exploiting data from currently being encoded frame. Figure 4.3 shows dependency between parts of the B_{CONST} component and quantization scale factor Q.



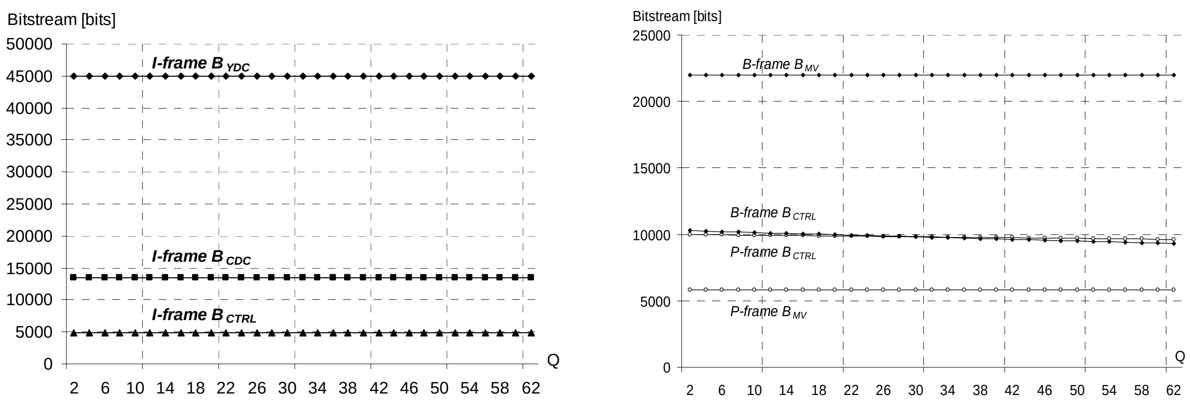
a) b)

Fig. 4.2 The independent from Q parts of the bitstream versus frame number. Components a) B_{MV}, B_{CTRL} (P- and B-frames) and b) B_{YDC}, B_{CDC} and B_{CTRL} (I=frames) for sequence of *Basket*. (MPEG-2 system with constant $Q=32$)

Let us consider the bitstream component that directly depends on the frame quantization factor *Q*:

$$B_{VAR}(Q) = B_{YV}(Q) + B_{CV}(Q) + B_{CBP}(Q) \tag{4.3}$$

where $B_{YV}(Q)$ and $B_{CV}(Q)$ denote the bits needed for encoding of the DCT coefficients (except the Intra DC ones) in the luminance and the chrominance, respectively. The $B_{CBP}(Q)$ is number of bits needed to encode field of *CodedBlockPatern* (CBP). That field occurs only in P- and B-frames and is encoded using VLC codes. The CBP field can be computed only after quantizing of DCT coefficient which means after the quantization scale factor *Q* is chosen. Hence, the computation of the bitstream for the CBP field is impossible. The value of the B_{CBP} has to be estimated.



a) b)

Fig.4.3 Constant parts of the bitstream versus quantization scale factor *Q* (MPEG-2 system) for firs frame from video sequence *Basket*, for a) I-frames and b) P- and B-frames.

The shortest VLC code for the CBP is 3 bits, and the longest is 9 bits for MPEG-2 system. In general, $C_{l\,MAX}$ is the length of the longest CPB Huffman codeword in some video coding system, hence:

$$B_{CBP} = MN \cdot C_{FT} \cdot C_{l\max} \cdot \left[1 - \left(\frac{Q}{Q_{MAX}}\right)\right],$$

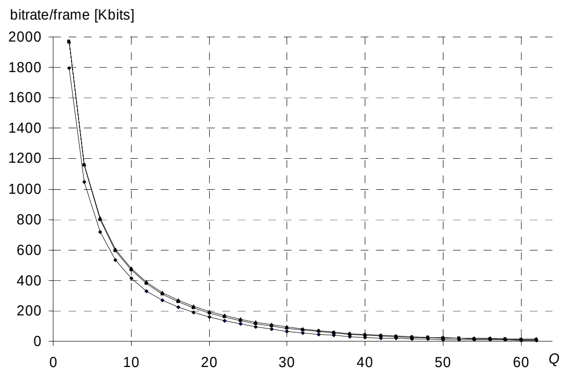
$$C_{FT} = \begin{cases} 0 & for\ I-frames \\ 0.5 & for\ P-frames \\ 1 & for\ B-frames \end{cases} \tag{4.4}$$

where MN is a macroblock number, C_{FT} is a constant dependent on type of frame, and Q_{MAX} denotes the maximum allowed value of quantizer scale factor Q (e.g. 62 for H.263 and 62 or 112 for MPEG-2). The length approximation of the CBP field by linear function is sufficient for estimation of total bitrate. The approximation error is always less then 0.5% due to small number of bits needed for encoding the CBP field in comparison to the whole bitstream needed for encoding the whole frame.

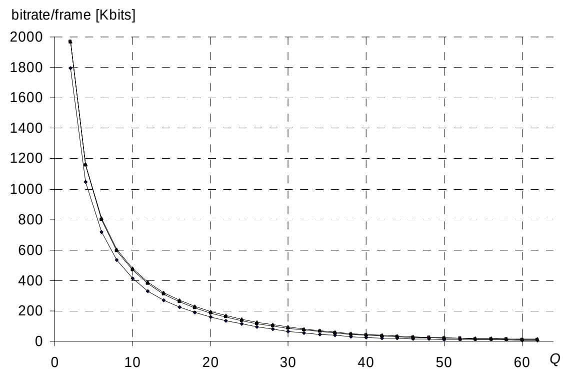
Table 4.1 Average values and standard deviation of bitstreams B_{CTRL}, B_{YDC}, B_{CDC}, B_{MV} (MPEG-2 system) obtained for encoded 128 frames of video sequence. In the P- and B-frames the Intra coded macroblocks are not taken into account.

Frame Type	B_{CTRL}		B_{YDC}		B_{CDC}		B_{MV}	
	Average [bits]	***Standard deviation [bits]***	***Average [bits]***	***Standard deviation [bits]***	***Average [bits]***	***Standard deviation [bits]***	***Average [bits]***	***Standard deviation [bits]***
Flower Garden								
I	14364	***13***	15279	***37***	15279	***37***	-	-
P	10075	***632***	-	-	-	-	9650	***296***
B	12121	***320***	-	-	-	-	22358	***889***
Cheer								
I	14383	***12***	43714	***832***	16872	***389***	-	-
P	5674	***361***	-	-	-	-	7724	***263***
B	11522	***311***	-	-	-	-	19614	***797***
Stefan								
I	14382	***11***	38330	***439***	12532	***146***	-	-
P	8842	***162***	-	-	-	-	11324	***698***
B	10682	***226***	-	-	-	-	18162	***1136***

Further, considering $B_{YV}(Q)$ and $B_{CV}(Q)$ components, the idea is to find a function $f(Q)$ which denotes $B_{YV}(Q)$ and $B_{CV}(Q)$, as well as best matches experimental data (Fig 4.4). On the Figure 4.4 one can see some exemplary curves.



a)



b)

Fig 4.4 Value of the bitstream for frame versus quantization factor Q. Example for a) luminance and b) chrominance components for different I-frames in video sequences *Basket* and *Cheer*

Those are monotonous diminish functions in the range of Q 2÷64. All the curves become very small and almost constant for high quantization values.

Because B_{CONST} values can be estimated, predicted or even computed before frame encoding therefore only $B_{VAR}(Q)$ component needs to be considered. The problem of the quantizer adjusting can be redefined as follows: for a given bit number B for a current frame, the quantize scale factor Q is calculated from:

$$B_{\mathrm{VAR}}(Q) = B - B_{CONST}. \tag{4.5}$$

This kind of approach needs a simple empirical model for the bitrate B_{VAR} as function of Q set for individual I-frames, P-frames and B-frames (Fig. 4.4). All those frames little differ in encoding results in sundry approximation parameters. We assume that such model should be valid for typical conditions of the coder operation (see: chapter 3).

For the sake of simplicity, the absence of adjustment of Q in the individual macroblocks is assumed.

4.2.1 Bitstream model

In order to find parametric function, the method of function fitting to experimental data is applied. First, one formula is denoted from different classes of functions like exponential, logarithmic and others which is the most suitable to experimental data. Next, the values of function parameters are estimated. The function which approximates experimental data subjectively well is selected by graph comparison. Because this method can give mistakenly results, the concordance of formula and experimental data has to be checked. It is done by the equalization method, which means that x axis and y axis are transfigured:

$$X = \varphi_X(x, y), \quad Y = \varphi_Y(x, y), \tag{4.7}$$

where (x,y) values are the original coordinates, the φ_X and φ_Y are transfigure functions. The X and Y are new transfigured coordinates. The assumption is that axes are related.

Such formulas are well-described in the literature [Fich97, Bron00, Bran99] and commonly used for a task like this. The goal is to find best parametric function which well-approximates experimental data and which has as few parameters as possible. It is done in a few stages.

- The function class which graphically matches the experimental data is chosen.
- The match is verified according to equalization method.
- The above two stages are repeated until the proper match is found.

For the experimental data obtained (Fig. 4.6a), it has been noticed that experimental curve can be approximated with hyperbolic functions (Fig. 4.6). In order to verify this observation the equalization method is to be applied. Therefore, y axis has been transfigured in the following way:

$$Y = a \cdot x / y, \tag{4.8}$$

and much simpler curve has been obtained (Fig.4.6). The *a* parameter is not so important being only a scale for better presentation of results.

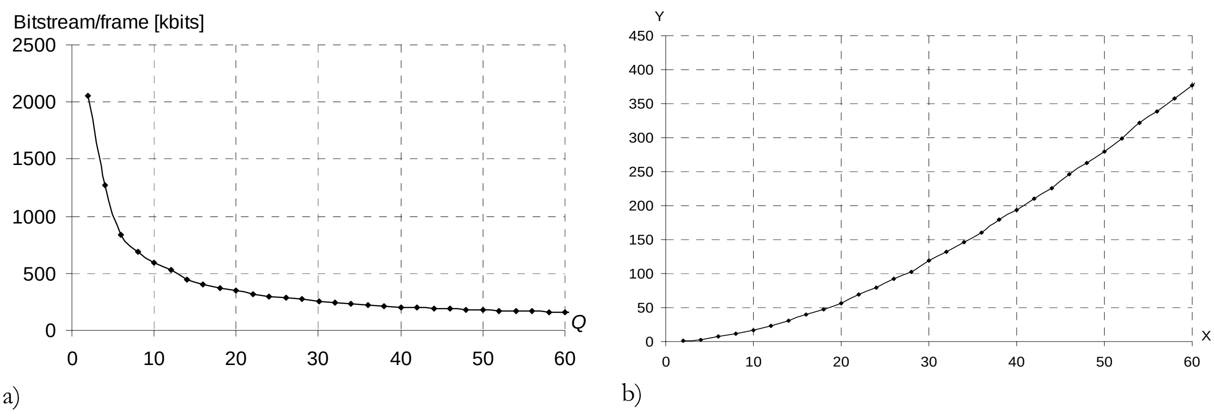
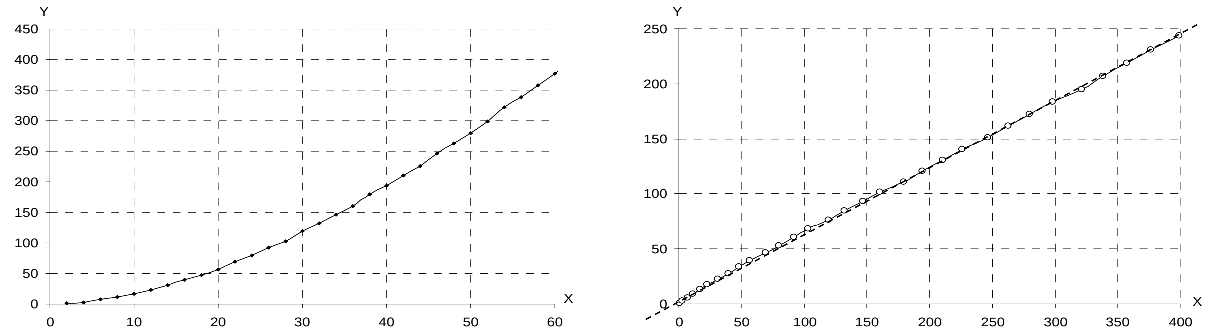


Fig.4.6 Example a) empirical curve before and b) after transfiguring of Y axis according to formula (4.10).

Subsequently, the *x* axis has been changed in order to obtain linear function. The conversion is as follows:

$$X = c \cdot x^b, \tag{4.9}$$

Figure 4.7 shows the curve obtained after transfiguring the X axis.



a) b)

Fig.4.7 Curve a) before and a) after transfiguring of X and Y axes according to formula 4.11.

The curve is near linear with acceptable deviation. Aggregating all transfigures the following formula is obtained:

$$y = \frac{a}{c \cdot x^b + d} \quad , \tag{4.10}$$

Where *a, b, c* and *d* are parameters of the approximation function. In order to increase flexibility of function the parameter *e* has been added. A new function is created:

$$B_x(Q) = \frac{a}{c(Q^b + e) + d} \quad , \tag{4.11}$$

where *a, b, c, d* and *e* are parameters of the function which have to be estimated. Figure 4.8 shows an exemplary approximation using proposed function.

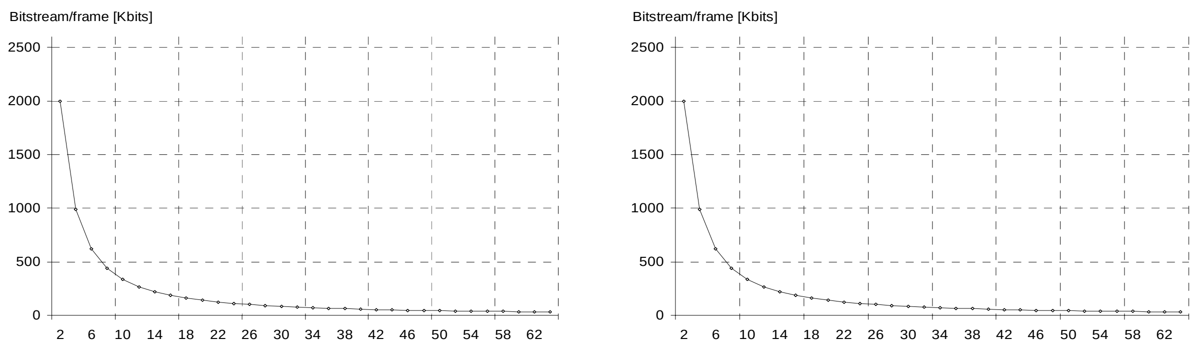


Fig. 4.8. The approximation function for exemplary parameters a=6*10[6], b=0.91, d=0.11, e=0.86 and a) c=2.05 or b) c=1.45.

The method of function fitting to experimental data turns out very efficient and easy. The equalization method confirmed that hyperbolic function fits the experimental data well. Obtained results of this approximation are very promising. As a next step, the parameters of function need to be found and the accuracy of approximation has to be checked.

4.3 Estimation of the model parameters

Parameters values can be estimated by minimization of approximation error over the whole interval of the allowed Q values (it is 2÷62 for H.263 and 2÷62 or 1÷112 for MPEG-2). The precision of experimental data B_e approximation is measured by

$$\varepsilon_B(Q) = \frac{\left|B_x(Q) - B_e(Q, a, b, c, d, e)\right|}{B_e(Q)} \cdot 100\%, \tag{4.12}$$

where $B_x(Q)$ denotes measured value and $B_e(Q)$ denotes approximated value. The minimization is performed as follows:

$$\min_{a,b,c,d.e} \max_{Q} \varepsilon_B(Q, a, b, c, d, e) , \tag{4.13}$$

To solve this problem the Quasi-Newton method of function minimization has been used [Fort95]. The figure 4.9 shows several curves of approximation function $B(Q)$ for various parameters c value. This is the main parameter because it inflects function allowing matching to the real curve and is correlated with image content the most.

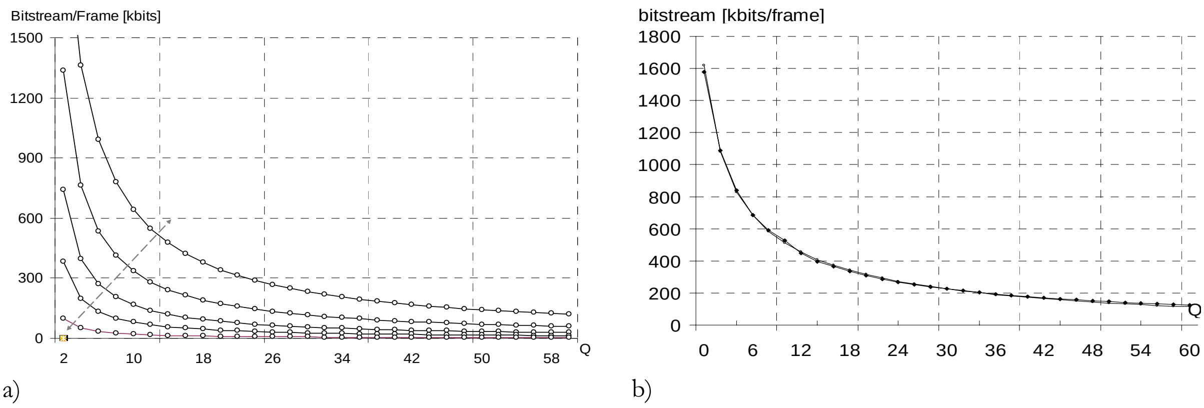
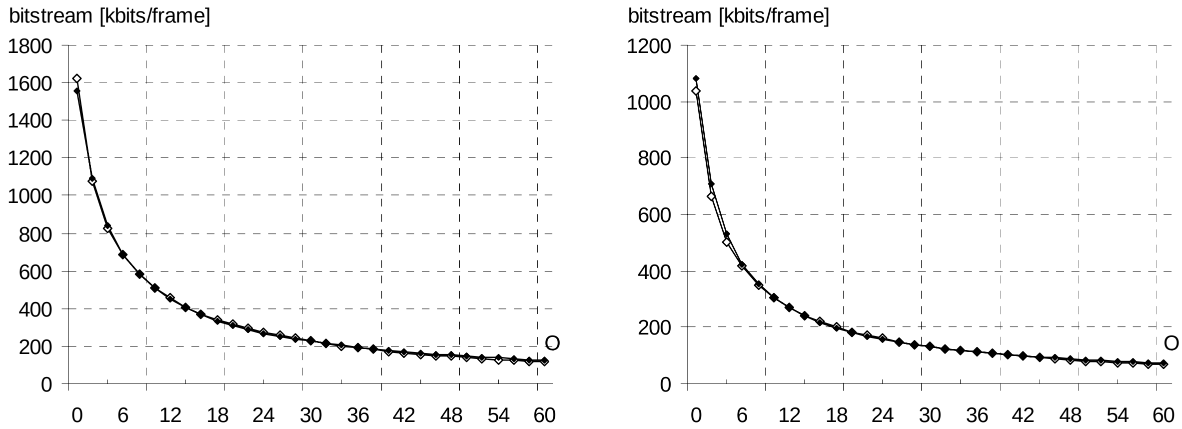


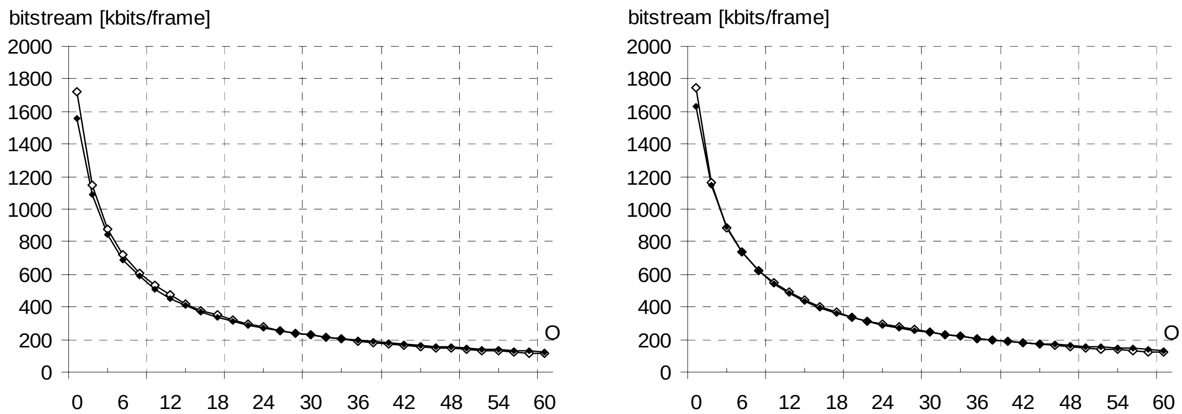
Fig. 4.9 Exemplary a) functions $B(Q)$ for several values of c parameter and b) approximation of experimental data for one frame of sequence *Basket.*

If the parameters are estimated for individual frames, the model $B_x(Q)$ approximates the experimental data $B_e(Q)$ so well that often it is difficult to distinguish the experimental and the modeled curve (Fig. 4.9b). The maximum approximation errors appear for small and large values of quantization scale factor Q. When Q is small many *RL*-pairs are *ESCAPE* coded which causes sudden bitstream increasing. Therefore, the experimental curve sometimes is not a continuous function.



a) b)

Fig. 4.10 Graph of curves of luminance bitrate approximation and experimental data for frames in sequences *a) Basket b) Football*

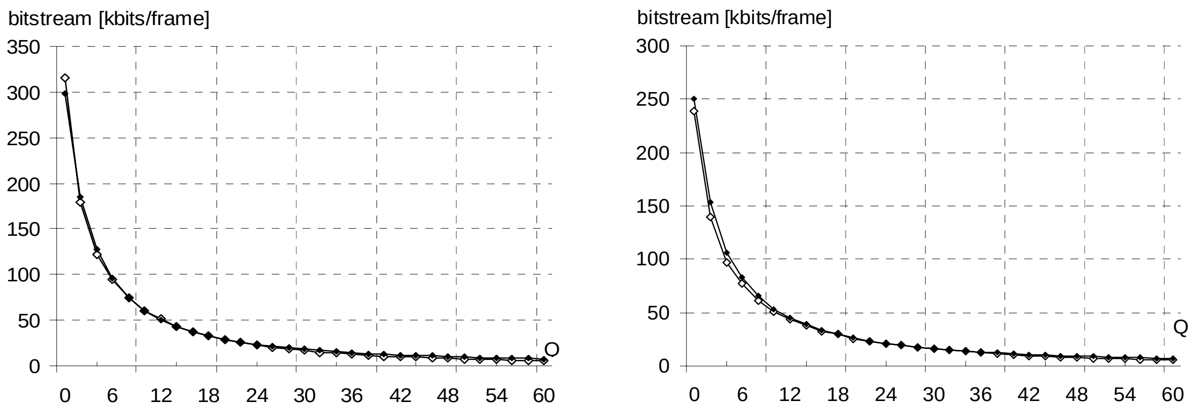


a) b)

Fig. 4.11 Graph of curves of luminance bitrate approximation and experimental data for frames in sequences a*) Cheer b) Mobile.*

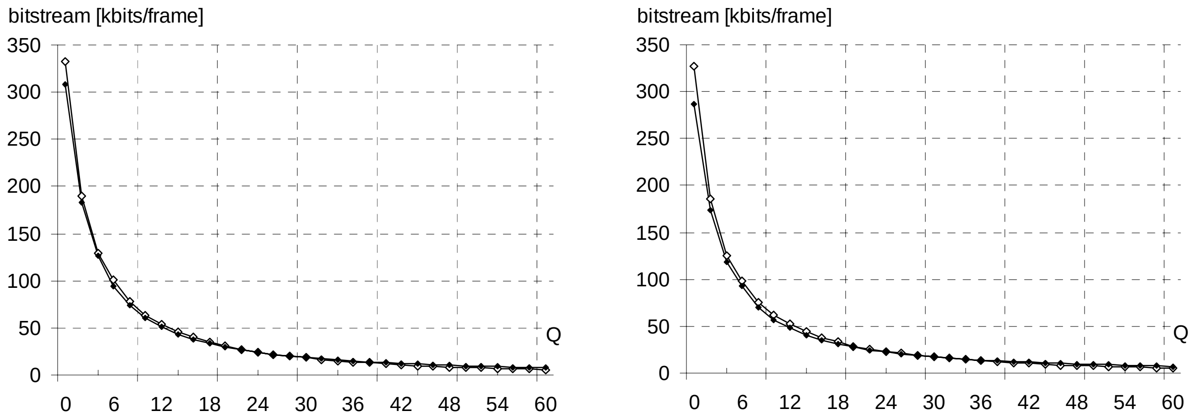
Furthermore, the maximum relative approximation error ε_B for large quantization scale factor Q is sometimes large because of low values of bitstream for such values of Q.

The bitrate for $Q=2$ can be ten times greater than for $Q=62$. It means that small absolute error of bitstream approximation for large values of Q gives large relative error (Tables 4.2 and 4.3).



a) b)

Fig. 4.12 Graph of curves of chrominance bitrate approximation and experimental data for frames in sequences *a) Basket b) Football.*



a) b)

Fig. 4.13 Graph of curves of chrominance bitrate approximation and experimental data for frames in sequences *a) Cheer b) Mobile.*

The value of all five parameters and obtained approximation error is shown in tables 4.2 and 4.3. For I-frame the average approximation error is lower then 3%. In case of P-frames this error is below 9%; such large value of average approximation error ε_B due to large approximation error of "the end" of the experimental curve, that is for largest values of quantization scale factor Q.

Table 4.2 Parameters for the luminance in I-frames.

Frame	Parameters					***Maximum ε_B***	***Average ε_B***
	$a/10^6$	***b***	***c***	***d***	***e***	[%]	[%]
			Flower garden				
0	5.7	0.91	2.06	0.11	0.87	**7.9**	**2.9**
32	6.4	0.91	2.05	0.11	0.88	**6.8**	**2.9**
64	6.5	0.91	2.05	0.11	0.89	**7.1**	**2.6**
96	3.3	0.86	2.28	0.10	0.86	**6.2**	**2.3**
			Cheer				
0	1.4	0.95	0.61	0.06	0.71	**7.6**	**2.8**
32	1.3	0.93	0.63	0.05	0.68	**8.2**	**2.8**
64	1.4	0.95	0.64	0.05	0.67	**8.0**	**2.9**
96	1.3	0.94	0.63	0.05	0.67	**8.3**	**2.9**
			Stefan				
0	1.9	0.89	0.59	0.05	0.38	**7.0**	**2.3**
32	1.1	0.88	0.66	0.05	0.44	**5.8**	**2.0**
64	1.1	0.86	0.67	0.05	0.38	**6.2**	**2.2**
96	1.1	0.85	0.67	0.05	0.37	**5.5**	**1.9**

Table 4.3 Parameters for the luminance in P-frames.

Frame	Parameters					***Maximum ε_B***	***Average ε_B***
	$a/10^6$	***b***	***c***	***d***	***e***	[%]	[%]
			Flower garden				
0	2.7	1.07	0.90	0.53	0.56	**13.8**	**4.9**
32	2.1	1.31	0.47	0.77	0.49	**15.6**	**8.2**
64	1.2	1.20	0.31	0.39	0.03	**13.7**	**8.5**
96	2.2	1.18	0.62	0.62	0.10	**14.7**	**8.7**
			Cheer				
0	2.3	1.24	0.56	0.56	0.47	**7.3**	**2.8**
32	2.1	1.14	0.63	0.66	0.41	**15.8**	**4.2**
64	2.0	1.26	0.58	0.70	0.42	**9.9**	**4.1**
96	2.0	1.13	0.61	0.64	0.41	**12.6**	**4.6**
			Stefan				
0	1.7	1.11	0.59	0.67	0.42	**4.2**	**2.2**
32	1.9	1.09	0.64	0.76	0.41	**10.6**	**5.6**
64	1.9	1.22	0.73	0.68	0.41	**13.8**	**5.6**
96	2.1	1.23	0.64	0.92	0.42	**8.6**	**4.3**

Pre-encoding would be required in order to find points $B_x(Q)$ for parameters estimation. Such pre-encoding is very costly and for the bitrate control, it would be very practical to have a model with only one parameter representing the frame content. Therefore, simplification of the global model has been introduced.

4.4 Single parameter model

We can notice that some parameters are rather constant or vary a little. Many experiments confirm that four of the five parameters can be fixed. Therefore, it was assumed that only the parameter *c* depends on frame content, while the other parameters are assumed to have general values that can be found by minimization of the approximation error ε_B over a set of frames from some training sequences.

Minimization of maximal approximation error is used to obtain the four parameters (*a, b, d* and *e)* of approximation curve. Parameter *c* was the only free parameter. Minimization was made over a few teaching images from the set of three teaching sequences. Next, these four obtained parameters are used in the experiments conducted on the remaining test video sequences in order to check the accuracy of modeling with only one free parameter.

According to experimental results, choosing *c* as the only free parameter simplifies the model, slightly worsening the optimization. The two sets of fixed parameters are obtained, for intra and non intra frames:

Table 4.4 Fixed parameters of the model for Intra and Inter mode of coding

Parameters	***a***	***b***	***d***	***e***
Intra:	$5 \cdot 10^6$	0.9	0.1	0.5
Non Intra:	$2 \cdot 10^6$	1.1	0.6	0.5

For such values new values of *c* have been founded. Table 4.5 presents approximation error. The approximation function with only one free parameter gives higher average errors but accuracy is still enough for bitrate estimation.

Table 4.5 Parameters for the luminance in a) I-frames and b) P-frames for four frames of sequences *Flower Garden*, *Cheer* and *Stefan*.

Frame	Parameter c	***Maximum*** ε_B	***Average*** ε_B
Flower garden			
0	1.98	**9.1**	**3.8**
32	1.96	**11.5**	**3.1**
64	1.99	**14.6**	**5.6**
96	1.94	**11.8**	**7.8**
Cheer			
0	0.42	**15.5**	**3.7**
32	0.42	**14.4**	**5.5**
64	0.51	**13.4**	**4.4**
96	0.48	**10.3**	**5.1**
Stefan			
0	0.45	**11.0**	**6.2**
32	0.44	**12.7**	**3.8**
64	0.49	**9.3**	**4.4**
96	0.56	**10.4**	**3.9**

a)

Frame	Parameter c	***Maximum***	***Average***
Flower garden			
0	0.68	**17.9**	**7.8**
32	0.49	**25.6**	**11.2**
64	0.70	**16.3**	**9.4**
96	0.66	**17.4**	**11.6**
Cheer			
0	0.68	**7.3**	**5.4**
32	0.59	**15.8**	**6.2**
64	0.55	**9.9**	**7.3**
96	0.58	**12.6**	**8.9**
Stefan			
0	0.71	**4.2**	**5.3**
32	0.80	**10.6**	**6.5**
64	0.69	**13.8**	**7.3**
96	0.73	**8.6**	**5.9**

b)

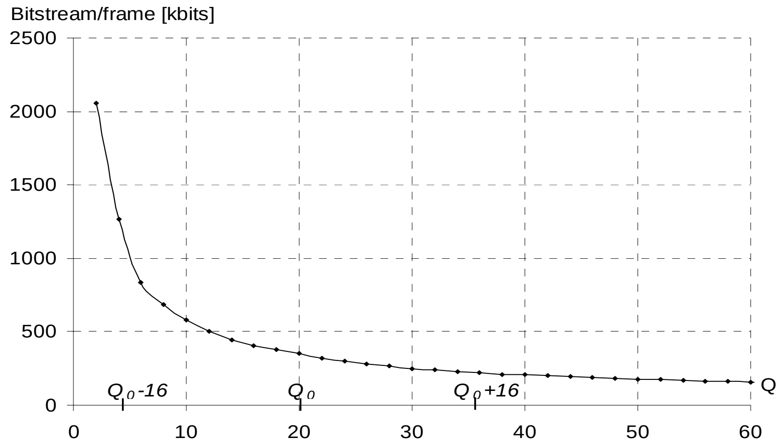


Fig. 4.14 Exemplary selected interval ΔQ for approximation by model. (Q_0 = 20 and ΔQ = 16)

Since the entire range of quantization scale factor Q is never used and this value is slowly changing from frame to frame, we can choose an interval of Q for which optimization is to be done. Hence, the optimization was performed in a window $Q_0 \pm \Delta Q$ for various Q_0 and ΔQ. The precision of approximation was measured using the parameter ε_B (formula 4.12) calculated for a given frame. This precision has been measured for a set of test sequences being different from that used to train the model.

Table 4.6. Results of approximation error for I-frames measured for several ΔQ intervals and several middle value of Q.

Quant	**Maximum ε_B [%]**			**Average ε_B [%]**		
	ΔQ=4	**ΔQ=8**	**ΔQ=16**	**ΔQ=4**	**ΔQ=8**	**ΔQ=16**
Flower garden – luminance						
Q_0=20	0.9	1.8	5.9	0.4	0.8	2.0
Q_0=32	0.6	1.1	2.6	0.4	0.4	0.8
Q_0=44	0.6	1.0	1.6	0.3	0.4	0.5
Cheer– luminance						
Q_0=20	0.9	3.6	9.4	0.5	1.3	3.3
Q_0=32	0.9	1.7	5.3	0.5	0.6	1.6
Q_0=44	0.8	0.9	2.3	0.3	0.3	0.6
Stefan– luminance						
Q_0=20	1.0	3.3	4.9	0.6	1.0	1.8
Q_0=32	1.0	2.2	4.9	0.4	0.6	0.9
Q_0=44	1.2	2.6	5.0	0.7	1.1	1.4
Flower garden - chrominance						
Q_0=20	0.9	3.6	10.9	0.4	1.1	3.9
Q_0=32	2.0	2.6	3.4	0.9	1.0	1.8
Q_0=44	2.4	1.7	3.07	1.1	0.7	0.9
Cheer - chrominance						
Q_0=20	4.0	7.6	21.3	1.9	2.5	3.6
Q_0=32	2.8	5.6	8.2	1.2	.2.1	3.6
Q_0=44	4.9	3.7	7.7	2.3	1.7	3.5
Stefan - chrominance						
Q_0=20	6.0	10.2	22.9	2.3	2.5	5.2
Q_0=32	4.0	7.3	18.2	1.3	2.0	4.3
Q_0=44	2.8	5.0	10.6	1.0	1.8	2.2

Table 4.7. Results of approximation error for P-frames measured for several ΔQ intervals and several middle value of Q.

Sequance	***Maximum*** ε_B [%]			***Average*** ε_B [%]		
	ΔQ=2	ΔQ=4	ΔQ=8	ΔQ=2	ΔQ=4	ΔQ=8
Flower garden - luminance						
Q_0=20	8.2	8.5	28.9	1.5	2.3	4.9
Q_0=32	6.9	8.1	16.9	1.3	3.1	4.9
Q_0=44	3.3	6.3	7.9	0.8	1.35	2.4
Cheer- luminance						
Q_0=20	7.5	19.3	29.9	1.7	2.9	5.5
Q_0=32	8.6	17.9	36.7	2.6	4.8	8.4
Q_0=44	8.5	13.6	19.9	2.4	3.9	5.9
Stefan- luminance						
Q_0=20	8.8	8.9	21.0	1.7	2.2	5.8
Q_0=32	10.9	9.3	16.6	2.5	2.8	4.0
Q_0=44	11.9	16.5	24.4	1.52	3.0	5.1
Flower garden - chrominance						
Q_0=20	14.4	29.5	63.6	4.9	5.5	8.8
Q_0=32	23.0	27.8	36.7	6.6	6.2	10.9
Q_0=44	15.6	24.2	40.9	3.6	8.1	11.8
Cheer- chrominance						
Q_0=20	.19.7	22.3	43.2	4.9	6.3	8.9
Q_0=32	20.5	25.4	34.8	3.7	6.5	8.7
Q_0=44	14.7	20.8	26.2	4.4	5.9	7.1
Stefan- chrominance						
Q_0=20	19.0	33.25	101.8	6.9	9.4	16.3
Q_0=32	39.8	33.7	43.2	7.1	8	12.9
Q_0=44	26.7	23.6	68.7	5.4	6.8	14.7

Table 4.8. Results of approximation error for B-frames measured for several ΔQ intervals and several middle value of Q.

Sequance	***Maximum*** ε_B [%]			***Average*** ε_B [%]		
	ΔQ=2	ΔQ=4	ΔQ=8	ΔQ=2	ΔQ=4	ΔQ=8
Flower garden - luminance						
Q_0=20	0.00	0.05	0.48	0.00	0.00	0.04
Q_0=32	0.07	0.17	0.14	0.01	0.03	0.03
Q_0=44	0.07	0.14	0.18	0.01	0.04	0.03
Cheer- luminance						
Q_0=20	0.00	0.02	0.07	0.00	0.00	0.02
Q_0=32	0.01	0.02	0.03	0.00	0.00	0.00
Q_0=44	0.00	0.01	0.06	0.00	0.00	0.04
Stefan- luminance						
Q_0=20	0.01	0.05	0.21	0.00	0.00	0.03
Q_0=32	0.02	0.04	0.12	0.00	0.00	0.02
Q_0=44	0.02	0.04	0.04	0.00	0.00	0.04

The experimental results (Tables 4.6, 4.7 and 4.8) indicate that narrowing the interval of quantization scale factor Q increases the accuracy of the model. Average error value of ε_B depends on Q_0 - the largest Q_0 values the largest average error εB. For I- and P-frames this error is about several percent (1-9% in most cases), but for B-frames the approximation error is below 0.1%.

4.5 Application

The encoder in CBR mode is considered. The experimental results are obtained for the MPEG-2 standard algorithm of video coding and the TM5 MPEG-2 coder has been used, but the similar approach is also useful for other hybrid video coders like H.263. It is assumed that no scene cut points exist in the GOP.

In order to find the best *Q* giving encoded bitstream closest to required the model parameters have to be calculated. There are three way finding them:

- to take them from the previous frame with or without modifications.
- to pre-encode image with certain Q_1 and estimate model parameters. Having one point $B_1(Q_1)$ we can easily compute *c* parameter (Fig. 4.15).
- Using the microscopic model (see chapter 5) more than one point $B_i(Q_i)$ can be determined (e.g. $B_1(Q_1)$ and $B_2(Q_2)$ - Fig. 4.15), and then *c* parameter (and also *a, b, d, e* if necessary) can be computed.

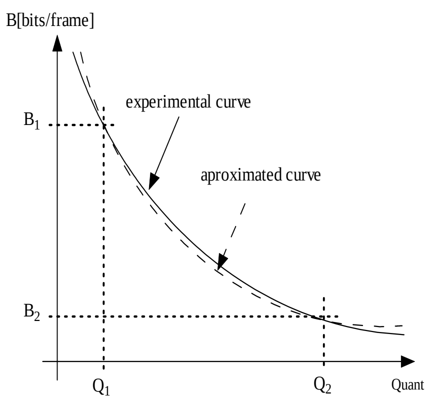


Fig. 4.15 Example parameter estimation by frame pre-encoding.

When prediction from previous encoded frame is used the *c* parameter can be determined as follows:

$$c = \frac{\dfrac{a}{B_{VAR_{prev}}} - d}{Q_{prev}{}^{b} + e} \tag{4.14}$$

then, the value of quantize scale factor *Q* for current frame can be computed using *c* parameter from the previous frame:

$$Q_{cur} = e^{b \cdot \log_{10}\left[\left(\frac{a}{B_{VAR}} - d\right)\cdot\frac{1}{c} - e\right]} \tag{4.15}$$

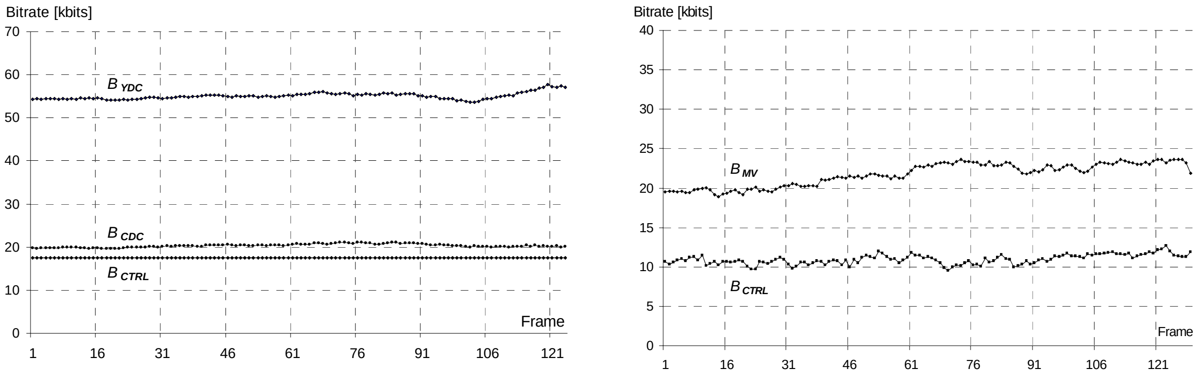


Fig.4.16. The bitstream components which are independent from *Q* in I-frames and P-frames of the test sequence *Basket*.

Values of constant parts of the bitstream can be easily predicted from the previous encoded frame because they vary only a little (Fig. 4.14). Table 4.9 shows how bitstreams B_{DC}, B_{CTRL}, B_{MV} are changing.

Table 4.9 The bitstream components for I, P and B frames that are independent from Q in various frames of three test sequences

Frame	bits/frame						
	I-frames			**P-frames**		**B-frames**	
Flower garden							
	B_{CTR}	B_{YDC}	B_{CDC}	B_{CTR}	B_{MV}	B_{CTR}	B_{MV}
1	14356	43895	14193	11120	9847	12044	19812
16	14355	45007	14837	10808	9887	11557	21735
32	14394	45694	14940	9784	9250	11950	20505
1-128	14364	45521	15279	10075	9650	12121	22385
Standard deviation [bits]	**13**	**527**	**37**	**632**	**296**	**320**	**889**
Cheer							
1	14383	43849	17010	7594	8310	10898	19985
16	14394	44196	17363	4848	7317	11566	21067
32	14395	43823	16936	7145	8198	11483	19890
1-128	14383	43714	16892	5674	7724	11522	19614
Standard deviation [bits]	**12**	**832**	**389**	**361**	**263**	**311**	**797**
Stefan							
1	14364	37783	12536	8870	10801	10477	18765
16	14372	37626	12370	8915	12122	10432	17597
32	14377	38173	12583	8879	11127	10688	16959
1-128	14382	38330	12532	8842	11324	10682	18162
Standard deviation [bits]	**11**	**439**	**146**	**162**	**698**	**226**	**1136**

These values can also be found by pre-encoding adequate data, for example DC coefficients. These DC coefficients are independent from *Q* and, after pre-encoding, can be buffered in order to avoid multiple encoding

4.5.1 Control algorithm

The proposed control algorithm exploits global model of bitstream. This algorithm sets the quantization scale factor Q in two stages:

Finding B_{CONST} , B and c parameter:

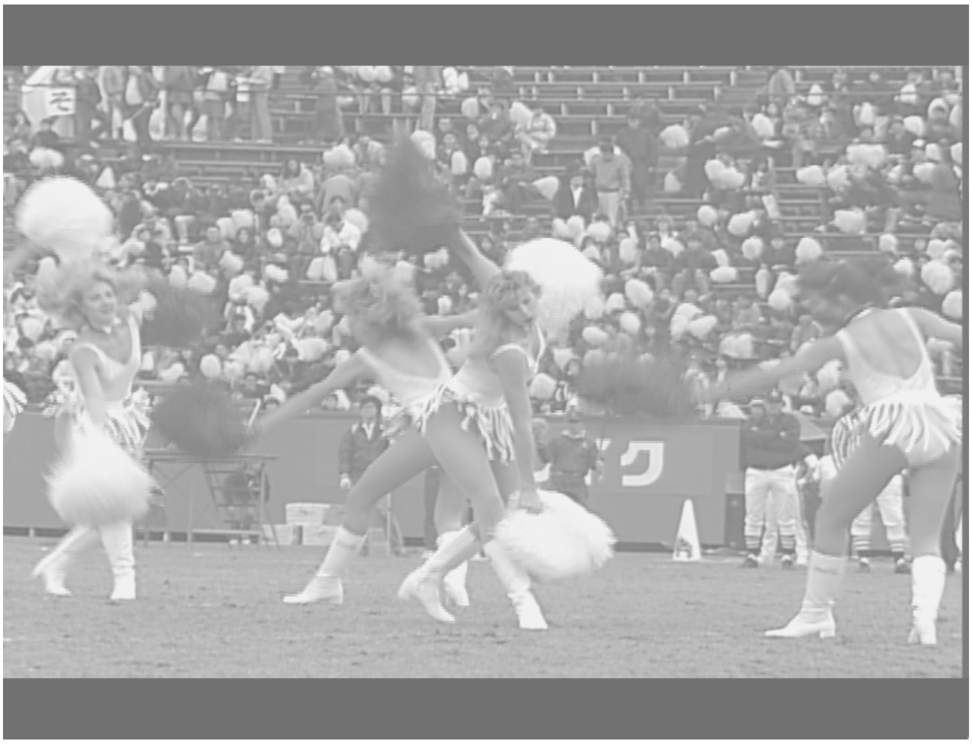
- All parts of the B_{CONST} bitstream are predicted from the previous encoded frame or are determined by pre-encoding or rather encoding with buffering. It means that invariable fields, headers, flags are encoded and buffered. When P- or B- frame is encoded the field B_{CBP} should be estimated (formula 4.4).
- Having the B_{CONST} bitstream, the number of bits for B_{VAR} is determined.
- Doing pre-encoding or using the microscopic model of bitstream (see chapter 5) the c parameter (or also a, b, d, e if possible) is estimated.

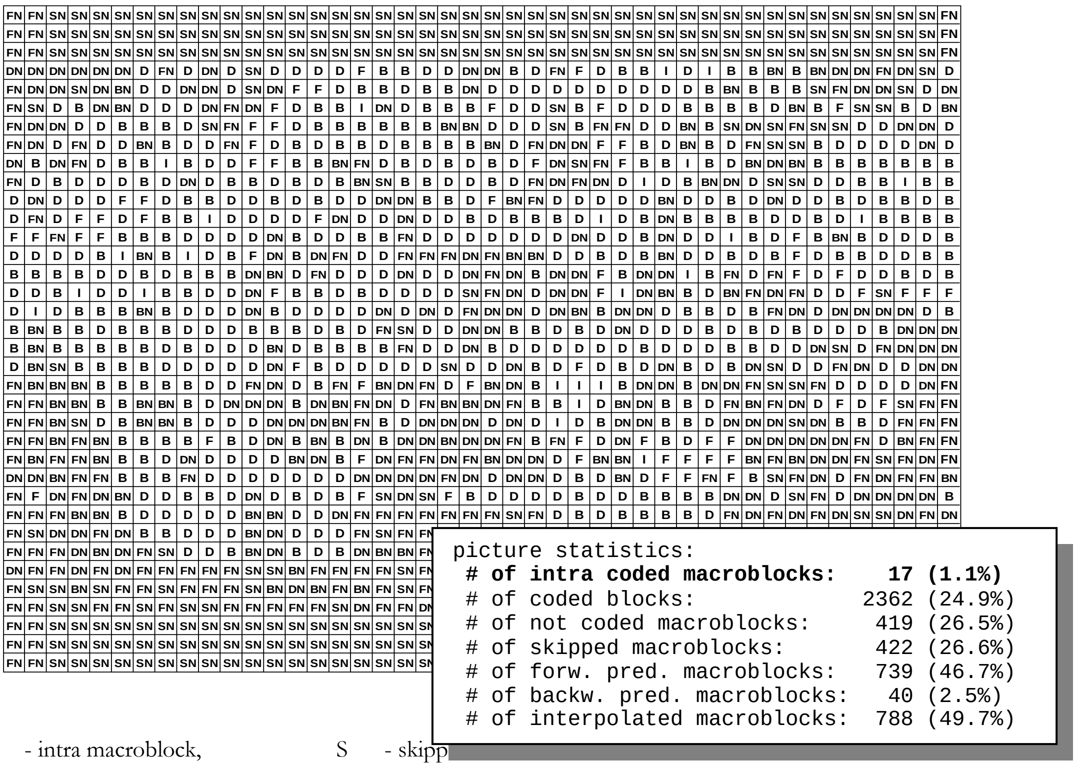
Finding quantization scale factor Q:

- The bitrate needed for luminance and chrominance component is approximated (formula 4.13).
- The global quantization scale factor Q is determined and clipped with saturation in required range.

After that, the normal encoding process is started and for DCT quantizing the global scale factor Q is used. On the figure 4.17 exemplary image from sequence *Cheer* and control information for the macroblocks are shown.

Parameter c is found for each image type (I, P, B) separately. In P-frames and B-frames there may occur intra macroblocks, but they are not taken into account in models of these frame types. However, this phenomenon does not disturb bitrate estimation as it is statistically insignificant. The average percentage of intra macroblocks in P-frame or B-frame is not greater then 2% (Fig. 4.17) for MPEG-2 system.





I - intra macroblock,	S - skipp
F - forward prediction	D - bidirectional prediction
B - backward prediction	N - no AC coefficients

Fig. 4.17 Macroblock coding type for exemplary B-frame from *Cheer* video sequence; top) encoded picture, bottom) map of macroblock types.

4.6 Results of experiments

The algorithm described above is implemented in reference TM5 MPEG-2 video coder (Test Model 5). The experiments have been made for a few test video sequences. The first 250 frames of each video test sequence have been used to experiments. The MPEG-2 video coder configuration is as follows:

- MainLevel@MainProfil.
- 4CIF input images resolution.
- The GOP structure ***I**BBB**P**BBB**P**BBB**P**BBB(**I..**)*
- The Constant Bitrate Mode (CBR) of coding.
- The VBV Buffer size: 1.9 Mbits.

The experiments have been done in two series, for progressive and interlaced mode of coding. The performance of MPEG-2 video coder will be measured in wide range of bitrates 3÷8 Mbits/sec.

Average bitrate and quality

Table 4.10 Comparison of control algorithms between default TM5 control algorithm and new based on global model (progressive mode of coding)

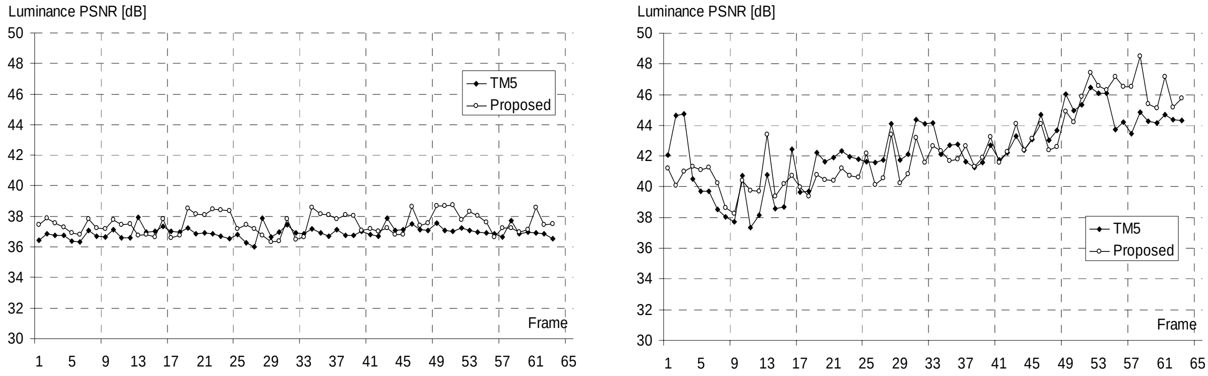
Bitrate [Mbits]	Bitrate [Mbits/sec]		PSNR [dB]		ΔPSNR (P_2-P_1) [dB]	Standard deviation σ of PSNR [dB]	
	Proposed	TM5	Proposed (P_1)	TM5 (P_2)		Proposed	TM5
Football							
3	2 929	2 904	36.76	36.76	**+0.00**	0.945	1.081
4	3 875	3 886	38.12	38.04	**+0.06**	0.989	1.012
5	4 756	4 874	39.19	39.01	**+0.18**	1.015	1.024
6	5 891	5 855	39.89	39.74	**+0.15**	1.098	1.146
7	6 887	6 836	40.47	40.27	**+0.20**	1.113	1.054
8	7 893	7 943	41.23	40.96	**+0.27**	1.231	1.123
Cheer							
3	2 945	2 904	34.76	34.59	**+0.17**	0.502	0.409
4	3 834	3 883	36.17	35.94	**+0.23**	0.554	0.387
5	4 837	4 861	37.20	36.94	**+0.26**	0.641	0.378
6	5 886	5 843	38.08	37.76	**+0.32**	0.740	0.475
7	6 893	6 825	38.86	38.45	**+0.41**	0.827	0.513
8	7 897	7 805	39.50	39.06	**+0.44**	0.912	0.663
Stefan							
3	2 934	2 922	40.78	40.83	-0.05	0.721	0.691
4	3 901	3 903	41.93	41.76	**+0.17**	0.721	0.764
5	4 934	4 876	42.73	42.47	**+0.26**	0.803	0.804
6	5 871	5 853	43.48	43.04	**+0.44**	0.579	0.864
7	6 685	6 834	44.02	43.51	**+0.51**	0.723	0.914
8	7 845	7 821	44.65	43.98	**+0.67**	0.791	1.001
Universal							
3	2.913	2 939	42.46	42.10	**+0.36**	1.745	1.784
4	3 951	2 933	43.84	43.32	**+0.52**	1.780	1.794
5	4 987	4 926	44.81	44.23	**+0.58**	1.855	1.883
6	5 933	5 921	45.62	44.98	**+0.64**	1.867	1.882
7	6 876	6 920	46.34	45.65	**+0.69**	1.645	1.861
8	7 834	7 952	46.90	46.19	**+0.71**	1.656	1.855
Warner							
3	2 890	2 904	40.28	39.71	**+0.57**	2.784	2.919
4	3 913	3 884	41.61	41.12	**+0.49**	2.698	2.700
5	4 833	4 854	42.71	42.08	**+0.63**	2.571	2.558
6	5 889	5 825	43.56	42.89	**+0.67**	2.434	2.502
7	6 917	6 821	44.35	43.66	**+0.69**	2.463	2.519
8	7 828	7 830	45.01	44.34	**+0.65**	2.359	2.538
Icon							
3	2 987	2 918	47.92	47.89	**+0.03**	2.763	2.758
4	3 925	3 894	48.90	48.81	**+0.09**	3.241	3.096
5	4 919	4 868	49.64	49.53	**+0.11**	3.310	2.998

The results for interlaced and progressive coding of video sequences in CBR mode are shown in table 4.10 and 4.11. Both default and proposed control algorithm achieves the required bitrate with similar accuracy. However, new control algorithm enables to obtain higher average PSNR of encoded video sequences. This gain equals about 0.1÷0.4 dB, and for higher bitrates it can achieve 0.8 dB. Both default and proposed control algorithm gives similar value of the PSNR variance. Differences of control algorithm performance between interlaced and progressive mode of coding are very small and depend on the sequence content. In general, proposed control algorithm increases efficiency of video coding.

Table 4.11 Comparison of control algorithms between Default TM5 control algorithm and new based on global model. (interlaced mode of coding).

Target bitrate [Mbits]	**Bitstrate [Mbits/sec]**		**PSNR [dB]**		**ΔPSNR (P_2-P_1) [dB]**	**Standard deviation of PSNR [dB]**	
	Proposed	**TM5**	**Proposed (P_1)**	**TM5 (P_2)**		**Proposed**	**TM5**
			Football				
3	2 934	2 904	36.86	36.76	**+0.10**	1.285	1.081
4	3 794	3 886	38.25	38.04	**+0.21**	1.362	1.012
5	4 713	4 874	39.30	39.01	**+0.29**	1.383	1.024
6	5 934	5 855	40.10	39.74	**+0.36**	1.467	1.146
7	6 932	6 836	40.61	40.27	**+0.34**	1.465	1.054
8	7 911	7 943	41.35	40.96	**+0.39**	1.450	1.123
			Cheer				
3	2 925	2 925	31.48	31.58	-0.10	0.677	0.517
4	3 945	3 920	33.15	33.00	**+0.15**	0.695	0.673
5	4 964	4 920	34.65	34.32	**+0.33**	0.751	0.900
6	5 938	5 919	35.98	35.33	**+0.65**	0.941	1.022
7	6 937	6 919	36.18	36.18	**+0.71**	0.932	1.209
8	7 933	7 918	36.97	36.97	**+0.72**	0.967	1.198
			Stefan				
3	2 91	2 950	38.08	37.95	**+0.13**	0.620	0.569
4	3 86	3 949	39.46	39.28	**+0.18**	0.848	0.784
5	4 97	4 950	40.40	40.17	**+0.23**	0.898	1.015
6	5 96	5 946	31.07	30.88	**+0.19**	0.982	1.199
7	6 97	6 945	41.86	41.55	**+0.31**	1.114	1.499
8	7 89	7 945	42.60	42.14	**+0.46**	1.106	1.455
			Universal				
3	2 923	2 982	30.78	30.33	**+0.45**	1.943	1.736
4	3 943	3 985	42.27	41.64	**+0.63**	1.960	1.752
5	4 954	4 990	43.21	42.54	**+0.67**	1.976	1.932
6	5 934	5 998	43.69	43.30	**+0.59**	1.934	1.939
7	6 959	7 007	44.74	43.95	**+0.79**	2.011	1.844
8	7 911	8 014	45.24	44.48	**+0.76**	2.086	1.863
			Warner				
3	2 909	2 962	38.03	37.69	**+0.54**	1.674	2.009
4	3 976	3 943	39.34	38.95	**+0.59**	2.616	2.492
5	4 970	4 955	40.10	39.49	**+0.61**	2.768	2.639
6	5 957	5 902	42.03	41.45	**+0.58**	2.665	2.380
7	6 938	6 941	42.75	42.05	**+0.70**	2.745	2.240
8	7 962	7 939	43.35	42.53	**+0.72**	2.756	2.335
			Icon				
3	2 933	2 967	46.21	46.16	**+0.05**	2.346	2.170
4	3 947	3 957	47.04	46.93	**+0.11**	2.327	2.431
5	4 965	4 953	47.77	47.64	**+0.13**	2.411	2.852

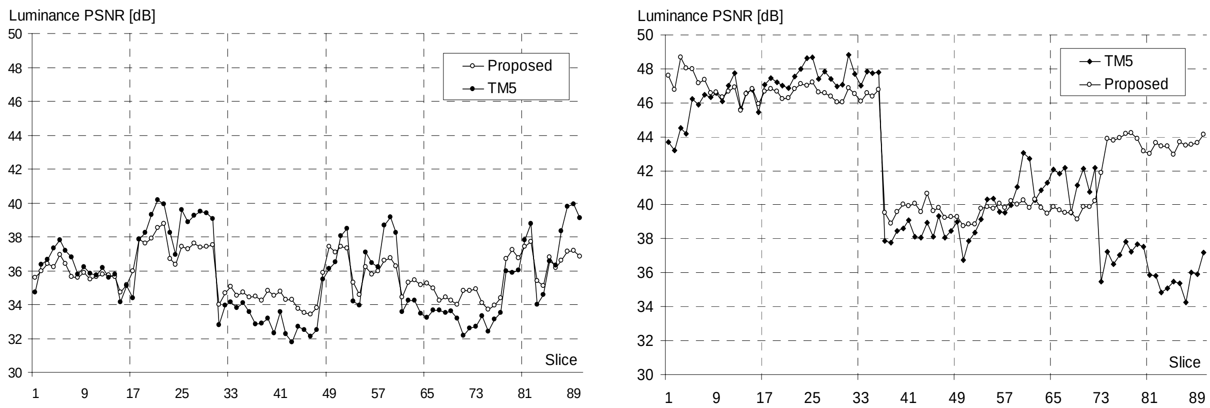
Qaality of encoded frames and slices



a) b)

Fig. 4.18 Image quality (PSNR) value versus frame number (for **progressive** MPEG2 encoding), for sequences a) *Cheer* and b) *Warner*

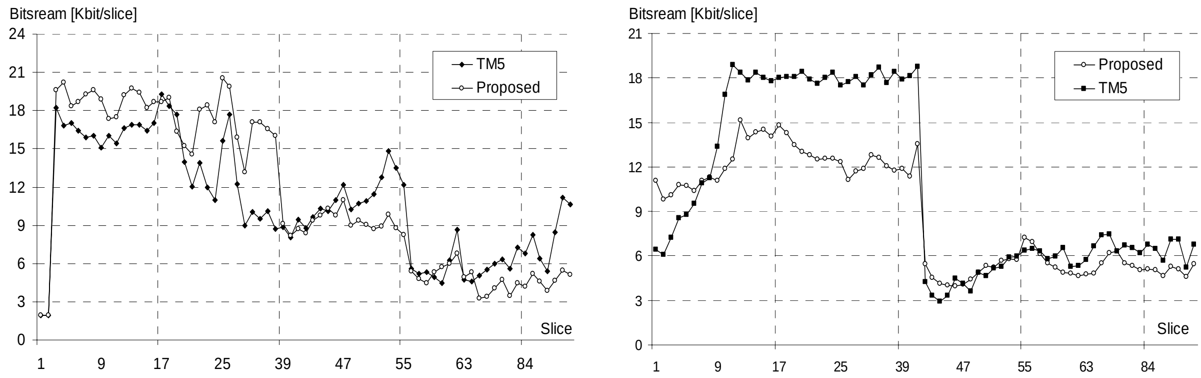
The figures above show quality of video sequences for both control algorithms. The respective plots are quite similar for various test sequences. Slightly better quality of encoded video sequence is achieved by coder with proposed control algorithm. Quality of each slice in encoded image is shown on the figures 4.19. One can see that the PSNR curve is much smoother when proposed control algorithm is applied. Due to this better average PSNR of encoded video sequence is achieved.



a) b)

Fig. 4.19 Graph of encoding sequence quality (PSNR) versus slice number (for **progressive** MPEG2 encoding) for sequences a) *Cheer* and b) *Warner*

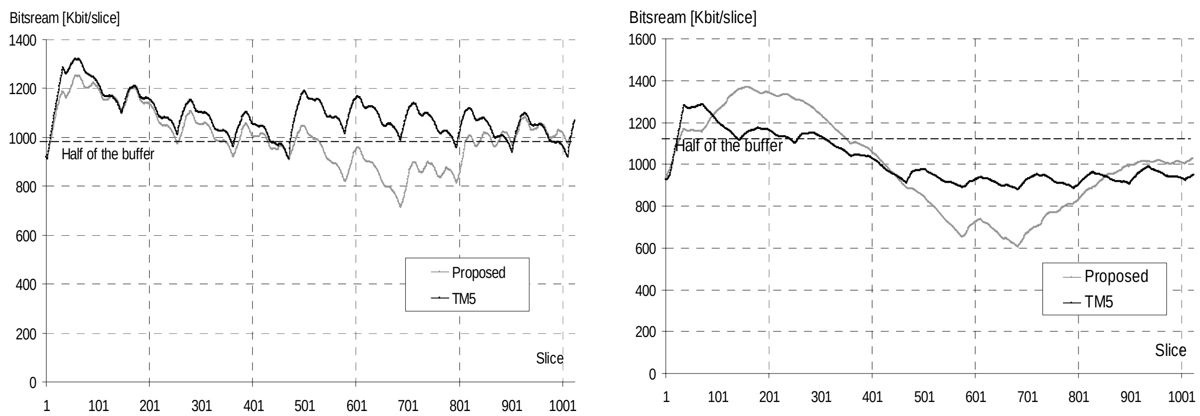
Short-term bitrate and buffer occupancy



a) b)

Fig. 4.20 Graph of the slice bitstream versus slice number (for progressive MPEG2 encoding) for sequences a) Cheer and b) Warner.

Figures 4.20 show the short-term value of the bitstream measured for each slice of macroblock. The proposed control algorithm allocates bits in a different way than default TM5 control algorithm. It results in increasing temporal variance of the bitstream, but causes increase in the variance of VBV buffer occupancy as well. Nevertheless, the VBV buffer is never overflowed or underflowed (Fig. 4.21) which means that proposed control algorithm makes the most of the VBV buffer and its available capacity, enabling to achieve better quality of encoded video sequence.



a) b)

Fig.4.21 Graph of the VBV buffer occupancy versus slice number (for **progressive** MPEG2 encoding) for sequences a) *Cheer* and b) *Warner*

Average value of c parameter

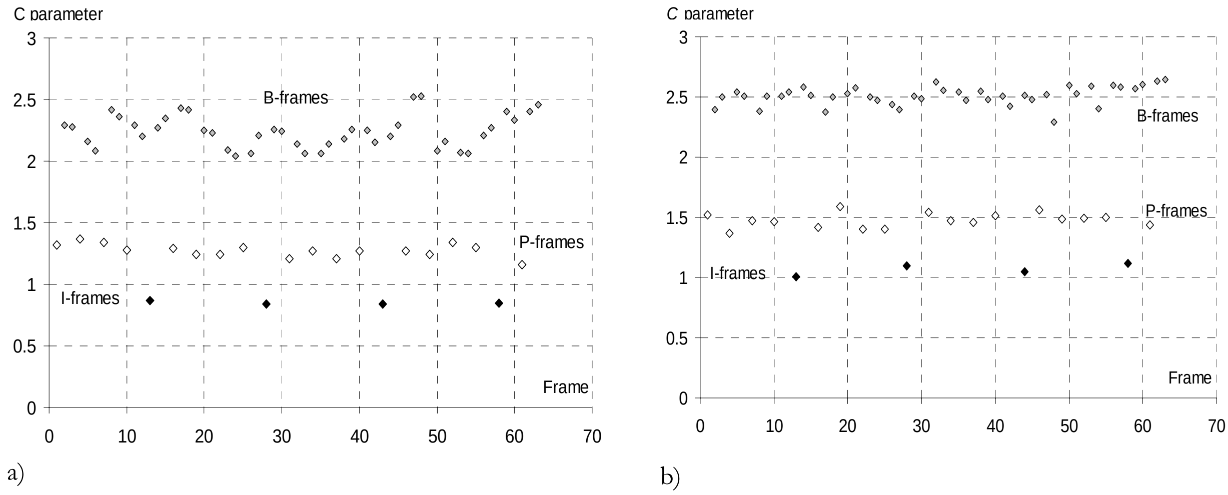


Fig. 4.22 Value of *c* parameter versus frame number (for **progressive** MPEG2 encoding), for sequences a) *Cheer* and b) *Warner*

Figures above show variance of the *c* parameter. For all the test video sequences the *c* parameter for I-frames is always close to value 1.0, mostly below 1.0. For the fragment of video sequence between two scene cut points *c* values varies a little. It means that under I-frame encoding the previous value of *c* parameter can be used.

For the other types of frames (P and B) predicting of the *c* value can be wrong. In such a case the frame pre-encoding or estimation using the microscopic model has to be applied.

4.7 Conclusions

In this chapter the author proposed a global model of video bitstream. The global model is simple but it is necessary to estimate several model parameters initially. The new control algorithm for hybrid video coders which exploits global model is presented as well. The experiments show high efficiency of proposed algorithm. It enables to achieve better video sequence quality for the same bitrates in comparison with default TM5 MPEG-2 control algorithm.

Chapter 5

Microscopic model of bitstream

5.1 Introduction

In the following chapter the author presents another original model of video bitstream which was already briefly reported by the author in [Lucz03] at Picture Coding Symposium 2003.

The main goal is to create more accurate bitstream model with such parameters that can be easily computed from image content. Therefore, this model is based on the histogram of the DCT coefficients analysis. All results of the model accuracy analysis will be presented for MPEG-2 and H.263 video coders. Moreover, experimental results of control algorithm will be given for the MPEG-2 video coder (MainLevel and MainProfile) and for 4CIF sequences both progressive and interlaced.

5.2 Coder modeling using histogram of DCT coefficients

As in the global model, the number of bits needed to represent DCT coefficients for luminance or chrominance in a frame is expressed as a function of quantization scale factor Q. Therefore, all the deliberations about bitstream partitioning are the same as before and the whole bitstream is:

$$B = B_{CONST} + B_{VAR}(Q). \tag{5.1}$$

$B_{VAR}(Q)$ is determined from the required value of bitstream B and estimated value of B_{CONST}.

$$B_{VAR}(Q) = B - B_{CONST}. \tag{5.2}$$

B_{CONST} is treated is the same way as in chapter 4 and can be estimated as already described in section 4.2 In the previous model the video coder was treated as a black box and, in this case, the value of bitstream $B_{VAR}(Q)$ is estimated using the analysis of coding process. Only B_{VAR} part of the bitstream will be considered. This part of the bitstream consists of luminance and chrominance bitstream components:

$$B_{VAR}(Q) = B_{YV}(Q) + B_{CV}(Q) + B_{CBP}(Q) \tag{5.3}$$

The $B_{CBP}(Q)$ component is estimated as in section 4.2 (see: chapter 4). The new approach is that the number of bits $B_{YV}(Q)$, $B_{CV}(Q)$ per frame is estimated from the histograms of the DCT coefficients, excluding Intra DC coefficients.

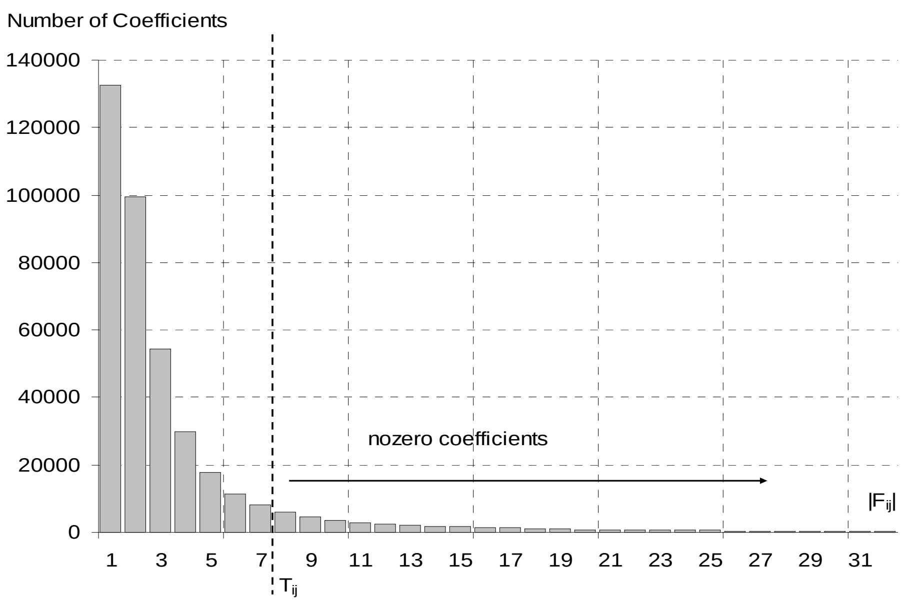


Fig. 5.1 Histogram of one AC DCT coefficient computed for the first frame of the *flower* sequence (4CIF).

Those DC coefficients are not quantized and do not depend on quantization scale factor Q. Since DCT computation is the very first stage of transform coding process, the DCT coefficients are known before the quantization scale factor Q is chosen.

Transformation is made in (8×8)-block giving 64 DCT coefficients. Each coefficient is quantized in the same way but with different quantization weight w_{ij}. Those quantization weights form quantization matrix predefined in each coding system (see chapter 2, point 2.3.3.2). Some video coding systems allow changing default quantization matrices. Therefore, a separate histogram for each DCT coefficient F_{ij} has to be computed (Fig. 5.1 which means that two-dimensional histogram H_{ij} is obtained.

For each component, i.e. the luminance and two chrominances, a two-dimensional histogram $H_{ij}(|F_{ij}|)$ is calculated for every DCT coefficient F_{ij}. Those DCT coefficients are rounded to the nearest integer and their maximum absolute value is 2048. Hence, 64 histograms of integer values in range 0÷2047 are obtained.

$$H = \begin{bmatrix} H_{00} & \cdots & H_{07} \\ \cdots & & \\ H_{70} & \cdots & H_{77} \end{bmatrix} \tag{5.4}$$

In this way, 64 histograms are calculated for each component in a P- and B-frame while 63 histograms are calculated for components from an I-frame due to Intra DC coefficients are not quantized, they are not taken into account when creating coefficients histogram,. Hence, H_{00} matrix element is provided to be a sequence of zeros. An $H_{ij}(|F_{ij}|)$ histogram expresses a probability distribution of the color component F_{ij} coefficient values within a given frame.

In video coder the DCT coefficients after quantization process are scanned and *RL* coded (see point 2.3.3.4). Each *RL*-pair denotes nonzero DCT coefficient. Such DCT-codeword (*RL*-pair) is coded with Huffman codes. The codes and their lengths are always normalized in the certain coding standard. In each system those codes are fixed and cannot be changed. The number of nonzero DCT coefficients can be obtained at a very first stage of the encoding process. As the number of *RL*-pairs is determined by the number of nonzero DCT coefficients, it is enough to estimate the number of the latter ones.

The Huffman codes are not defined for all the *RL*-pairs (see point 2.3.3.2). Such pairs with an undefined Huffman code are encoded using *ESCAPE* coding (Fig. 5.2 and 5.3÷5.6). When the mentioned coding mode is used the *RL*-pair is directly sent in a binary form without additional encoding. Therefore, the two sets of *RL*-codewords have to be determined: *VLC* coded and *ESCAPE* coded.

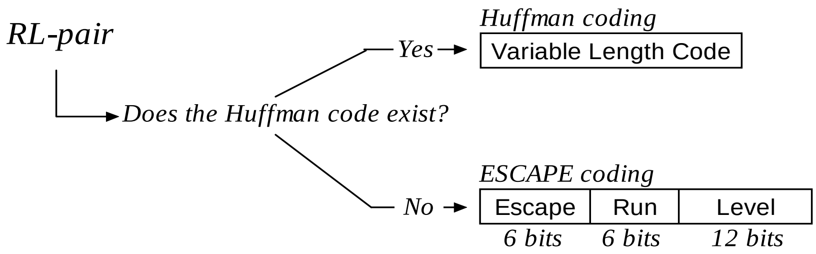


Fig. 5.2 Coding of a *RL*-pair.

The number of bits for DCT coefficients (excluding Intra DC ones) plus the number of bits needed for the EOB codes (codes for *End of block*) is equal to $B_{YV}(Q)$ or $B_{CV}(Q)$ for luminance and chrominance respectively. The number of *RL*-code words is equal to the number of nonzero DCT coefficients. The number of nonzero coefficients depends on quantization scale factor Q. Therefore, the values of $B_{YV}(Q)$ and $B_{CV}(Q)$ can be modeled statistically as functions of Q.

Gray fields on the figures 5.3÷5.6 denote *RL*-pairs encoded with Huffman codes whilst remaining pairs are *ESCAPE* coded. In case of H.263 system the *End of Block* flag is included. Such *RL*-pairs are marked with the darker background.

Run	Level									
	1	2	3	4	5	6	7	8	9	10
0	70614	26910	14658	12215	8314	6035	4617	3834	3809	2931
1	13160	2295	742	452	224	141	104	64	61	31
2	4676	385	81	32	14	12	4	3	2	3
3	2206	88	7	1	1	3	1	0	0	0
4	1229	30	10	2	0	0	0	0	0	0
5	660	15	7	0	0	3	0	0	0	0
6	333	4	1	0	0	1	1	0	0	0
7	276	3	0	0	0	13	5	1	0	0
8	158	2	0	0	0	0	0	0	0	0
9	80	0	0	0	0	0	0	0	0	0
10	52	0	0	0	0	0	0	0	0	0
11	43	9	0	0	0	0	0	0	0	0
12	10	0	0	0	0	0	0	0	0	0
13	6	0	0	0	0	0	0	0	0	0
14	9	0	0	0	0	0	0	0	0	0
15	4	0	0	0	0	0	0	0	0	0
16	3	0	0	0	0	0	0	0	0	0
17	1	0	0	0	0	0	0	0	0	0
18	1	0	0	0	0	0	0	0	0	0
19	2	0	0	0	0	0	0	0	0	0

a)

Run	Level									
	1	2	3	4	5	6	7	8	9	10
0	7320	2341	1000	432	217	91	63	23	19	8
1	2656	592	206	75	43	25	13	6	4	1
2	1227	143	25	12	0	0	0	0	0	0
3	880	99	7	1	0	0	0	0	0	0
4	491	40	4	1	0	0	0	0	0	0
5	313	19	6	1	0	0	0	0	0	0
6	192	4	0	0	0	0	0	0	0	0
7	196	3	0	0	0	0	0	0	0	0
8	103	0	0	0	0	0	0	0	0	0
9	57	0	0	0	0	0	0	0	0	0
10	37	0	0	0	0	0	0	0	0	0
11	18	0	0	0	0	0	0	0	0	0
12	12	0	0	0	0	0	0	0	0	0
13	14	0	0	0	0	0	0	0	0	0
14	2	0	0	0	0	0	0	0	0	0
15	1	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0
17	2	0	0	0	0	0	0	0	0	0
18	2	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0

b)

Fig. 5.3 An exemplary histogram of (r,l) pairs for luminance of first frame in the test sequence *Basket* (4CIF sequence) for Intra quantization mode for a) Q=2 and b) Q=16 (MPEG-2 system).

Run	Level 1	2	3	4	5	6	7	8	9	10
0	35065	16364	11277	6899	5345	3732	3034	2547	2166	1461
1	12762	4145	2061	1006	697	392	312	195	170	86
2	6106	1488	563	224	116	75	56	37	14	8
3	3344	621	196	79	44	17	11	4	12	3
4	2168	318	90	38	14	15	7	2	1	1
5	1408	196	46	18	4	4	3	0	0	0
6	997	127	28	14	1	3	1	1	0	0
7	797	101	23	5	2	3	0	1	0	0
8	586	57	13	2	3	0	1	0	0	1
9	517	45	14	3	0	2	0	0	0	0
10	391	28	7	3	1	0	0	0	0	0
11	330	22	2	1	0	1	0	0	0	0
12	215	12	3	0	0	0	0	0	0	0
13	159	8	4	0	0	0	0	0	0	0
14	82	3	0	0	0	0	0	0	0	0
15	55	0	0	0	0	0	0	0	0	0
16	27	0	0	0	0	0	0	0	0	0
17	34	0	0	0	0	0	0	0	0	0
18	27	0	0	0	0	0	0	0	0	0
19	15	0	0	0	0	0	0	0	0	0

a)

Run	Level 1	2	3	4	5	6	7	8	9	10
0	2192	582	249	111	64	36	22	27	15	17
1	625	93	28	10	0	1	0	0	0	0
2	347	31	4	1	0	0	0	0	0	0
3	229	19	4	0	0	0	0	0	0	0
4	214	4	0	0	0	0	0	0	0	0
5	99	3	0	0	0	0	0	0	0	0
6	48	0	0	0	0	0	0	0	0	0
7	41	0	0	0	0	0	0	0	0	0
8	39	0	0	0	0	0	0	0	0	0
9	33	1	0	0	0	0	0	0	0	0
10	19	0	0	0	0	0	0	0	0	0
11	2	0	0	0	0	0	0	0	0	0
12	1	0	0	0	0	0	0	0	0	0
13	1	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0
15	1	0	0	0	0	0	0	0	0	0
16	3	0	0	0	0	0	0	0	0	0
17	1	0	0	0	0	0	0	0	0	0
18	1	0	0	0	0	0	0	0	0	0
19	1	0	0	0	0	0	0	0	0	0

b)

Fig. 5.4 An exemplary histogram of (*r*,*l*) pairs for chrominance of first frame in the test sequence *Basket* (4CIF sequence) for Inter quantization mode for a) *Q*=2 and b) *Q*=16 (MPEG-2 system).

Run	Level 1	2	3	4	5	6	7	8	9	10
0	54348	29584	18354	12388	8808	6436	4918	3776	2916	2532
1	20088	8076	3884	2194	1376	922	628	400	332	260
2	9874	3154	1094	532	312	192	88	88	58	16
3	5914	1540	560	246	102	76	32	30	18	10
4	3910	802	276	122	74	26	20	8	16	4
5	2908	576	212	110	52	16	18	10	6	4
6	2210	424	88	44	12	2	8	2	2	0
7	1884	368	102	28	10	8	12	6	0	4
8	1554	266	52	22	8	0	6	4	0	2
9	1252	222	60	40	8	16	0	10	0	0
10	962	108	40	6	0	2	2	2	2	0
11	790	76	12	6	6	14	6	14	2	0
12	586	62	10	2	0	0	0	0	0	0
13	528	22	10	0	0	0	0	0	0	0
14	314	14	2	0	0	0	0	0	0	0
15	184	2	4	0	0	0	0	0	0	0
16	142	2	0	0	0	0	0	0	0	0
17	102	0	0	0	0	0	0	0	0	0
18	62	2	0	0	0	0	0	0	0	0
19	68	0	0	0	0	0	0	0	0	0

a)

Run	Level 1	2	3	4	5	6	7	8	9	10
0	6194	2530	1167	531	290	172	84	65	50	29
1	3068	915	373	147	56	39	23	10	13	5
2	1487	323	84	39	18	1	1	3	1	1
3	886	166	52	21	15	8	2	0	0	0
4	606	86	32	7	5	0	0	0	0	0
5	476	91	22	8	3	0	2	1	0	0
6	380	44	17	3	1	0	0	1	0	0
7	313	45	6	4	1	0	0	0	0	0
8	255	28	3	4	1	0	1	0	0	0
9	175	17	3	0	0	0	0	0	0	0
10	138	5	2	0	0	0	0	0	0	0
11	161	6	2	1	0	0	0	0	0	0
12	68	4	0	0	0	0	0	0	0	0
13	40	5	2	0	0	0	0	0	0	0
14	38	0	0	0	0	0	0	0	0	0
15	16	0	0	0	0	0	0	0	0	0
16	19	0	0	0	0	0	0	0	0	0
17	9	0	0	0	0	0	0	0	0	0
18	9	0	0	0	0	0	0	0	0	0
19	9	0	0	0	0	0	0	0	0	0

b)

Fig. 5.5 An exemplary histogram of (*r*,*l*) pairs for luminance of first frame in the test sequence *Basket* (4CIF sequence) for Intra quantization mode for a) *Q*=2 and b) *Q*=16 (H.263 system).

Run	Level 1	2	3	4	5	6	7	8	9	10
0	20616	10824	6546	4177	2785	2012	1530	1151	884	697
1	7918	2967	1494	850	469	325	201	150	99	62
2	3995	1058	429	180	102	54	28	22	9	10
3	2465	613	188	82	33	20	12	10	3	1
4	1738	325	107	51	19	11	9	0	0	4
5	1307	308	76	29	17	12	4	2	1	5
6	974	130	37	8	7	5	4	0	1	0
7	833	139	24	13	14	2	2	1	0	1
8	689	74	23	8	3	1	1	1	0	2
9	552	95	29	10	5	0	3	1	0	0
10	454	54	11	3	2	1	0	0	0	0
11	283	22	9	15	14	2	2	0	0	0
12	256	13	3	2	1	0	0	0	0	0
13	167	16	1	0	0	0	0	0	0	0
14	125	9	1	0	0	0	0	0	0	0
15	70	3	0	0	0	0	0	0	0	0
16	54	1	0	0	0	0	0	0	0	0
17	43	1	0	0	0	0	0	0	0	0
18	22	1	0	0	0	0	0	0	0	0
19	23	1	0	0	0	0	0	0	0	0

a)

Run	Level 1	2	3	4	5	6	7	8	9	10
0	4682	1678	657	316	156	80	56	29	14	5
1	2328	671	219	76	43	13	21	2	1	1
2	1152	212	70	20	5	5	2	2	1	0
3	692	128	40	20	8	1	0	0	0	0
4	467	78	11	5	1	0	0	0	0	0
5	363	64	18	3	1	1	0	0	0	0
6	303	38	6	0	1	2	0	0	0	0
7	270	20	8	0	0	0	0	0	0	0
8	222	14	3	1	0	1	0	0	0	0
9	146	9	1	0	0	0	0	0	0	0
10	93	7	2	0	0	0	0	0	0	0
11	91	7	1	0	0	0	0	0	0	0
12	53	3	0	0	0	0	0	0	0	0
13	39	3	0	0	0	0	0	0	0	0
14	22	1	0	0	0	0	0	0	0	0
15	8	0	0	0	0	0	0	0	0	0
16	9	0	0	0	0	0	0	0	0	0
17	9	0	0	0	0	0	0	0	0	0
18	7	0	0	0	0	0	0	0	0	0
19	9	0	0	0	0	0	0	0	0	0

b)

Fig. 5.6 An exemplary histogram of (*r*,*l*) pairs for luminance of first frame in the test sequence *Basket* (4CIF sequence) for Inter quantization mode for a) *Q*=2 and b) *Q*=16 (H.263 system).

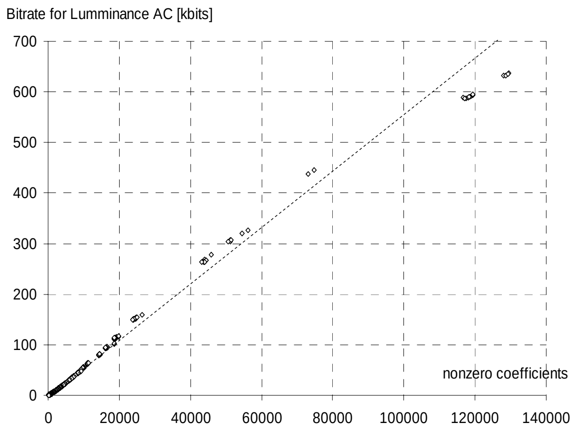
After quantization with a dead zone around zero, the coefficient F_{ij} remains nonzero if

$$F_{ij} \geq T_{ij} \quad \text{where threshold } T_{ij} = T(Q) \tag{5.5}$$

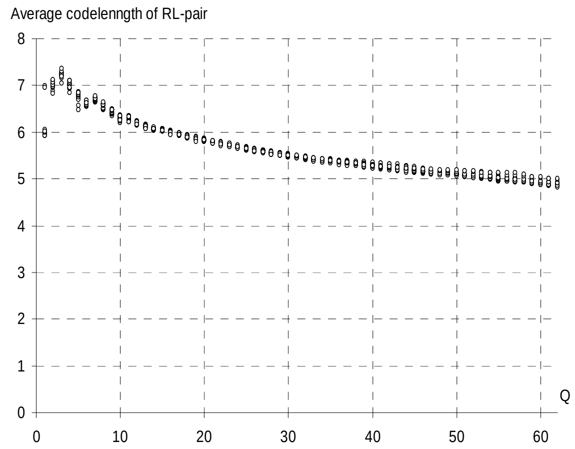
where T_{ij} denote threshold for *(i,j)*-th histogram (Fig. 5.1). The threshold function $T(Q)$ depends on quantization function in certain coding system and weighting matrices W. The number of bits $B_{\mathrm{VAR}}(Q)$ can be simply estimated as

$$B_{\mathrm{VAR}}(Q) \approx C \cdot \sum_{i,j} \sum_{|F_{ij}|=T_{i,j}}^{2048} H_{ij}(|F_{ij}|), \tag{5.6}$$

where summation is performed over DCT coefficients F_{ij} that exceed respective thresholds T_{ij} (Fig. 5.1). The parameter C denotes average Huffman code length for *RL*-pairs. The good accuracy is achieved for medium values of Q and typical value C=6.05 (average code length from MPEG-2 Huffman codes table). Nevertheless, the accuracy may be insufficient for larger range of Q. It is because the value of C actually depends on Q, sequence type and content (Fig. 5.7).



a)



b)

Fig. 5.7 a) The bitrate versus number of non-zero coefficients and b) the average Huffman code length for *RL*-pairs versus quantization scale factor Q.

This model can be transformed into a more accurate version. The improvement is based on observation of the properties of the probability density $p(r,l)$ of pairs (r,l) = (*Run*, *Level*). The majority of the most significant values of $p(r,l)$ are along $r = 0$ and $l = 0$ axes (Fig. 5.3÷5.6).

The number of bits needed for encoding component (luminance or chrominance) can be estimated as follows:

$$B_{\mathrm{VAR}}(Q) \approx \cdot \sum_{i,j} B_{ij}(Q) + 2 \cdot N_B , \tag{5.7}$$

where N_B stands for the number of coded blocks and expresses the number of bits for the *End of block* codes (if it exists in certain system). The $B_{ij}(Q)$ is defined as follows:

$$B_{\mathrm{ij}}(Q) \approx \sum_{l=|F_{ij}|=T_{ij}+1}^{40} C_l \cdot H_{ij}\left(|F_{ij}|\right) + 24 \cdot \sum_{|F_{ij}|=T_{i,j}+Esc}^{2048} H_{i,j}\left(|F_{ij}|\right) \ , \tag{5.8}$$

where the values of constants C_l for $l>4$ (Fig.5.2 and 5.3) are close to the code lengths for the *RL*-pairs equal to (0,*l*) and *Esc* means the lowest value of *Level* which is *ESCAPE* encoded. The first term expresses the bits for nonzero DCT coefficients encoded by Huffman codes, the second term stands for the bits related to *ESCAPE* codes and respective numerical values of *r* and *l*. The *Esc* value depends on coding system (*Esc*=41 for MPEG-2 and *Esc*=13 for H.263).

5.2.1 The model parameters estimation

The values of C_l for $l \leq 4$ have been estimated from probability distributions for pairs *(r,l)* (Fig.5.7 and 5.8).

$$C_{\mathrm{l}} = \sum_r \left[p(r,l) \cdot C(r,l)\right], \tag{5.9}$$

where *C(r,l)* is a code length for pair *(r,l)*. The average codeword obtained for level *l* is constant and independent from the image content.

Those functions are very similar for different frames and even for different sequences, thus can be treated as invariable. For both MPEG-2 and H.263 systems those curves are very similar. For estimation of the first four values of C_l the average curves of probability are used.

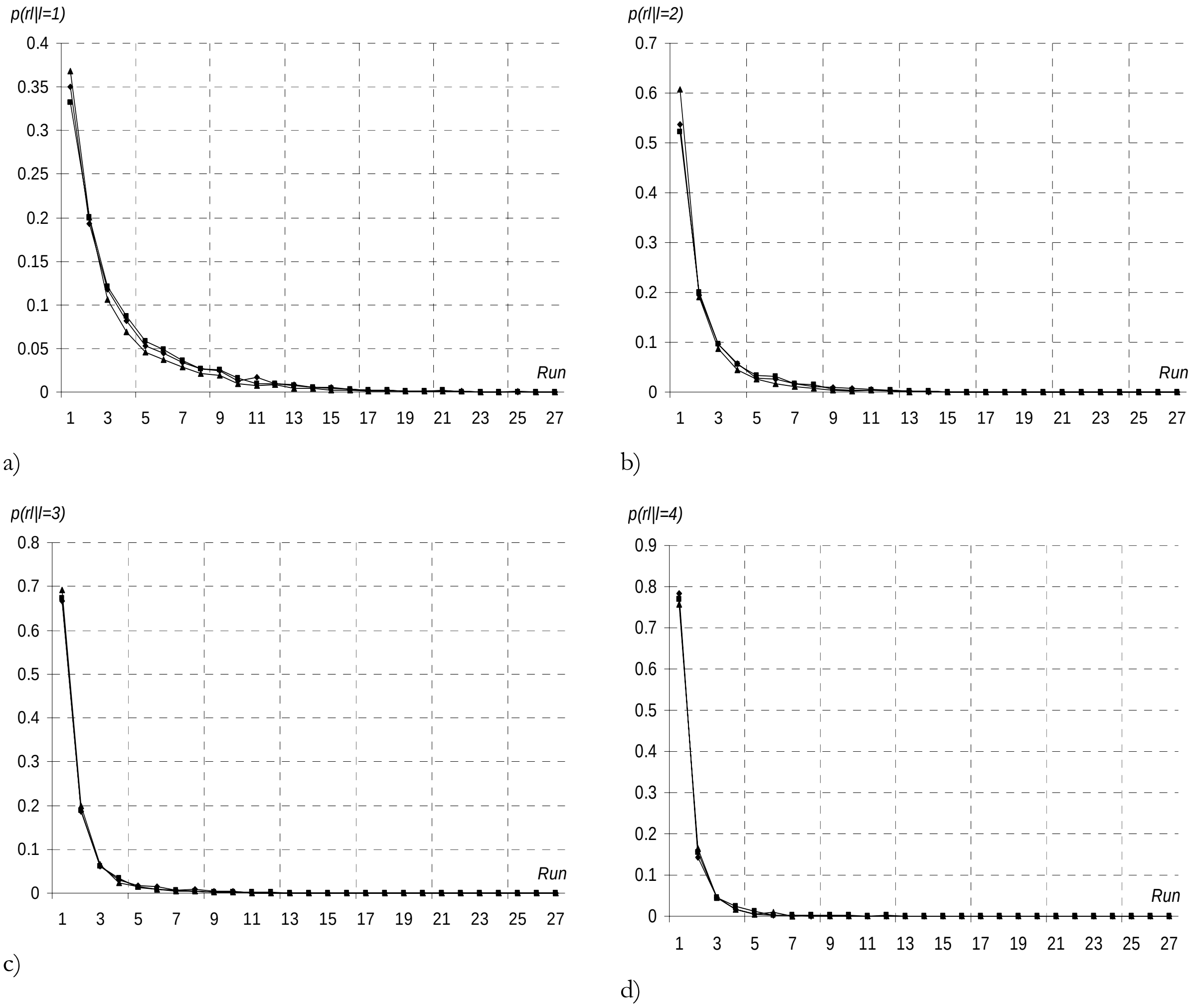


Fig. 5.8 Probability distributions of pairs *(r,l)* for various I-frames of sequence *Basket* for a) *Level*=1, b) *Level*=2, c) *Level*=3 d) *Level*=4 (MPEG-2 system).

Constants $C_5 \div C_{14}$ for MPEG-2 system or $C_5 \div C_6$ for H.263 system have been computed as an average code length of codes for pairs *(r,l)* with $r=0$ and $r=1$, what is done as follows:

$$C_l = p(0,l) \cdot C(0,l) + p(1,l) \cdot C(1,l) \tag{5.10}$$

Remaining C_l values ($C_{14} \div C_{40}$ for MPEG-2 system or $C_7 \div C_{12}$ for H.263 system) are set equal to code length of pair *(r,l)* with $r=0$.

$$C_l = C(0,l) \tag{5.11}$$

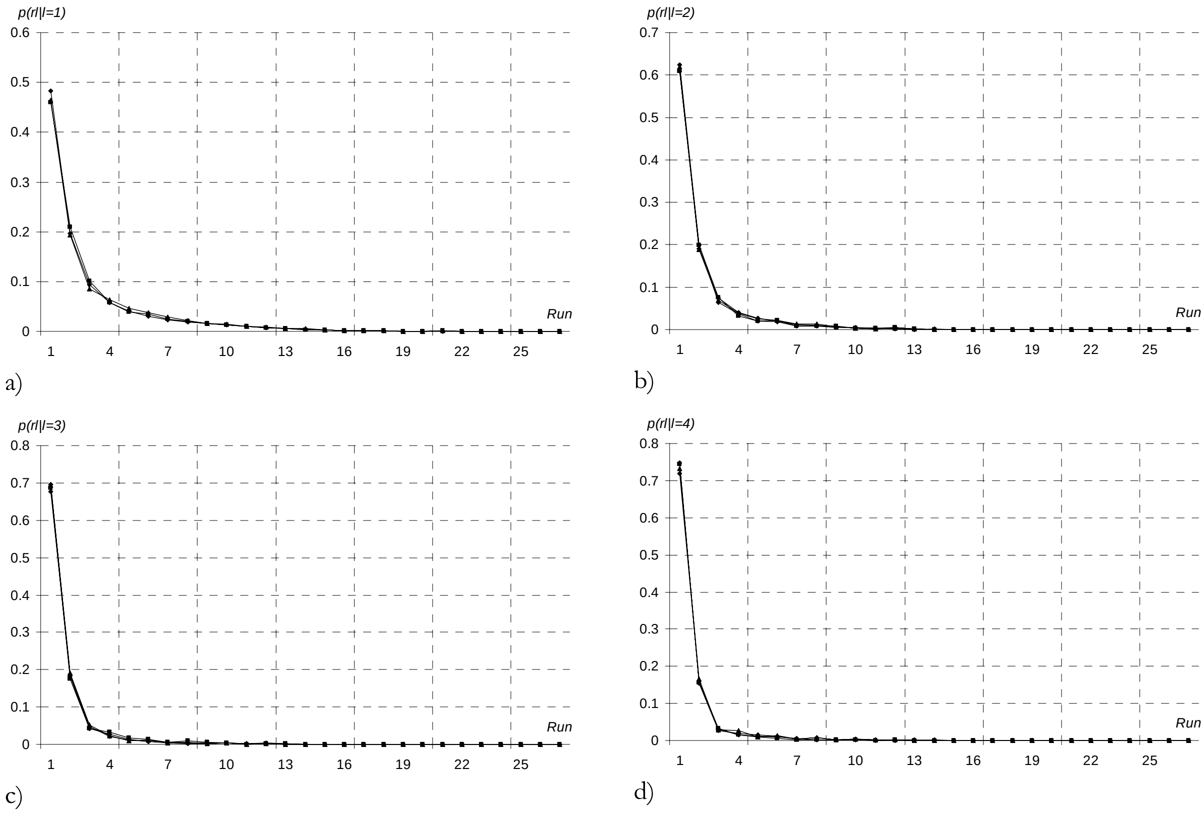


Fig. 5.9 Probability distributions of pairs *(r,l)* for several I-frames of sequence *Basket* for a) *Level*=1, b) *Level*=2, c) *Level*=3 d) *Level*=4 (H.263 system).

All the test video sequences were used to create the probability distribution *p(r,l)* of *RL*-pairs. Computed C_l values are shown in table 5.3. Having that average codelength the $B_{VAR}(Q)$ can be estimated from formula 5.7 and 5.8. Experimental results for estimation error of the number of bits necessary for encoding *RL*-pair with $l\leq 4$ is shown in the table 5.1. For the greater *l* the number of *RL*-pairs constitutes a smaller fraction of all *RL*-pairs, hence the contribution of corresponding estimation errors to the overall estimation error decreases as *l* increases.

Table 5.1 Maximum and average estimation error of average codelength C_l for all test sequences over whole range of quantization scale factor Q.

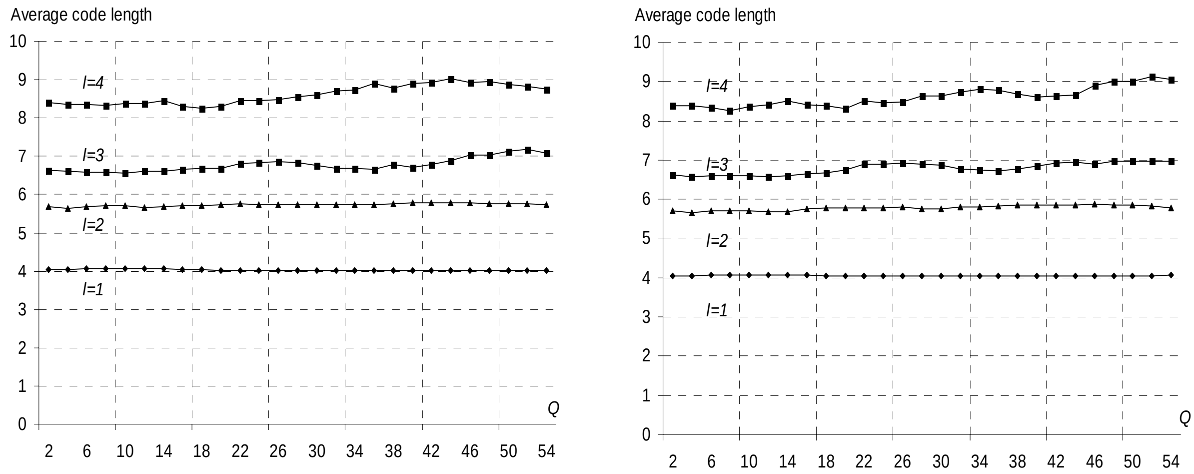
Level in RL-pair		***(r,1)***	***(r,2)***	***(r,3)***	***(r,4)***
ΔC_l	**Maximum**	2.3%	4.6%	8.7%	9.4%
	Average	1.1%	2.6%	4.9%	5.4%

For example, for the set of *RL*-pairs with $l=4$ the average approximation error is 5.4%, but number of *RL*-pairs in this set is below 6% of all *RL*-pairs. Hence, this contribution into overall approximation error does not exceed 1% (Table 5.1 and 5.2). The exemplary percentage of RL-pairs with certain *l* is shown in the table 5.2.

Table 5.2 .RL-pairs with certain l as a fraction of all RL-pairs for two test sequences *Basket* and *Cheer*.

Quantization scale factor Q	***Percentage of RL-pairs with l≤4***			
	Pairs (r, 1)	***Pairs (r,2)***	***Pairs (r,3)***	***Pairs (r,4)***
	[%]	[%]	[%]	[%]
8	45.12	16.25	8.68	5.49
16	49.87	16.35	8.25	4.93
24	52.90	16.06	7.65	4.35
32	54.66	15.59	6.94	3.86
40	55.51	15.14	6.50	3.43
48	56.22	14.44	5.92	2.79
56	56.29	13.76	5.44	2.27

On the figure 5.10 the changes of actual average codeword length (for *l≤4)* versus quantization scale factor *Q* are shown. For different frames and different sequences these figures are very similar. These average codeword lengths are independent from quantization scale factor *Q*, therefore they are treated as constant values.



a) b)

Fig. 5.10 Average code length for *RL*-pairs with *Level*=1, 2, 3 and 4 versus quantization scale factor *Q* for 16-th frame of sequence a) *Basket* and b) *Cheer*.

The table 5.3 shows values of C_l for Intra frame mode of coding. Note that there exist two sets of model parameters C_l: the first one for I-frames, and the second one for P- and B-frames. Each of them has to be estimated separately. These tables of average code length scan be used directly in control algorithm in order to estimate $B_{VAR}(Q)$.

Table 5.3 The C_l parameter values for all allowed *Level* values. Tables for I-frames in a) MPEG-2 system and b) H.263 system. Gray fields denote values for *Level*≤4.

l	C_l	*L*	C_l	*l*	C_l	*l*	C_l
1	4.0	11	13.2	21	15.0	31	16.0
2	5.6	12	14.1	22	15.0	32	16.0
3	6.7	13	14.1	23	15.0	33	16.0
4	8.5	14	14.1	24	15.0	34	16.0
5	9.5	15	15.0	25	15.0	35	16.0
6	9.5	16	15.0	26	15.0	36	16.0
7	11.5	17	15.0	27	15.0	37	16.0
8	13.2	18	15.0	28	15.0	38	16.0
9	13.2	19	15.0	29	15.0	39	16.0
10	13.2	20	15.0	30	15.0	40	16.0

a)

L	C_l	*l*	C_l
1	4.2	11	12.0
2	6.4	12	12.0
3	8.0	-	-
4	8.8	-	-
5	9.5	-	-
6	10.4	-	-
7	10.0	-	-
8	11.0	-	-
9	11.0	-	-
10	12.0	-	-

b)

5.3 Accuracy of model

The values of constants C_l (Table 5.3) have been estimated for a set of training video sequences for MPEG-2 encoder with default quantization matrices and with the first set of Huffman codes. These C_l values have been used for checking estimation accuracy of the bitstream value $B_{VAR}(Q)$. The estimation error $\varepsilon_B(Q)$ is defined as follows:

$$\varepsilon B(Q) = \frac{|B_e(Q) - B_x(Q)|}{B_x(Q)} \cdot 100\%, \tag{5.12}$$

where the $B_x(Q)$ is the measured value of bitstream resulting from encoding of DCT coefficients (excluding Intra DC), and $B_e(Q)$ is the estimated value of bitstream. Table 5.3 shows the average and maximum estimation error for small, medium and large value of quantization scale factor Q (16, 32 and 48). The maximum $\varepsilon_B(Q)$ is always below 11% in the whole range of Q for all test video sequences. Moreover, the average $\varepsilon_B(Q)$ is below 3.5% and is smaller for larger Q values, and for those values does not exceed 2.5% (Fig. 5.6). The model is quite accurate in the whole range of quantization scale factor Q.

Table 5.4 Maximum and average estimation error computed for exemplary three sequences (Basket, Cheer and Warner) for three values of quantization scale factor Q.

Frame Type	*Q=16*		Q=32		Q=48	
	average ε_B	*maximum ε_B*	*average ε_B*	*maximum ε_B*	*average ε_B*	*maximum ε_B*
Basket						
I	1.21	7.98	1.03	8.19	1.12	8.78
P	1.84	6.67	1.41	7.53	1.52	7.02
B	2.01	6.45	1.86	7.69	1.92	8.59
Cheer						
I	1.89	6.65	1.72	5.18	1.60	7.25
P	2.21	8.16	2.56	6.54	2.01	6.72
B	2.64	8.51	2.98	7.93	2.32	8.83
Warner						
I	2.31	5.47	2.29	4.83	1.70	8.92
P	2.98	6.52	2.64	5.22	1.89	9.45
B	3.34	6.10	3.17	8.35	2.03	10.02

Figure 5.11 shows results of AC DCT bitstream estimation results for luminance and chrominance components separately. Those results are obtained by encoding of 250 frames of sequence *Basket*. For other test sequences results are very similar with the same maximum error of bitstream estimation. For the quantization scale factors Q which result in substantial bitstream, the approximation error is insignificant when compared with the bitstream difference corresponding with two neighboring Q values.

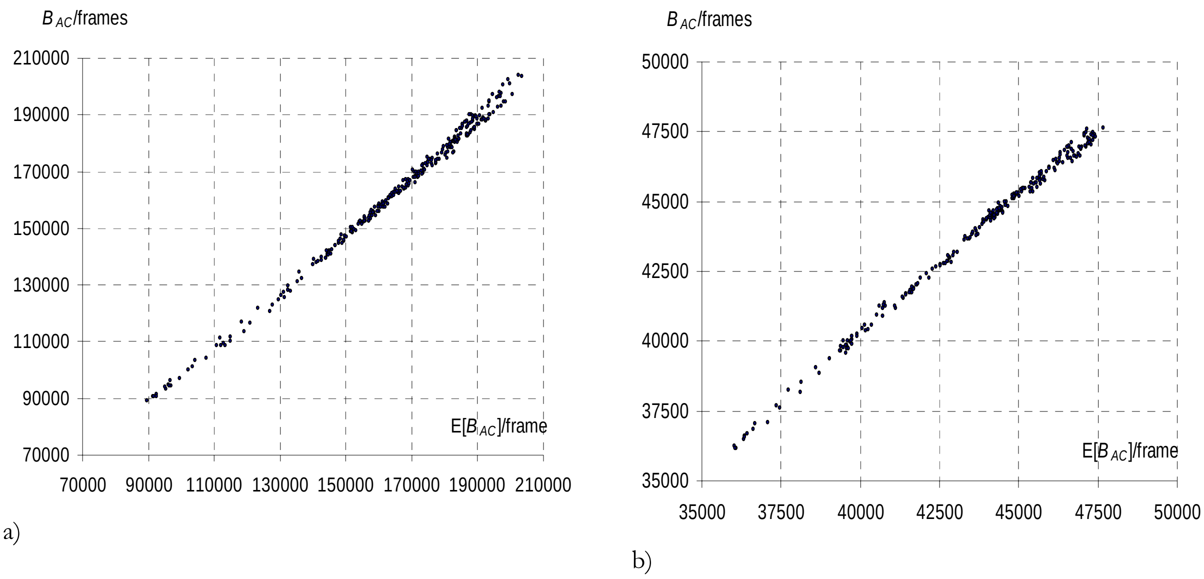


Fig. 5.11 Measured bitstream B_{AC} of the AC luminance coefficients versus $E[B_{AC}]$ estimated bitstream from (5.7) for the test sequence *Basket* in the 4CIF resolution for a) luminance component and b) chrominance component.

5.4 Application to the bitrate control

This model is applicable for the whole frames as well as for individual slices. In this work we use it for determining Q value for the whole frame. The implementation is tested on MPEG-2 video coder exploiting the software of MPEG-2 Test Model 5 [TM5]. Only control algorithm in this software has been changed.

The proposed algorithm requires some initial computations. At the beginning of encoding process the threshold values are determined. These threshold values are computed for given weights matrices W and for given quantization scale factors Q. The values of thresholds T_{ij} for $F_{ij} \geq 1$ (nonzero coefficients) and for $F_{ij} \geq Esc$ must be determined. Subsequently, for each encoded frame the following algorithm is applied:

- First step: computing histogram of DCT coefficients (note that all hybrid coders need to compute DCT transformation in order to encode image or image prediction error, therefore DCT computing is not an additional cost of this control algorithm.).

- Second step: estimating $B_{VAR}(Q)$ for certain quantization scale factor Q (exploiting histogram of DCT coefficients and C_l values with formula 5.7).

- Third step: choosing quantization scale factor Q value (B_{CONST} is determined by predicting or pre-encoding. Subsequently, the total B value for several quantization scale factors Q is estimated. The quantization scale factor Q which gives estimated bitstream closest to the required bitstream is chosen).

The proposed control algorithm is also suitable for H.263 coder. In H.263 coder the quantization function is very simple, therefore it is easy to derive the threshold function. The threshold function obtained is as follows:

$$T_{ij}(Q) = l \cdot 2 \cdot q \tag{5.13}$$

Due to complexity of quantization process in MPEG-2 the threshold function is more complicated. It has to be derived from three-staged quantization process.

Threshold function:

The derivation of threshold function for MPEG-2 system is presented below. In this system the quantization function is as follows:

$$F_{ij}^{'} = s \cdot \left\lfloor \frac{(y_{ij} + d_{ij})}{2 \cdot Q} \right\rfloor; \tag{5.14}$$

where s is the sign of the coefficient and y_{ij} and d_{ij} are defined as follows:

$$y_{ij} = \left\lfloor \frac{32 \cdot F_{ij} + \frac{w_{ij}}{2}}{w_{ij}} \right\rfloor \tag{5.15}$$

where F_{ij} is the (i,j)-th DCT coefficient and w_{ij} is the quantization weight. Subsequently, the d_{ij} term is computed:

$$d_{ij} = \left\lfloor \frac{3 \cdot Q + 2}{4} \right\rfloor; \tag{5.16}$$

this term is responsible for the appropriate rounding of float values to integer and introduces dead zone into quantizer. Finally, the quantized DCT coefficient is obtained. It needs to be noted that this quantization process is performed in three steps. First, the quantization by w_{ij} is performed and subsequently, quantized with quantization scale factor Q. Those quantized values are always rounded towards zero. The quantization function can be described as follows:

$$F_{ij}^{'} = f_q(Q, w_{ij}, F_{ij}) \tag{5.17}$$

The threshold is the lowest F_{ij} value of DCT coefficient which will be nonzero after quantization with certain quantization scale factor Q and quantization weight w_{ij}. In general, the threshold determines minimal F_{ij} value of DCT coefficient which gives value above certain value l after quantization process. Hence:

$$F_{ij}^{'} \geq l, \tag{5.18}$$

After substituting the function 5.15 into the formula 5.16 the inequality is obtained:

$$l \leq f_q(Q, w_{ij}, F_{ij}) \tag{5.19}$$

thus:

$$l \leq \left\lfloor \left(\left\lfloor \frac{32 \cdot F_{ij} + \frac{w_{ij}}{2}}{w_{ij}} \right\rfloor + \left\lfloor \frac{3 \cdot Q + 2}{4} \right\rfloor \right) \cdot \frac{1}{2 \cdot Q} \right\rfloor, \tag{5.20}$$

Threshold function is derived after finding the minimal F_{ij} satisfying the above inequality:

$$T_{ij}(Q) = \left\lceil \frac{w_{ij} \cdot l \cdot 2 \cdot q - w_{ij} \cdot \left\lfloor \frac{(3 \cdot Q + 2)}{4} \right\rfloor - \frac{w_{ij}}{2}}{32} \right\rceil, \tag{5.21}$$

where l denotes required *Level* value after quantization. It means that threshold function detects the DCT coefficients which will be grater then l value after quantizing with quantization scale factor Q. Threshold values T_{ij} are determined before frame is encoded. If quantization matrices do not change, it will be possible to determine the threshold value even before video sequence encoding process starts. It enables fast detection of DCT coefficients which will be zero after quantization. Exploiting this feature the quantization process can be significantly accelerated by the elimination superfluous operations (of three-staged quantization of the DCT coefficient resulting in zero values).

5.5 Results of experiments for MPEG-2 coder

For experiment purposes the set of test video sequences was used. This set consists of 50 Hz progressive sequences such as: *Basket*, *Flower Garden*, *Mobile*, *Funfair*, *Football*, *Cheer*, *Bus*, *Stefan*, and 25 Hz progressive sequences *Icon*, *Warner*, *Universal*. The Test Model 5 MPEG-2 encoder was used as a reference coder. The proposed control algorithm has been implemented in this reference software. Therefore, all sequences were encoded with the same coder with both default and proposed control algorithm. Standard GOP structure **I**BB**P**BB**P**BB**P**BB(**I**..) was used. In each experiment the 250 frames from every test sequences were encoded. Sequences were encoded in both progressive and interlaced mode of coding.

Average bitrate and quality

Table 5.5 Experimental results for default TM5 control algorithm and new based on DCT histogram analysis – interlaced mode of coding.

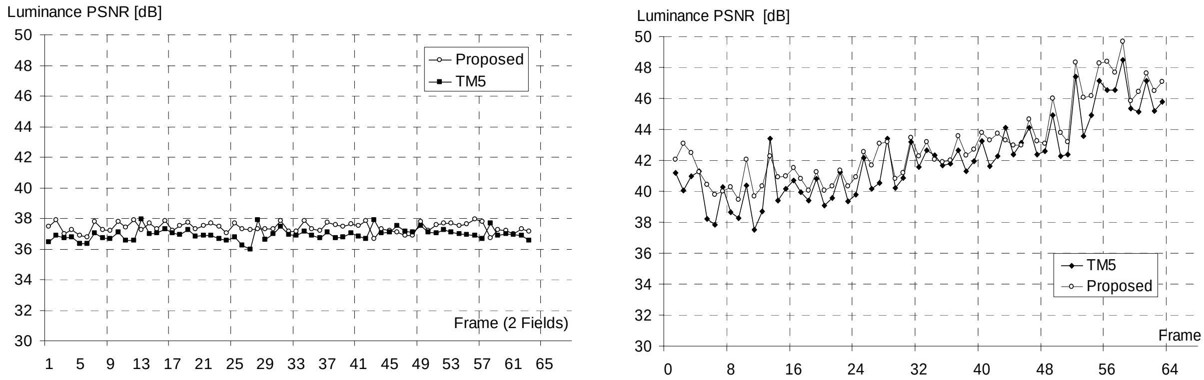
Bitrate [Mbits]	Bitrate [Kbits]		PSNR [dB]		ΔPSNR (P_1-P_2) [dB]	Standard deviation σ [dB]	
	Proposed	TM5	Proposed (P_1)	TM5 (P_2)		Proposed	TM5
Football							
3	2 955	2 972	36.79	36.76	**+0.03**	0.928	1.081
4	3 915	3 948	38.18	38.04	**+0.14**	0.973	1.012
5	4 837	4 866	39.26	39.01	**+0.25**	1.019	1.024
6	5 878	5 933	39.92	39.74	**+0.18**	1.263	1.146
7	6 913	6 901	40.58	40.27	**+0.31**	1.045	1.054
8	7 824	7 853	41.35	40.96	**+0.39**	1.118	1.123
Cheer							
3	2 918	2 904	34.86	34.59	**+0.27**	0.260	0.409
4	3 894	3 883	36.25	35.94	**+0.31**	0.268	0.387
5	4 890	4 861	37.39	36.94	**+0.45**	0.318	0.378
6	5 859	5 843	38.31	37.76	**+0.55**	0.358	0.475
7	6 850	6 825	39.00	38.45	**+0.55**	0.380	0.513
8	7 842	7 805	39.83	39.06	**+0.77**	0.421	0.663
Stefan							
3	2 916	2 922	41.01	40.83	**+0.18**	0.721	0.691
4	3 892	3 903	42.15	41.76	**+0.39**	0.721	0.764
5	4 879	4 876	43.11	42.47	**+0.64**	0.803	0.804
6	5 841	5 853	43.83	43.04	**+0.71**	0.579	0.864
7	6 781	6 834	44.40	43.51	**+0.89**	0.723	0.914
8	7 784	7 821	45.07	43.98	**+1.09**	0.791	1.001
Universal							
3	2.908	2 939	42.67	42.10	**+0.57**	1.745	1.784
4	3 923	2 933	44.00	43.32	**+0.68**	1.780	1.794
5	4 938	4 926	45.03	44.23	**+0.80**	1.855	1.883
6	5 839	5 921	45.76	44.98	**+0.78**	1.867	1.882
7	6 844	6 920	46.46	45.65	**+0.81**	1.645	1.861
8	7 900	7 952	47.04	46.19	**+0.85**	1.656	1.855
Warner							
3	2 923	2 904	40.48	39.71	**+0.77**	2.784	2.919
4	3 879	3 884	41.85	41.12	**+0.73**	2.698	2.700
5	4 889	4 854	42.92	42.08	**+0.86**	2.571	2.558
6	5 895	5 825	43.72	42.89	**+0.83**	2.434	2.502
7	6 933	6 821	44.50	43.66	**+0.84**	2.463	2.519
8	7 838	7 830	45.10	44.34	**+0.76**	2.359	2.538
Icon							
3	2 922	2 918	47.90	47.89	**+0.01**	2.500	2.758
4	3 910	3 894	48.96	48.81	**+0.15**	2.556	3.096
5	4 874	4 868	49.69	49.53	**+0.16**	2.560	2.998

Results for interlaced and progressive coding of video sequences in CBR mode are shown in table 5.5 and 5.6. Both default and proposed control algorithm achieves required bitrate with the same accuracy. However, coder with the new control algorithm obtains higher average PSNR of encoded video sequences. This gain is about 0.2÷0.5 dB but for higher bitrates it can achieve 1 dB. Thus, the microscopic model enables to achieve even better results then global model (chapter 4). Moreover, variance of the PSNR is slightly lower then that for default control. Differences of control algorithm performance between interlaced and progressive mode of coding are very small and depend on sequence content.

Table 5.6 Experimental results for default TM5 control algorithm and new based on DCT histogram analysis – progressive mode of coding.

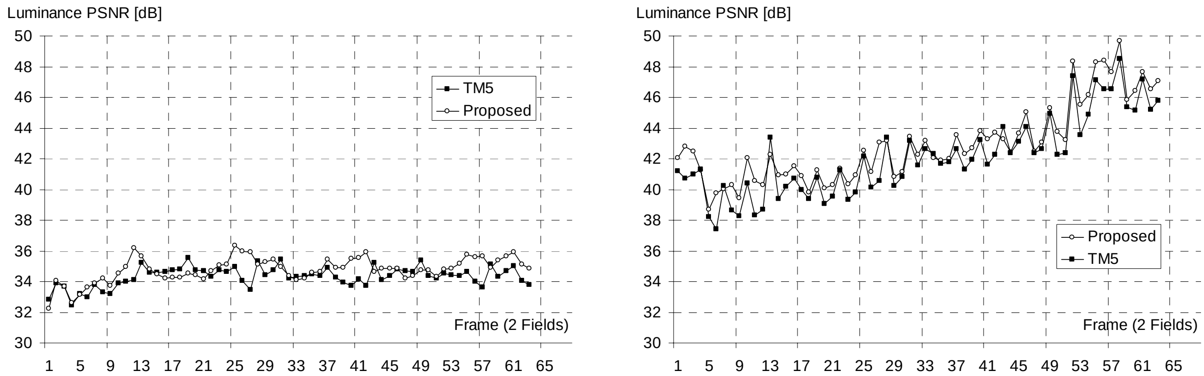
Bitrate [Mbits]	Bitstrate [Kbits]		PSNR [dB]		ΔPSNR (P_1-P_2) dB]	Standard deviation σ	
	Proposed	TM5	Proposed (P_1)	TM5 (P_2)		Proposed	TM5
Football							
3	2 916	2 904	36.85	36.76	**+0.09**	0.888	1.081
4	3 893	3 886	38.22	38.04	**+0.17**	0.838	1.012
5	4 783	4 874	39.29	39.01	**+0.27**	0.873	1.024
6	5 847	5 855	40.11	39.74	**+0.36**	1.032	1.146
7	6 827	6 836	40.80	40.27	**+0.53**	1.024	1.054
8	7 926	7 943	41.51	40.96	**+0.54**	1.130	1.123
Cheer							
3	2 964	2 925	31.68	31.58	**+0.10**	0.500	0.517
4	3 953	3 920	33.37	33.00	**+0.37**	0.554	0.673
5	4 941	4 920	34.67	34.32	**+0.35**	0.641	0.900
6	5 946	5 919	35.82	35.33	**+0.48**	0.740	1.022
7	6 934	6 919	36.56	36.18	**+0.38**	0.827	1.209
8	7 916	7 918	37.42	36.97	**+0.45**	0.912	1.198
Stefan							
3	2 958	2 950	37.72	37.95	*-0.23*	0.620	0.569
4	3 946	3 949	39.14	39.28	*-0.14*	0.848	0.784
5	4 934	4 950	40.38	40.17	**+0.21**	0.898	1.015
6	5 908	5 946	41.31	30.88	**+0.43**	0.982	1.199
7	6 917	6 945	42.16	41.55	**+0.61**	1.114	1.499
8	7 843	7 945	43.16	42.14	**+1.02**	1.106	1.455
Universal							
3	2 964	2 982	41.02	30.33	**+0.69**	1.738	1.736
4	3 946	3 985	42.37	41.64	**+0.73**	1.689	1.752
5	4 950	4 990	43.47	42.54	**+0.93**	1.808	1.932
6	5 974	5 998	44.22	43.30	**+0.92**	1.821	1.939
7	6 981	7 007	44.90	43.95	**+0.95**	1.836	1.844
8	7 909	8 014	45.41	44.48	**+0.93**	1.672	1.863
Warner							
3	2 965	2 962	38.56	37.69	**+0.87**	1.873	2.009
4	3 958	3 943	39.89	38.95	**+0.94**	2.171	2.492
5	4 954	4 955	40.76	39.49	**+0.73**	2.372	2.639
6	5 947	5 902	42.22	41.45	**+0.77**	2.315	2.380
7	6 915	6 941	42.91	42.05	**+0.86**	2.328	2.240
8	7 949	7 939	43.51	42.53	**+0.98**	2.363	2.335
Icon							
3	2 966	2 967	46.21	46.16	**+0.10**	2.069	2.17
4	3 927	3 957	47.00	46.93	**+0.07**	2.308	2.431
5	4 943	4 953	47.79	47.64	**+0.15**	2.500	2.852

Quality of encoded frames and slices



a) b)

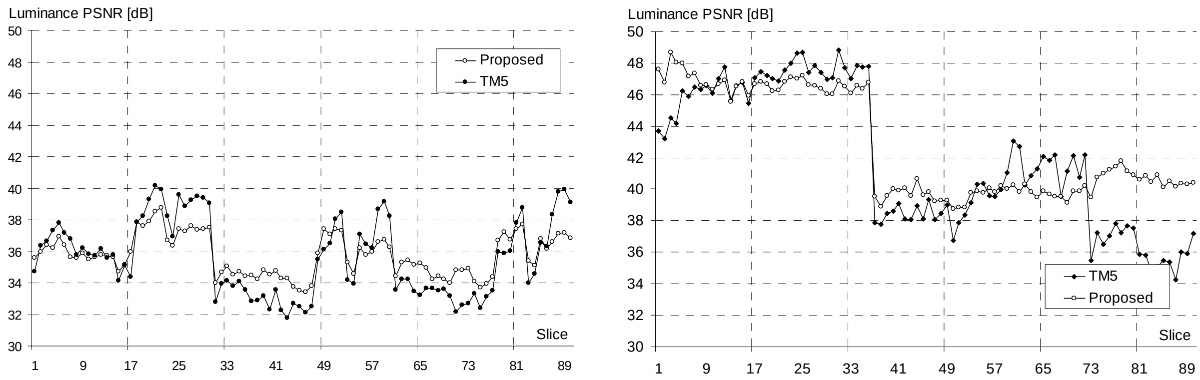
Fig. 5.13 Sequence quality (PSNR) versus frame number (for **progressive** MPEG2 encoding), for sequences a) *Cheer* and b) *Warner*



a) b)

Fig. 5.14 Sequence quality (PSNR) versus frame number (for **interlaced** MPEG2 encoding), for sequences a) *Cheer* and b) *Warner*

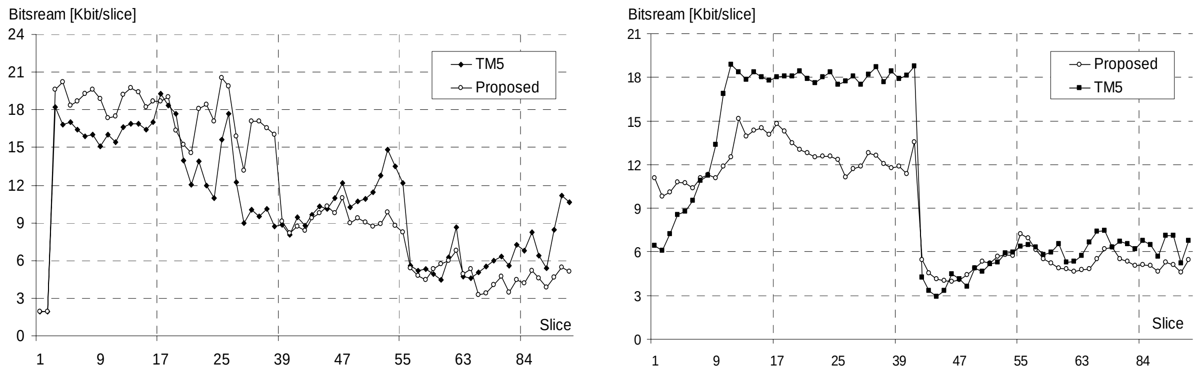
The figures above show these differences of performance between progressive and interlaced mode of coding graphically. Differences are very small and become visible after encoding many frames. The Figures 5.15 show each slice of macroblock of encoded frames quality. When coder with the proposed control algorithm is applied, the PSNR curve is much smoother, hence quality of image pieces is equalized giving better subjective quality. Due to that, better average PSNR of encoded video sequence is achieved.



a) b)

Fig. 5.15 Graph of encoding sequence quality (PSNR) versus slice number (for **progressive** MPEG2 encoding) for sequences a) *Cheer* and b) *Warner*

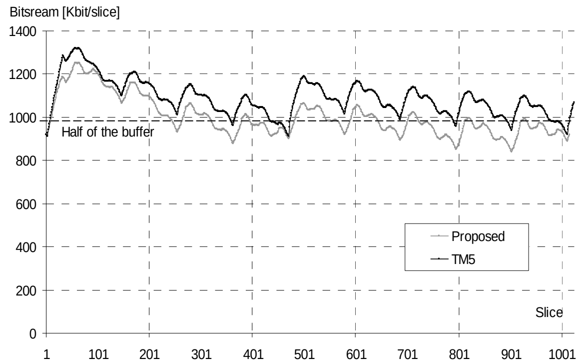
Short-term bitrate and buffer occupancy



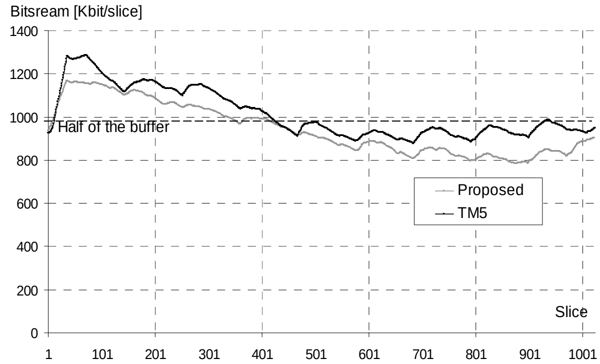
a) b)

Fig. 5.16 Graph of the slice bitstream versus slice number (for **progressive** MPEG2 encoding) for sequences a) *Cheer* and b) *Warner*

Figure 5.16 shows bitstream necessary for encoding consecutive rows of macroblocks (slices). Variation of slice bitstream is greater when coder uses proposed control algorithm, and due to that VBV buffer is more efficiently used what results in better image quality. Despite greater variations of slice bitstream the VBV buffer is never underflowed or overflowed.



a)



b)

Fig. 5.17 Graph of the VBV buffer occupancy versus slice number (for **progressive** MPEG2 encoding) for sequences a) *Cheer* and b) *Warner*

Variable bandwidth of transmission channel

When available bitrate of channel vary the control algorithm should dynamically adjust video encoding bitrate by fast changing of encoding parameters. Such algorithm has to be fast enough in order to avoid buffer overflow or underflow under rapid changes of a transmission channel bandwidth. Below the experimental results of coding with variable bitrate channel are presented.

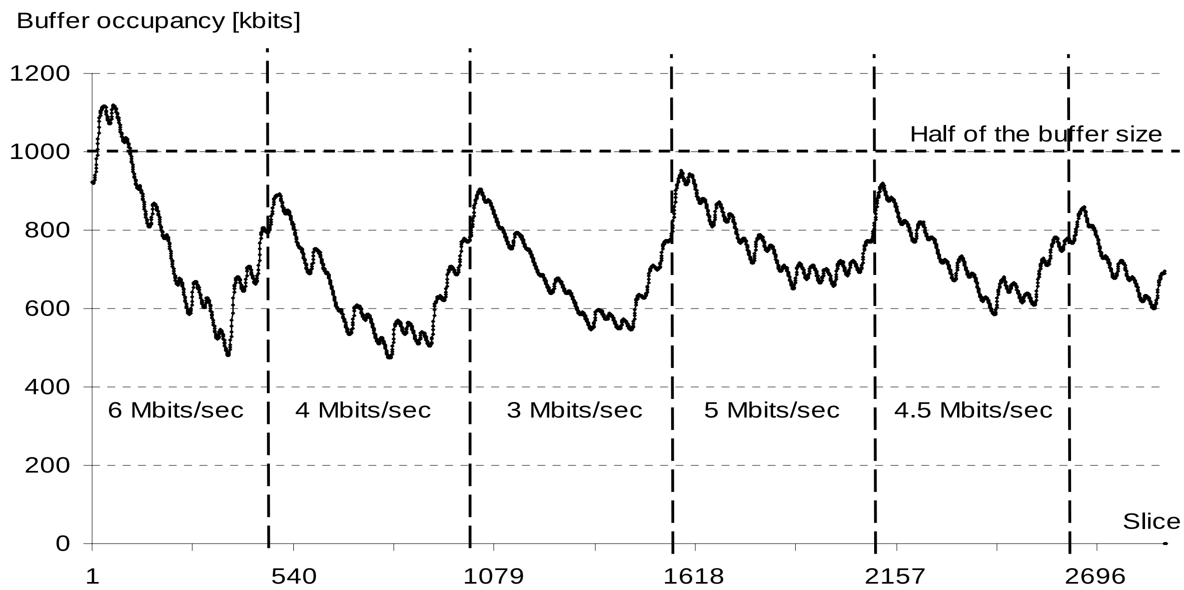


Fig. 5.18 Graph of the VBV buffer occupancy versus slice number. This experimental result is obtained for sequence *Basket* encoded with variable bitrate channel.

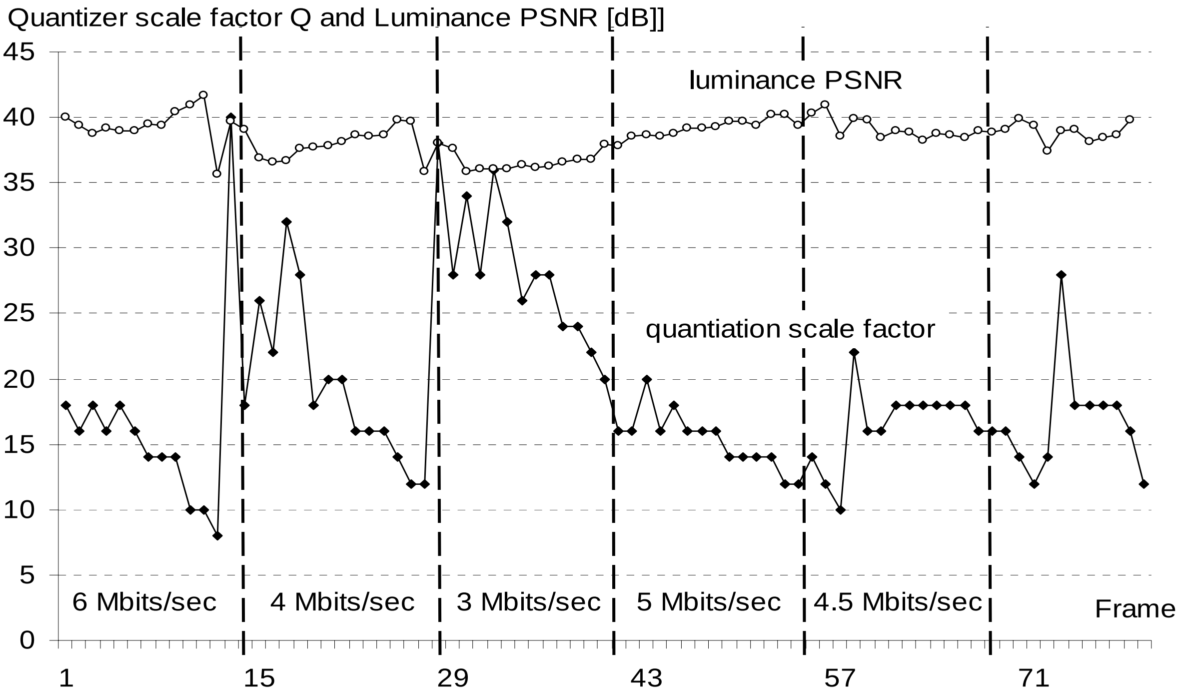


Fig. 5.19 Graph of the quantization scale factor Q value and luminance PSNR for consecutive frames of sequence *Basket* encoded with variable bitrate channel.

The control algorithm reacts to rapid changing of the channel bitrate with very slight delay. This delay is equal to time between two points (in time) when the quantization scale factor Q is set. In this experiment quantization scale factor Q was calculated for the whole frame, which means that reaction delay is equal to time of frame encoding i.e. 40 ms. Moreover, if the control algorithm were applied for slice level (instead of frame level) the reaction delay would be decreased to about 1.1 ms. Presented experimental results indicate that the proposed control algorithm can be use in video hybrid coders which work with variable bitrate channels over communication networks, especially for wireless networks.

Average length of codeword

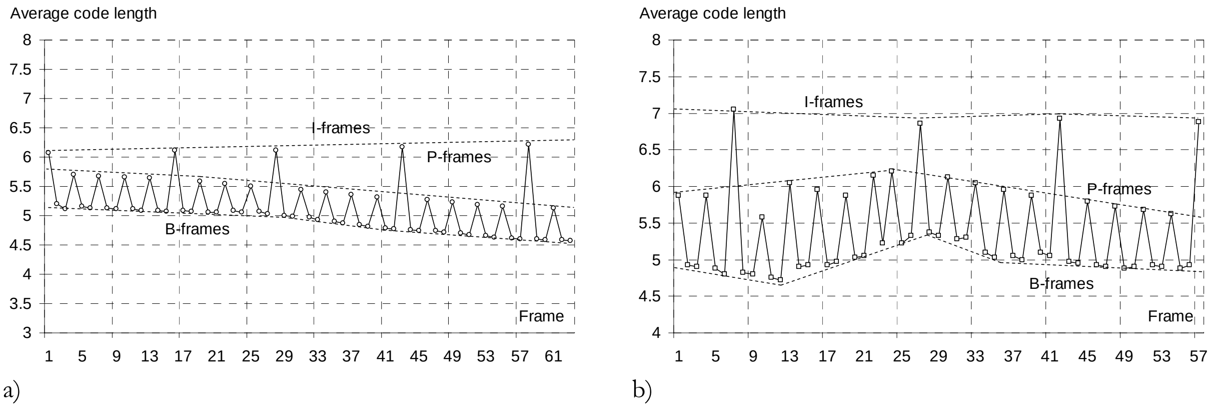


Fig. 5.20 Average code length for AC DCT coefficients encoding versus frame number (for **progressive** MPEG2 encoding), for sequences a) *Cheer* and b) *Warner*

On the figure 5.20 the average code length of *RL*-pair is shown. There are different lengths for I-, P- and B-frames. Dashed line on these figures connects points of the same type frames (I-, P- or B-frames). That line shows changes of *C* through encoded sequence, and how the average codeword length depends on image content and mode of frame coding as well as differs for various sequences.

Computational cost of control algorithm:

Table 5.7 Computational profits: T_Q gross profit, T_H cost, T_{OVERAL} net profit for two bitrates 4 and 6 Mbits/sec for video sequences *Basket, Cheer* and *Warner.*

Frame	4 Mbits/sec			6 Mbits/sec		
	ΔT_Q[%]	ΔT_H[%]	$\Delta T_{OVERALL}$[%]	ΔT_Q[%]	ΔT_H[%]	ΔT_{OVERAL}[%]
Basket						
I	19.88	-0.1	19.78	10.16	-0.1	10.16
P	4.32	-0.1	4.22	2.91	-0.1	8.73
B	3.62	-0.1	3.52	3.31	-0.1	26.48
Average profit per GOP			**5.05**			**3.78**
Cheer						
I	18.87	-0.1	18.77	8.73	-0.1	8.63
P	3.37	-0.1	3.27	2.36	-0.1	2.26
B	3.21	-0.1	3.11	2.81	-0.1	2.71
Average profit per GOP			**4.35**			3.09
Warner						
I	13.67	-0.1	13.57	12.15	-0.1	12.05
P	6.54	-0.1	6.44	4.18	-0.1	4.08
B	7.79	-0.1	7.69	5.50	-0.1	5.40
Average profit per GOP			**7.97**			**5.62**
For all video test sequences						
Average profit per GOP			**4.93**			**3.84**

For computation cost measurement the three measures are defined as follows:

$$\Delta T_Q = \frac{T_C - T_T}{T_C} \cdot 100\%, \tag{5.21}$$

where T_C – total time of frame encoding, T_T – total time of frame encoding by coder with modified quantization function, and factor T_H:

$$\Delta T_H = \frac{T_C - T_{CH}}{T_C} \cdot 100\%, \tag{5.22}$$

where T_{CH} – total time for frame encoding including time of the histogram creating. The overall cost/profit factor $T_{OVERALL}$ is defined as follows:

$$\Delta T_{OVERALL} = \frac{T_{CH} - T_T}{T_C} \cdot 100\%, \tag{5.23}$$

The computational cost increase due to application of control algorithm does not exceed 0.1%. The application of threshold function to eliminate superfluous operations of quantization brings savings of about **3÷5%** of total computational time per GOP (Table 5.7). If computational cost of motion estimation is not taken into account the savings will grow up to **20%** for each mode of coding (and also per GOP).

5.6 Conclusions

The proposed model of the bitstream can be used in control algorithms for video hybrid coders such as H.261, H.263, MPEG-1 and MPEG-2. The mentioned coders have similar structure and functionalities and this is why the conclusions for MPEG-2 are valid for all of them. The author confirms that by some experiments with H.263 video coder. It is also proved that control algorithm using proposed microscopic bitstream model works properly both for smaller resolution (CIF, QCIF) and higher resolution sequences (16CIF). Additionally, it is possible to modify quantization scale factor value Q for each slice of image. Hence, coder can adjust its parameters to varying channel bandwidth very fast.

Numerous experiments prove that the proposed control algorithm is versatile and can be used to control the video hybrid coder in various environments and with various sets of constraints. It is worth to be noting that described algorithm can be applied in real-time video hybrid coders working in wireless networks (e.g. mobile phones).

The advantages of new control algorithm:

- low computational complexity of the algorithm itself
- used in a coder enables overall computational cost reduction
- ability of controlling coder slice after slice makes it well suited for coders working with channels in which bitrate changes rapidly

Chapter 6

Scalable coder control

6.1 Problem definition

When coder is working in scalable mode the new problem appears: how to allocate bits between two or more layers of bitstream? And how to adjust quantization scale factor Q in each layer in order to get an adequate quality in the layers?

In general, scalability means that coder produces a bitstream partitioned into layers [Hask97, Giro97, Mack02, Blas02, Doma00e, Chim00]. These layers represent various qualities or various spatial and/or temporal resolutions of encoded video. In this chapter the two-layer scalable coder will be considered [Doma99c].

6.2 Two-layer scalable coding

In his dissertation [Mack02] S. Mackowiak presents original proposal of video coder for spatio-temporal scalability. His coder is based on hybrid DCT-based video coders like MPEG-2. The low resolution base layer bitstream is fully compatible with the MPEG-2 standard [Doma99a]. The whole structure exhibits high level of compatibility with individual functional blocks of MPEG-2 encoders and the enhancement bitstream exploits the MPEG-2 bitstream semantics and syntax, with some modifications

[Doma00b, Doma00d, Doma00f]. The video sequence structure is that of MPEG-2. The sequence consists of Group of Pictures (GOPs) that are defined as content access units.

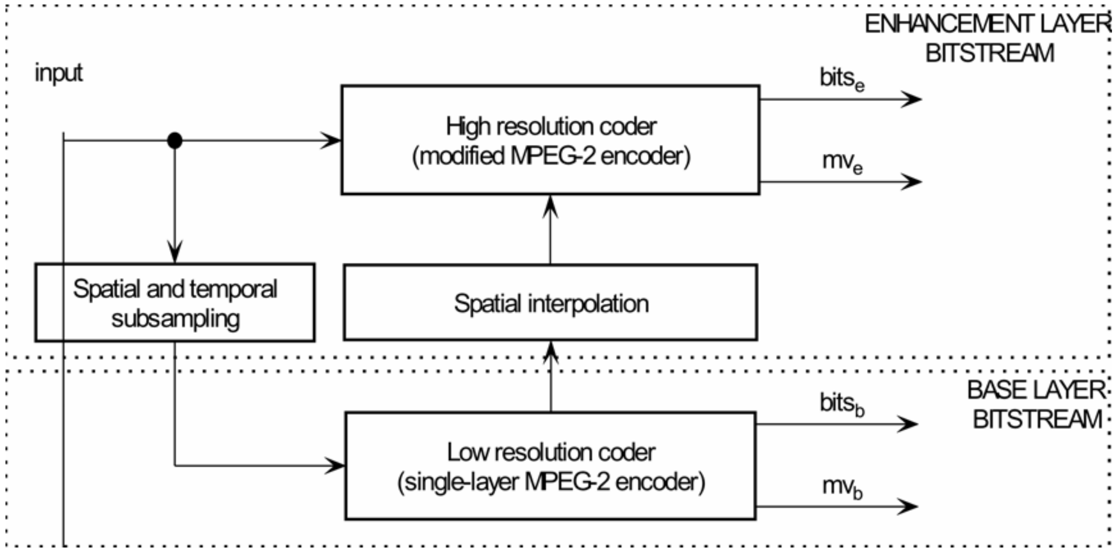


Fig.6.1. Structure of the two-layer scalable coder [Mack02].

The scalable encoder consists of two motion-compensated hybrid encoders (Fig. 6.1), which encode a video sequence and produce two bitstreams corresponding with two different levels of spatial and temporal resolution. In that structure, it is also possible to encode a video sequence without temporal decomposition.

The basic structure of a group of pictures (GOP) consists of I- P- and B-frames (Fig. 6.2). The variant with 3 B-frames between two consecutive I- or P-frames has been chosen because of simple temporal decimation with factor 2.

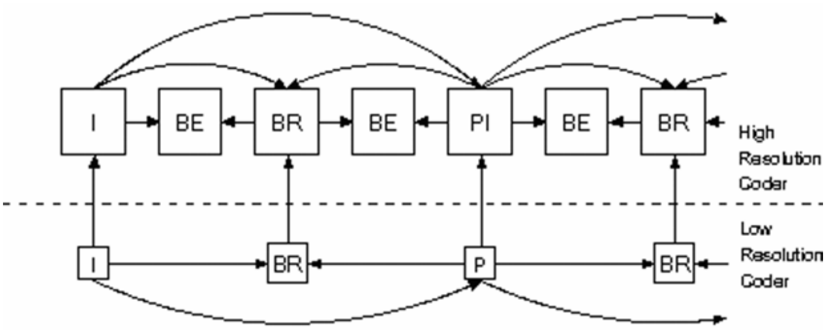


Fig. 6.2. Selected GOP structure with P- and B-frames in low and high resolution bitstreams.

In the proposed encoder, the temporal resolution reduction is achieved by partitioning the stream of B-frames: every second frame is skipped in the low resolution encoder. There are two types of B-frames, i.e. BE-frames that exist only in the high resolution sequence and BR-frames that exist in both sequences (Fig. 6.2). For the high resolution video sequence the number of B-frames between two consecutive I- or P-frames has to be even. Moreover, PI-frames, i.e. frames which are encoded without motion vectors exist in the enhancement layer

The motion-compensated predictor employed in the high resolution layer uses new prediction for B-frames. 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. 6.2).

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: every second frame is not included intthe base layer. The bitstream produced in the base layer is described by MPEG-2 standard syntax. The proposed encoder applies the independent motion compensation loops in all layers.

The author of this dissertation creating the control algorithm for mentioned scalable coder proposes the original, 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.

6.2.1 Improved prediction of BR-frames

First improvement is proposed for BR-frames, those which are B-frames represented in both layers (Fig. 6.2). Each macroblock in a high resolution BR-frame can be predicted from the following reference macroblocks (Fig. 6.4):

- previous reference macroblock PM (I- or P-frame),
- next reference macroblock NM (I- or P-frame),
- interpolated reference macroblock IM (BR-frame).

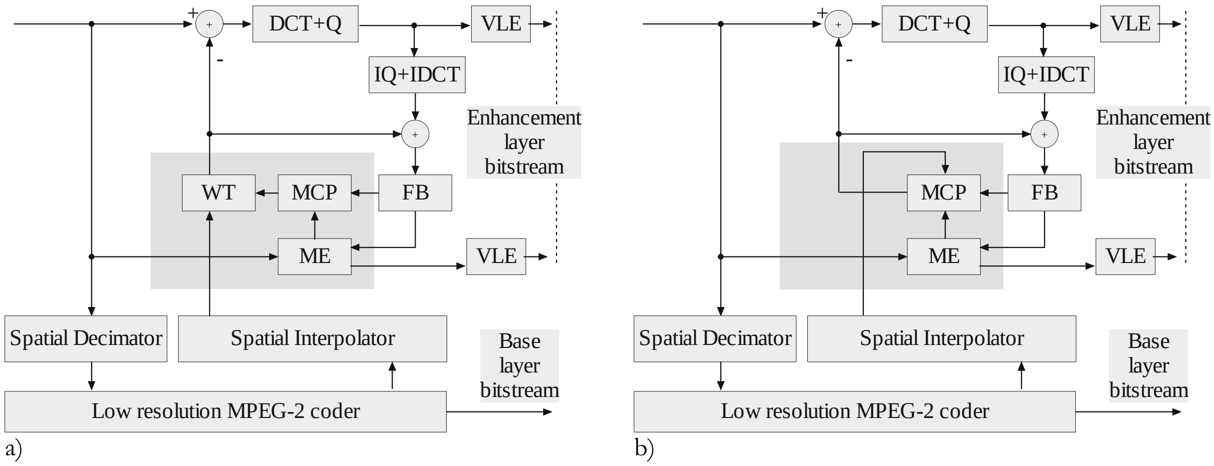
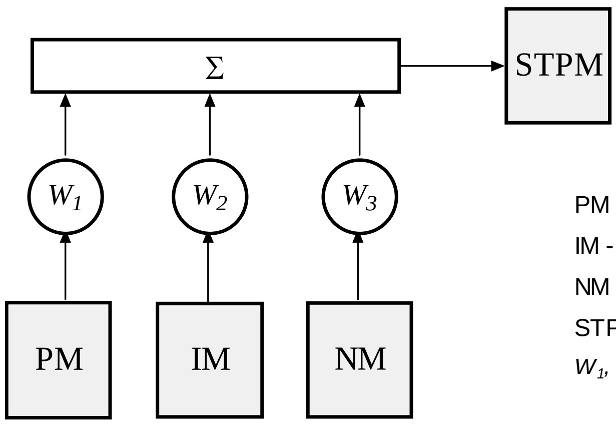


Fig.6.3. The encoder structure with improvements. The gray field indicates the difference between a) standard and b) improved prediction blocks.

The improvement of the standard MPEG-2 prediction within a single layer consists in a different decision strategy (Fig. 6.3). In the improvement, the best prediction/interpolation is chosen from all three possible reference frames: previous, future and interpolated one (Fig. 6.4).



PM - macroblock from previous frame

IM - macroblock from interpolated frame

NM - macroblock from next frame

STPM - spatio-temporal predicted macroblock

W_1, W_2, W_3 - weights

W_1	W_2	W_3	***Description of Prediction Mode***
1	0	0	forward prediction from the previous reference frame
0	1	0	interpolation from the current base layer frame
0	0	1	backward prediction from the next reference frame
1/2	1/2	0	the average of the forward prediction and interpolation
0	1/2	1/2	bi-directional prediction from the previous and next reference frames
1/2	0	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. 6.4. Proposed spatio-temporal weighted prediction in spatial scalability

The data both from the previous and the next reference macroblock PM and NM are motion-compensated, and data from the current reference macroblock are up-ampled in the 2-D space domain giving interpolated macroblock. The best-suited reference macroblock or average of two or three reference macroblocks is chosen according to the criterion of smallest prediction error.

$$STMP = w_1 \cdot PM + w_2 \cdot IM + w_3 \cdot NM \tag{6.1}$$

The structure of MPEG-2 encoder with improved prediction is presented in Fig. 6.3. That kind of prediction requires transmitting an additional bit per macroblock to identify the selected mode of prediction [Doma00d].

6.2.2 Improved 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. 6.5).

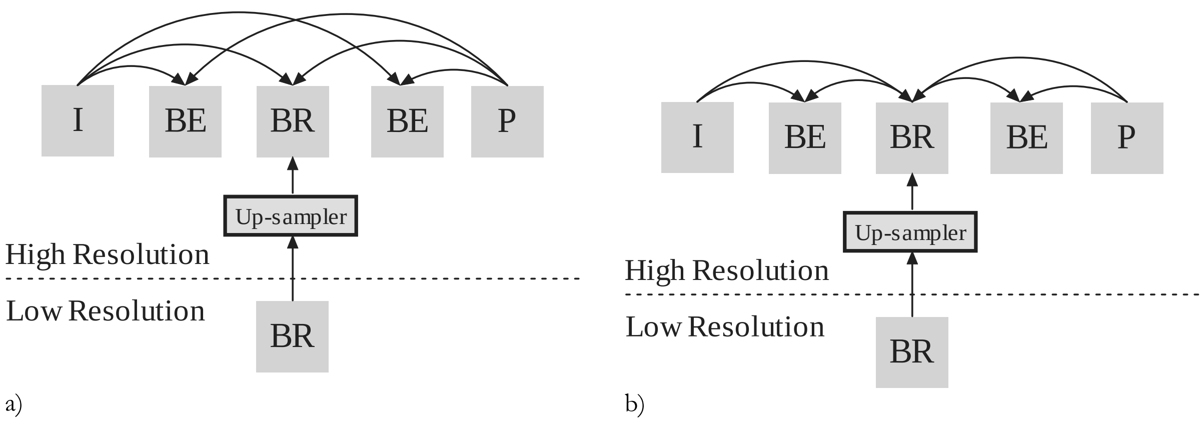


Fig. 6.5 Improved prediction of BE-frames. Figure indicates the difference between a) standard and b) improved prediction.

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. 6.5b). Therefore, higher correlation between the currently encoded BE-frame and the reference frame is achieved due to decreasing time difference. This modification reduces the number of bits allocated to BE-frames compared to the standard scalable MPEG encoder.

6.3 Control algorithm of scalable coder

The sequence consists of Group of Pictures (GOPs) that are defined as content access units. Figure 6.2 presents the structure of scalable GOP. The base layer works with sequences decimated in time and the enhancement layer works with full resolution images. It is assumed that small fluctuations of bitrate in both layers are allowed but overall bitrate should be constant.

$$B_{BASE\ LAYER} + B_{ENHANCEMENT\ LAYER} = CONST \tag{6.2}$$

Output VBV buffer makes both layers dependent. Despite layer bitrate fluctuations, the average layer bitrate should be constant. The author uses the proposed global model (see: chapter 4) for control algorithm in the constant bitrate mode of operation. The bitrate value B of AC DCT coefficients is computed according to:

$$B = \frac{a}{c \cdot (Q^b + e) + d} \tag{6.3}$$

for simplification optimal and fixed values of *a, b, e* and *d* parameters are chosen. For modified scalable video MPEG-2 coder founded parameters are shown in table 6.1.

Table 6.1 Fixed values of parameters *a, b, d* and *e* for Intra and Inter mode of coding

Coding mode	***Parameters of model***			
	a	***b***	***d***	***e***
Intra (I)	$5\ 10^6$	0.8	-0.4	0
Inter (P,B)	$5\ 10^6$	1.25	-0.4	0

The *c* parameter can be computed using previous parameters:

$$c = \frac{\dfrac{a}{B_{VAR\ prev}} - d}{Q_{prev}{}^{b}} \tag{6.4}$$

where: $B_{VAR\,prev}$ denotes the number of bits in the previous frame; Q_{prev} is the value of Q used in the previous frame. The quantization factor for current frame is computed according to formula:

$$Q_{cur} = e^{b \cdot \log_{10}\left[\left(\frac{a}{B_{VAR}} - d\right) \cdot \frac{1}{c}\right]} \tag{6.5}$$

where B_{VAR} denotes required bitrate for encoding AC DCT coefficients in current frame and Q_{CUR} denotes calculated quantization factor for the current frame.

6.4 Application

Control algorithm is applied in both layers of MPEG-2 scalable coder implementation which was developed by Sławomir Maćkowiak [Mack02]. Algorithm is used separately for each layer the only link between which is output VBV buffer. The control algorithm works in three steps.

- The c parameter is estimated. That parameter depends on the number of bits used for encoding AC DCT coefficients in previous frame and on quantization scale factor Q_{PREV} used in previous frame.
- The required B_{VAR} value is calculated. That value depends on the buffer fullness, bitstream B_{CONST}, and required bitrate B.
- The actual Q is computed from formula (6.3) and ΔQ is determined.
- ΔQ is clipped and added to previous global quantization scale factor Q.

Small bitrate fluctuations in each layer are accepted but overall bitrate is set constant. In order to limit the fluctuations of the consecutive images quality, the value of $(Q_{prev} - Q_{cur})$ is clipped (Fig. 6.4) to the range [-qc,…,qc] and added to Q_{prev}. The clipping function is defined as follows:

$$\Delta Q = \left\lfloor arctg((Q_{prev} - Q_{cur}) \cdot qc) \right\rfloor \tag{6.6}$$

where Q_{prev} is the value of Q used in the previous frame, Q_{cur} is calculated quantization parameter for the current frame and qc is maximum allowed change of quantization scale factor Q.

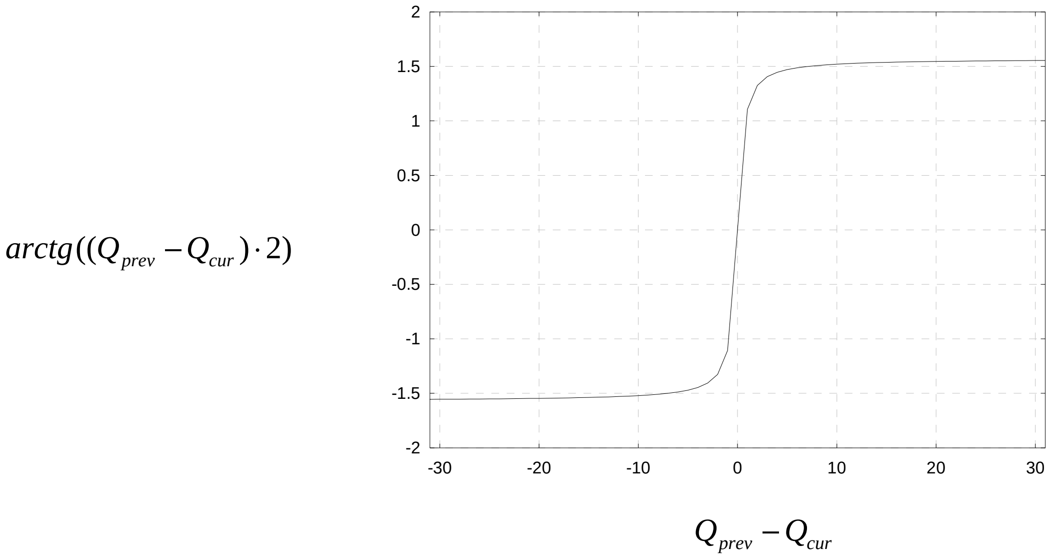


Fig. 6.4 The ΔQ clipping function for exemplary value of qc=2.

Parameter c, used in a global bitstream model, is calculated separately for each layer, and is updated for every frame. The difference between demanded and actual base layer bitrate value affects the enhancement layer. The greater bitstream is produced in the base layer, the smaller bitrate is offered to the enhancement layer. At the beginning of each GOP bits are allocated to the base and enhanced layers according to user's demands.

6.5 Results of experiments

For experiment purposes the set of progressive test video sequences such as: *Basket*, *Flower Garden*, *Mobile*, *Funfair*, *Football*, *Cheer*, *Bus*, *Stefan* has been used. The scalable MPEG-2 encoder was used as a reference coder. Proposed control algorithm has been implemented. The GOP structure **I**-BE-BR-BR-**P**-BE-BR-BE-**P**-BE-BR-BE-**P**-BE-BR-BE-(**I**..) has been used. In every experiment the 250 frames from each test sequences were encoded (125 of base layer and 250 of enhancement layer).

Improvement of BR- and BE-frames prediction:

Experimental results obtained from the set of video 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. 6.7). In particular, the interpolation from low resolution images allow more efficient predictive coding of macroblocks, which would be intra coded otherwise.

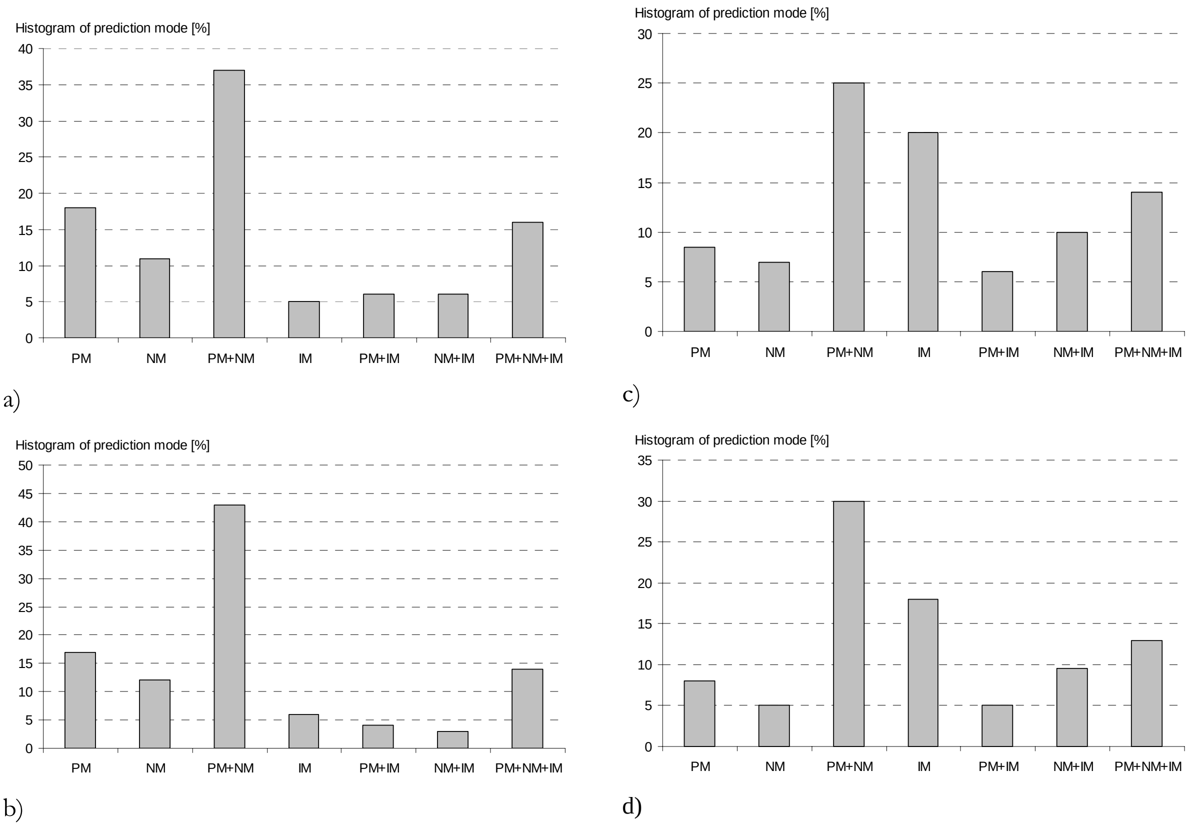
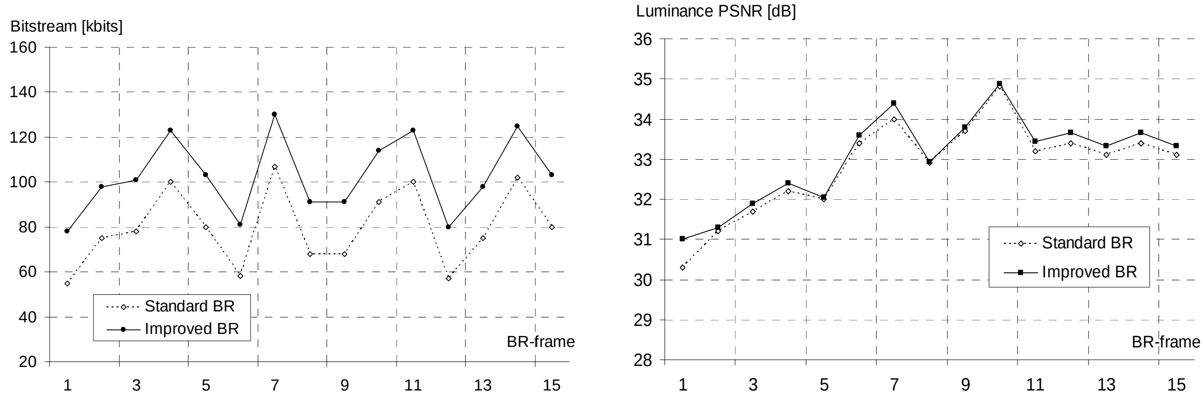


Fig. 6.7. Percentage of macroblocks predicted using individual reference frames or their averages for test sequence a) *Flower garden* encoded with 5 Mbit/sec, b) *Flower garden* encoded with 8.5 Mbit/sec, c) *Funfair* encoded with 5 Mbit/sec and d) *Funfair* encoded with 8.5 Mbit/sec.

The application of improved prediction leads to lower bitrates and higher PSNR, compared to standard prediction (Fig. 6.8). That improvement reduces the average size of a BR frame by about 6÷10% (average 7.5%), compared to spatially scalable coding defined in the MPEG-2 standard. Moreover, the new scheme of BE frames prediction causes the bitstream reduction or quality improvement in BE frames. Eventually, the BE and BR frame prediction improvement has an impact on improving quality of the whole sequence and overall bitrate reduction.

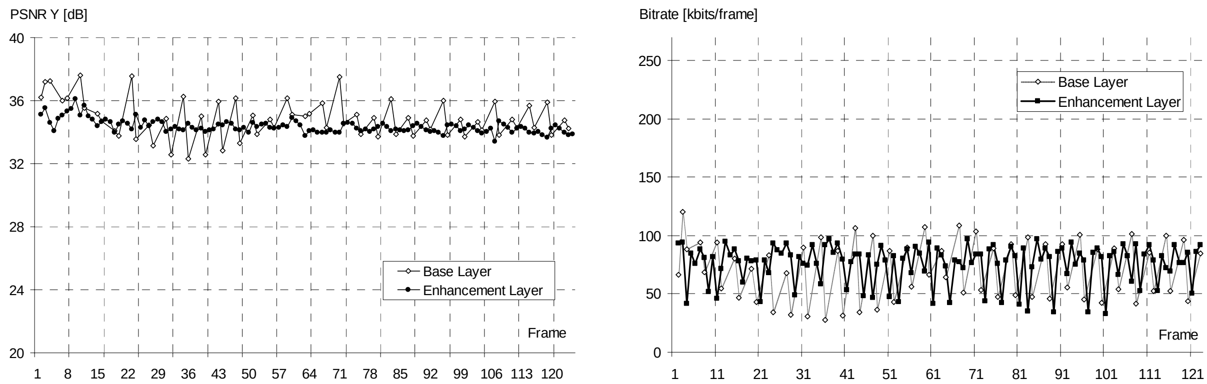


a) b)

Fig. 6.8. Graph of a) bitrate and b) PSNR for improved and standard prediction of BR-frames.

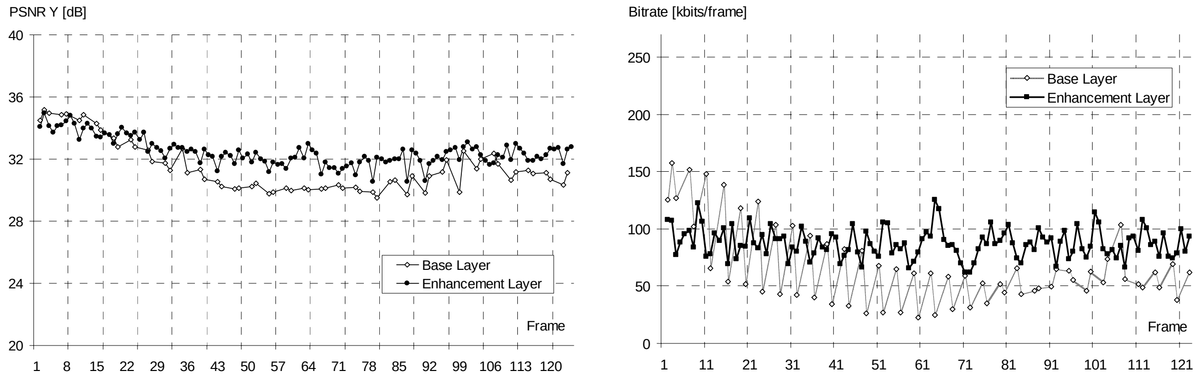
Average bitrate and quality:

The quality of video sequence (Figures 6.4a and 6.5a) in the enhancement layer (full resolution images) varies only a little. Such stable quality of encoded images results in better subjective quality. For some video sequences the quality of base layer sequence (CIF resolution) substantially varies. Both layers are coupled, hence bitrate variations in enhancement layer affects bitrate variation in base layer. As the enhancement layer is usually given a higher bitrate, the small relative bitrate variation in it results in great relative bitrate variation in a base layer. This is how the above mentioned quality variations in a base layer originate.



a) b)

Fig. 6.4 Figure a) shows quality of the encoded video sequence for base layer and enhancement layer. On the b) figure the bitstream per frame is shown. Those results are obtained for the *Cheer* sequence (scalable MPEG-2 with proposed control algorithm).



a) b)

Fig. 6.5 Figure a) shows quality of encoded video sequence for base layer and enhancement layer. On the b) figure the bitstream pre frame is shown. These are results are obtained for the sequence *Funfair* (scalable MPEG-2 with proposed control algorithm).

For most tested video sequences these variances of base layer quality are below 0.8 dB. The qualities of video sequences in both layers are similar (Tab. 6.2).

The bitstream produced by the scalable coder varies in layers but overall bitstream (the bitstreams of layers sum) varies only a little. Under experiments the VBV buffer has never been overflowed or underflowed. Experimental results indicate that target bitrate in layers is achieved with good/enough accuracy (Tab. 6.2). The ε_B error is defined as follows:

$$\varepsilon_B = \frac{|B_T - B_A|}{B_T} \cdot 100\%, \tag{6.7}$$

where B_T is the target bitrate and B_A is the obtained bitrate under encoding. Experimental results show that the average ε_B is less then 3% for each layer.

Table 6.2 Results for test sequence encoding in scalable mode. Table shows the achieved bitrates and average quality in both layers.

	Required bitrate [kbits/sec]			
	Base Layer	**Enhanced Layer**	**Base Layer**	**Enhanced Layer**
	2 Mbits/sec	3 Mbits/sec	4 Mbits/sec	3 Mbits/sec
Football				
Bitrate [Mbits/sec]	2.07	2.98	3.12	5.14
PSNR [dB]	29.60	29.95	31.12	32.10
Funfait				
Bitrate [Mbits/sec]	2.02	2.95	3.17	5.09
PSNR [dB]	31.02	31.30	32.30	32.90
Flower Garden				
Bitrate [Mbits/sec]	1.93	3.11	3.01	5.17
PSNR [dB]	29.76	29.61	32.40	32.56

6.6 Conclusions

Numerous experiments have been done in order to test the efficiency of scalable coding and to check the new control algorithm. All experimental results prove that proposed improvements of B-frame coding (BR and BE) significantly increase the efficiency of scalable coding. Moreover, the additional experiments indicate that new order of B-frame coding (BE-frame improvement) can be applied in standard (non scalable) hybrid coders giving either better quality or lower bitrate of encoded sequence.

Proposed control algorithm works well and can be used in scalable hybrid coders. It is simple and bitrates achieved in layers are very close to the required one.

Chapter 7

Results and conclusions

7.1 Main results of the dissertation

The author creates two new bitstream models of hybrid video coders which have no equivalent described in literature. The new control algorithms have been proposed as well. The numerous experiments with MPEG-2 (non-scalable and scalable) and H.263 coders with Standard Digital Television (SDTV) video signals have been done by the author.

Bitstream models which describe the bitstream as a function of quantization scale factor are proposed. These models achieve high bitstream estimation accuracy which was thoroughly examined during numerous experiments. The models mentioned above were exploited when designing the video coder control algorithms. Application of these algorithms improves video coder efficiency.

The thesis is: “there exist simple models B=f(Q) that can be identified either by a coding experiment or doing transform coefficients calculations and such models can be efficiently applied to bitrate control”. This thesis was proved in numerous experiments, in which the proposed coding techniques were used.

The main original results of the dissertation are:

- A new original global bitstream model (chapter 4).
    - Mathematical global model of bitstream is presented. This is a simple five- or one-parameter model.
    - Five-parameter model accuracy is experimentally examined (for MPEG-2 coders). It is proved the accuracy of the proposed model is good in the entire interval of quantization scale factor Q - about 3% for Intra-frames and 9% for Inter-frames on average.
    - One-parameter model is experimentally examined (for MPEG-2 coders). The five-parameter model simplification worsens the accuracy, however, it is still sufficient for video coder control algorithms. The model accuracy can be improved by narrowing the interval of quantize scale factor Q. Such one-parameter model turns out very useful in coder control algorithms.

- A new original microscopic bitstream model (chapter 5).
    - Bitstream model exploiting DCT coefficients histogram is presented (for MPEG-2 and H.263 systems).
    - Model accuracy is experimentally examined. The experimental results prove that the microscopic model is characterized by high bitstream estimation accuracy achieving bitstream estimation error below 3% on average. Such a model enables to create a very efficient control algorithm.

- The efficient control algorithm for a hybrid video coder with both proposed original bitstream models incorporated is proposed.
    - Video coder control algorithm applying one-parameter global model is implemented and tested (for MPEG-2 systems). According to the experimental results coder applying the global model achieves quality gain up to 0.8 dB compared to coder with standard control algorithm TM5. The control algorithm exploiting this model is simple and its computational cost is low.

- Video coder control algorithm applying microscopic model is implemented and tested (for MPEG-2 system as an example). The experimental results indicate that applying the microscopic model for control increases the coding efficiency, giving quality gain even up to 1 dB, and reduces quality fluctuations. Omitting the excessive quantization enabled by this model application reduces computational cost of overall coding process.

- The method of reducing the complexity of MPEG-2 coding algorithm computation is presented. Proposed method can give savings up to 4% percent of total computational time and up to 20% when time of motion estimation process is not taken into account.

The other original achievements of the dissertation are:

- Improving the efficiency of scalable coding mode (chapter 6).
  - The new improved BR-frames and BE-frames prediction is proposed.
  - BR- and BE frames improvements for the scalable video coder are implemented and tested (for modified scalable MPEG-2 system). It is evident from experiments that improved BE- and BR-frames prediction decreases total bitrate by 8% on an average.

- Creating an effective control algorithm for scalable video coder
  - Proposing of control algorithm for scalable video coder with global model of bitstream incorporated.
  - The new control algorithm for scalable video coder is implemented and tested (for modified scalable MPEG-2 system). The scalable coder with the proposed control algorithm keeps the required bitrate very accurately and stably.

As a result of applying the proposed bitstream models in video coder control algorithms an efficiency improvement is achieved, which is experimentally proved by comparison with video coders using default control algorithms (TM5). Moreover, using the author's control algorithm exploiting the microscopic model, the overall computational cost of sequence encoding decreases.

The presented new solution of BE- and BR-frames prediction improves scalable coding efficiency, and the proposed control algorithm efficiently controls the two-layer scalable video coder.

All the conclusions and achievements presented in this dissertation are obtained after thorough and careful experimental work was done. For experiments 20 test sequences of 250 4CIF frames each were taken. These sequences were processed with video-coders with several sets of coding parameters and both in progressive and interlaced mode of coding.

7.2 Future developments

Since much more sophisticated hybrid coders (like AVC/H.264) are developed proposed bitstream models must be adapted. The initial research conducted by the author shows that the proposed models can be employed to create control algorithms of the new coders. However, to retain the high accuracy of the above, they must be modified with respect to the new phenomena being the result of the implemented modifications.

New approach to improve efficiency of the video coder control algorithms and to speed them up is using metadata during the encoding process. Metadata describe visual material in detail and can be used for coder control on each level of coding. It enables to omit superfluous operation in process of coding as well as can help to predict the best encoding mode. It is supposed the author will continue his research in this field.

References

Author’s contributions

[Blas02] Ł. Blaszak, M. Domański, A. Łuczak, S. Maćkowiak, “Spatio-Temporal Scalability in DCT-based Hybrid Video Coders”, ISO/IEC JTC1/SC29/WG11 MPEG2002/M8672 Klagenfurt, July 2002.

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

[Doma00b] M. Domański, A. Łuczak, S. Maćkowiak , “Spatio-Temporal Scalability Using Modified MPEG-2 Predictive Video Coding”, Proceedings of European Signal Processing X, pp. 961-964, Tampere 2000.

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

[Doma00d] M. Domański, A. Łuczak, S. Maćkowiak, “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] M. Domański , A. Łuczak, S. Maćkowiak, “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] M. Domański , A. Łuczak, S. Maćkowiak, “Ulepszony koder skalowalny dla systemu MPEG-2”, Systemy i technologie telekomunikacji multimedialnej STM 2000, pp. 153-158, Łódź, 2000.

[Doma02] M. Domański, A. Łuczak, “Modeling of Bitstreams Produced by Hybrid Video Coders” Proceedings of ISCE'2002, September 23-26, Erfurt, Germany 2002

[Doma02b] M. Domański, A. Łuczak, “Simple Global Model of an MPEG-2 Bitstream”, Proceedings of IEEE International Conference on Multimedia and Expo ICME'2002, August 26-29, Switzerland, Lausanne 2002.

[Doma98] M. Domański, A. Łuczak, S. Maćkowiak, R. Świerczyński, “Hybrid Coding of Video with Spatio-Temporal Scalability Using Subband Decomposition”, Signal Processing IX: Theories and Applications, pp. 53-56, Typorama, 1998.

[Doma98b] M. Domański, A. Łuczak, S. Maćkowiak, R. Świerczyński, “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] M. Domański, A. Łuczak, Maćkowiak S., R. Świerczyński, “MPEG-Based Scalable Video Codec with 3-D Subband Decomposition”, Bulletin of the Polish Academy of Sciences Technical Sciences, Vol. 47, No. 3, pp. 247-253, Warszawa, 1999.

[Doma99a] M. Domański, A. Łuczak, Maćkowiak S., Świerczyński R., U. Benzler, “Spatio-Temporal Scalable Video Codecs with MPEG-Compatible Base Layer”, Proceedings of Picture Coding Symposium, pp. 45-48, Portland, 1999.

[Doma99b] M. Domański, A, Łuczak, S. Maćkowiak, “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.

[Lucz03] A. Łuczak, M. Domański, “MPEG Video Bitrate Control Using Coder Modeling”, Picture Coding Symposium, Saint Malo, France, 2003.

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

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

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

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

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

References

[Arch00] C. Archer, T. Leen, "Adaptive Transform Coding as Constrained Optimization", IEEE Transactions on Image Processing, pp. 925-931, 2000.

[Bass96] A. Basso, I. Dalgic, F. Tobagi, C. van den Branden Lambrecht, "Study of MPEG-2 Coding Performance Based on a Perceptual Quality Metric", Proceedings of Picture Coding Symposium '96, Australia, pp. 263--268, March 1996.

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

[Bran99] S. Brandt, "Analiza Danych, metody statystyczne i obliczeniowe", tłumaczył Lech Szymanowski, Wydawnictwo Naukowe PWN, wydanie 2, Warszawa 1999.

[Bron00] I. N. Bronsztejn, K.A. Siemiendiaje, „Matematyka, poradnik encyklopedyczny", Wydawnictwo Naukowe PWN, wydanie 18, Warszawa 2000.

[Chia97] T. Chiang, Y. Q. Zhang, "A New Rate Control Scheme Using Quadratic Rate Distortion Model", IEEE Transactions on Circuits and Systems for Video Technology, vol. 7, no. 1, pp. 246-250, September 1997.

[Chim00] A. Chimienti, M. Conti, E. Gregori, M. Lucenteforte, R. Picco, "MPEG-2 Sources: Exploiting Source Scalability for an Efficient Bandwidth Allocation", Mutimedia Systems, vol. 8, pp. 240-255, 2000.

[Choi94] J. H. Choi, D. C. Park, "A Stable Feedback Control of The Buffer State Using Lagrange Multiplier Method", IEEE Transactions on Image Processing, vol. 3, pp. 546-558, September 1994.

[Cote98] G. Cote, B. Erol, M. Gallant, and F. Kossentini, "H.263+ Video Coding at Low Bit Rates", IEEE Transactions Circuits and Systems for Video Techniques, vol. 8, no. 7, pp. 849--866, November 1998.

[Cote98] G. Cote, M. Gallant, B. Erol, F. Kossentini, "H.263+: Video Coding at Low Bit Rates", Special issue of the IEEE Transactions on Circuits and Systems for Video Technology on Interactive Multimedia, 8(5), September 1998.

[Dalg95] I. Dalgic and F. Tobagi, "Constant Quality Video Encoding", Proceedings of IEEE ICC'95, pp.1-7, Seattle, WA, June 1995.

[Dawo98] A. M. Dawood, M. Ghanbari, "Mpeg Video Model Based on Scene Content", IEE Electronic Letters, vol. 34, pp. 868-870, April 1998.

[Dawo98a] A. M. Dawood, M. Ghanbari, "MPEG Video Modeling Based on Scene Description", Proceedings of the IEEE International Conference on Image Processing ICIP'98.

[Ding96] W. Ding, B. Liu, "Rate Control of MPEG-2 Video Coding and Recording by Rate-Quantization Modeling", IEEE Transactions on Circuits and Systems for Video Technology, vol. 6, pp. 12-20, June 1996.

[Erol98] B. Erol, M. Gallant, G. Cote, F. Kossentini, "The H.263+ Video Coding Standard: Complexity and Performance", Proceedings of IEEE Data Compression Conference, Snowbird, Utah, March 1998.

[Faer97] N. Faerber B. Girod, E. Steinbach, "Performance of the H.263 Video Compression Standard", Journal of VLSI Signal Processing: Systems for Signal, Image, and Video Technology, Special Issue on Recent Development in Video: Algorithms, Implementation and Applications, 1997.

[Flie00] M. Flierl, T. Wiegand, B. Girod, "A Video Codec Incorporating Block-Based Multi-Hypothesis Motion-Compensated Prediction", Proceedings of the SPIE Conference on Visual Communications and Image Processing, vol. 4067, pp. 238-249, Perth, Australia, June 2000.

[Flue96] P. Fleury, J. Reichel, and T. Ebrahimi, "Image Quality Prediction for Bit Rate Allocation", Proceedings of IEEE International Conference on Image Processing, Lausanne, Switzerland, pp. 339-342, Oct. 1996.

[Fort95] Z. Fortuna, B. Macukow, J. Wąsowski, „Metody Numeryczne", Wydawnictwo Naukowo Techniczne, wydanie 3, Warszawa 1995.

[Fros01] O. Frossard, P. Verscheure, "Joint source/FEC Rate Selection for Quality-Optimal MPEG-2 Video Delivery", IEEE Transactions on Image Processing 10(12), pp. 1815-1825, 2001.

[Ghar96] H. Gharavi, W. Y. Ng, "H.263 Compatible Video Coding and Transmission", First Int. Workshop on Wireless Image/Video Communication, Loughborough, pp. 115-120, 1996.

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

[Giro95] B. Girod, E. Steinbach, N. Faerber, "Comparison of the H.263 and H.261 Video Compression Standards", in Standards and Common Interfaces for Video Information Systems, K.R. Rao, editor, Critical reviews of optical science and technology, Philadelphia, Pennsylvania, vol. 60, pp. 233-251, Oct. 1995.

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

[Giro98a] B. Girod, N. Färber, K. Stuhlmüller, "Internet Video Streaming", IEEE 10th Tyrrhenian Workshop on Digital Communications (Multimedia Communications), Ischia, September, 1998.

[Giro99a] B. Girod, N. Färber, K. Stuhlmüller, "Internet Video-Streaming", In Multimedia Communications, F. De Natale and S. Pupolin, Eds., Springer, ISBN 1-85233-135-6, pp. 547-556, 1999.

[Gray98] R. M. Gray, D. L. Neuhoff, "Quantization", IEEE Transactions on Information Theory, vol. 44, no. 6, October 1998.

[Grec03] C. Grecos, J. Jiang, "On-Line Improvements of the Rate-Distortion Performance in MPEG-2 Rate Control", IEEE Transactions on Circuits and Systems for Video Technology, vol. 13, no. 6, pp. 519-528, 2003.

[Hamd97] M. Hamdi, J. W. Roberts, Pierre Rolin, "Rate Control for VBR Video Coders in Broad-Band Networks", IEEE Journal on Selected Areas in Communications, vol. 15, no. 6, August 1997.

[Hask97] B. Haskell, A. Puri, A. N. Netravali, "Digital Video: An Introduction to MPEG-2", New York: Chapman & Hall, 1997.

[Held96] G. Held, T. R. Marshall, "Data and Image Compression: Tools and Techniques", John Willey and Sons Ltd., New York, 1996.

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

[Hsia00] S. T. Hsiang, J. W. Woods, "Embedded Video Coding Using Motion Compensated 3-D Subband/Wavelet Filter Bank", Packet Video Workshop, Sardinia, Italy, May 2000.

[ISO00] ISO/IEC JTC 1/SC 29/WG 1, ISO/IEC FCD 15444-1: Information Technology – JPEG 2000 image coding system: Core coding system, WG 1 N 1646, March 2000.

[ISO00a] ISO/IEC 14496-2/FPDAM4, “Coding of Audio-Visual Objects, Part-2 Visual, Amendment 4: Streaming Video Profile,

[ISO00b] ISO/IEC 14496-2:2001, “Information technology -- Coding of audio-visual objects -- Part 2: Visual”, July 2000.

[ISO01] ISO/IEC International Standard 13818, “Information Technology - Generic Coding of Moving Pictures and Associated Audio Information”.

[ISO02] ISO/IEC JTC1/SC29/WG11 MPEG2002/N4584, “Experimental Conditions for Wavelet Coding Exploration”, March 2002.

[ISO93] ISO/IEC 11172-2, “Information technology -- Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbits/sec - Part 2: Video”, 1993.

[ISO94] ISO/IEC IS 13818-2 / ITU-T Rec. H.262, “Generic Coding of Moving Pictures and Associated Audio, part 2: Video”, November 1994.

[ITUT93] Recommendation ITU-T J.81, “Transmission of component-coded digital television signals for contribution-quality applications at the third hierarchical level of ITU-T Recommendation G.702”, 1993.

[ITUT93b] ITU-T Recommendation H.261 – p×64 kbits/sec video coding standard for audiovisual services . 1993.

[ITUR94a] 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.

[ITUT98b] ITU-T Rec. H.245, “Infrastructure of Audiovisual Services – Communication procedures - Control protocol for multimedia communication”, September 1998.

[Katt95] J. Katto, M. Ohta, “Mathematical Analysis of MPEG Compression Capability and its Applications to Rate Control”, Proceedings of the International Conference on Image Processing'95, vol. 2, pp. 2555-2558, Washington D.C., Oct. 1995.

[Kees96] G. Keesman, “Improvement of MPEG-2 Video Compression Based on Rate Distortion Concepts”, Proceedings of the European Workshop on Image Analysis and Coding for TV, HDTV and Multimedia Applications, Rennes, February 1996.

[Kim00] T. Kim, B. Roh, J. Kim “An accurate bit-rate control for real-time MPEG video encoder”, Signal Processing: Image Communication 15, pp. 479-492 Elsavier, 2000.

[Kim01] Y. Kim, Z. He, K. Mitra, "A Linear Source Model and a Unified Rate Control Algorithm for H.263/MPEG-2/MPEG-4", IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), May 2001.

[Kim97] T. Kim, B. Roh, J. Kim, "An Accurate Bitrate Control for Real-Time MPEG Video Coder", Signal Processing: Image Communication, vol. 7, pp. 479-492, 2000.

[Kim97a] B.-J. Kim, W. Pearlman, "An Embedded Wavelet Video Coder Using Three-Dimensional Set Partitioning in Hierarchical Trees (SPIHT)," in Proceedings of Data Compression Conference, pp. 251--260, 1997.

[Kim99] W. J. Kim, J. W. Yi, S. D. Kim, „A bit Allocation Method Based on Picture Activity for Stall Image Coding", IDEE Transactions on Image Processing, vol. 8, no. 7, pp. 974-976, July 1999.

[Krun97a] M. Krunz and S. K. Tripathi, "On The Characterization of VBR MPEG streams", Proceedings of the ACM SIGMETRICS'97, May 1997.

[Laza94] A. Lazar, G. Pacifici and D. E. Pendarakis, "Modeling Video Sources for Real-Time Scheduling", Multimedia Systems, pp. 253-266, 1994.

[Lee01] S. Lee, M. S. Pattichis, A. C. Bovik, "Foveated Video Compression with Optimal Rate Control", IEEE Transactions on Image Processing, vol. 10, no. 7, July 2001.

[Lee03] J. W. Lee, A. Vetro, Y. Wang, and Y. S. Ho, "Bit Allocation for MPEG-4 Video Coding With Spatio-Temporal Tradeoffs", IEEE Transactions on Circuits and Systems for Video Technology, vol. 13, no. 6, pp. 488-502, 2003.

[Lee96] W. Y. Lee, J. B. Ra, "Fast Algorithm for Optimal Bit Allocation in Rate-Distortion Sense", Electronics Letters, vol. 32, pp. 1871-1873, September 1996.

[Lee97] W. Y. Lee, S. Kwon, "A Scene Adaptive Bitrate Control Method in MPEG Video Coding", Proceedings of Visual Communication and Image Processing.

[Lee97a] J. Lee, "Rate-Distortion Optimized Motion Smoothing for MPEG-2 Encoding", in IEEE ICIP-97, Santa Barbara, pp. 45-48, Oct. 1997.

[Lee97a] W. Y. Lee, J. B. Ra, "A Fast Algorithm for Optimal Bit Allocation", Proceedings of SPIE Visual Communication and Image Processing, pp. 167-175, January 1997.

[Lee98] J. Lee, "Rate-Distortion Optimization of Parameterized Quantization Matrix for MPEG-2 Encoding", Proceedings of the IEEE International Conference on Image Processing ICIP'98,

[Libh95] W. Li, Y. Bhaskazan, M. Kunt, "Very Low Bitrate Video Coding with DFD segmentation", Signal Processing: Image Communication, vol. 7, pp. 419-474, 1995.

[Lino95] L.-J. Lin, A. Ortega, C.-C. J. Kuo, "A Gradient-Based Rate Control Algorithm with Applications to MPEG Video", IEEE International Conference on Image Processing, Washington, October 1995.

[Lino96] L.-J. Lin, A. Ortega, C.-C. J. Kuo, "Cubic Spline Approximation for Rate and Distortion Functions for MPEG video", IS&T/SPIE's Symposium on Electronic Imaging: Science & Technology, San Jose, January 1996.

[Lino98] L.-J. Lin, A. Ortega, "Bit-Rate Control Using Piecewise Approximated Rate-Distortion Characteristics", IEEE Transactions on Circuits and Systems for Video Technology , vol. 8, no. 4, pp. 446-459 , August 1998.

[Liu01] D. Liu, E. I. Sara, W. Sun, "Nested Auto-Regressive Processes for MPEG-Encoded Video Traffic Modeling", IEEE Transactions on Circuits and Systems for Video Technology. Vol. 11, no. 2, February 2001.

[Lou97] L. Luo, C. Zou, Z. He, "A New Algorithm on MPEG-2 Target Bit-Number Allocation at Scene Changes", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 7, No. 5, October 1997.

[Mack02] S. Maćkowiak, "Scalable Coding of Digital Video", Doctoral dissertation, Poznań 2002.

[Mall98] S. Mallat, "Analysis of Low Bit Rate Image Transform Coding", IEEE Transactions on Signal Processing, vol. 46, no. 4, pp. 1027-1042, April 1998.

[Masr01] M. A. Masry, S. Hemami, "An Analysis of Subjective Quality in Low Bit Rate Video", IEEE International Conference on Image Processing, Thessalonica, Greece, October 2001.

[Masr03] M. A. Masry, S. S. Hemami, "CVQE: A Continuous Video Quality Evaluation Metric for Low Bit Rates", Proceedings Human Vision and Electronic Imaging 2001, San Jose, CA, January 2003.

[Ohm01] J.-R. Ohm, M. van der Schar, "Scalable Video Coding", Tutorial Material SM.T3, IEEE International Conference in Image Processing ICIP'2001, Thessaloniki, Greece, 2001.

[Orte95] A. Ortega, M. Khansari, "Rate Control for Video Coding Variable Bit Rate Channels with Applications to Wireless Transmission", In Proceedings of the Second IEEE International Conference on Image Processing ICIP'95, Washington, October 1995.

[Orte96] A. Ortega, "Optimal Bit Allocation under Multiple Rate Constraints", Proceedings of the Data Compression Conference, DCC'96, pp. 349-358, Snowbird, April 1996.

[Orte98] A. Ortega, C. Hsu, „A Lagrangian Optimization Approach to Rate Control for Delay-Constrained Transmission Over Burst-Error Channels", Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 1998.

[Park96] S. Park, Y. Lee, H. Chang, "A New MPEG-2 Rate Control Scheme Using Scene Change Detection", ETRI Journal, vol. 18, number 2, July 1996.

[Pazo98] C. M. D. Pazos and M. Gerla, "A Rate Based Back-Pressure Flow Control for the Internet", Proceedings of Conference on High Performance Networking (HPN). Vienna, Austria - 1998.

[Riba99] J. Ribas-Corbera, S. Lei, "Rate Control in DCT Video Coding for Low-Delay Communications", IEEE Transaction on Circuits and Systems for Video Technology, vol. 9, no. 1, February 1999.

[Ryu00] H. Ryu, Y. Miyanaga, K. Tochinai, "Self-Organized Edge Detection for an Image Compression", IEEE International Symposium on Circuits and Systems, Vol. 4, pp. 625-628, May 2000.

[Salo00] D. Salomon, "Data Compression", Springer-Verlag, New York 2000.

[Sawg96] Y. Saw, P. M. Grant, and J. M. Hannah, "Feed-forward buffering and rate control based on scene change parameters for MPEG video coder", Proceedings of European Signal Processing Conference, vol. II, pp. 727-730, September, 1996.

[Sawg97] Y. Saw, M. Grant, J. Hannh, B. Mulgrew, "Nonlinear Predictive Rate Control For Constant Bit Rate Mpeg Video Coders", IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '97), vol. 4, April 21 - 24, 1997.

[Sawg98] Y. Saw, M. Grant, J. Hannh, B. Mulgrew, "Video Rate Control Using a Radial Basis Function Estimator for Constant Bit-Rate MPEG Coders", Signal Processing: Image Communication 13, pp 183-199, 1998.

[Schu97] G. M. Schuster, A. K. Katsaggelos, "A Video Compression Scheme with Optimal Bit Allocation among Segmentation, Motion, and Residual Error" IEEE Transaction on Image Processing, vol. 6, no. 11, November 1997.

[Schu99] G. M. Schuster, G. Melnikov, A. K. Katsaggelos, "A Review of The Minimum Maximum Criterion for Optimal Bit Allocation Among

Dependent Quantizers", IEEE Transactions on Multimedia Signal Processing, vol. 1, pp. 3-17, March 1999.

[Seim01] G. Siemek, "The Coefficient Based Rate Distortion Model for the Low Bit Rate Video Coding", Proceedings of Computer Analysis of Images and Patterns Conference, pp. 118-124, 2001.

[Serv98] S. D. Servetto, C. I. Podilchuk, "Stochastic Modeling and Entropy Constrained Estimation of Motion from Image Sequences", Proceedings of IEEE International Conference on Image Processing, Chicago, October 1998.

[Siem01] G. Siemek, "The Coefficient Based Rate Distortion Model for the Low Bit Rate Video Coding", Computer Analysis of Images and Patterns Conference, LNCS 2124, pp. 118-124, 2001

[Siko99] T. Sikora, "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.

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

[Smoo96] S. R. Smoot and L. A. Rowe, "Study of DCT Coefficient Distributions", Proceedings of the SPIE Symposium on Electronic Imaging, San Jose, CA, vol. 2657, January 1996.

[Soda99] I. Sodagar, H. J. Lee, P. Hatrack and Y. Q. Zhang, "Scalable Wavelet Coding for Synthetic/Natural Hybrid Images", IEEE Transactions of CSVT, vol. 9, no. 2, pp. 244-254, March 1999.

[Stre97] J. Streit, L. Hanzo, "Fixed-rate Video Codecs for Mobile Radio Systems," European Transactions in Telecommunications (ETT), vol. 8, no. 6, pp. 551-572, Nov-Dec 1997.

[Stuh00] K. Stuhlmuller, N. Farber, M. Link, B. Girod, "Analysis of Video Transmission over Lossy Channels", IEEE Journal on Selected Areas in Communications, vol. 18, no. 6, pp. 1012-- 1032, June 2000.

[Sull98] G. J. Sullivan, T. Wiegand, "Rate-Distortion Optimization for Video Compression", IEEE Signal Processing Magazine, vol. 15, pp. 74-90, November 1998.

[Teoh94] P. Teo, D. Heeger, "Perceptual Image Distortion", Proceedings of the First IEEE International Conference on Image Processing, vol. 2, pp. 982--986, 1994.

[Test5] "Test Model 5", MSSG MPEG-2 video software, www.mpeg.org/MPEG/MSSG/tm5.

[Tura00] D. S. Turaga, T. Chen, Book Chapter "Fundamentals of Video Coding: H.263 as an example", Compressed Video over Networks, The Signal Processing Series, Marcel Dekker, 2000.

[Wang00] L. Wang, "Rate Control for MPEG Video Coding", Signal Processing: Image Communication 15, pp. 493-511, 2000.

[Wang02] Q. Wang, Z. Xiong, F. Wu, S. Li, "Optimal Rate Allocation for Progressive Fine Granularity Scalable Video Coding", IEEE Signal Processing Letters. Vol.9 no. 2, February 2002.

[Wang02a] R. Wang, M. Valla, M. Y. Sanadidi, and M. Gerla, "Using Adaptive Rate Estimation to Provide Enhanced and Robust Transport over Heterogeneous Networks", Proc. 10th IEEE International Conference on Network Protocols, Paris, France, 12 - 15 November, 2002.

[Wang95] L. Wang, "Rate Control for MPEG Video Coding", Proceedings of SPIE The International Society for Optical Engineering v 25011/1, Society of Photo-Optical Instrumentation Engineers, Bellingham, WA, USA, pp. 53-64, 1995.

[Wats94] A. B. Watson, "Perceptual Optimization of DCT Color Quantization Matrices", Proceedings of the IEEE International Conference on Image Processing, Austin, November 1994.

[West99] P. H. Westerink, R. Rajagopalan and C. A. Gonzales, "Two-Pass MPEG-2 Variable-Bit-Rate Encoding", IBM Journal of Research and Development, vol. 43, no. 4, 1999.

[Wieg00] T. Wiegand, M. Flierl, "A Video Codec Incorporating Block-Based Multi-Hypothesis Motion-Compensation Prediction", Proceedings of the SPIE Conference on Visual Communications and Image Processing, Perth, Australia, June 2000.

[Wieg01] T. Wiegand, B. Girod, "Lagrange Multiplier Selection in Hybrid Video Coder Control," Proc. IEEE Int. Conf. Image Proc., pp. 542-545, October 2001.

[Wieg96] T. Wiegand, M. Lightstone, D. Mukherjee, T. G. Campbell, S. K. Mitra, "Rate-Distortion Optimized Mode Selection for Very Low Bit Rate Coding and the Emerging H.263 Standard", IEEE Transactions Circuits and Systems for Video Technology, vol. 6, no. 2, pp. 182-190, Apr. 1996.

[Wieg97] T. Wiegand, X. Zhang, B. Girod, "Block-Based Hybrid Video Coding Using Motion-Compensated Long-Term Memory Prediction", Proceedings of the Picture Coding Symposium, Berlin, Germany, pp. 153-158, September 1997.

[Wieg99] T. Wiegand, X. Zhang, and B. Girod, "Long-Term Memory Motion-Compensated Prediction", IEEE Transactions on Circuits and Systems for Video Technology, vol. 9, pp. 70-84, Feb. 1999.

[Wien03] M. Wien, "Integer Transforms for Variable Block Size Transform Coding", Proceedings of Picture Coding Symposium, Saint Malo, 2003

[Wink98] S. Winkler, "Perceptual Quality Metric for Digital Color Images", International Conference on Image Processing, Chicago, Illinois, USA, October 4-7, 1999.

[Wink99] S. Winkler, "Issues in Vision Modeling for Perceptual Video Quality Assessment", Signal Processing, vol. 78, pp. 231--252, Oct. 1999.

[Wuwo01] M. Wu, R. A. Joyce, H. Wong, L. Guan, S. Kung, "Dynamic Resource Allocation via Video Content and Short-Term Traffic Statistics", IEEE Transactions on Multimedia, vol. 3, no. 2, pp. 186-199, June 2001.

[Yang00] Y. Yang, S. S. Hemami, "Generalized Rate-Distortion Optimizations for Motion-Compensated Video Coding", IEEE Transactions Circuits and Systems for Video Techniques, September 2000.

[Yang97] Y. Yang, S. S. Hemami, "Rate-Constrained Motion Estimation and Perceptual Coding" Proceedings IEEE International Conference on Image Processing, Santa Barbara, CA, October 1997.

[Yang98] K. H. Yang, A. Jacquin "Real Time Implementation of Rate-distortion Optimized Coding Mode Selection for H.263 Video Coders", Proceedings of the IEEE International Conference on Image Processing ICIP'98,

[Yuzh01] Y. Yu, J. Zhou, Y. Wang, Ch. Chen, "A Novel Two-Pass VBR Coding Algorithm for Fixed-Size Storage Application", IEEE Transactions on Circuits and Systems for Video Technology, vol. 11, pp.345-356, 2001.

[Yuzh98] Y. Yu, J. Zhou. Y. Wang, "A Fast Effective Scene Change Detection and Adaptive Rate Control Algorithm", Proceedings of the IEEE International Conference on Image Processing, 1998.

Annex A

Testing environment

A.1. Software coders

All experiments in this work related to MPEG-2 have been done with the reference coder Test Model 5. Software sources of this coder are available on MPEG organization site (www.mpeg.org) [Mpeg96].

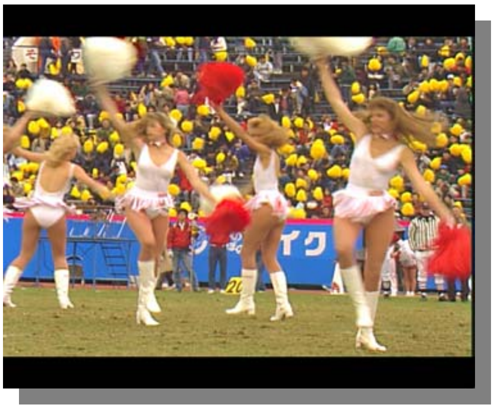
Experiments with H.263 coder and scalable coder have been done with coders implemented by S. Maćkowiak and the author. The construction of the verification model of mentioned video coders is based on universal software library [Lucz98] which was created in the Laboratory of the Division of Multimedia Telecommunications and Radioelectronics of Poznań University of Technology (av-lib.multimedia.edu.pl). Operation of the scalable coder and H.263 coder was verified during several tests, consisting in comparing the selected blocks of the encoder with the MPEG-2 verification model [Mpeg96] and TMN-2 (H.263 Test Model Number 2) respectively. 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].

A.2. Test sequences

Experiments have been done for the set of test video sequences which are presented below. The luminance format of these video sequences is progressive BT.601 [ITUR], i.e. 704×576 pixels and 25/50 frames per second. The chrominance subsampling scheme is 4:2:0.

Basket

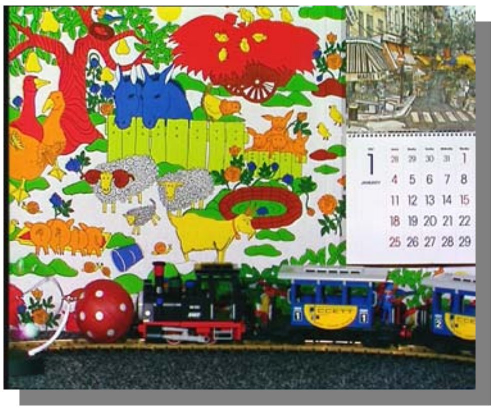
Bus



Cheer

Football

Flower Garden



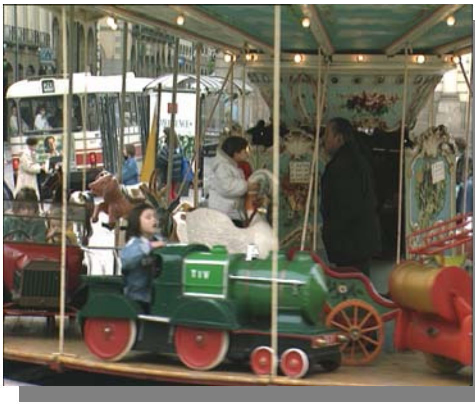
Mobile

Stefan

Warner

Universal

Icon


Funfair


Container