Stellingen

behorende bij het proefschrift

Variable Bit Rate Compressed Video

door P.J. van der Meer

3 maart 1997

1. Bij het comprimeren van een digitaal videosignaal kan een hogere compressiefactor behaald worden door een variabele bit rate (VBR) toe te staan. (Dit proefschrift)

2. De MPEG standaard zou meer geschikt zijn voor compressie met constante kwaliteit als de kwantisatie-karakteristiek voor voorspelde macroblokken geen dode zone zou bevatten.
(Dit proefschrift, hoofdstuk 3).

3. Effening van de bitstroom in de interface tussen video-encoder en transmissienetwerk voorkomt verlies van data bij het multiplexen van VBR video.
(Dit proefschrift, hoofdstuk 4 en 5)

4. Het gebruik van complexe modellen voor de variabiliteit van een VBR videobitstroom op encoder-specifieke niveaus is zinloos.
(Zie referenties [85,101])

5. Omdat uit een gegeven gecomprimeerde MPEG bitstroom alle door de encoder gebruikte technieken af te leiden zijn, is het zinloos om de "source code" van een software encoderachter te houden.

6. Variabele Lengte Coderings (VLC) tabellen voor Constante Bit Rate (CBR)videocompressiealgoritmen zouden ontworpen moeten worden aan de hand van "worst case" data.

7. Bij thuiswerken kan er zowel een teveel als een tekort aan afleiding ontstaan.

8. Gezien de recente fusies van veel telecommunicatiebedrijven wekt een gedwongen aandelenverkoop omwille van concurrentie in deze tak bevreemding.

9. Ondanks jarenlange standaardisatie-activiteiten belemmert het gebrek aan compatibiliteit tussen MPEG encoders en decoders van verschillende fabrikanten de snelle invoer van digitale televisie.

10. De analogie van intelligente telecommunicatie-netwerken met de autosnelweg loopt spaak omdat op de autosnelweg de "pakketten" bestuurd worden door zichzelf intelligent noemende "headers".

11. Bij schadevergoeding voor immateriële schade kan men beter spreken van schadevergelding.

12. Om de aantrekkelijkheid van voetbalwedstrijden in het knock-out systeem te verhogen, dienen de eventueel beslissende strafschoppen vooraf genomen te worden.

Variable Bit Rate Compressed Video

Patrick van der Meer

Variable Bit Rate Compressed Video

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus Prof.dr.ir. J. Blaauwendraad,
in het openbaar te verdedigen ten overstaan van een commissie,
door het College van Dekanen aangewezen,
op maandag 3 maart 1997 te 13.30 uur
door

Patrick Johannes VAN DER MEER,

elektrotechnisch ingenieur,

geboren te Roelofarendsveen.

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

Promotiecommissie:

Rector Magnificus	
Prof.dr.ir. J. Biemond,	TU Delft, promotor
Dr.ir. R.L. Lagendijk,	TU Delft, toegevoegd promotor
Prof.dr.ir. E. Backer,	TU Delft
Dipl.-Ing. W. Bachnick,	Deutsche Thomson Brandt
Prof.dr.ir. A.J. van de Goor,	TU Delft
Prof.dr.ir. F.W. Jansen,	TU Delft
Prof.dr. F.C. Schoute,	TU Delft

Published and distributed by:

Delft University Press
Mekelweg 4
2628 CD Delft
The Netherlands

Telephone: +31 15 2783254
Fax: +31 15 2781661

CIP-DATA KONINKLIJKE BIBLIOTHEEK, DEN HAAG

Meer, P.J. van der

Variable Bit Rate Compressed Video / P.J. van der Meer
Delft : Delft University Press. - Ill.
Thesis Technische Universiteit Delft. - With ref. - With summary in Dutch.
ISBN 90-407-1424-X
NUGI 832
Subject headings: Video Compression, Variable Bit Rate, MPEG, Traffic Shaping, Traffic Modelling, Multiplexing, Error Concealment.

Contents

Summary

In the past decades, a wide variety of different visual services have come into existence. Each of these services used to have a medium for transmission or storage specific to that particular service. Because of a recent shift from analogue representation to digital representation of information, the situation arose that existing analogue media were used for the transmission and storage of new digital services. Further, new digital media emerged on which all kinds of digital services can be transmitted or stored. Hence, the available bandwidth will have to be distributed among different services as efficiently as possible.

The robust nature of digitized information has as a drawback in that much more bandwidth is required to transmit them compared with the original, analogue signals. Especially video signals contain al lot of redundant information. By using video compression this information is removed so that digital video uses less bandwidth than analogue video.

In the field of video compression for transmission and storage, the digital signal processing world meets the telecommunication world in a supply and demand situation. The telecommunication world delivers bandwidth, the digital signal processing world uses the bandwidth as efficiently as possible. Better results in reducing the costs for video services can be achieved by the two fields working more closely together.

Because of the nature of traditional telecommunication channels, most video algorithms were designed to generate a constant bit rate (CBR) output stream. The quality of the video compressed by these algorithms, however, varies. Since the quality assessed by a human observer is determined by the worst parts of the video, CBR compression does not compress optimally. For constant quality, the transmission or storage medium should allow a variable bit rate (VBR). This thesis describes the interdependent issues concerned with the transmission and storage of VBR compressed video. It shows how advantage can be taken of the compression efficiency of VBR compressed video despite the CBR nature of telecommunication channels.

The concept of VBR video compression differs from the concept of CBR video compression. In CBR compression, the quality is optimized by distributing the available

bits over the different pictures and picture regions. In VBR video compression, compression is optimized, while the distortion introduced by compression is not visible under normal viewing conditions. To do this, the non-linear relation between distortion and visual quality needs to be studied. Each video compression technique can adjust its parameters so that the distortion does not exceed a certain visibility level. In MPEG, the quantizer scaling factor and weighting matrices can be chosen according to this principle. Compression, however, can be improved by exploiting the effect that distortion is often masked by the contents of the video. By using this effect, it is possible to obtain 20 % more compression without introducing visible distortion.

The low mean bit rate of VBR compressed video is beneficial if the source is recorded on a randomly accessible device, such as an optical disc. If the intended storage device is not randomly accessible, or if the video has to be transmitted, a conversion from the VBR stream to a CBR stream is needed. This can be done by stuffing the bit stream with dummy bits, or by multiplexing several VBR streams onto a single CBR stream with additional stuffing. In the latter case, advantage can be taken of the low mean bit rate if the bit rate of the resulting CBR stream is lower than the sum of the peak bit rates of the individual VBR streams. In this case, however, a loss of information may occur if the sum of the instantaneous bit rates of the individual sources exceeds the output CBR bit rate. Consequently, a trade-off between efficient use of the CBR channel and the probability of information loss will have to be made.

Multiplexing VBR sources onto a CBR channel can be performed either within the network or, if the network does not support this facility, within the interface between the sources and the network. In both cases, resources will have to be allocated in advance in order to guarantee a specified maximum amount of information loss while obtaining a high network load. The allocated bandwidth will be larger than the mean bandwidth, but lower than the peak bandwidth of the VBR source. Consequently, the generated traffic will have to be described by parameters which will have to be monitored to verify whether each source conforms to those parameters. The video compressor will have to take the characteristics of the traffic into account as well. Using parameter control functions, it will have to adjust the quality when too many bits are generated. Hence, a good description of the traffic is essential to prevent excessive loss of quality and to obtain a high network load.

In addition to the peak and mean bit rate of a VBR video source, a parameter indicating the *burstiness* is needed for an accurate description of VBR video traffic. It is shown that the techniques used in adapting the source to the network have a large impact on the traffic parameters, in particular the burstiness. By applying smoothing, the burstiness can be reduced, leading to VBR traffic that is much easier to predict.

The performance of multiplexing, which can be expressed by the probability of loosing a cell at a certain network load, depends on the characteristics of the traffic. To predict this performance, many complex models of VBR video traffic are found in literature. This thesis shows that if smoothing is applied in the adaptation to the network, the traffic merely reflects the input video. Therefore, models of the contents of the video are needed to predict the multiplexing performance of smoothed VBR video sources. Further, it is shown that the performance is significantly improved with respect to multiplexing non-smoothed sources.

If VBR sources are transmitted or stored by using statistical multiplexing, there is always a probability that information is lost. Therefore, a concealment of these losses is needed in order to reduce the visual effects of these losses. This can be done by applying layered coding, in which the data resulting from a VBR source is split. The most vital information is transmitted or stored in a low bit rate CBR channel. Using this technique, a good concealment can be achieved at the costs of a small inefficiency ($< 1\%$).

The techniques discussed in this thesis have led to the description of a layered reference VBR MPEG coder with a network adaptation incorporating smoothing and optional parameter control functions. The traffic resulting from this coder has a low burstiness and mean bit rate, while the peak bit rate is in the range of the bit rate of a CBR coder generating a similar quality of reconstructed video. The advantages of the low mean bit rate can be achieved on a randomly accessible storage device, or by multiplexing followed by transmission or storage on CBR media. An efficient use of the CBR channel can be achieved because of the low burstiness and the predictable nature of the VBR video traffic.

List of Abbreviations

AAL	ATM Adaptation Layer
AC	Auto-Correlation
ACF	Auto-Correlation Function
AR	Auto Regressive
ATM	Asynchronous Transfer Mode
B-ISDN	Broadband Integrated Services Digital Network
B-picture	Bi-directionally interpolated picture
CAC	Call Admission Control
CBR	Constant Bit Rate
CD	Compact Disc
CD-I	Compact Disc Interactive
CD-ROM	Compact Disc Read Only Memory
CD-video	Compact Disc video
CLP	Cell Loss Priority
CoV	Coefficient of Variation
CSF	Contrast Sensitivity Function
DC-coefficient	Coefficient determining the DC-level
DC-level	Direct Current level, here: average level of a block
DCT	Discrete Cosine Transform
DFD	Displaced Frame Difference
DPCM	Differential Pulse Code Modulation
DNSPP	Discrete Non-Stationary Periodic Process
DSPP	Doubly Stochastic Point Process
DVD	Digital Versatile Disc
EOB	End Of Block
GCRA	Generic Cell Rate Algorithm
GFC	Generic Flow Control
GOP	Group of Pictures
HDTV	High Definition Television
HEC	Header Error Control

HVS	Human Visual System
IDCT	Inverse Discrete Cosine Transform
I-picture	Intra coded picture
ISDN	Integrated Services Digital Network
ISO	International Standards Organisation
ITU	ITU Telecommunication Standardization Sector
ITU-T	International Telecommunication Union
JND	Just Noticeable Difference
LBC	Local Band-limited Contrast
MBS	Maximum Burst Size
MLBC	Masked Local Band-limited Contrast
MMPP	Markov Modulated Point Process
MOR	Mean Overload Ratio
MPEG	Moving Pictures Experts Group
MPEG-1	MPEG International standard for coding of moving pictures and associated audio for digital storage media up to about 1.5 Mbit/s
MPEG-2	MPEG International standard for generic coding of moving pictures and associated audio
MSD	Maximum Scene Duration
MSE	Mean Squared Error
MTI	Mean Traffic Intensity
MTF	Modulation Transfer Function
NNI	Network-Network Interface
OSI	Open System Interconnection
OSI model	Open System Interconnection reference model
PAR	Peak to Average Ratio
PC	Personal Computer
PCM	Pulse Code Modulation
PDF	Probability Density Function
PEM	Perceptual Error Measure
PMF	Probability Mass Function
P-picture	(Forward) Predicted picture
PRM	Protocol Reference Model
PSTN	Public Switched Telephone Network
PTI	Payload Type Indicator
QoS	Quality of Service
SAR	Standard deviation to Average Ratio
SNR	Signal to Noise Ratio
SDTV	Standard Definition Television
TAT	Theoretical Arrival Time
TE	Threshold Elevation

TV	Television
UNI	User Network Interface
UPC	Usage Parameter Control
VBR	Variable Bit Rate
VC	Virtual Channel
VCI	Virtual Channel Identifier
VCR	Video Cassette Recorder
VHS	Video Home System
VLC	Variable Length Code
VP	Virtual Path
VPI	Virtual Path Identifier
WMSE	Weighted Mean Squared Error
WWW	World Wide Web

Chapter 1

Introduction

Among the many different technologies developed in the 20th century, the gathering, processing and distribution of information is by far the most important. With the discovery of electricity a new era arose in which the distances between points on the globe appeared much smaller due to the speed at which electric signals travel. Pictures and picture sequences became a most important representation of information because of the great impact visual pictures have on human beings.

Many different visual services have come into existence, each one developed for or itself developing its own medium for transmission or storage. Today there is a complicated infrastructure in which video signals are transmitted and stored in a wide variety of different media as depicted in Figure 1.1.

Over the years there has been a shift from analogue representation to digital representation of information. This is because of the high robustness and flexibility of digital signals and systems. Because of this flexibility, existing networks can be used for new services, and new networks can be developed to support all kinds of digital services.

Due to the robust nature of digital signals, they require much more bandwidth than the original analogue signals. Especially in video applications, the costs of transmission and storage of (raw) digitized video are very high. Traditionally, research to reduce these costs is carried out in two fields. The video compression field develops algorithms to reduce the amount of information to be transmitted. The telecommunication field investigates the various media to reduce the costs for transmission of a certain amount of information.

The video compression field interacts with the telecommunication field in a supply and demand kind of way. The video compression field looks at the available medium to store or transmit the video. It then develops compression algorithms that generate

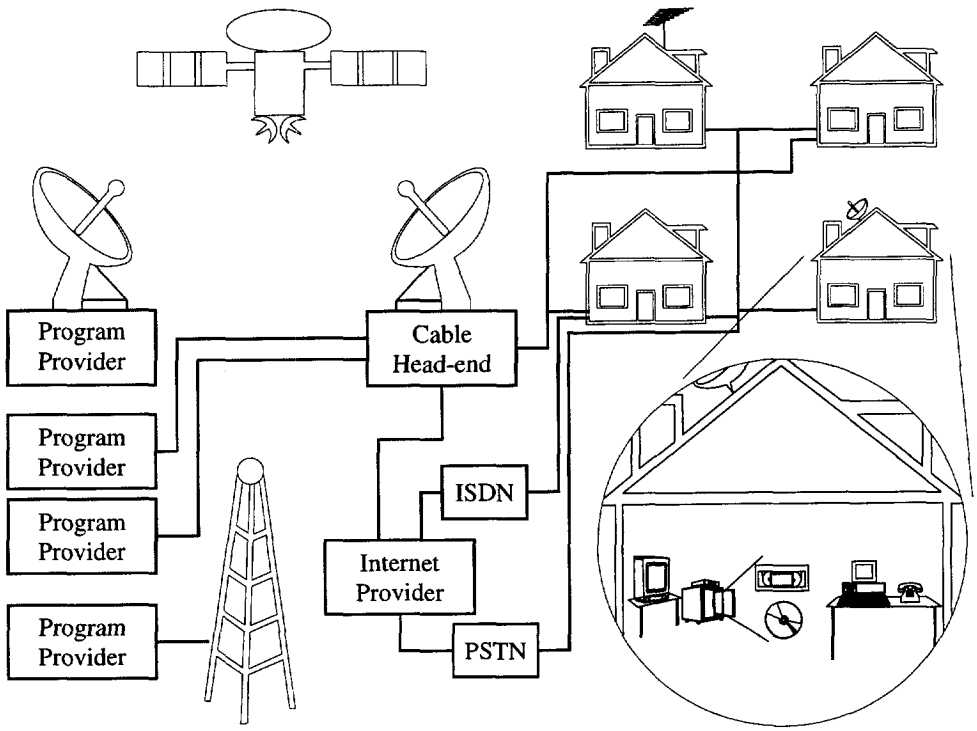


FIGURE 1.1: INFRASTRUCTURE FOR VIDEO SERVICES: TELEVISION SIGNALS ARE TRANSMITTED VIA THE ETHER, SATELLITE, AND CABLE NETWORKS; INTERNET VIA TELEPHONE, ISDN, AND CABLE; VIDEOPHONE VIA ISDN. STORAGE AT HOME IS POSSIBLE WITH TAPES, COMPUTER HARD-DISCS AND IN THE NEAR FUTURE WITH OPTICAL DISCS.

data rates that conform to the demands of those media. The telecommunication field looks at the data rates generated by video compression algorithms. Next, it designs new media conforming to these data rates. Better results in reducing the costs for video services can be achieved by two fields working more closely together, for instance by using the techniques discussed in this thesis.

Because of the nature of traditional telecommunication channels, most video algorithms were designed to generate a constant bit rate (CBR) output stream. The quality of the video compressed by these algorithms, however, varies. Since the quality assessed by a human observer is determined by the worst parts of the video, CBR compression does not compress optimally. For constant quality, the transmission or storage medium should allow a variable bit rate (VBR). This thesis describes

the interdependent issues concerned with the transmission and storage of VBR compressed video. It shows how advantage can be taken of the compression efficiency of VBR compressed video despite the CBR nature of telecommunication channels.

To identify the possibilities for VBR video, we will have to take a closer look at the whole process of transmission and storage of digital video. The International Standards Organization (ISO) has standardized a layered structure for all data communication, called the open system interconnection (OSI) reference model. This model, which is described in Section 1.1, shows the separation of transmission layers, located in the network and research for which is carried out in the telecommunication field, and user layers, located in the terminal and in the case of digital video handled by the video compression field.

Section 1.2 introduces the fundamentals of digital video compression, indicating the possible techniques to be used in the higher layers of the OSI model in the case of video applications. Section 1.3 describes the concept of variable bit rate (VBR) in contrast to constant bit rate (CBR) compression and the advantages VBR has over CBR. Finally, Section 1.4 outlines the structure of this thesis.

1.1 The ISO OSI Reference Model

The compression of digital video is only one of the techniques used to store or transmit a video signal. Some of these techniques, such as cryptography and error protection, are used to provide the demands a user specifies concerning the type and quality of the video. Others, like channel modulation, are needed to conform to the medium available to transmit or store the video. In general, the techniques to be used are chosen according to a trade-off which has to be made between costs and quality.

To reduce the complexity of data (including video) communication terminals, the different techniques are usually organized in several layers, the number and contents of which are different for each medium. Each layer has the task to offer services to higher layers and thus to hide the details of how these services are implemented. Each layer on a transmitting or recording terminal communicates with the same layer in the receiving or playback terminal. Therefore, the data offered by layer n conforms to certain rules, the **protocol** of layer n.

As an initial step towards international standardization of data-communication protocols, the International Standards Organization (ISO) has developed a reference model for Open Systems Interconnection, the ISO OSI Reference Model [1]. Although developed for computer networks, it also applies to other communication media and to storage media as well.

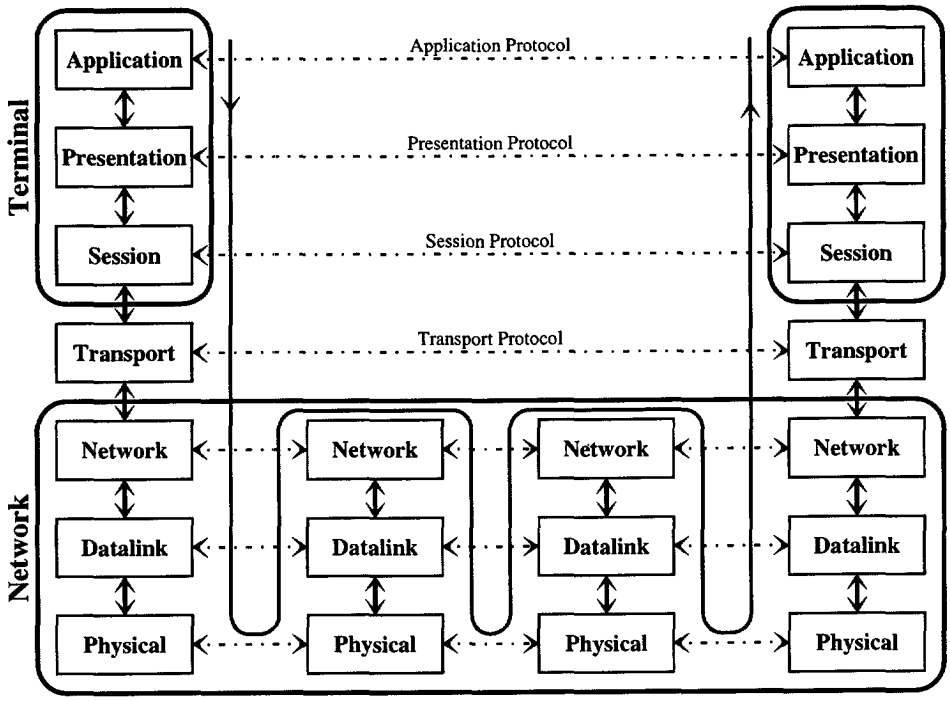


FIGURE 1.2: DATA COMMUNICATION CONFORM TO THE OSI MODEL.

The division of the different functionalities into a number of layers is not straightforward. The number of layers should be large enough to prevent different functionalities from being put into the same layer. It should be small enough, however, to avoid inefficient architectures. Further, the amount of information to be transmitted between the layers should be as small as possible. The OSI model is composed of seven layers and is shown in Figure 1.2.

The top layer in the OSI model is the *application* layer. It deals with the representation of user data elements used for each application. Different users may have their data in a different format. It is the task of the application layer to present the data in such a way that it can be handled by the application layer at the receiving side.

The *presentation* layer deals with the syntax of the data to be transmitted. It translates the data into a bit stream according to demands the user may impose. It thus includes cryptography and authentication as well as data compression functionalities. The video compression techniques are part of this layer. Below the presentation layer, the syntax of the data is no longer interpretable and the data is considered as a bit stream only.

The *session* layer takes care of establishing and maintaining sessions of data transmission. When a dialogue is held on a one-way channel, the session layer decides who is talking and who is listening. In a transmission service, it takes the decision when to transmit the data. Hence, the synchronization must also be handled by this layer, so that it can decide when to chop off transmission.

The *transport* layer forms the interface between network and terminal. It organizes the bit stream according to the network type and takes care of error correction to provide a certain end-to-end user quality of service. To do this, it may use several protocols, like error correction codes or re-transmission. Below this layer, the bit stream cannot be identified as such, and the treatment of the data units only depends on how they are generated here.

The *network* layer is where the data is transmitted. Network oriented protocols, such as routing, are handled here. Thus, the network layer provides a connection path between a pair of transport entities (end-points or intermediate nodes). The costs of transmission are also calculated here, based on the amount of data transmitted and the route through the network (there may be cheap and expensive links from a source to a destination).

The task of the *data-link* layer is to provide a transmission channel to the network layer which appears to be error free. Often, this task is performed by splitting the data into frames, so that resynchronization can be performed at the beginning of each frame. Also, the use of confirmation frames may increase the reliability of the service at the expense of extra delay.

Finally, the *physical* layer transmits the bit stream over the communication channel. It takes care of modulation, and time and frequency multiplexing to convert the bit stream into electric, magnetic or optic signals. Also, some protocols need to be defined on how a connection is set up and terminated, for instance, with unique bit patterns.

The OSI model applies to any kind of network and application. The implementation of each layer, however, will differ with each network and with each application. The lower three layers are the transmission layers while the upper three layers are application specific. The transport layer is responsible for interconnecting the network to the application.

Although the layered concept isolates the different implementations from each other, the choices to be made in higher layers depend on certain implementations of the lower layers and vice versa. In the case of video services, the support of timing requirements and variable bit rate transmission on the lower layers may influence the implementation of the application specific layers.

1.2 Digital Video Compression

When an analogue signal is digitized, this means that samples are taken from the original signal, and that these samples are represented by a finite number of representation levels. According to the Nyquist theorem [2], the signal has to be sampled with a sampling distance T less than or equal to the reciprocal of twice the bandwidth W of that signal:

$$T \leq \frac{1}{2W}. \tag{1.1}$$

The meaning of (1.1) can also be reversed. Given a sampling distance T it determines the maximum bandwidth W of the input signal. A low-pass filter can be applied prior to the digitizing process so that the input video does not contain higher frequencies. The number of samples with which a picture is represented (determined by the sampling distance and the size of the captured scene) is called the *resolution* of that picture. It is clear that in some applications, such as video-phone, a lower resolution is acceptable than in others, like HDTV. Hence, the application layer in the OSI model determines the resolution of the pictures.

After digitizing, the samples are called pixels (picture elements) and are usually represented with 8 bit words, reaching 256 different representation levels. One black-and-white picture of standard definition television (SDTV) resolution (576 lines with 720 pixels per line) thus requires over 3.3 million bits (415 thousand bytes). Considering that moving colour television requires three colours (red, green and blue) and a picture frequency (temporal resolution) of 25 Hz, this yields an enormous bit rate of about 250 Mbit/s. Table 1.1 indicates the resolution and corresponding uncompressed bit rates for some applications together with the targeted bit rates for transmission and storage. From this, it is clear that compression is required, which is carried out in the presentation layer of the OSI model.

Application	**Resolution**	**Uncompressed**	**Compressed**
HDTV	1440x1024x50	2 Gbit/s	40 Mbit/s
SDTV	720x576x25	250 Mbit/s	5-10 Mbit/s
CD-I, video-CD	360x288x25	60 Mbit/s	1.5 Mbit/s
video-phone	180x144x10	6 Mbit/s	64 kbit/s

TABLE 1.1: RESOLUTION AND BIT RATES FOR SOME VIDEO APPLICATIONS.

Within a picture sequence, consecutive pictures are highly correlated. Also, neighbouring pixels within a picture have this property. Video compression techniques try to make use of this so-called *redundancy* by applying an intelligent representation of

the data. This can be done, for instance, by predicting data from surrounding pixels. Next, the statistics of the resulting data can be exploited by Variable Length Codes (VLCs) applying short code words to frequently occurring events and longer code words to less frequent events. These techniques allow for a perfect reconstruction of the original data, and are therefore called *compression* techniques.

In most applications, compression alone is not sufficient for cost effective transmission or storage. In such a case, additional *reduction* of data is necessary. In this case, data that is irrelevant to the perception of the pictures is omitted. An example of data reduction is quantization, where a continuous range (or large amount) of values is projected onto a limited number of representation levels. Because of the loss of information, data reduction is irreversible: the original signal cannot be reconstructed perfectly. The distortion in the reconstructed video depends on the amount of data reduction applied. In many situations, it is difficult to distinguish data reduction from data compression techniques. Especially in video applications, the term compression is often used when compression in combination with reduction is meant [3].

The minimum bit rate to be achieved with data compression when no distortion is allowed, is given by the entropy H of the source. When a certain degree of distortion is allowed, the optimal performance of a video compression scheme is theoretically given by the *rate-distortion function* $R(D)$ [4]. In other words, $R(D)$ is the minimum amount of information needed to reconstruct the video source with a given distortion D. A typical rate-distortion function is sketched in Figure 1.3.

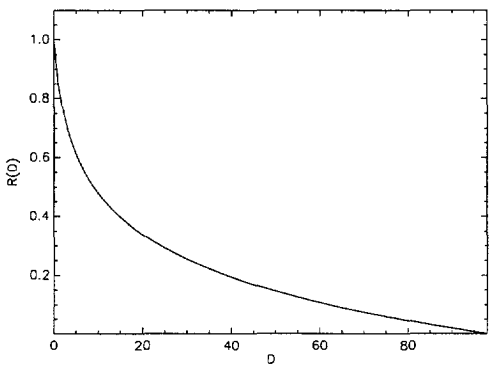


FIGURE 1.3: A TYPICAL RATE-DISTORTION FUNCTION $R(D)$.

Practical results for the rate-distortion function are only known for a few special cases [4]. For video, it is usually not possible to calculate this function. Nevertheless, much insight can be gained from a plot of the distortions at different rates for a

given compression scheme and input video. In this case, we will use the term rate-distortion *curve*.

When we compare different rate-distortion curves of a given compression scheme for different pictures, we see that they differ considerably. The curve of a slowly changing low detailed scene is positioned lower than that of a highly active or detailed scene. Hence, when either rate or distortion is fixed, the other one will vary. A varying quality, resulting from a varying distortion, is not attractive from an applications point of view and thus a variable rate transmission channel would be required. The transmission layers in the OSI model, however, mostly offer real-time services with constant bit rates only. Hence, CBR compression was adopted by most of the video compression algorithms.

1.3 CBR versus VBR Compression

Most of the transmission and storage media supporting real-time services do not allow for variable bit rates. Hence, research performed in video compression is traditionally based on generating a constant-bit-rate (CBR) output bit stream. By using compression techniques, however, the number of bits per part of a picture or picture sequence varies over time. To generate a constant output bit rate, CBR coders thus apply an output buffer in conjunction with a bit rate control algorithm. The buffer stores the bits produced by the encoding algorithm before they are transmitted, thereby removing fluctuations in the bit stream. To prevent buffer over- and underflow, control algorithms increase the data reduction (and thereby decrease the quality) when the buffer is full and decrease it when the buffer is empty [5].

In the rate-distortion function of Figure 1.3, we should consider that in a CBR coder the rate is fixed. The objective of the coder is to get a distortion that is as small as possible, the function indicating the theoretical minimum. The distortion measure is usually some kind of average of the squared pixel distortions in a picture. Hence, best results are achieved when the distortions in all parts of the picture are in the same order of magnitude. To accomplish this, advanced bit rate control algorithms apply bit allocation techniques to allocate more bits to detailed parts of the picture [6, 7].

Although good bit allocation algorithms limit the short term fluctuations of quality in CBR compressed video, long term fluctuations cannot be avoided because of the different amount of detail and action in different video scenes. For each video application, however, a desired quality can be identified. The quality in each part of the sequence should not, or only very rarely, drop below this desired quality. Hence, when a target bit rate is chosen for compression, it should be high enough to encode the most detailed and active scene expected in that application with the desired

quality. In less active scenes, the quality will be higher than the desired quality. Most of this *extra* quality, however, will not be assessed as such.

When we assume that the medium does tolerate a varying rate of the output bit stream, a different situation arises. The purpose of the compression algorithm in this case is to compress as much as possible, taking into account the desired quality. Looking at the rate-distortion function again, the problem of VBR compression seems not to be much different from CBR compression. Instead of a fixed bit rate, in VBR compression the distortion is fixed [8]. It is not trivial, however, to find a distortion measure that reflects the quality perceived by a human observer [9, 10]. Although this is a problem in CBR compression as well, it is less critical there, since in most scenes there the quality will be higher than the desired quality. Thus, in VBR compression, it is even more essential to have a good quality metric than in CBR.

The advantage of VBR over CBR compression can be seen from the bit rate and quality plots in Figure 1.4. Figure 1.4a shows the desired quality and the quality obtained by CBR compression. We see that this quality is higher almost everywhere, except for a small part of the sequence, where the activity is very high. Figure 1.4b shows the bit rate of the CBR source together with the bit rates of a VBR source generating the desired quality along the whole sequence. The bit rate of this source is lower in all parts of the sequence, except in the small part where the bit rate is higher in order to reach the desired quality where the CBR coder failed to reach it. Hence, much less bits are needed to represent the video sequence with a certain quality when VBR compression is used. Further, the desired quality can also be reached in scenes which have more detail than the scenes for which the CBR compression was dimensioned.

The choice to use CBR or VBR compression depends on the implementation of the transmission layers in the OSI model. It is, in fact, the transport layer that needs to adapt the bit stream generated by the presentation and session layer to the characteristics required by the transmission layers. When these layers do not support the transmission of VBR bit streams, the streams will have to be converted into CBR bit streams. When they do support VBR, the bit streams are probably multiplexed with other VBR bit streams to share a CBR link in the network. In both cases, the variability of the VBR bit stream determines how efficient the network bandwidth can be allocated.

1.4 Outline

The advantage of VBR compression with respect to CBR as discussed in Section 1.3 is agreed upon by many researchers. The applicability of VBR compression, however, depends on the capability of the transmission and storage channels to support

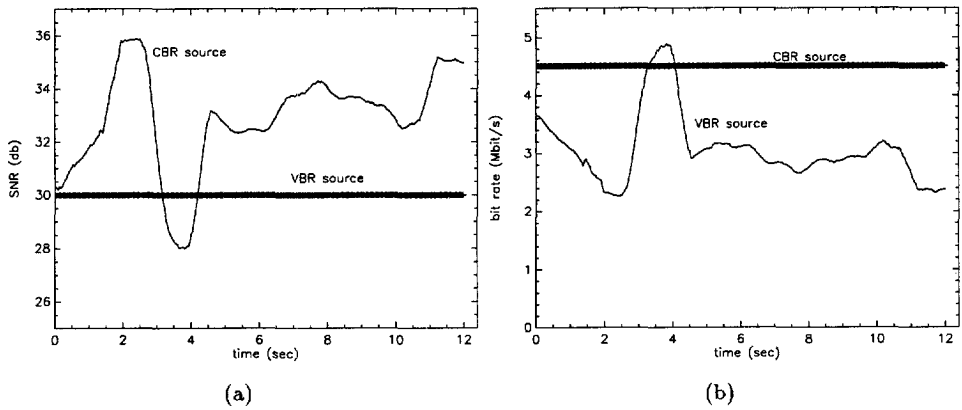


FIGURE 1.4: ILLUSTRATION OF (A) QUALITY AND (B) BIT RATE OF A CBR AND VBR SOURCE.

variable bit rates. In other words, the implementation of the transmission layers in the OSI model has a considerable impact on the compression capabilities in the presentation layer of that model. Chapter 2 discusses the different technologies applied in these layers for transmission and storage media. Emphasis is placed on the capabilities to support VBR video.

As discussed in Section 1.3, the concept of VBR compression is different from the concept of CBR compression. In VBR compression, the objective is to obtain a constant quality of reconstructed pictures, while compression is optimized. Chapter 3 deals with the definition of quality and the possible compression techniques to be used. In fact, it deals with the higher layers in the OSI model, which are application specific.

Chapter 4 discusses the possible implementations of the transport layer in the OSI model. This layer represents the actual interface between the application-specific layers and the network layers. It is shown that the bit rate characteristics depend on the implementation of this layer. In practical situations, these characteristics will have to be specified in advance. In order to conform to these characteristics, a control algorithm needs to be applied. In Chapter 4 some possible implementations of such a control are discussed, in which knowledge about the implementation of the transport layer is incorporated.

In most cases, the advantages of VBR can be exploited by multiplexing multiple sources on one channel. The performance of multiplexing depends on the characteristics of each source. Chapter 5 describes some ways to model these characteristics

and to use these models to predict the multiplexing performance. Also, the influences of the different interface implementations discussed in Chapter 4 are analyzed and verified by experiments.

When the total instantaneous bit rate of the sources to be multiplexed exceeds the channel rate, data will be lost. Chapter 6 describes some ways to cope with this loss, together with some techniques to prevent the most vital information from being lost.

The techniques discussed in this thesis lead to the design of a robust constant-quality reference codec with proper network adaptation, leading to a variable, but highly predictable bit rate. Chapter 7 concludes with a discussion on the applicability of this codec in present and future transmission and storage media.

Chapter 2

Transmission and Storage Possibilities for VBR Video

The reason to prefer VBR over CBR compression is that the compression performance for video is better when a variable bit rate is allowed. In particular, the mean bit rate is much lower than the bit rate of a comparable constant bit rate source. Traditionally, however, transmission and storage media are designed to support constant bit rates. This chapter gives an historical overview of the different media. It focuses on the capabilities to support real-time services and in particular VBR video. It is not only the support of VBR that is important, but also the capability to profit from the advantages of VBR video.

2.1 Transmission Media

The development of transmission networks has not only been pushed by technology, but also by the type of service requested. When technology made it possible to transmit signals over the air, radio and television were invented. Subsequently, the demand for more television channels was the reason for the development of cable television. Further, cable networks were created parallel to the existing telephone network because of the differences between television and telephone. Television is a distributive one-way service using a large bandwidth while telephone is a communicative two-way service using a small bandwidth. Similarly, the need for computers to interchange information started the development of computer networks.

The idea to use existing networks for emerging new services is rather recent. It started in the computer world, where PCs at home could be connected with laboratory PCs over telephone lines by using modems. The newest idea in this respect is to develop a so-called information super-highway, a world-wide network infrastructure with a multitude of access points of large capacity over which a wide variety of services is offered.

The characteristics and quality of data transmission depends on the implementation of the different layers in the OSI model. The physical layer is of particular importance to the data rates and robustness of the signals. Both analogue and digital signals can be transmitted over a number of different physical media. Although the physical carrier of digital signals is always analogue, we distinguish analogue from digital transmission by the treatment of the signal in the physical layer.

Analogue transmission means that the analogue signal is transmitted without regard to its content; it may represent analogue or digital data. In either case, the signal will attenuate after a certain distance. To achieve longer distances, the signal is amplified several times along the transmission path. Unfortunately, an amplifier boosts the noise as well. Hence, the signal becomes increasingly distorted along the line. For analogue data such as voice and analogue TV, this leads to a graceful degradation of perceived quality. In practice, for these signals, quite a lot of distortion can be tolerated before the distortion becomes annoying. For digital data, however, cascaded amplifiers introduce more and more errors, leading to intolerable losses.

Digital transmission, in contrast, treats the signal differently. Instead of amplifiers, it uses repeaters to achieve longer transmission distances. At a repeater, the digital data is recovered from the analogue carrier and a new analogue carrier is generated from this data. Hence, the noise imposed by the different links between the repeaters is not cumulative.

The minimum required distance needed between repeaters in digital transmission is closely related to the medium being used and the data rates to be transmitted over the medium. In general, the higher the data rate on a channel is, the more the damage undesirable noise can do, which is described in the Shannon theory [11]. Table 2.1 shows the data rates, bandwidth and required repeater spacing to reach acceptable reliability for digital data (an error probability of 10^{-8} to 10^{-12}) on the most important guided media (media where the signal is guided by some kind of wire)[12].

For non-guided media (transmission through air or vacuum) other considerations apply. To provide point-to-point transmission, a directional beam is required. In general, however, at low frequencies the signal propagates in all directions from an antenna. Hence, signals in the range 30 MHz to 1 GHz are used for broadcasting only and are referred to as radio waves. In contrast, with microwave frequencies, covering a range of about 2 to 40 GHz, highly directional beams are possible, and

thus these signals can be used for long distance terrestrial and satellite point-to-point transmission of data rates in the order of 100 Mbit/s.

Transmission Medium	Total Data Rate	Bandwidth	Repeater Spacing
Twisted pair	4 Mbit/s	250 kHz	2-10 km
Coaxial Cable	500 Mbit/s	350 MHz	1-10 km
Optical Fibre	2Gbit/s	2 GHz	10-100 km

TABLE 2.1: TRANSMISSION CHARACTERISTICS OF GUIDED MEDIA

The data-link layer of the OSI model may reduce the number of errors by retransmitting data frames which are damaged. In real-time services, however, the delay of these retransmitted frames would be too long for transmission. Hence, for these services this layer is usually ignored. In this case, the errors will have to be handled by higher layers in the OSI model. In the rest of this thesis, we assume that the choices made in the physical layer (analogue or digital transmission, repeater/amplifier spacing etcetera) guarantee an error behaviour that is robust enough for the application under consideration.

The differences between real-time services, such as audio and video, and other (computer) data is the origin of differences in the technology used in the different networks. In the OSI model, these differences can be found in the network layer. Where networks for real-time services often use circuit switching techniques, the nature of (computer) data services has led to packet switching techniques for data transmission. The following sections describe these techniques and discuss their capabilities to transmit real-time audio-visual services. Also, the Asynchronous Transfer Mode (ATM) is described. ATM is the technique proposed to be used on the information super-highway, the Broadband Integrated Services Digital Network (B-ISDN). Hence, it should be capable of transmitting a wide range of different services, including VBR video.

2.1.1 Circuit Switched Channels

For analogue real-time audio and video services, the delay of information sent should be constant over time. A guarantee of this can be given when a fixed physical channel is available at all times. For distributive services, like cable television, one link is needed from the service provider to each costumer, as is illustrated in Figure 2.1a. In a communication network a link is needed between any two customers. Figure 2.1b shows that this leads to a total of $\frac{n(n-1)}{2}$ wires if n customers were involved. To

avoid such an excessive amount of wires, the available resources are allocated by circuit switching techniques.

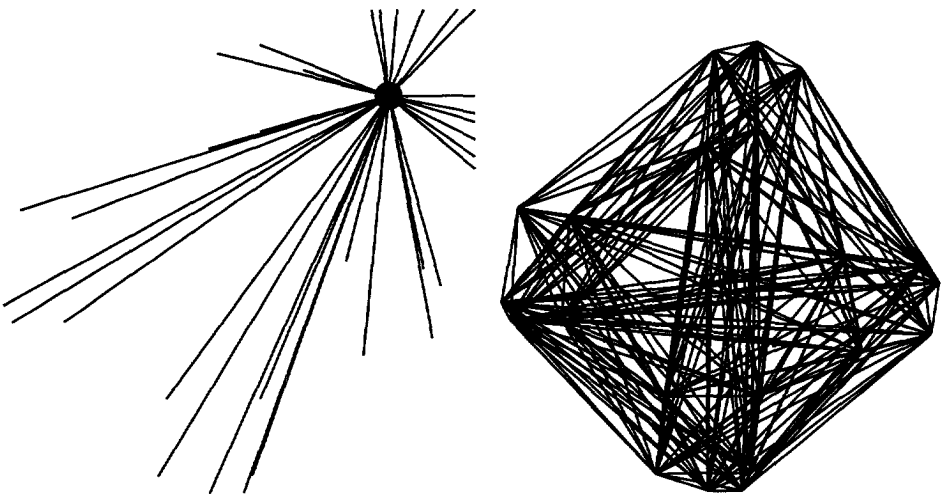


FIGURE 2.1: TOPOLOGY OF SIMPLE (A) DISTRIBUTION AND (B) COMMUNICATION NETWORKS.

In the early days of telephone history, all the local users were connected to a *local phone exchange centre*, which was in charge of interconnecting any two customers on demand. In chronological order, the switching was carried out by manual, mechanical, and electronic procedures. By doing this, only n links between the customers and the phone centre had to be used.

To connect remote customers with each other, the local phone exchange centres had to be connected. Instead of connecting them two by two, a new level in the network hierarchy was introduced by connecting each local phone exchange centre to a new common exchange centre. By doing this again and again on a larger scale, an entire exchange network came into existence which is based on a thorough hierarchical structure as shown in Figure 2.2. Each level in that structure corresponds to a geographical region. Such an organization reduces the connection costs drastically at the expense of switching costs. Since the communication paths are established by the switching at connection set-up, the resources are allocated for the duration of the connection.

A network topology as shown in Figure 2.2 is also practical for distribution services since many customers live next to each other while the distance to the distribution

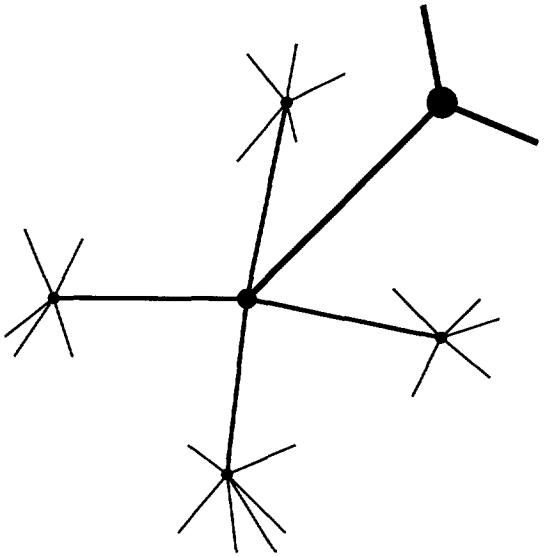


FIGURE 2.2: TOPOLOGY OF A HIERARCHICAL NETWORK.

centre is relatively large. Further, since different customers may require different services, circuit switching is employed to deliver the different services to the customers. In this case, the switching is semi-permanent and only changed when a costumer requires a different service.

In the conventional Public Switched Telephone Networks (PSTN), the customers are connected by single lines with a limited bandwidth. The first multi-purpose digital network (ISDN, Integrated Services Digital Network) evolved from these networks in the eighties and is based on the circuit switching of basic channels with a capacity of 64 kbit/s. By using multiple channels, however, it is possible to allocate a dynamic range of capacities [13]. To transmit real-time audio-visual data over such a network, the service provider determines the bandwidth needed to support the desired quality of the service. The network then decides whether a connection can be made supporting this bandwidth. Since the allocated bandwidth is fixed, constant bit rate coding needs to be applied. As discussed in Chapter 1, the bandwidth should be chosen to be high enough to generate an acceptable quality in the scenes with high activity.

VBR compressed video can only be transmitted by circuit switched networks when the bit rate is stuffed in the transport layer so that the total output bit rate is constant. This concept is shown in Figure 2.3a, where the generated bit rate and the channel bit rate are plotted. The stuffing bits are indicated by the shaded region

between the VBR bit rate an. Although this yields an inefficient use of bandwidth, the loss is limited, assuming that encoding has been carried out such that the peak bit rate of a VBR source (and thus the required channel bit rate) resembles the bit rate of a comparable CBR source.

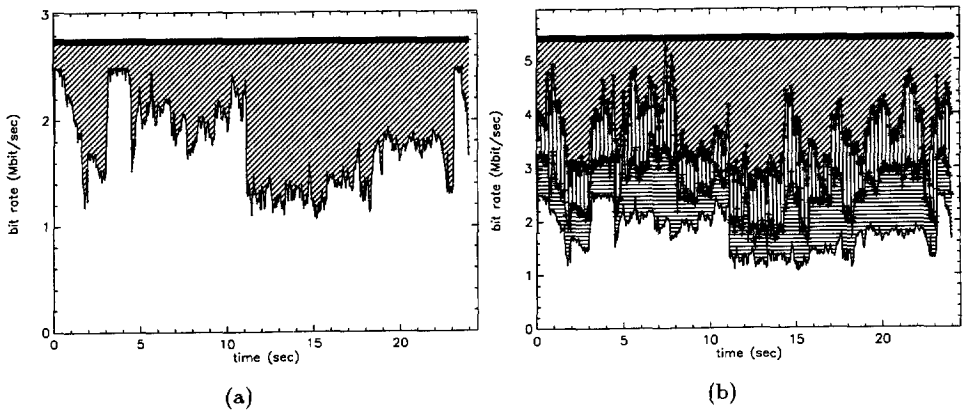


FIGURE 2.3: TRANSMISSION OF VBR SOURCES ON A CBR LINK. (A) THE BIT RATE OF A SINGLE SOURCE IS STUFFED TO GENERATE A CBR STREAM. (B) MULTIPLE SOURCES ARE MULTIPLEXED AND STUFFED TO GENERATE A CBR STREAM.

The advantages of VBR over CBR can only be exploited when multiple sources are multiplexed in the network layer. One practical application of this is the distribution of multiple television channels on one outgoing link at the cable head-ends. Then, the bit rate of the CBR link can be lower than the sum of the peak bit rates of the individual VBR sources, which is shown in Figure 2.3b, where the sum of the bit rates from 1, 2 and 3 VBR sources are plotted, together with channel bit rate, which is only twice the peak bit rate of a single source. The individual bit rates of the 3 sources are the non-shaded, horizontally shaded, and vertically shaded surfaces, respectively. The amount of stuffing bits is indicated by the diagonally shaded region. This concept is only useful if the sources are statistically independent. Further, the bandwidth of the CBR link should be chosen according to a trade-off between efficient use of the network and tolerable losses caused by overflow.

2.1.2 Packet Switched Channels

Traditionally, packet switched channels are used in networks for data communication. With the transmission of data, no constraints are imposed on the delay, or variations in the delay (delay jitter). There is, however, the demand that all data arrives at

the receiving end without errors. Further, the duration of a data communication session is relatively long, but the data is transmitted in bursts. Hence, there is a demand to share the available resources more efficiently than is the case with circuit switching techniques.

To share resources in packet switched networks, the data of a source is split into packets which are sent in the next free time slot of the output link [1]. In each network node the packets are routed from an input to an output link where they are remultiplexed with packets of other sources. One of the objectives of the network layer in packet switched networks is to maximize the efficiency of the network, defined as the amount of packets correctly delivered by the network divided by the total capacity of the network.

When the number of packets offered to the network increases, the efficiency should also increase. However, when it exceeds the capacity of the network, packets are lost due to the limited buffer sizes in the network nodes. Although retransmission protocols compensate for these losses, too many losses and retransmissions may lead to congestion, thus causing a drop in efficiency, as is illustrated in Figure 2.4 [1].

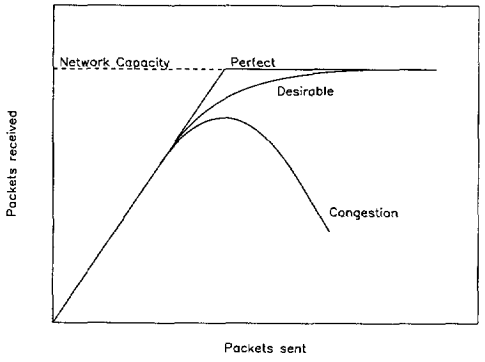


FIGURE 2.4: DROP OF EFFICIENCY CAUSED BY CONGESTION.

The most important way to improve network efficiency and thereby avoid congestion is by applying effective routing algorithms. To transmit a packet from source to destination, many different routes through the network are possible. Routing algorithms can be divided into non-adaptive and adaptive algorithms. In non-adaptive algorithms the route from point i to point j is fixed and calculated off line. Adaptive algorithms change their routing decisions according to the changes in the load and the time-varying topology of the network.

[1] In some circuit switched networks the data is also split into packets. In that case, however, fixed time slots are allocated for these packets.

Adaptive routing algorithms can be divided into three classes: centralized, isolated and distributed algorithms. Centralized algorithms try to gather information from the whole network to make optimal decisions. Isolated algorithms operate individually on each node and use only local information, like the length of the queues on every output line. Distributed algorithms make use of both local and global information by exchanging information between the network nodes.

Optimal routing algorithms can rearrange the traffic so that the occurrence of congestion is minimized. However, they cannot prevent the occurrence of congestion. To do this, several congestion control algorithms can be implemented, based on the discarding of packets, flow control, admission control and allocation of resources [12]. These strategies limit the maximum efficiency of the network. For instance, when all resources are allocated in advance, packet switching becomes as inefficient as circuit switching. The amount of congestion control applied therefore depends on the type and purpose of the network.

Since each packet may be routed differently and retransmission of lost packets may occur, no guarantee can be given with respect to the delay. Surely, the order in which packets are received may differ from the order in which they are sent. Although this concept of packet switching networks is not suitable for real-time services, there are two possibilities to transmit audio-visual data over these networks. The first is by off-line transmission and playback, the second by accepting the limitations of the network.

With off-line transmission and playback, the audio and video are transmitted as data files. Lost packets are handled by retransmission protocols. At the receiving side, the complete files have to be stored and delayed packets arriving in the wrong order are reorganized to complete the original data files. After receiving the complete files, playback is performed off line. The objective of video compression in this case is to compress as well as possible to reduce the size of the data files to be transmitted. Hence, it is mostly VBR compression techniques that are applied here. This type of video transmission is performed by different sites for the world wide web (WWW) browsers on the Internet.

When not enough memory is available to store the complete files, or when a communication service is required, real-time transmission over packet switched networks has to be performed. The limitations of the network will then have to be taken into account in the encoder design. Usually, neither CBR nor VBR is applied, but the algorithm adjusts its coding techniques and temporal resolution according to the available bandwidth [14]. The pictures are encoded with constant quality, while the output rate is adjusted to the network load. Since the retransmission protocols for data transmission are replaced by real-time protocols [14, 15], a relatively large amount of packet loss will have to be dealt with. Nevertheless, for some applications an acceptable quality can be reached. When more stringent properties for real-time

transmission are desired, an allocation of network resources is necessary. To be able to share resources in this case, other network architectures and protocols are needed [16].

2.1.3 The Asynchronous Transfer Mode

As a switching technique to support a wide range of services on a B-ISDN, the Asynchronous Transfer Mode (ATM) has been proposed and subject to standardization by the ITU-T [17, 18]. It is based on a trade-off between circuit and packet switching and is therefore discussed separately here.

The organization of the ATM Protocol Reference Model (PRM) as shown in Figure 2.5 is similar to the ISO OSI reference model. As in other networks designed for the real-time transmission of services, the data-link layer is omitted since ATM deals with high data rates and complicated data-link protocols would hamper the transmission speed. The ATM layer corresponds to the OSI network layer, where the packets are routed, switched and multiplexed. The support and adaptation of the different services takes place in the equivalent of the OSI network layer, the ATM Adaptation Layer (AAL).

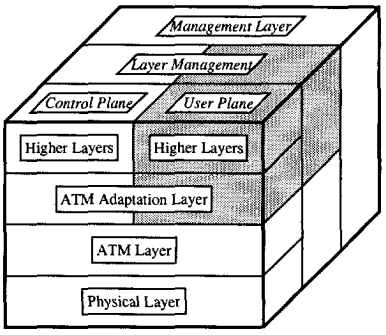


FIGURE 2.5: THE B-ISDN ATM PROTOCOL REFERENCE MODEL.

An additional division into planes refines the description of the protocol. The user plane provides the information flow. The control plane provides the signalling flow, used to set up, control and release the connections. The management plane is divided into two parts, one managing the layer functions, and one managing the co-ordination between the different planes.

ATM is based on the switching of fixed-size packets called ATM cells, which are 53 octets long, 5 containing header information, 48 for data. The switching occurs via a fixed route through the network, selected during set up. Hence, the header does not contain the address of the destination, but only information about which

virtual channel (VC) it belongs to. Thus, the cells can be switched fast, without complicated routing decisions for each individual cell, and the order in which the cells are sent is preserved. This concept makes the real-time transmission of video services possible.

The virtual channel a cell belongs to is identified by the Virtual Channel Identifier (VCI) field in the header. In addition, the Virtual Path Identifier (VPI) identifies the virtual path (VP) the virtual channel belongs to. A virtual path is a semi-permanent connection between end-points of the network. The switching of complete virtual paths is easier and faster, permitting simple VP switches besides the more complicated VC switches, as is shown in Figure 2.6.

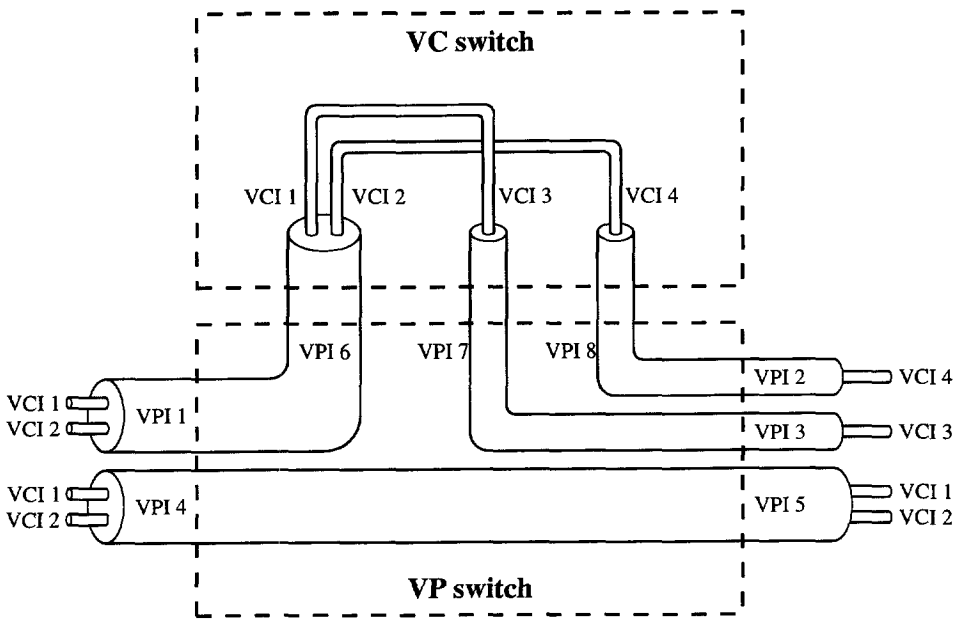


FIGURE 2.6: VIRTUAL PATH AND VIRTUAL CHANNEL SWITCHING IN ATM.

The structure of an ATM cell header at a User Network Interface (UNI) is shown in Figure 2.7. At a Network-Network Interface (NNI), the Generic Flow Control (GFC) field, which is used by the access flow control mechanism at the UNI, is also used for the VPI since the number of virtual paths to a user is smaller than the number of virtual paths in a network node. The Payload Type Indicator (PTI) field indicates whether the payload contains data or maintenance information and whether congestion has been experienced by the cell along its way through the network. Both maintenance and congestion information can be used by the network

to improve the resource management. When the Cell Loss Priority (CLP) bit is set by either user or network, the cell will be discarded when congestion occurs to make space for other cells. The HEC field is a check sum to detect and correct bit errors in the cell header.

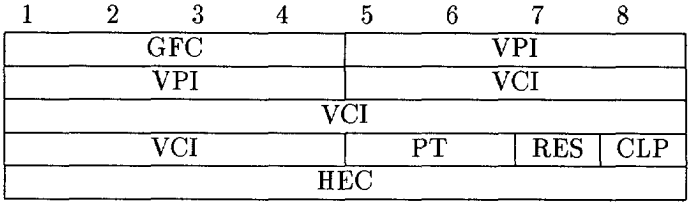


FIGURE 2.7: THE STRUCTURE OF AN ATM CELL AT THE UNI.

In order to select a route through the network at connection set-up, the network has to decide whether enough bandwidth is available to support the desired traffic. This decision is made by the Call Acceptance Control (CAC) and it is based on the traffic descriptions and Quality of Service (QoS) parameters of all services currently present on the network. It is the objective of the CAC to achieve a network load that is as high as possible, while the QoS of all user contracts can be guaranteed. Since in an ATM switch the cells of different sources are treated individually, all sources may suffer from a network overload, not only the newly admitted ones.

When a path able to support the traffic at the requested QoS is found, a contract specifying QoS and traffic characteristics is made. The network guarantees the QoS on the condition that the characteristics of the traffic adhere to the contract. To monitor the behaviour of each source, a Usage Parameter Control (UPC) is active where the source enters the network. When this network function detects a violation of the user contract, it punishes the user by discarding the cells, or by setting the CLP bit in the header, causing an early discarding in the case of congestion in a network node.

The CAC and UPC have to work together in close co-operation. When a certain UPC function (or a combination of functions) is standardized, vendors may apply such a control in their video encoder design. These encoders will optimize the quality of the reconstructed pictures by using as many cells as allowed by the UPC function(s). Since the quality in easy scenes will reach above the level where users become insensitive to further picture refinement, these algorithms are called greedy [19]. The network load estimate of the CAC function has to be based on these greedy algorithms, since this is the worst case situation.

Currently proposed CAC and UPC functions are mainly based on the peak bit rates of VBR sources. If greedy algorithms are assumed using these functions, no compression gain is achieved with the VBR concept described in Chapter 1. Hence, users should be encouraged to use non-greedy algorithms instead. This can be done by using more complicated UPC algorithms, or by using a tariffing system, which rewards the user for a low mean cell rate [19].

2.2 Storage Media

In the field of storage media, a distinction can be made between the storage of real-time audio-visual services and the storage of (computer) data. Because of the nature and extensive bandwidth requirements of real-time audio-visual services, traditional storage media for these services are mostly based on magnetic tape devices. Conversely, data storage is often performed on randomly accessible devices, mostly magnetic and optical disks. Recently, however, the capacity and throughput of these devices has increased enough to support real-time services as well.

Since errors in storage devices cannot be retransmitted, no difference with respect to the robustness required for data and real-time services can be identified in storage media. Again we assume that the error behaviour of the physical layer is robust enough for the application.

The storage of real-time audio-visual services includes two cases. First, there is a need to store real-time audio-visual data in a home environment. In this generic recording case, the data is usually received via a network, and has been encoded for that network. Hence, the data is not adapted to the storage device. The aim of the storage device is to act as a common network node. This means that at playback, the characteristics of the bit stream should be identical to those of the bit stream received at recording time.

Second, we should consider specific recording, where the data is encoded for one specific recording device. In this case, all the features of that device can be taken into account to optimally adapt to the device. Extra functionalities can be included in this case, for instance to support fast search and editing capabilities. Specific recording can also be applied in the home environment by transcoding: decoding the data before encoding for storage. This option may be attractive to include the extra functionalities, but it also produces additional distortion of the data. In the following sections, generic and specific recording of real-time audio-visual services are discussed for disk and tape storage devices.

2.2.1 Disk Storage

The development of disk storage devices began with magnetic devices for personal computers. The capacity of these devices has increased from less than 1 Mbyte in the early days up to several Gbytes in the newest systems. Parallel to this technology, the Compact Disc (CD) was developed to be the carrier for digital audio. The success of the CD extended to the computer and entertainment environment by introducing CD-ROM and CD-I(nteractive), and CD-video as a specific application of CD-I. Although these systems are successful, a need for more capacity and higher data rates to support higher quality video has arisen. This has led to the introduction of the DVD (Digital Versatile Disk) which has a larger storage capacity and higher throughput rates, as shown in Table 2.2. The throughput of hard-disk and CD drives shown here is fixed, while the DVD may operate at varying speeds. Further, although a writable CD-ROM is being developed, it will probably not be possible to write CD-Is, while DVD aims at replacing the VCR, to be able to record video in future generations of DVD-players.

Device	Capacity	Throughput	Read/Write
Hard-disk	0.2-2 GByte	10 Mbit/s	R/W
CD-I/ROM	680 MByte	(n*)1.44 Mbit/s	R
DVD	4.7 GByte	10 Mbit/s	R(/W)

TABLE 2.2: CHARACTERISTICS OF SOME DISK STORAGE DEVICES.

An advantage of disk devices is the random accessibility. This property makes it very easy to include fast search capabilities for digital video by skipping sections and waiting for the next fully encoded picture, which is possible for both specific and generic recorded video. Editing can also be implemented easily, by dividing the video into small parts that can be edited separately.

Where CD-I's only operate on a fixed bit rate of 1.44 Mbit/s [2], a DVD may operate at varying rates. Hence, VBR video recording is possible on a DVD, although a certain peak bit rate will have to be observed. For DVD-specific recording of a movie, often a two-pass VBR technique is applied: in the first pass the complete movie is analysed in order to predict the complexity. The second pass encodes the movie at a constant quality and thus a variable bit rate. Thus, the complete movie will fit almost exactly on one disk, while the quality over the whole movie is (nearly) constant.

[2] The different speeds of CD-ROM drives are not used in CD-I or CD-video since it would reduce the playing time of a single CD.

2.2.2 Tape Storage

Using magnetic tape for the storage of real-time services has been a topic from the early days of magnetic recording to date. First, only voice and audio recording were considered for which longitudinal techniques were applied. These techniques write the data on one or more tracks along the tape. Video signals, however, require roughly 300 times the bandwidth of audio signals, which could not be recorded with the existing longitudinal techniques. Thus, a new technique, called helical-scan recording, was developed for the video recorder. Later, both longitudinal and helical-scan techniques have been extended to record digital signals [20].

In a helical-scan recorder, the magnetic heads are mounted on a rotary head wheel inside a cylindrical drum, which has the tape helically wrapped around it. Thus, the tracks written on tape have a small angle with respect to the tape direction and they cross the width of the tape gradually. A high track density is achieved by applying azimuth recording, where the bits of consecutive tracks are written with different orientation so that there is little cross-interference when the tracks are read.

In analogue helical-scan video recording, as applied in the VHS home systems, each track contains the information for one picture and a certain position on the track corresponds with a certain part of the picture. Hence, when the speed of the tape is increased while the speed of the drum is fixed, data from multiple tracks is read which leads to a fairly well recognisable picture. To get similar search modes with digital recording, each picture segment has to be encoded with a fixed number of bits. Hence, for tape-specific recording, a feed-forward rate-control technique is implemented in the video compression algorithm [21]. Another constraint arising from the demand for search modes is that it should be possible to decode each segment stand-alone, which means that no prediction from previous pictures can be used in the video compression algorithm. Since much compression in video coding results from this prediction, often a compromise is made, and only every other picture is predicted [21].

When we wish to use a tape recorder for generic video services recording, we encounter two difficulties. First, the support of fast search modes needs to be considered [22]. Second, an adaptation for VBR sources is needed. Although the tape in both longitudinal and helical-scan recording may run at different speeds, it is not feasible to vary this speed to correspond to the characteristics of the bit stream to be recorded. Hence, the recording and playback rate of tape recorders is always fixed. In principle, VBR recording would still be possible using a start/stop mechanism in co-operation with a buffer. However, the requirements on the size of the buffer and on the mechanics of the recorder are very high, especially for use in a consumer product.

As in VBR transmission on circuit switched channels, one possibility to record a VBR source on a CBR tape recorder is by using stuffing. In this scenario the recorder

runs at the peak bit rate of the VBR source. Each track is filled with the available VBR data and then stuffed with dummy bits. Clearly, no advantage of the VBR concept is obtained in this case.

To make use of the low mean bit rates of VBR sources, it is possible to record more than one VBR source, in other words to apply multiplexing on tape. This requires a labelling of the data to distinguish the different sources, but the bit rate of the tape recorder in this case can be smaller than the sum of the peak rates of the recorded sources. It is obvious that a higher efficiency is reached with more sources. The choice of the recorder rate should be a trade-off between tolerable losses due to overload and efficient tape use.

2.3 Conclusions

Many transmission and storage media are traditionally based on constant bit rate (or constant bandwidth) sources. New media like the ATM network and the DVD, however, identify the advantages of VBR video and make use of them. In the infrastructure of the near future, an interaction between old and new media is inevitable. Although transcoding of data can be performed to conform to the characteristics of traditional media, in many cases it is much more efficient to leave the encoded data in its original form. Fortunately, there are some possibilities to transmit and store variable bit rate sources on constant bit rate media. Stuffing does not make use of the advantages of VBR over CBR sources. As long as the peak bit rate of VBR sources resembles the bit rate of comparable CBR sources, this may not be a problem. However, it is more attractive to use multiplexing to make use of VBR advantages. When multiplexing is applied, a trade-off will have to be made between the efficient use of the channel and the probability of channel overflow resulting in information loss.

Chapter 3

Variable Bit Rate Video Compression

The concept of VBR compression differs from that of CBR compression: where CBR compression aims at distributing the available bits as efficiently as possible over the video data, VBR compression aims at encoding the video at a constant quality. Although no restrictions on the output bit rate are imposed, the compression should still be maximized. This chapter describes some techniques to apply constant-quality video compression and goes on to the description of a reference VBR codec, the bit rate characteristics of which are considered in the following chapters. The objective of this chapter is not to develop the ultimate VBR coder, but to describe the techniques and difficulties in constant-quality compression, and to develop a coder that is representative for a wide class of VBR coders, especially in bit rate properties.

The quality of compressed video is not defined explicitly and depends on three factors. First, it is obviously related to the amount of distortion introduced in the compression stage. Second, it depends on the response of the human visual system. Finally, the conditions under which the video is viewed play an important role. The objective of constant-quality video compression is to have not only a constant, but also an acceptable quality of reconstructed video. Different applications, however, have different requirements with respect to the quality. Videophone applications require lip synchronization, standard television applications assume a certain viewing distance from which no distortion should be visible, and in HDTV the enhanced resolution should allow a user to zoom in on the details in a picture without quality degradation. Further, if we assume that the video is stored by a user, no distortion should be perceived in different playback modes, such as slow motion and still

picture. To be able to distinguish different levels of quality performance, we define constant-quality as perceptually lossless under certain viewing conditions. Changing the viewing conditions thus changes the quality requirements. In this chapter, normal viewing conditions for standard definition television (SDTV) are assumed, since this is the most important class of video application.

In the past, a variety of video compression algorithms has been developed, all aiming at reducing the costs of video transmission. Each of these algorithms applies its own techniques to remove redundancy and to control the output bit rate or quality. In sub-band coding, for instance, bits are allocated to each frequency sub-band [23]. The application of constant-quality (or constant-distortion) compression implies that enough bits will have to be allocated to ensure the specified minimum quality (or maximum distortion) [8]. In vector quantization, achieving a constant quality not only yields the design of a codebook that provides enough properly spaced vectors to guarantee the prescribed quality, but also an efficient updating of codebooks and address maps to optimize compression [24]. Hence, besides a definition of quality, each compression algorithm also requires a specific strategy for constant-quality compression.

In making a choice for the video compression algorithm to be used in constant-quality compression, we consider a wide applicability of the techniques and thus a widely used algorithm. Although recent advances in integrated circuit technology allow the implementation of most video compression algorithms in real-time hardware, the high costs of these codecs can only be reduced by a standardization of video compression techniques. Also, the problem of interoperability of equipment from different manufacturers can only be solved this way. In the last decade, standardization activities in the ISO-IEC JTC1/SC29/WG11 Moving Pictures Experts Group (MPEG) have led to the development of standards for codecs at throughput rates of 1.5Mb/s and up to 40Mb/s, which are called MPEG-1 and MPEG-2, respectively [25]. Although the first one is mainly geared towards digital storage media in a multi-media environment and the second has a much wider area of applicability, the techniques used in both standards are similar. Since the support of, for instance, interlacing in MPEG-2 may be superfluous in the future because of the better performance of progressive scan systems, we do not consider the impact of these extensions of the basic techniques, which are part of the MPEG-2 standard. Therefore, we consider the possibilities to apply constant-quality compression within the MPEG-1 video compression standard.

In this chapter, we develop a constant-quality MPEG-1 coder. An introduction of MPEG-1 is given in Section 3.1, with emphasis on the quantization stage, which is responsible for the trade-off between distortion and bit rate. To apply optimal constant-quality compression, it is necessary to know how quality is assessed by human observers. Section 3.2 describes some results of the ongoing research in the field of human quality assessment of compressed video. Section 3.3 describes the

techniques to perform perceptual compression leading to our reference constant-quality MPEG-1 codec, described in Section 3.4.

3.1 The MPEG-1 Video Compression Standard

The MPEG-1 standard is divided into five parts. Apart from the video [26] and audio [27] parts, which specify the coded representation of the video and audio components, the system [28] part specifies the structure to multiplex audio and video data and the means of representing the timing information needed to synchronize the sequences in real-time playback. Parts 4 and 5 of the standard are for conformance testing and software simulation, respectively. Here, we look only at the video part, since this part contains most of the data and since the fluctuations in video data rate dominate the fluctuations of the total output rate. In Sections 3.1.1-3.1.4 various aspects of the standard are described. Section 3.1.5 focuses on the quantization stage, where the trade-off between distortion and bit rate occurs. Section 3.1.6 describes how quantization parameters can be used to control bit rate or quality.

3.1.1 Basic Coding Scheme

The MPEG-1 video compression algorithm is based on the hybrid coding scheme shown in Figure 3.1 [29, 30]. In this scheme temporal DPCM is applied to remove the temporal redundancy. The use of motion information to improve the prediction is essential here. The motion between a previously encoded reference picture and the current picture is estimated [31, 32] and a motion compensated prediction is made. In general, the entropy of the prediction error, also called the displaced frame difference (DFD) is much smaller than the entropy of the original picture. Therefore, spatial encoding techniques, represented by the "Intra Coder" in Figure 3.1, result in a much lower bit rate if they are performed on the DFD than on the original picture.

3.1.2 Syntax

The standardization within MPEG concerns the syntax of the encoded video data, not the encoding algorithm itself. In fact, through this, the standard specifies the requirements for a decoder to decode the incoming data. Each encoder must generate a data stream which can be decoded by a decoder that meets these requirements. Obviously, standardization of the syntax largely defines the encoding algorithm. However, there are still many possibilities for optimization of the encoder performance by, for instance, a better motion estimation or quantization control.

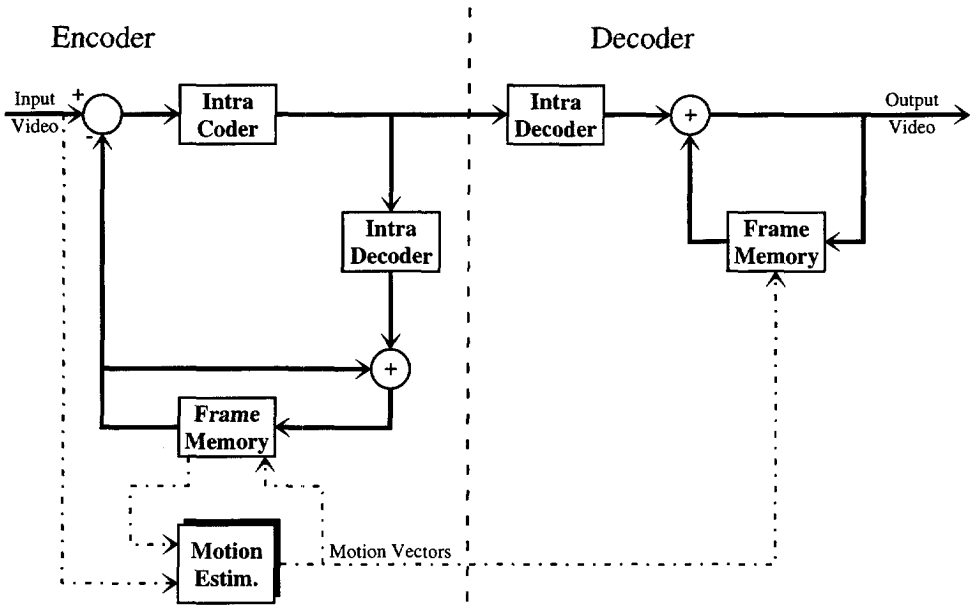


FIGURE 3.1: HYBRID CODING SCHEME

The syntax of the MPEG video bit stream has a layered structure, as is shown in Figure 3.2. Each layer contains encoded parameters which define the interpretation of underlying layers. These parameters are transmitted in the header of each layer, after which the structures of the underlying layers are sent.

On top of the syntax hierarchy is the video sequence layer. The parameters of the sequence are defined in this layer: the picture frequency, the dimensions and aspect ratio of a picture, the bit rate, and the required decoder buffer size. In addition, the quantization matrices, which are discussed later on in this section, can be transmitted in this layer. When they are not included, the standard matrices are used.

The layer below the sequence layer is the Group Of Pictures (GOP) layer. A picture sequence is divided into groups of pictures to provide access points from which decoding can be started in the sequence. Therefore, a GOP always starts with an intra-coded picture, which does not need a temporally referenced picture to be decoded.

Pictures are the primary coding units which are defined in the Picture layer. A picture is subdivided into a number of slices. Each slice consists of a number of macroblocks in raster scan order. The size of a slice is not standardized and different slices in one picture may have different sizes. In the bit stream, each slice, and every

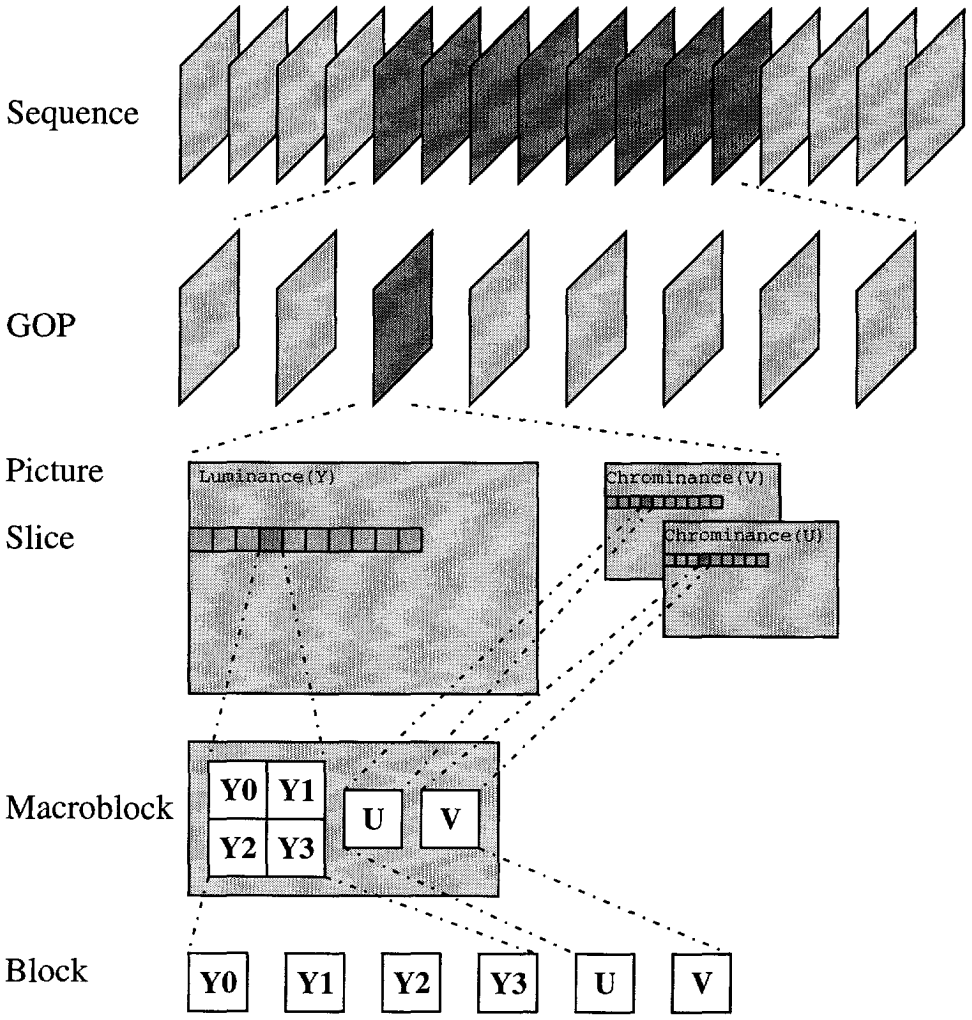


FIGURE 3.2: THE LAYERED SYNTAX OF MPEG.

layer above the slice, starts with a synchronization word. Hence, the slice layer is the lowest layer on which a decoder may re-synchronize when synchronization is lost due to an error in the bit stream.

The macroblock is the basic element which is used for motion compensation. Hence, motion information is transmitted in this layer. A macroblock contains a block of 16x16 luminance pixels and a block of 8x8 pixels for each chrominance (colour-difference) component. Finally, a macroblock is divided into six 8x8 blocks: four luminance and two chrominance blocks. A block is the basis for the DCT-based spatial redundancy reduction techniques.

3.1.3 Picture Types

In order to provide random access in the bit stream and because of the significant bit rate reduction offered by motion compensation, three different picture types are defined in MPEG: Intra-pictures (I), Predicted pictures (P) and Bi-directionally interpolated pictures (B). I pictures are coded independently of other pictures. They provide access points in the bit stream but require relatively many bits. P pictures are predicted from a preceding I or P picture and will be used as a reference for future P pictures and past and future B pictures. B pictures may be predicted from a preceding and a following I or P picture. On the macroblock level, a decision is made as to whether the prediction should be made from the preceding or the following picture, thereby dealing with occluded areas, or by averaging between both predictions, which reduces the sensitivity to noise. B pictures provide the highest amount of compression and are never used as a reference.

The relation between I, P, and B pictures in a GOP is illustrated by an example in Figure 3.3. The dashed B pictures indicate the use of *open* GOPs. These pictures require reference pictures from two different GOPs. When they are omitted, the term *closed* GOP is used. Due to the use of B pictures, which require a future reference, the transmission order is different from the display order. The future references to which B pictures refer are always transmitted before the actual B pictures.

3.1.4 Spatial Redundancy Reduction

The intra coder shown in Figure 3.1 uses the spatial redundancy in pictures to compress the video signal. In MPEG, it is based on the 8x8 Discrete Cosine Transform (DCT) [33] as is shown in Figure 3.4. This operation decorrelates the coefficients and concentrates the energy in the lower frequencies. Further, because of the frequency-dependent response of the human eye, it is possible to apply a coarser quantization to the higher DCT coefficients. Hence, the DCT in conjunction with the quantization stage provides a high compression capability.

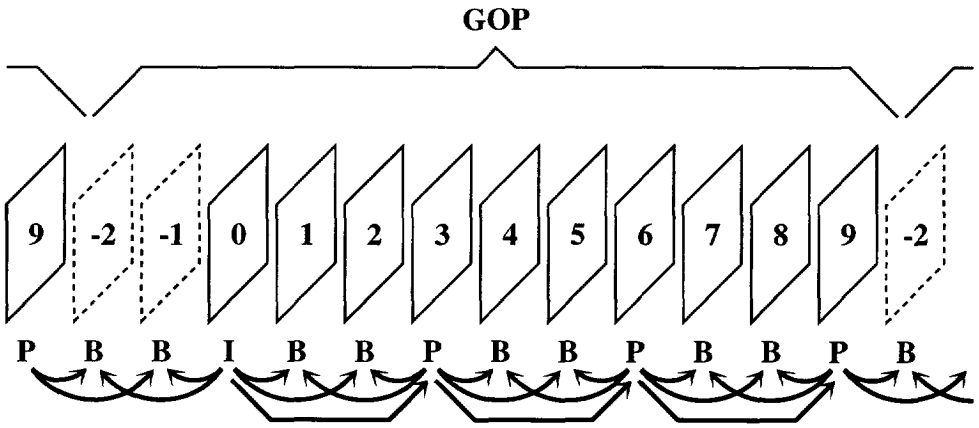


FIGURE 3.3: A GOP WITH ITS THREE PICTURE TYPES IN DISPLAY ORDER

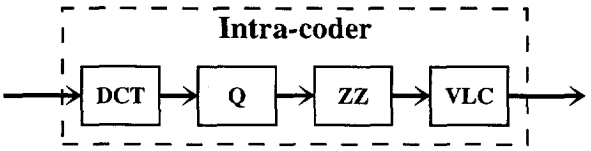


FIGURE 3.4: INTRA-CODER IN MPEG WITH DCT TRANSFORMATION, QUANTIZATION (Q), ZIGZAG SCANNING (ZZ) AND VARIABLE LENGTH CODING (VLC).

By quantization, a certain amount of information is permanently lost, which causes distortion in the reconstructed pictures. By adjusting the number of quantization levels, the output bit rate of the encoder can be controlled. By allowing fewer quantization levels, the bit rate is reduced at the cost of a higher distortion. Hence, the quantization stage is where the trade-off between bit rate and quality is performed.

After quantization, the data is scanned according to the pattern in Figure 3.5. For each non-zero coefficient, the number of zeros preceding it is counted. The run-level events created this way are entropy encoded, using variable length codes (VLCs) to make use of frequently occurring events, and transmitted sequentially. As a result, the occurrences of zeros provide an efficient compression.

3.1.5 Quantization

As the quantization stage determines the trade-off between quality and bit rate, we take a closer look at it. A quantizer is characterized by its decision intervals I_k and

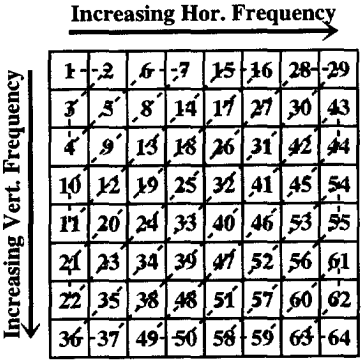


FIGURE 3.5: ZIGZAG SCANNING OF DCT COEFFICIENTS.

reconstruction levels y_k. For I pictures in MPEG, the quantization is based on a uniform quantizer which is shown in Figure 3.6a and characterized by:

$$y_k = 2kS, \qquad k = -Max..., -2, -1, 0, 1, 2, ...Max, \tag{3.1}$$

and

$$(2k-1)S < I_k < (2k+1)S. \tag{3.2}$$

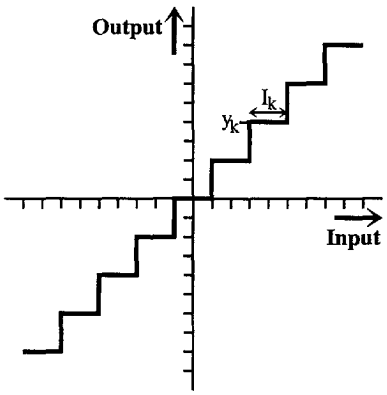
Here S denotes the quantizer step size and k is the representation level with range Max, determined by the VLC tables. The quantization of I pictures in MPEG only differs from a uniform quantizer in that the exact reconstruction levels of MPEG allow no even values, which has been found to prevent the accumulation of mismatch errors [26].

To exploit the limited sensitivity of the human eye to higher frequencies, a weighting matrix $m(u, v)$ can be used prior to quantization. As a result, S becomes dependent on the spatial frequency (u, v) of the DCT coefficient:

$$S(u, v) = \frac{m(u, v)}{16} Q. \tag{3.3}$$

The quantizer scaling factor Q is the parameter that controls the bit rate and the amount of quantization noise.

The best weighting matrix $m(u, v)$ to be used, if there is one, depends on many external parameters, such as the characteristics of the intended display, the viewing distance and the amount of noise in the input signal. Therefore, the matrix can be chosen by the encoder and, as was stated earlier, is transmitted in the sequence layer. When no matrix is transmitted, the decoder assumes a standard matrix, which is based on a coarser quantization of higher frequencies, based on the characteristics of the human visual system. This matrix is shown in Figure 3.6b.



(a)

Increasing Hor. Frequency

Increasing Vert. Frequency

8	16	19	22	26	27	29	34
16	16	22	24	27	29	34	37
19	22	26	27	29	34	34	38
22	22	26	27	29	34	37	40
22	26	27	29	32	35	40	48
26	27	29	32	35	40	48	58
26	27	29	34	38	46	56	69
27	29	35	38	46	56	69	83

(b)

FIGURE 3.6: (A) QUANTIZER CHARACTERISTIC AND (B) STANDARD WEIGHTING MATRIX FOR I PICTURES.

Because of the re-ordering of the data into run-level events, the occurrence of zeros provides an efficient compression. To exploit this property even more, a dead zone as shown in Figure 3.7a is introduced in the quantizer characteristic for P and B pictures. This yields for the reconstruction levels y_k:

$$y_k = 2kS + sign(k)S, \quad sign(k) = \begin{cases} -1, k < 0 \\ 0, k = 0 \\ 1, k > 0 \end{cases} \tag{3.4}$$

and decision intervals I_k:

$$\begin{cases} 2kS < I_k < 4kS, k \neq 0 \\ -2S < I_k < 2S, k = 0. \end{cases} \tag{3.5}$$

Because of this dead zone, more coefficients are clipped to zero and the bits gained this way allow a finer quantization. In most situation, the bits are more effectively used this way.

Apart from a difference in quantizer characteristics for intra- and predicted pictures, the MPEG standard also uses a different weighting matrix for predicted pictures. It was argued that the prediction error to be quantized contains mostly high frequency coefficients, since these are not easily predictable. When the same weighting is applied as in I pictures, most of this information is lost. In order to prevent this, for predicted pictures, another standard weighting matrix is applied, which yields a less coarse quantization of high frequency coefficients. This standard weighting matrix is shown in Figure 3.7b.

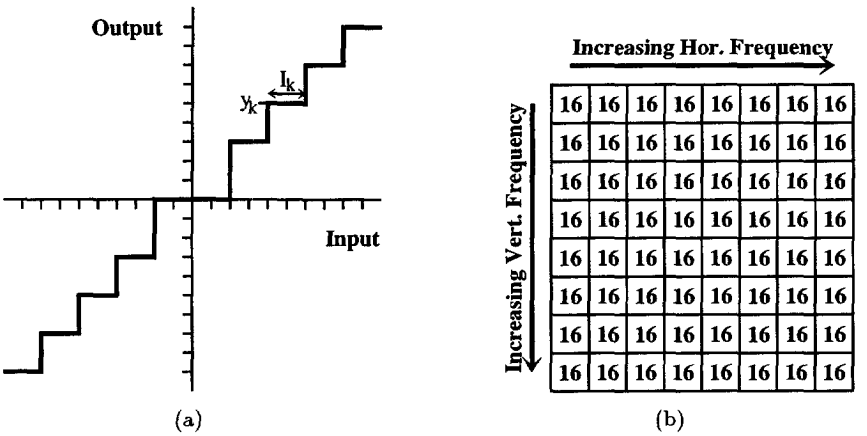


FIGURE 3.7: (A) QUANTIZER CHARACTERISTIC AND (B) STANDARD WEIGHTING MATRIX FOR P AND B PICTURES.

3.1.6 Control Parameters

Bit rate and distortion of a compression algorithm not only depend on the compression techniques used, but also on the input material. As a consequence, the rate-distortion curves for different pictures differ considerably. Further, in some parts of a picture more distortion is allowed for a certain visual quality than in other parts. Hence, there is a need for parameters to control the bit rate and quality.

To ensure a constant output bit rate, most CBR coders use a buffer in conjunction with a bit rate control technique that changes the quantizer scaling factor Q in (3.3), according to the complexity of the video scene. Therefore, Q can be seen as a global control parameter, used to control the output bit rate. Other global parameters are the weighting matrices, which control the relation between the quantization accuracies in the different coefficients. These matrices can only be changed at the sequence layer. The quantizer scaling factor, however, can also be changed at each macroblock and can thus be used as a local control parameter as well, to adapt the quantization according to local picture contents.

In conventional MPEG encoders, the quantizer scaling factor is the only local parameter used for adaptive quantization. It is, however, possible to change the quantizer characteristic in MPEG in other ways as well. Since only the decoding stage is standardized, only the reconstruction levels y_k of the quantizer, as described in (3.1) and (3.4), are fixed. The placement of decision intervals I_k is up to the encoder. A simple and effective way to exploit this freedom is by using a technique called thresholding: after quantization, some of the non-zero quantized coefficients are clipped to zero,

according to a certain criterion. Such a criterion may be, for instance, the maximum amount of distortion to be introduced. This maximum may be defined globally, but also locally, when it is determined according to the local picture contents.

In [34] thresholding is performed to optimize the compression for a certain bit rate. After quantization, a Lagrangian strategy is used to determine the subset of quantized coefficients that minimizes the total Lagrangian costs. This subset is transmitted and the rest of the coefficients are removed (thresholded to zero). The Lagrangian costs are expressed by:

$$J(\lambda) = D(C, \tilde{C}) + \lambda R(\tilde{C}). \tag{3.6}$$

Here, the distortion D depends on the original (non-quantized) coefficients C and the chosen set of quantized coefficients $\tilde{C}$. The thresholding of a coefficient enlarges the distortion but reduces the bit rate R needed for encoding the set of coefficients. Whether the Lagrangian costs are enlarged by the thresholding of a coefficient depends on the Lagrange multiplier λ, which is in fact the slope of the rate distortion curve.

The strength of using the Lagrange multiplier lies in the property that R and D are independent for different blocks. As a result, the subset of coefficients for a picture can be determined by using the same strategy in each block for the fixed slope λ. This can be seen from the argument that bit allocation is performed best when the R-D slope is equal in each block of the picture. Apart from finding the optimal set of quantized coefficients in a block, which can be solved, for instance, with the dynamic programming algorithm proposed in [34], the problem is reduced to determining λ so that the total rate or distortion for the entire picture is fixed. Hence, λ can be seen as an alternative global parameter, controlling rate and distortion.

Each implementation of an MPEG encoder has to make a choice between the local and global parameters to optimize quality and bit rate. In CBR compression, the global parameters are changed according to the complexity of the scene to generate a constant output bit rate. Local parameters can be used to re-allocate the bits according to the local picture contents. In VBR compression, a global parameter is kept constant for encoding. This approach is called open-loop VBR coding, and usually the quantizer scaling factor Q is used as the global parameter [35], with standard weighting matrices. In addition, however, thresholding can be applied using the Lagrange multiplier λ as global parameter [36]. In VBR compression, local parameters can be used to optimize compression while the quality is kept constant. To do this, we need to know how quality is assessed by human observers. Section 3.2 deals with human quality perception and Section 3.3 describes the techniques to use the global and local parameters for constant-quality compression.

3.2 Quality of Compressed Video

The quality of a video sequence after reconstruction depends on the amount of distortion introduced in the compression stage. The relation between distortion and the visual quality assessed by a human observer, however, is not straightforward. A lot of research has already been performed to find the relation between distortion and the visual quality of a compressed picture [9,10,37-64]. Nevertheless, video quality is still far from understood. Here, only a limited number of effects are discussed, modelled and exploited for compression.

The quality as assessed by human observers can be measured by subjective tests, where the video to be judged is shown to a large number of human subjects, who rate it on a five-point scale [65]. This suggests that the human visual system assesses the video globally. However, when a subject is asked why he rated the video as he did, he will point at several impairments in the video, which indicates local assessment. Here, we first look at how quality is assessed locally, before we discuss how these observations can be put into a global quality metric.

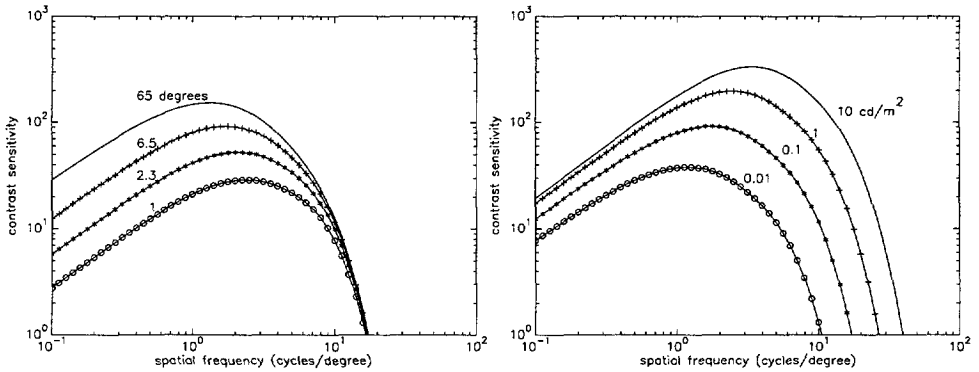
To describe how local quality is assessed by human observers, a model of the human visual system can be made, from which the quality as perceivedby a human observer can be predicted. Alternatively, in a video compression environment, it may be more useful to determine when quantization noise becomes visible. In the following sections, these methods are discussed separately, although many commonalities can be identified.

3.2.1 Modelling the Human Visual System

When a picture is electrically or optically transferred from one picture formation stage to another, the different spatial frequency components of the picture suffer from different responses. In general, the amplitude of the high frequency waves is reduced compared with the amplitude of the low frequency components. In fact, not the amplitude itself is of importance, but the amplitude divided by the average value, which is called the *modulation* or, in pictures, the *contrast*. The response of a system to different frequencies is called the Modulation Transfer Function (MTF) of that system. Since the human visual system responds differently to different spatial frequencies, it is desirable to describe this behaviour in an MTF. Because of the complex nature of the HVS, however, the MTF cannot be measured. What can be measured is the *contrast discrimination*, the ability of the eye to observe differences in the contrast. These measurements lead to a Contrast Sensitivity Function (CSF), characterizing the properties of the human visual system.

In literature, many measurements of the CSF can be found, as well as mathematical approximation formulas. Only the model in [37], however, takes the dependence on

background luminance, as described in [38], and display size (or viewing distance), found for instance in [39], into account. This model is shown in Figure 3.8a for a background luminance of 0.1 cd/m^2 and angular display sizes of 1, 2.3, 6.5, and 65 degrees. It is obvious that the sensitivity increases with increasing display size or decreasing viewing distance. Figure 3.8b shows the CSF for an angular display size of 6.5 degrees and varying background luminance of 0.01, 0.1, 1 and 10 cd/m^2. Here, the CSF not only increases with increasing background luminance, but the peak also shifts to higher frequencies.



(a) (b)

FIGURE 3.8: CONTRAST SENSITIVITY FUNCTIONS VARYING (A) ANGULAR DISPLAY SIZE AND (B) BACKGROUND LUMINANCE.

Apart from the spatial frequency sensitivity itself, experiments showed that the HVS treats stimuli of different orientations or frequencies independently. The human visual system appears to contain band-pass filters with a bandwidth of 1 octave and an orientation of about 30 degrees [40]. The distortion in different bands is not cumulative: the visibility of distortion in a frequency orientation band is independent of the distortion in other bands. Further, the sensitivity is higher in horizontally and vertically oriented bands than in diagonal bands.

The contrast sensitivity as function of the spatial frequency can be seen as a global property. The dependency on background luminance, however, is a local property, since the background luminance depends on the location on the screen. The phenomenon where the sensitivity of the human eye depends on local picture contents is called *masking*. The masking effects most often referred to are luminance, edge, and temporal masking.

Figure 3.8b shows that the contrast sensitivity increases with increasing luminance. When a picture is displayed, however, we are interested in the sensitivity in the

luminance of the picture as a function of the background grey value in the picture. The exact curve of the luminance masking depends on the monitor that converts the video signal, which is expressed in grey values k ranging from 0 to 255, into screen luminance L. Assuming that a correction for the γ characteristic of the display has already been applied, this should be a linear function. Figure 3.9a shows the luminance sensitivity function according to different background luminance values with the following conversion from k to L:

$$L = 0.00025(k + 15) \quad cd/m^2. \tag{3.7}$$

As different displays have different conversions, different curves for luminance masking are found in literature [62, 63, 55], but they all have in common that the sensitivity of the human eye has a peak at mid-grey levels.

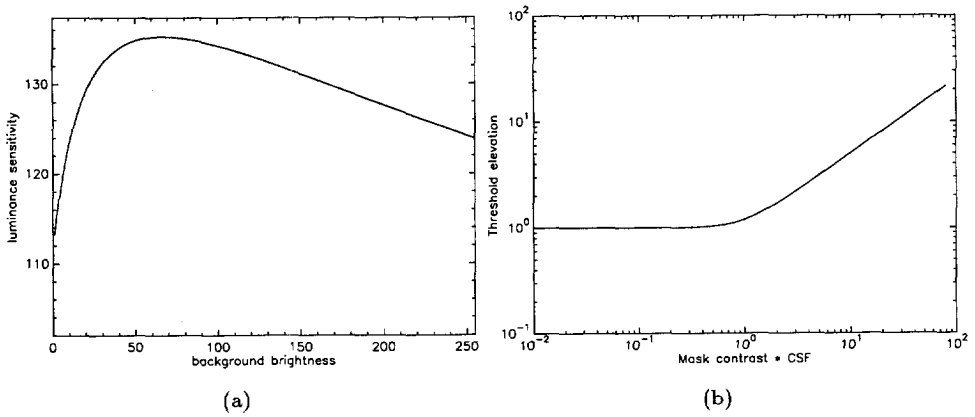


FIGURE 3.9: (A) LUMINANCE SENSITIVITY AS A FUNCTION OF BACKGROUND GREY VALUE. (B) THRESHOLD ELEVATION AS A FUNCTION OF LOCAL CONTRAST.

Edge masking is the phenomenon where the contrast sensitivity is smaller in the neighbourhood of an edge. This effect, however, very much depends on the size, shape and duration of the stimuli [41]. In general, maximum masking occurs at the exact position of the edge, and the effect decreases as the distance to the edge increases. Although edge masking was first identified for true edge patterns, it has been shown to exist in each frequency band. It expresses itself as an elevation of the visibility threshold of a frequency orientation band when the contrast in the original band exceeds a certain value [42], as is shown in Figure 3.9b.

Equivalent to edge masking, *temporal masking* results in a drop of the sensitivity near temporal discontinuities. Experiments have shown that the detection threshold of

narrow lines increases fourfold at temporal discontinuities [43]. Further, the masking is found to be much greater when mask and stimuli have the same polarity. The fact that masking is greatest for local spatial configurations indicates that masking occurs in normal video at scene changes.

3.2.2 Visibility of DCT Quantization Noise

The quality of a compressed picture after reconstruction depends on the kind of distortion introduced by the compression stage. In the previous section it was argued that the sensitivity of the human eye to distortion depends on the frequency of the distortion. When the compression algorithm, however, does not decompose the input signal into pure frequency components, the visibility of the distortion cannot easily be determined.

The DCT transform used in MPEG composes the picture in some kind of frequency components. In [44], the difference between sine waves and the DCT are captured in a transformation in order to use the CSF for the visibility of DCT distortions. This can be enhanced by using orientation tuning as was found in [48]. The visibility threshold for general DCT basis functions on a uniform mid-level background luminance are found in [45] for different pixel sizes. The base functions of the DCT, however, do not reflect pure frequencies. Especially for the higher spatial frequencies, several DCT coefficients occur in the same band and vice versa, one DCT coefficient may appear in several different frequency bands of the HVS [45].

Similar to the measurements of the CSF, the exact visibility threshold for each DCT coefficient can be obtained psychophysically by measuring under certain viewing conditions the smallest value that yields a visible signal [46, 47]. These measurements can be generalized to other viewing conditions (mean luminance, viewing distance, pixel sizes, etc.) by means of a model [48, 49]. Although some DCT coefficients may overlap in a single frequency band of the HVS, this model is extensively used for compression purposes. Masking effects other than luminance masking, however, are not included in the model.

Edge masking, as described in the previous section, influences the visibility of distorted DCT coefficients. Because of DCT coding, however, the distortion caused by quantization is spread over a block. Consequently, although the distortion on the exact edge can be masked, the same quantization may cause a distortion that is visible a few pixels away from the edge. In fact, around moving objects in front of a homogeneous background, this results in one of the most annoying artefacts in reconstructed video, known as mosquito noise.

In a textured region of a picture, a large number of edges are present. Consequently, the distortion in a textured DCT block is spread over several edges thus preventing the visibility of the distortion a few pixels away from an edge. This effect is called

texture masking, although it is caused by edge masking. Figure 3.10 shows the picture *car* encoded with constant quantizer scaling factor $Q = 8$. While the distortion is visible around the edges of the car, it is hardly visible in the textured region in the upper left part of the picture.

FIGURE 3.10: EDGE VS. TEXTURE MASKING IN BLOCK-BASED CODING. IN THE PICTURE *car* ENCODED WITH CONSTANT QUANTIZER SCALING FACTOR $Q = 8$ THE DISTORTION IS VISIBLE AROUND THE EDGES, BUT IS MASKED IN THE TEXTURE.

When the distortion in the DCT coefficients is not masked by texture, it becomes visible either as a loss of resolution or as artefacts. Unsharp edges and blur are examples of the first, mosquito noise and blocking of the second.

3.2.3 Quality Metrics

To quantify the quality of compressed video after reconstruction, an objective quality metric, expressed as a function of the original and the reconstructed picture, is desired. In terms of compression, the quality depends on the amount of distortion introduced by the compression process. Therefore, the most widely used quality measures are based on the variance of the reconstruction error, also called the Mean Squared Error (MSE):

$$MSE = \frac{1}{N_r * N_c} \sum_{x=0}^{N_c-1} \sum_{y=0}^{N_r-1} (p(x,y) - \hat{p}(x,y))^2, \tag{3.8}$$

where p and $\hat{p}$ denote the original and reconstructed pictures with N_r rows and N_c columns, respectively. From this measure, two kinds of Signal-to-Noise Ratio (SNR) can be calculated:

$$SNR_p = 10 \log \frac{255^2}{MSE}, \tag{3.9}$$

and

$$SNR_v = 10 \log \frac{\sigma_s^2}{MSE}. \tag{3.10}$$

where 255 is the peak intensity, assuming that the original picture was quantized with 256 levels (8 bits per pixel), and σ_s^2 is the variance of the original picture. Hence, SNR_p is often referred to as the *peak-to-peak SNR* in contrast with SNR_v, which may be called *variance SNR*.

MSE and SNR are both simple mathematical measures which reflect the distortion in a picture. However, they do not reflect the frequency distribution of the errors and since the response of the human visual system is frequency dependent, they cannot reflect human quality perception. To compensate for this, the MSE can also be calculated in the Fourier domain [50] by virtue of the Parseval theorem. Then, a weighting of the errors in the frequency domain can easily be applied:

$$WMSE = \frac{1}{N_r * N_c} \sum_{u=0}^{N_c-1} \sum_{y=0}^{N_r-1} (w(u,v)(P(u,v) - \hat{P}(u,v)))^2, \tag{3.11}$$

where $w(u,v)$ denote the weighting coefficients according to the CSF of the human eye and $P(u,v)$ are the Fourier transform coefficients of the picture for frequency (u,v). When the calculation is not performed in the Fourier but in the DCT domain, the WMSE can be calculated per block. Then, $P(u,v)$ is replaced by the DCT coefficients $c(u,v)$, and $w(u,v)$ are the visibility thresholds for the individual DCT coefficients. The WMSE of a picture is derived by averaging over the blocks.

The frequency response is only one of the characteristics that should be taken into account. Unfortunately, only global characteristics of the HVS are incorporated in

an analytically tractable measure like the MSE. Metrics that include the masking effect of the human visual system are much more complicated and can be subdivided into two classes: metrics based on the human visual system (HVS) and metrics based on visual artefacts [51].

Metrics Based on Modelling of the HVS

To estimate picture quality assessment by a human observer, the HVS can be modelled [52, 53, 10]. First, the pictures are split into similar frequency orientation bands as in the HVS. Then, the contrast is calculated from the pixel values in these frequency bands. The contrast can be very easily defined as Weber contrast [40] for stimuli that are symmetric relative to the background luminance:

$$C_W = \frac{\Delta L}{L}. \tag{3.12}$$

Here ΔL is the luminance difference in the pattern and L is the background luminance. For practical pictures, however, this definition cannot be used, because the background luminance cannot be easily defined. In this case, a modified version of the definition can be used, where a low-pass version of the picture serves as the background luminance [54].

In [10], the resulting Local Band-limited Contrast (LBC) pictures are normalized to the visibility threshold. When the LBC in pixel (x, y) equals one, that particular frequency and orientation is just visible in that pixel, taking into account light and frequency sensitivity. Other models of the HVS found in literature use a similar strategy, expressing the signals in just noticeable differences ($JNDs$) [55]. Mostly, edge and texture masking are incorporated by adjusting the visibility threshold, according to the original picture. The visibility of the distortion is then measured in each pixel of each frequency band. In [10], it is measured by the masked LBC difference $\Delta MLBC_{k,l}(x, y)$:

$$JND_{k,l}(x, y) = \Delta MLBC_{k,l}(x, y) = \frac{LBC_{k,l}(x, y) - LBC^*_{k,l}(x, y)}{TE_{k,l}(x, y)} \tag{3.13}$$

where $LBC_{k,l}(x, y)$ and $LBC^*_{k,l}(x, y)$ are the LBCs of the original and the distorted pictures, respectively. $TE_{k,l}(x, y)$ is a threshold elevation function as shown in Figure 3.9, which increases as the envelope of the LBC of the original signal exceeds 1, indicating a just noticeable fluctuation in the frequency orientation band under consideration.

At this stage, the visibility of the difference between the original and the distorted picture is measured for each pixel in each frequency orientation band. To combine these local measures into a global measure of the perceptual quality for the whole picture is the most difficult part of the model. A popular way to do this is by

assuming a pooling of errors by means of the β-norm or Minkowski metric [56]. In this metric, the visibility of errors is pooled by probability summation in the following way:

$$PEM = \left(\sum_{x,y,k,l} |JND_{k,l}(x,y)|^{\beta} \right)^{\frac{1}{\beta}} . \tag{3.14}$$

The exponent β determines the degree of pooling. If $\beta = 1$, the pooling is a linear summation of absolute values. When $\beta = 2$, the errors combine quadratically, as in the MSE. When $\beta = \infty$, the pooling rule becomes a maximum norm: only the largest error matters. Pooling can be performed over the frequency components in a picture to find an expression for the visibility of distortion in a pixel, and over the pixels in a picture to find a global measure for the quality of that picture. In both cases, a β of about 4.0 has been observed [56].

A more general way to pool the distortions is proposed in [10], and is based upon a vector norm:

$$PEM = \left(\sum_{x,y} \left| \sum_{k,l} |JND_{k,l}(x,y)|^{\alpha} \right|^{\beta} \right)^{\gamma} . \tag{3.15}$$

Here α, β and γ are constants that influence how the different responses combine in the quality metric. They can be derived by maximizing the correlation between PEM and results from subjective tests. In [10] such tests resulted in the values $\alpha = 1.5$, $\beta = 1.0$ and $\gamma = 0.33$.

Metrics Based on Visual Artefacts

Another approach to find an expression on the quality of distorted pictures is by identifying typical artefacts in compressed video. These artefacts may include false contours, blocking, blurring, lost motion energy and added motion. Based on subjective tests, a linear or non-linear combination of these artefacts forms the quality metric, which is then defined on the five-point scale used in subjective tests [65].

A pioneer in this area is Miyahara [9] with a method that extracts five features from which a simple model is computed, tuned to maximize correlation with subjective tests. The features reflected random errors with perceptual weighting, block disturbances and structured errors caused by the picture contents, such as ringing. In [57] a metric was developed based on three features, only one of which reflects the spatial distortion. The other features reflect changes in motion energy, namely, picture skipping and jerky motion. Instead of using a linear combination of the extracted features, in [58] a non-linear back-propagation neural network is used to combine the different features in a quality metric.

In [51], it is shown that the quality metric described in [57] is not suitable for evaluation of MPEG-2 encoded video, because of the design for low bit rates and the lack of capturing DCT coding artefacts as blocking and mosquito noise. Consequently, if a quality metric based on visual artefacts is used, it should be adapted to the coding techniques used in the compression algorithm.

Discussion

In our philosophy, distortion is assessed locally and a local artefact is either visible or not visible. The expression of the distortion relative to the just noticeable difference is therefore very useful. The combination of local measures into a global measure, however, may lead to inconsistencies. When a large distortion is visible at a certain location, while the rest of the picture is not distorted, a global metric either averages over the picture, leading to a high picture quality, or looks at the maximum distortion, leading to a low picture quality. In neither case is the metric capable of identifying the problem. Nevertheless, for some purposes, as in comparing the results of two CBR coding schemes, a global metric is needed. The following section discusses the different techniques found in literature for constant-quality coding, either based on a global metric or on local visibility thresholds.

3.3 Constant-Quality Compression

The remaining question is that of how a compression algorithm can be designed to encode the pictures with a constant perceived quality. Moreover, the objective of the algorithm should be to minimize the bit rate while a certain quality is maintained. In this section, three methods of constant-quality compression are described, a feedback mechanism based on a global quality metric, a method controlling the maximum distortion in a coding coefficient based on the CSF of the human eye, and a method that adapts the coding algorithm locally to exploit the masking effects.

3.3.1 Feedback Quality Control

When the quality in constant-quality compression is determined by a global quality metric, this metric should be used in the encoding algorithm. The main problem is that the video will have to be encoded first, before the quality can be measured. This problem is similar to the problem in CBR compression, where the resulting bit rate is only known after encoding. To be able to use the quality metric, a feedback mechanism equivalent to the ones in most CBR schemes can be used [58].

In Figure 3.11, a feedback scheme to encode video streams at a constant quality q_{target} is shown. The quality $\hat{q}$ is measured for each encoded picture, and the difference $q_{target} - \hat{q}$ is fed back to adjust a global control parameter. In MPEG, it is mostly the quantizer scaling factor Q that is used. Where the metric proposed in [58] considers spatial quality only, in [59] a similar scheme is used with a metric that incorporates temporal degradation of the video as well, which causes extra complexity. Further, the number of past pictures that are incorporated in the quality estimation of the current picture has to be determined very carefully. Too few pictures may cause an over-reaction and too many an instability of the algorithm. The scheme becomes even more complicated when I, P, and B pictures are considered in MPEG [60].

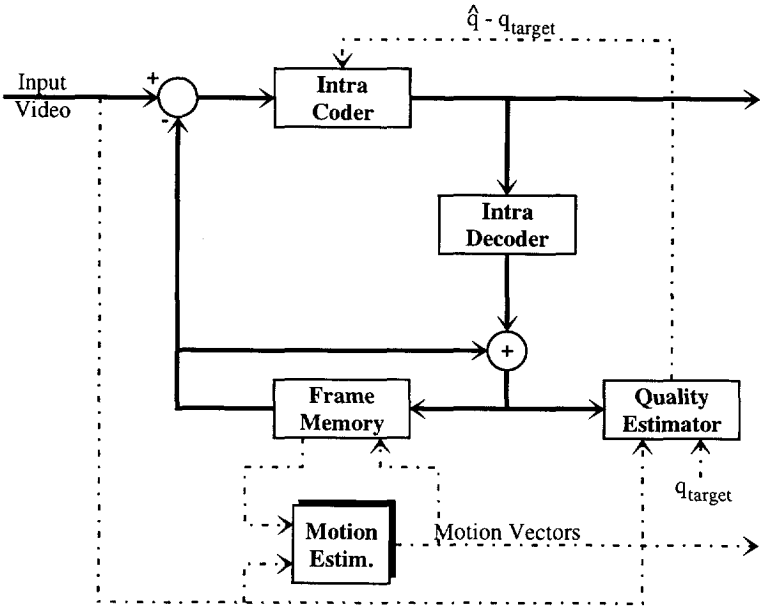


FIGURE 3.11: BLOCK DIAGRAM OF A FEEDBACK CONSTANT-QUALITY ENCODER.

A feedback quality control scheme suffers from the same drawback as a feedback rate control scheme: although the target is reached on average, no guarantee can be given about the quality or rate of a single picture. Especially at scene changes, the algorithm needs several pictures to stabilize the quality around the target. Further, the algorithm uses a global quality metric, while quality is assessed locally.

3.3.2 Maximum Coefficient Distortion

When we consider that quality is assessed locally, constant quality can be defined as perceptually lossless under specific viewing conditions. Dependent on the compression algorithm, a maximum distortion to be allowed in each coding unit can be derived from these viewing conditions. When, for instance, spatial DPCM is used, the maximum slope in intensity is fixed by the quantization, and it should match the maximum slope-overload distortion derived from the viewing conditions. In PCM, the visibility of false contours should determine the number of representation levels of the quantizer. When DPCM and PCM are used in sub-band coding, the quantizers can be adjusted for each sub-band, based on the CSF of the human visual system. In MPEG, we can derive the maximum distortion to be allowed for each DCT coefficient. By choosing the global parameters of an MPEG coder properly, a distortion lower than the maximum can be guaranteed for each DCT coefficient, which leads to perceptually lossless compression under the assumption that the visibilities of distortion from different coefficients are independent.

The quantizer scaling factor and weighting matrices control the coarseness of the quantization in MPEG. The maximum distortion introduced by a quantizer depends on the quantizer step size, derived from the scaling factor and weighting matrices, and on the quantizer characteristic. For I pictures, a uniform quantizer characteristic as described in (3.1) and (3.2) is used, yielding a maximum distortion equal to the step size:

$$D^c_{max}(u,v) = S(u,v). \tag{3.16}$$

For P and B pictures, the motion compensated prediction has to be taken into account. Since this temporal DPCM loop can be performed in the DCT domain [66], the difference between an original coefficient $c(u,v)$ and its reconstruction $\hat{c}(u,v)$ depends on the difference in the prediction error $c_e(u,v)$ and its reconstruction $\hat{c}_e(u,v)$, and not on the prediction $c_p(u,v)$:

$$\left.\begin{aligned} c(u,v) &= c_p(u,v) + c_e(u,v) \\ \hat{c}(u,v) &= c_p(u,v) + \hat{c}_e(u,v) \end{aligned}\right\} c(u,v) - \hat{c}(u,v) = c_e(u,v) - \hat{c}_e(u,v). \tag{3.17}$$

However, since the quantizer in P and B pictures has a dead zone, the maximum distortion of the quantized prediction error depends on the size of this error, and thus on the prediction itself. When the decision levels are chosen as in (3.5), the maximum distortion occurs at the edge of the dead zone:

$$D^c_{max}(u,v) = 2S(u,v). \tag{3.18}$$

Consequently, we see that to ensure a maximum distortion for a DCT coefficient, a finer quantization is needed for the DCT coefficients in P and B pictures than for those in I pictures. However, since only the reconstruction levels in MPEG are

fixed, the maximum distortion can be reduced by choosing the decision interval for reconstruction level 0 to be as large as the decision interval for reconstruction level 1:

$$-1.5S < I_0 < 1.5S, \tag{3.19}$$

which reduces the maximum distortion to:

$$D^c_{max}(u,v) = 1.5S(u,v). \tag{3.20}$$

As mentioned in Section 3.2.2, the visibility threshold of a quantization error is determined for each DCT coefficient for various viewing conditions [48, 49]. When we choose the maximum distortion for each DCT coefficient to be equal to its corresponding threshold for the specific viewing conditions by means of the weighting matrix and quantizer step size, the individual distortions should not be visible according to the model. Since some DCT coefficients occur in the same band of the HVS, there is a small probability that a combination of distortions will be visible. Usually, this chance is neglected [45].

Ensuring a maximum distortion below the visibility threshold by means of the global control parameters allows perceptually lossless compression under the chosen viewing conditions. However, not all properties of the human visual system are accounted for and, thus, compression is not optimized. The algorithm can be improved by using the local parameters to exploit the masking effects.

3.3.3 Local Adaptive Coding

The generation of output video with constant quality is only one of the objectives in VBR compression. The other objective is to optimize compression with respect to that quality. To do this, the masking effects mentioned in Section 3.2.1 and Section 3.2.2 can be taken into account to optimize the compression algorithm locally. This is only possible when the compression algorithm divides the picture, or the frequency bands of the picture, into local parts prior to coding. Then, the local parameters in each part of the picture depend on its contents.

Since MPEG divides the pictures into blocks before encoding, there are several ways to optimize the compression algorithm according to local scene contents. It is, for instance, possible to change the quantizer scaling factor Q per macroblock, although this yields an extra overhead, since the value of Q will have to be transmitted each time it changes. Another possibility is to use thresholding, changing the decision level I_0 for each DCT coefficient, or defining a maximum distortion for the whole block according to the masking elements present.

One of the masking effects that can be used is that of the local luminance. It is relatively easy to use this effect by letting the quantizer step size depend not only

on the frequency of the DCT coefficient, but also on the DC level of the block. Since this is not implicitly standardized in MPEG, it is only possible to include it explicitly per macroblock.

Edge masking cannot be exploited in a DCT coding scheme because of the annoying artefacts around objects in front of a homogeneous background. Texture masking, however, can be exploited. Based on these observations, the blocks in a picture can be classified into three classes [64]: "flat", "textured", and "edge-containing" blocks. The allowable distortion depends on the class of the block and the background luminance.

Temporal masking can be exploited by allowing a coarser quantization at scene changes. This is particularly interesting since at scene changes no reference to an earlier, resembling transmitted picture can be made. Therefore, either a peak in the bit rate may occur in VBR sources or a drop in quality in CBR sources. The latter is often masked by temporal masking and, thus, a drop in quality can also be applied in VBR sources to prevent the peak.

3.4 Reference Codec

As described in the previous section, a number of techniques can be applied in constant-quality VBR MPEG compression. In this section, our reference codec is described. The bit rates resulting from this codec are used in the next chapter. Although we do not target the ultimate constant-quality codec, a combination of the techniques described in Section 3.3.2 and Section 3.3.3 is used in an attempt to optimize compression. The strategy to realise our reference codec is shown in Figure 3.12. In succession, the choices for GOP structure, control parameters, global parameters, and local adaptation are described in the following sections.

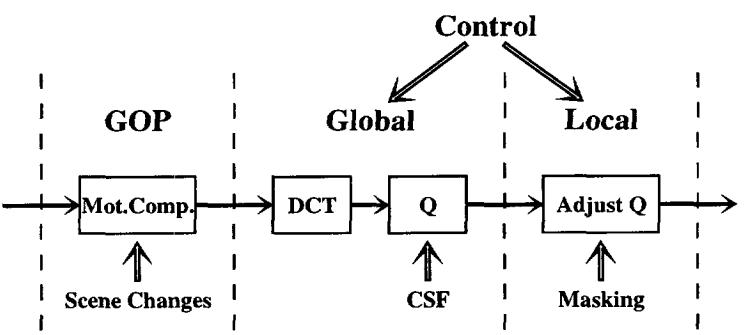


FIGURE 3.12: STRATEGY USED IN THE REFERENCE VBR MPEG CODEC.

3.4.1 GOP Structure

First, an appropriate choice of GOP structure will have to be made. For optimal compression, the best choice would include many B pictures and a long GOP size in order to limit the amount of expensive I pictures. There are, however, also other considerations, like the random accessibility of the bit stream and the limited delay required in real-time transmission of video. Hence, a closed GOP containing 8 B and 4 P pictures is chosen, which is common for real-time applications.

At the occurrence of a scene change, a prediction of the first picture of the new scene is not possible. Hence, when this picture is coded as a P or B picture, compression is poor. The encoding algorithm will have to position the beginning of a GOP on such a scene change. There are two ways to do this. First, in [67], a simple measure of the spatio-temporal correlation in each picture is calculated. This parameter is calculated before the encoding of each picture, by taking the average difference between adjacent pixels in the difference picture between the current picture and the previous picture. This measure can be used to predict the occurrence of a scene change. Second, another way to detect scene changes is based on pixel histograms [68]. In our codec, the first technique is used.

3.4.2 The Control Parameter

Before encoding can be performed, a criterion has to be defined as to which encoding is to be performed. In CBR coding, the criterion is to have a constant output bit rate. In VBR coding, a constant quality is desired. In our reference codec quality is defined as perceptually lossless for normal TV viewing conditions. With this definition, the global and local parameters are determined in the following sections.

An additional requirement in a practical VBR video coder is to be able to prevent a violation of the traffic parameters by means of a control algorithm. For such an algorithm, it is necessary to have one global parameter which can be changed to reduce the output bit rate. Consequently, the global and local parameters should be scaled according to this parameter.

Although it seems logical in a constant-quality coder to use the quality as the control parameter, this is not practical, since it is necessary to be able to change the control parameter during the encoding of a picture. Consequently, the only possible control parameter is the quantizer scaling factor Q. The global and local parameters as described in the following sections are therefore defined relative to Q.

3.4.3 Determination of Global Parameters

Before constant-quality encoding is possible, the desired quality has to be decided upon. It can be determined by the viewing conditions and application. Here, normal TV quality is considered, and we have deduced that, although a viewing distance of six times the screen height is normally assumed, the minimal viewing distance is four times the screen height. With this, we have derived the visibility threshold for each DCT coefficient from the CSF [45].

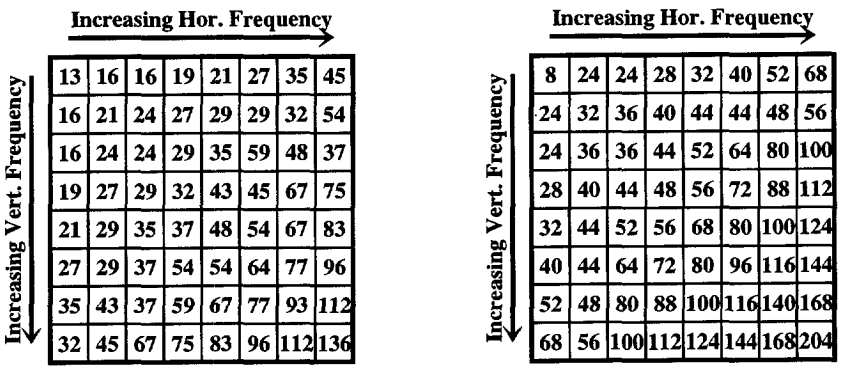


FIGURE 3.13: MPEG WEIGHTING MATRICES FOR A MAXIMUM DISTORTION BELOW THE VISIBILITY THRESHOLD FOR (A) I PICTURES AND (B) PREDICTED PICTURES.

To ensure a quantization with a maximum distortion lower than the visibility threshold, the quantizer scaling factor and quantization matrices can be used. When the scaling factor is fixed as $Q = 4$, the matrix for a uniform quantizer characteristic as in I pictures is shown in Figure 3.13a. Note that the quantization of the DC coefficient in I pictures cannot be changed in MPEG-1. For P and B pictures, a dead zone is standardized in MPEG, so that a finer quantization is needed to ensure a maximum distortion lower than the visibility threshold. In the reference codec, the decision level I_0 is adapted according to (3.19) to reduce this effect, the resulting weighting matrix is shown in Figure 3.13b.

In our experience, no artefacts are visible when the proposed quantizer scaling factor and weighting matrices are used for a viewing distance of four times the screen height. When this distance is decreased, and the viewing angle is thus increased, some *local* artefacts may become visible, as is illustrated in Figure 3.14.

(a) (b)

FIGURE 3.14: RESULTS OF CODING WITH PERCEPTUALLY LOSSLESS QUANTIZATION (A) NO ARTEFACTS ARE VISIBLE AT NORMAL VIEWING ANGLE (B) LOCAL ARTEFACT AT INCREASED VIEWING ANGLE.

3.4.4 Local Adaptation

As can be seen from Figure 3.14b, artefacts become visible only locally when the viewing angle is increased. This indicates that masking effects allow more distortion in most parts of the picture and compression can thus be increased locally. To do this, first, a classification into "flat" "edge" and "textured" blocks is obtained. Second, the techniques to optimize compression are determined based on the classification.

Block Classification

The classification scheme is an extension of the one used in [64] and is shown in Figure 3.15.

As a first step in block classification, the local activity is measured for each pixel in the picture, using changes of the luminance values in the neighbourhood of that pixel. Four operator matrices, which are shown in Figure 3.16, are used to measure structure in the horizontal, vertical, and both diagonal directions. The local activity of a pixel at position (x, y) is defined as:

$$A_{x,y} = \max_{k=1,\ldots,4}(|A_{x,y}(k)|), \tag{3.21}$$

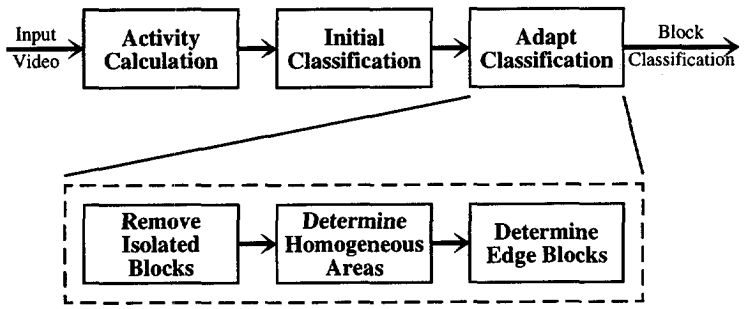


FIGURE 3.15: ALGORITHM TO CLASSIFY THE BLOCKS IN A PICTURE.

0	0	0	0	0
0	1	0	-1	0
0	1	0	-1	0
0	1	0	-1	0
0	0	0	0	0

0	0	1	0	0
0	1	0	0	0
1	0	0	0	-1
0	0	0	-1	0
0	0	-1	0	0

0	0	0	0	0
0	1	1	1	0
0	0	0	0	0
0	-1	-1	-1	0
0	0	0	0	0

0	0	1	0	0
0	0	0	1	0
-1	0	0	0	1
0	-1	0	0	0
0	0	-1	0	0

FIGURE 3.16: OPERATOR MATRICES TO DETERMINE THE ACTIVITY OF A PIXEL IN FOUR DIRECTIONS.

with

$$A_{x,y}(k) = \sum_{x=1}^{5}\sum_{y=1}^{5} p(x-3+i, y-3+j), m_{i,j}(k), \tag{3.22}$$

where $p(x, y)$ denotes the luminance value at position (x, y) and $m_{i,j}(k)$ the elements of the matrix $m(k)$. As in [64], pixels with an activity lower than 32 are classified as "flat", lower than 64 as "low textured", lower than 128 as "high textured" and the remaining ones as "edges". The results of this classification for the picture *soccer* can be seen in Figure 3.18. In this figure, "flat" pixels are white, "edge" pixels are black and "low" and "high structured" pixels are light and dark grey, respectively. The original picture is shown in Figure 3.17.

After calculation of the activity per pixel, an 8x8 block is considered "flat" if it contains "flat" pixels only, it is considered "edge-containing" when it contains both "edge" and "flat" pixels and the rest is considered to be "textured". We have added one more class to distinguish "low textured" from "high textured" regions, which contain "edge" pixels. This initial block classification is shown in Figure 3.19 for the picture *soccer*.

FIGURE 3.17: ORIGINAL PICTURE *soccer*.

After the classification based on the activity per pixel, there are "edge-containing" blocks and isolated "flat" blocks within "textured" regions. Since these are surrounded by "textured" blocks, texture masking prevents artefacts from being visible in these blocks as well as in the "textured" blocks themselves. Hence, "flat" and "edge-containing" blocks are re-classified as "textured" blocks when none of their neighbours are "flat".

When there are no isolated "flat" blocks left, the picture is, in fact, divided into "flat" and "textured" regions, separated by edges. "low textured" blocks, however, do not contain "edge" pixels and should therefore be part of one of the other regions. A recursive region-growing algorithm is applied to the "flat" blocks to extend the "flat" regions with the "low textured" blocks connected to it. Then, the last step in Figure 3.15 is to determine the "edge" blocks, which surround the "textured" regions. Figure 3.20 shows the final classification of the blocks in the picture *soccer*.

Techniques to Increase Compression

With the resulting classification, compression can be increased based on the difference in the visibility of distortion in the different classes of blocks. There are two methods to do this: first, thresholding can be applied to remove small coefficients. Second, a coarser quantization can be applied by using a larger quantizer scaling factor Q. For thresholding, a maximum distortion needs to be defined to determine which coefficients are to be thresholded. This maximum can be defined

FIGURE 3.18: THE ACTIVITY PER PIXEL.

either per DCT coefficient, assuming that the distortions in different coefficients are independent, or per block, assuming that a pooling of the errors occurs.

In the reference codec, we aim at thresholding only a few small coefficients per block. In "textured" blocks, more coefficients could be thresholded, based on the masking effects, but in that case it is more efficient to use a coarser quantization. We determine a maximum distortion per block, assuming error pooling. The maximum distortion needs to be chosen such that only a few coefficients can be thresholded. To do this, we consider the distortion that is typically introduced by quantization and thresholding, respectively.

Since we would like to take the CSF of the human visual system into account, and since the distortion added by thresholding a DCT coefficient should be considered, the WMSE in the DCT domain is used, with the weighting matrix shown in Figure 3.13a. Consequently, the maximum WMSE introduced by uniform quantization with the same matrix and quantizer scaling factor Q is given by:

$$WMSE_{max}^{quant} = Q^2. \tag{3.23}$$

To be able to determine a maximum value for the WMSE per block in each class, we should consider the statistical distribution of DCT coefficients, which is peaked around zero. In decision intervals other than I_0, however, it is approximately uniform, and the mean square of the quantization error thus equals $\frac{Q^2}{3}$. Hence, the

FIGURE 3.19: INITIAL CLASSIFICATION BASED ON ACTIVITY.

WMSE in a block after quantization depends on the number of coefficients quantized to zero [45] and a typical expression is given by:

$$WMSE_{typ}^{quant} = \frac{(64 - Z)}{64} \frac{Q^2}{3}, \tag{3.24}$$

where Z is the number of coefficients in a block that are quantized to zero. Thresholding a small coefficient (i.e. that was quantized to level 1) introduces a distortion that is typically Q^2 larger and thus the WMSE after thresholding can be expressed as:

$$WMSE_{typ}^{thres} = \frac{1}{64} \left((64 - Z) \frac{Q^2}{3} + TQ^2 \right), \tag{3.25}$$

where T is the number of thresholded coefficients. From this expression and considering how many coefficients in "flat", "textured", and "edge-containing" blocks are quantized to zero and how many are allowed to be thresholded, the maximum distortion can be derived for each class.

Obviously, the WMSE is much lower in "flat" blocks than in "textured" and "edge-containing" blocks, since only a small number of coefficients (here we assume $64 - Z \leq 5$) are not quantized to zero. Further, we assume that only one or two small coefficients in "flat" blocks can be thresholded. Hence, the maximum distortion in "flat" blocks is determined by:

$$WMSE_{max}^{flat} = \frac{1}{64} \left(\frac{5}{3} Q^2 + (1 + \frac{(lum - 128)^2}{128^2}) Q^2 \right), \tag{3.26}$$

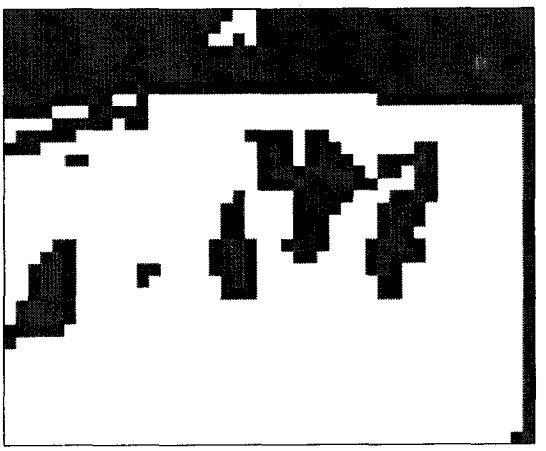


FIGURE 3.20: FINAL BLOCK CLASSIFICATION.

where *lum* is the mean grey value in the block, so that approximately one coefficient may be thresholded in mid-grey luminance and two in very dark and bright areas, where luminance masking occurs.

Similarly, for "edge-containing" blocks only one coefficient may be thresholded when the background is dark or bright, and since the number of non-zero coefficients is higher (here we assume $64 - Z \leq 15$) the maximum WMSE becomes:

$$WMSE_{max}^{edge} = \frac{1}{64}\left(\frac{15}{3}Q^2 + \frac{(lum - 128)^2}{128^2}Q^2\right), \tag{3.27}$$

where *lum* is the background luminance, measured as the mean grey value of the "flat" pixels in the block.

Finally, in "textured" blocks much more distortion is allowed. Since there are relatively many coefficients in these blocks with amplitude larger than twice the quantizer scaling factor, thresholding alone will not optimize compression. Hence, we propose to increase Q in these blocks:

$$Q_{texture} = Q + 2. \tag{3.28}$$

Further, to allow the thresholding of maximally 4 small coefficients, the maximum WMSE is chosen as:

$$WMSE_{max}^{texture} = \frac{1}{64}\left(\frac{15}{3}Q_{texture}^2 + 4Q_{texture}^2\right). \tag{3.29}$$

Discussion

In the local adaptation techniques described above, texture and luminance masking are taken into account. Temporal masking is not considered, since the occurrence of a peak at scene changes is very typical for VBR sources and should therefore be considered in the following chapters.

3.4.5 Results

The reference codec as described in this section is designed for perceptually lossless compression at specific viewing conditions. By using the proposed weighting matrices and quantizer scaling factor in MPEG-1, a maximum distortion for each DCT coefficient can be guaranteed, which is sufficient for this objective. The local adaptation is used in order to optimize compression by making use of masking effects. The gain in compression performance can easily be calculated by comparing the bit rates generated by the codec with and without local adaptation.

In addition to the results with and without local adaptation, we also present the effects of adapting the decision level I_0 in the quantizer characteristic of P and B pictures. As stated before, it is our opinion that no dead zone should be used in this characteristic, but since the MPEG standard includes this, the adaptation is made. The results presented in Table 3.1 show that for our codec the standard dead zone would cause an increase in bit rate of 15% with respect to a uniform quantizer characteristic (compare "Uniform" with "Standard MPEG"). This is caused by our objective to guarantee a maximum distortion for the DCT coefficients. To do this, the quantization will have to be finer with a dead zone than without, which causes the extra bits. By adapting the decision level I_0, the loss in efficiency is reduced to 10% (compare "Uniform" with "Adapted").

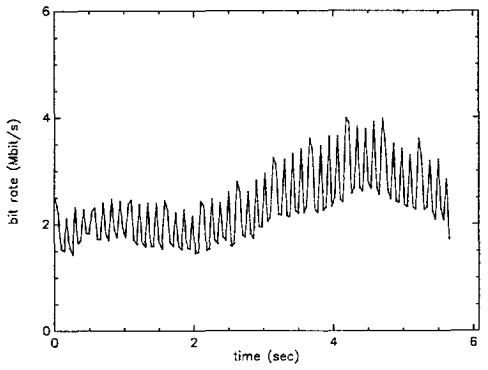
TABLE 3.1: MEAN BIT RATES IN MBIT/S IN THE SEQUENCE *car* FOR DIFFERENT QUANTIZERS IN P AND B PICTURES AND FOR DIFFERENT LOCAL ADAPTATIONS.

Quantizer Characteristic	Local Adaptation		
	No	Thres	Thres + Q
Uniform	2.677	2.490	2.135
Standard MPEG	3.092	3.046	2.470
Adapted	2.927	2.845	2.345

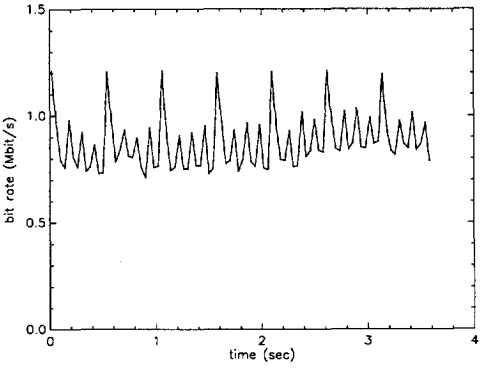
Our proposed local adaptation technique uses both thresholding and a coarser quantization in "textured" blocks. Thresholding is very effective for small coefficients,

but large ones are still quantized with a fine step size. When there are many of those coefficients, as is the case in "textured" blocks, the gain in bit rate is larger when using a coarser quantization than the gain when using thresholding only. Table 3.1 shows this by comparing the results when using no local adaptation ("No") with the results when using thresholding ("Thres"), and the results when using the proposed combination of thresholding and a coarser quantization in "textured" blocks ("Thres + Q"). The total gain of the local adaptation is as large as 20%, which is considerable.

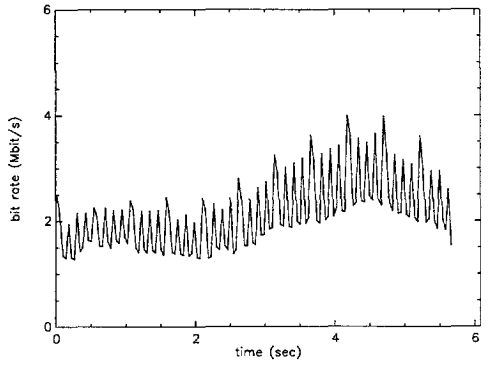
Figure 3.21 shows some bit rate curves to evaluate the bit rates generated by a VBR MPEG encoder and the effects of the different adaptations on the variability of the bit stream. From the periodic structure in the curves, the different compression capabilities for I, P, and B pictures can be deduced. The use of a dead zone in P and B pictures in MPEG instead of a uniform quantizer affects the bit rates in these pictures negatively, which can be seen by comparing curves (a) and (c). From comparing the curves from the *car* sequence with those from the *miss* sequence, we conclude that the local adaptation reduces the bit rate, especially in highly detailed scenes. This can be explained by considering that less detail yields less masking effects which can be exploited by the local adaptation. Nevertheless, the variability caused by the different activity and detail in different scenes remains unchanged, and is representative for a VBR video source.



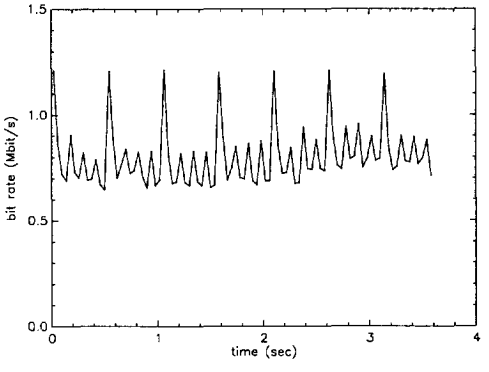
(a) Sequence *car*, reference codec



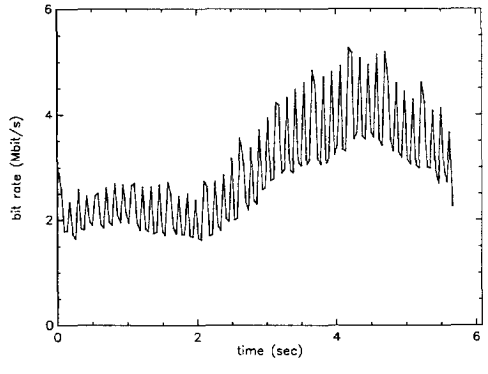
(b) Sequence *miss*, ref. codec



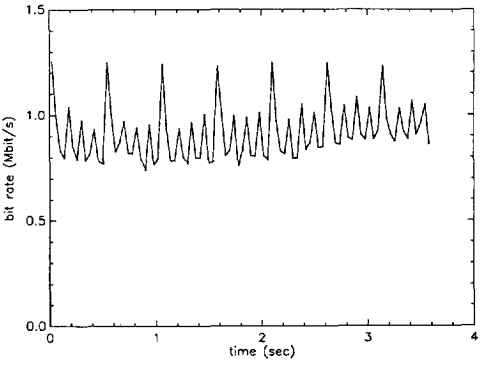
(c) *car*, uniform quantization



(d) *miss*, uniform quantization



(e) *car*, no local adaptation



(f) *miss*, no adaptation

FIGURE 3.21: BIT RATE CURVES FOR SEQUENCES *car* AND *miss*.

Chapter 4

Network Adaptation for VBR Video

The transport layer in the OSI model provides the interface between the application-specific and the medium-specific layers. Its tasks depend on both the application and the type of medium. When a large number of services need to be considered, as in the ATM network, several different implementations of the transport layer, or ATM Adaptation Layer (AAL), can be identified [18].

For real-time services like video and audio, the main tasks of the transport layer are the provision of a common reference clock and the protection of the data from network-specific errors. In synchronous networks, a clock is shared among the network and the encoders and decoders, which enables them to synchronize with each other. In asynchronous networks, the clock will have to be explicitly recovered at the receiver side by time stamps provided in the transport layer, or implicitly, by using the decoder buffer filling level [17, 69]. Network-specific errors, like cell loss in ATM, can be handled by using error-correcting codes and cell interleaving in the transport layer [69].

In the case of VBR video, another part of the adaptation concerns the bit rate characteristics. If the medium-specific layers accept VBR sources, these characteristics will have to be specified in advance. They depend, however, on how the bit stream is offered to the network, i.e. the network adaptation. If the medium-specific layers accept constant bit rate sources only, the transport layer has to convert the VBR source(s) to a CBR source using stuffing and multiplexing. For an efficient allocation of bandwidth, the bit rate characteristics will have to be specified there as well. This chapter focuses on these characteristics, before multiplexing is discussed in Chapter 5.

In networks that support variable bit rate services, like ATM, the amount of data to be sent by each user is not unlimited. As discussed in Chapter 2, a contract with the network is negotiated in which the user specifies the traffic characteristics, such as the peak cell rate, the mean cell rate, and the burstiness, indicating the variability of the cell rate. In the same contract, the network specifies the quality of service parameters such as cell loss rate, delay and delay jitter introduced by multiplexing and switching in the network. The number of sources to be admitted by the network and the cost to the user will depend on the specified traffic characteristics. To determine the traffic characteristics of a VBR source, Section 4.1 discusses the description of traffic in ATM, focusing on the traffic as generated by our reference VBR MPEG coder, described in Chapter 3. Section 4.2 describes some smoothing techniques which aim at improving the traffic characteristics in order to minimize the cost of transmission. Finally, Section 4.3 discusses how the traffic parameters which are defined in the network contract can be controlled.

4.1 Traffic Description

A description of the traffic on ATM networks is needed for two reasons. First, it is necessary to predict the network load in order to determine the probabilities of network failure. Second, the traffic generated by each source should be monitored in order to prevent network failure. For the first purpose, a model of broadband traffic is used and discussed in Section 4.1.1. Then, Section 4.1.2 describes the interpretation of the model in the case of VBR MPEG. Finally, the resulting traffic parameters are discussed in Section 4.1.3.

4.1.1 A Model of Broadband Traffic

The description of broadband traffic in general is divided into three time scales [70, 71], which are depicted in Figure 4.1. The durations displayed in this figure are only a rough indication, showing that the measures of duration differ one or more orders of magnitude. At *connection* level, the layer above the AAL (session layer in OSI terminology) demands the establishment and termination of connections. Here, the line being high indicates that a connection is made. The activities at *burst* level can only occur when this is the case. The AAL transmits a burst of traffic when data is available. The transmission of cells within a burst is indicated by the *cell* level. The spacing of the different cells depends on the implementation of the AAL.

different types of network failure can be identified with this model. First, at connection level, call blocking may occur when the call admission control decides not to allow more sources on the network. At burst level, blocking occurs when the aggregate instantaneous cell rates of all sources transmitting a burst exceeds the capacity

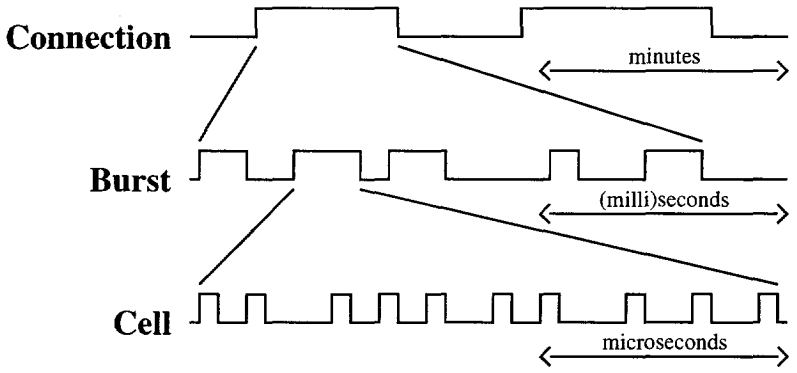


FIGURE 4.1: TIME SCALES IN BROADBAND TRAFFIC.

of the link. Finally, at cell level, blocking may occur due to the limited length of the queues in network nodes. Evidently, these blocking probabilities influence each other. For instance, if the call admission control is very strict, less blocking will occur at the burst level at the cost of more blocking at the connection level.

4.1.2 Interpretation of VBR MPEG

In a video encoder, the compressed picture data becomes available at fixed moments in time. Dependent on the implementation of a real-time encoder, these moments are at the beginning of a new picture or part of a picture. When we take a look at the MPEG standard, these parts may be slices or macroblocks. The model shown in Figure 4.1 can therefore be interpreted by assuming that the burst level in that model coincides with one of the layers in the MPEG hierarchy. The AAL will then determine how the cells carrying the compressed video of that layer are distributed over the burst interval. If the AAL uses *bufferless packetizing* [69], which means that the ATM cells are launched onto the network immediately after they are filled, the characteristics of the traffic in a burst are determined by the hardware implementation of the MPEG encoder. Here, we assume that the cells are buffered over a layer in the MPEG hierarchy before they are transmitted.

In general, the transmission of cells in a burst can be either at the speed of the output link, the peak cell rate as agreed upon in the contract with the network, or uniformly over the burst interval. The latter solution is possible since it is known in advance when the following burst of data will arrive. The rate of transmission in this case is calculated by

$$R_{out} = \frac{\#cells}{T_{burst}}. \tag{4.1}$$

where T_{burst} is the duration of a burst, equal to the time between the processing of two successive units in a layer of MPEG. The three methods are shown in Figure 4.2.

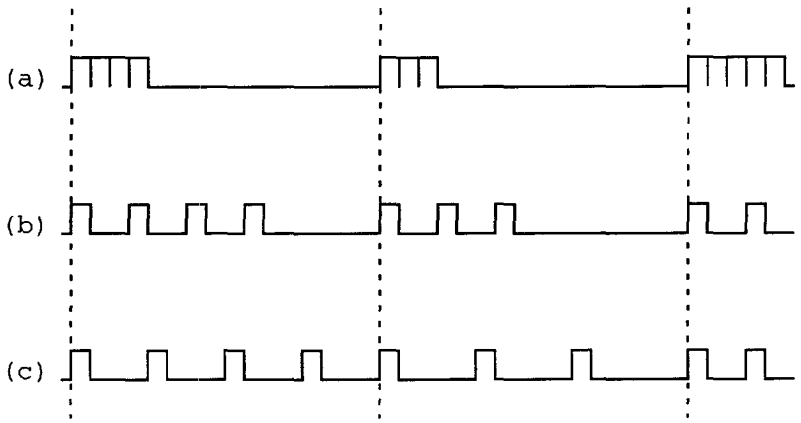


FIGURE 4.2: TRANSMISSION OF CELLS (A) AT OUTPUT LINK RATE, (B) AT PEAK CELL RATE (C) UNIFORMLY DISTRIBUTED.

It is clear that the uniform distribution leads to the largest cell inter-arrival times and thus smallest instantaneous cell rates. Further, with a uniform distribution at cell level, the characteristics of the VBR source are determined by the burst cell rates $y(k)$, while the bursts are not separated by periods where no cells are transmitted. From these cell rates, the MPEG structure is visible, dependent on which layer is chosen to coincide with the burst level in the transmission model. Figure 4.3 shows the characteristics for the slice and picture layer.

Evidently, buffering over higher layers in the MPEG hierarchy prior to a uniform distribution over the burst interval yields a smaller peak cell rate and burstiness at the costs of increased delay and larger buffers. The extreme case of distribution over the sequence layer means that the whole sequence is transmitted at a constant cell rate and stored at the receiver before playback. As a reference to the size of the buffers and delay in real-time video traffic, a comparison can be made with a constant bit rate MPEG coder. In this case, a buffer size of 10% of the bit rate is used in typical situations. At a picture rate of 25 Hz, this yields buffering over 2 to 3 pictures. Hence, the highest layer to coincide with the burst level is the picture layer, which we will assume from here on.

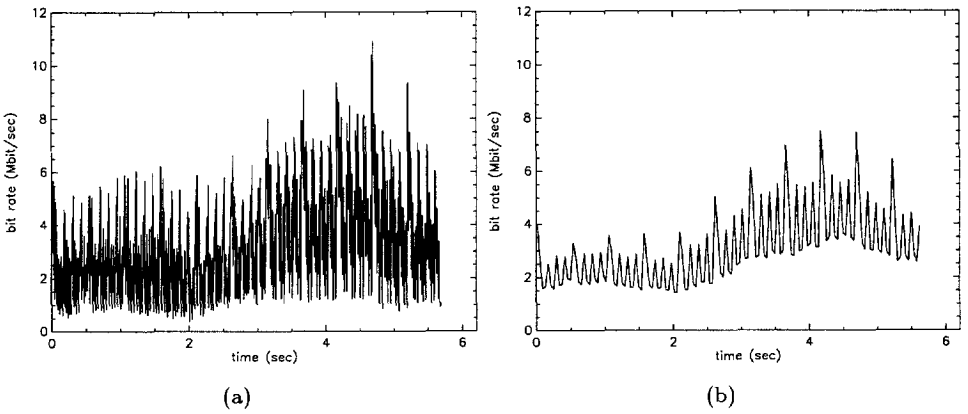


FIGURE 4.3: CELL RATES PER (A) SLICE AND (B) PICTURE IN CONSTANT-QUALITY MPEG.

4.1.3 Traffic Parameters

In order to predict the behaviour of statistical multiplexing in the network, some parameters concerning the characteristics of the traffic will have to be defined in the network contract. During the connection, the network checks whether the source behaves according to these parameters. Consequently, the algorithm that checks a parameter, in fact, determines its description. For the peak and mean cell rates, such an algorithm is proposed for standardization in ATM. For a good description of the traffic, however, a third parameter is needed to describe the cell rate fluctuations, also called burstiness. In the following, the proposed algorithms and definitions of peak and mean cell rate are described and some possibilities for defining a burstiness parameter are given.

Peak and Mean Cell Rates

The peak cell rate of a connection can be defined at cell level by the inverse of the minimum inter-arrival time T between two consecutive cells, also called peak emission time. Since the ATM network is a slotted transmission medium, the possible values of this peak cell rate are constrained to be the reciprocal of an integer number of cell slot times δ. Further, consecutive cells may suffer from a variation in the delay, caused by multiplexing in the network nodes. Hence, traffic conformance is defined in terms of a generic cell rate algorithm (GCRA), which takes the cell delay variation tolerance τ into account [72]. The GCRA depends on two parameters, the increment I, which should be chosen as the peak emission time T and the limit L, which defines the cell delay variation tolerance τ. There are two equivalent versions of the GCRA,

the virtual scheduling algorithm and the continuous-state leaky bucket algorithm. Since they are equivalent, we only describe the first.

The virtual scheduling algorithm updates a theoretical arrival time ($TAT(i)$), which is the arrival time $t_a(i-1)$ of the previous cell plus the increment I. If

$$t_a(i) \geq TAT(i) - L, \tag{4.2}$$

the cell is conforming, otherwise the cell is non-conforming. If

$$TAT(i) - L < t_a(i) < TAT(i), \tag{4.3}$$

then

$$TAT(i+1) = TAT(i) + I, \tag{4.4}$$

otherwise

$$TAT(i+1) = t_a(i) + I. \tag{4.5}$$

Figure 4.4 shows an example of the delay variation tolerance allowed by a GCRA with L small relative to I, which is typical for a peak rate control algorithm.

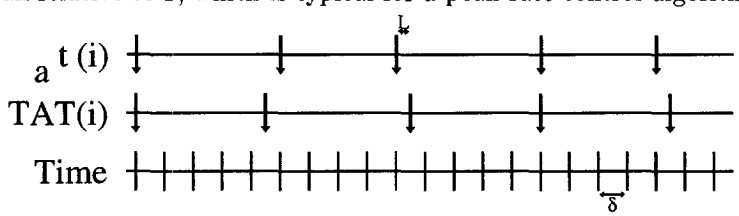


FIGURE 4.4: INFLUENCE OF THE GCRA ON THE POSSIBLE CELL ARRIVALS, EXAMPLE FOR A PEAK CELL RATE CONTROL ($I = 4.5\delta, L = 0.5\delta$).

The mean, or average, cell rate of a connection can be defined at connection level as the total number of transmitted cells divided by the duration of the connection. However, it is only possible for the network to check the conformance of a source to such a mean cell rate at the end of the connection. To enable the network to describe and monitor the future cell flow more accurately, a sustainable cell rate (or its inverse, the sustainable emission time T_s) is defined together with a burst tolerance parameter τ_s. Again, the conformance to these parameters can be checked with a GCRA algorithm as described above. In this case, however, $L = \tau_s$ is large relative to $I = T_s$, as shown in Figure 4.5.

If the peak cell rate of a connection is determined by a GCRA algorithm with $I = T$ and $L = 0$ and the sustainable cell rate with a GCRA algorithm with $I = T_s$ and $L = \tau_s$, a maximum burst size MBS can be determined when transmitting at peak cell rate [72]:

$$MBS = \left\lfloor 1 + \frac{\tau_s}{T_s - T} \right\rfloor, \tag{4.6}$$

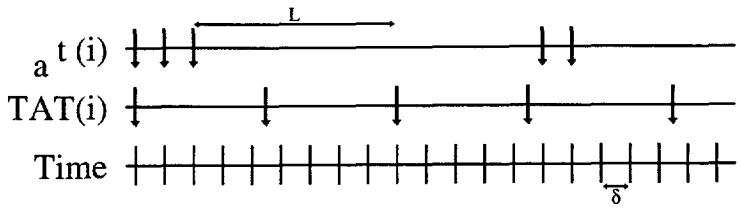


FIGURE 4.5: INFLUENCE OF THE GCRA ON THE POSSIBLE CELL ARRIVALS, EXAMPLE FOR A SUSTAINABLE CELL RATE CONTROL ($I = 4.5\delta, L = 7\delta$).

where $\lfloor x \rfloor$ stands for the integer part of x. In the interpretation of VBR MPEG, the duration of a burst is a picture period and, thus, with $MBS = \frac{R_{link}}{25}$, where R_{link} is the cell rate of the ATM link, we find the requirement on the burst tolerance τ_s from (4.6):

$$\tau_s = (\frac{R_{link}}{25} - 1)(T_s - T). \tag{4.7}$$

This expression, however, only holds when the sustainable cell rate is chosen as the mean cell rate in high detailed scenes. As an example, consider the transmission of the source shown in Figure 4.2b on an ATM link with $R_{link} = 20,000 cells/s$. Then, $T = 1$ and $T_s = 2$, corresponding to a peak cell rate of $20,000 cells/s$ ($\approx 8 Mbit/s$) and a sustainable cell rate of $10,000 cells/s$ ($\approx 4 Mbit/s$), and $\tau_s = 800$. In this concept, however, no advantage is taken of the low bit rate of VBR sources in less detailed scenes. Then, the sustainable cell rate should be chosen as the mean cell rate over multiple scenes, for instance $T_s = 3$ and the maximum duration of a peak should be the maximum duration of the scene, for instance, 5 seconds or more, yielding for the burst tolerance $\tau_s > 32,000$.

A higher burst tolerance will make it more difficult for the network to predict future traffic and thus to allocate resources. Therefore, the costs will increase when the burst tolerance is higher. However, a lower burst tolerance forces the user to define a high sustainable cell rate to support its traffic, thus yielding an even worse allocation of resources and the temptation to use greedy cell production algorithms. Therefore, the costs of a burst tolerance high enough to define the sustainable cell rate in a VBR video source at the long-term mean cell rate should be much lower than the costs of a higher sustainable cell rate with a low burst tolerance.

In order to reduce the costs for a high burst tolerance, the network should be able to predict future traffic more accurately. Therefore, another traffic parameter is needed to reflect the fluctuations of the cell rates in time. Currently, there are no policing algorithms for such a parameter as yet, but there are some possibilities to describe the burstiness of VBR traffic.

Burstiness

The ratio between peak and mean cell rate indicates how *bursty* a source is. Therefore, this peak to average ratio (PAR) is often used as the burstiness of the source. However, measuring burstiness this way provides no information about the fluctuations in the cell rate of the source in time, which is of importance to the behaviour at multiplexing. Therefore, a burstiness measure should have three desirable properties [73]. First, it should not only yield statistical values of the signal, but also reflect the fluctuations in time. Second, the measure should be able to evaluate the statistical multiplexing effect. Third, the measure should allow easy modelling of VBR video sources.

According to the transmission model, the burstiness on both cell and burst level can be considered. On the cell level, it concerns the fluctuations in cell inter-arrival times within a burst. On the burst level, it concerns the occurrences and durations of bursts and the mean cell rate in a burst. If the cells within a burst are uniformly spread over the burst interval, and consecutive bursts are not separated by intervals in which no cells are transmitted, only the fluctuations in burst cell rates $y(k)$ need be considered. In the following, however, $y(k)$ can also be seen as the reciprocal of the inter-arrival time between cells k and $k-1$.

To provide more information about the statistical behaviour of the cell rates, the standard deviation to average ratio (SAR) of the cell rates is often used. To evaluate the buffering of an $n+1$ sequence of cell rates, the SAR can be extended to the coefficient of variation $CoV(n)$ [73]:

$$CoV(n) = \frac{\sqrt{(n+1)Var[y_b(n,m)]}}{E[y_b(n,m)]}, \tag{4.8}$$

where

$$y_b(n,m) = \sum_{k=m}^{n+m} y(k). \tag{4.9}$$

Although $CoV(n)$ describes the variation in a sequence, it does not represent the fluctuations in time. Therefore, the autocorrelation function $AC(n)$ is often used:

$$AC(n) = \frac{E[(y(k)-M)(y(k+n)-M)]}{Var[y(k)]}. \tag{4.10}$$

This function expresses the temporal behaviour, but it is difficult to quantify the burstiness from it.

Recently, some new burstiness parameters have been proposed which represent the degree of overload of the allocated bandwidth [74, 75]. For this purpose, a different definition of a burst is used, namely the interval in which the cell rate is greater than

the mean cell rate of the source. As opposed to the earlier definition of a burst, the inter-burst, defined as the period between two bursts, now does contain cells. From this definition of bursts and inter-bursts, the mean traffic intensity (MTI) [74] and the mean overload ratio (MOR) [75] are derived as measures of burstiness. The MTI is defined as the average between the cell rate in the burst and the cell rate in the inter-burst preceding it. The MOR is defined as the average of the ratio between the number of cells transmitted in a burst and the expected number of cells, based on the mean cell rate. Although both measures reflect the behaviour at statistical multiplexing better than the conventional measures, they do not incorporate the fluctuations of cell rates within a burst or inter-burst.

For a characterization of the fluctuations in time, an adaptation of the auto-correlation function can be used, where a time average is considered. Since consecutive cell rates are considered, we propose to express the burstiness as the mean deviation between these cell rates. Then, the burstiness in a sequence of N bursts transmitted in a time period T_{seq} can be defined as:

$$\beta = \frac{1}{T_{seq}} \sum_{k=1}^{N} (y(k) - y(k-1))^2 . \tag{4.11}$$

The advantage of this burstiness parameter in the case of VBR MPEG is that it makes no difference whether $y(k)$ are the cell rates in a burst or the reciprocal of the inter-arrival times. Consequently, it can be used in the modelling of VBR sources. Although a direct link between the multiplexing effect and β cannot be given, sources with a smaller β are less bursty and are therefore expected to provide a higher multiplexing gain. Thus, all three requirements mentioned previously are satisfied and β will be used in the rest of this chapter to reflect the burstiness of a source. As an example, the burstiness for the slice, picture, and GOP layer in MPEG as in Figure 4.2a, b, and c is 208.4, 99.0, and 1.2, respectively.

4.2 Smoothing of VBR MPEG

Although much research has been carried out in the field of variable bit rate traffic, not many papers address the influence of the network interface (transport layer in OSI, AAL in ATM) on the bit rate characteristics. This may be due to the fact that ATM networks support any variability in traffic rates, which makes the use of complex interfaces, in principle, superfluous. Further, it is often stated that the gain achieved by smoothing algorithms cannot compete with the extra buffer requirements and delay introduced by them. Hence, bufferless packetizing is often assumed, leading to high burstiness.

In most video coders, however, a delay of several picture periods is normal because of the re-organization of data, like the use of B pictures in MPEG. In distributional services, this is no problem whatsoever, and even in communicational services, a delay of up to 0.3 seconds (8 picture periods) is still acceptable [76]. The buffer requirements imposed by smoothing are also common, particularly if the encoder must be able to produce CBR bit streams as well. The following sections describe some techniques to apply smoothing in order to reduce the burstiness of a VBR MPEG source.

4.2.1 Cell Spacing

It is well known that video traffic resulting from bufferless packetizing has a periodic nature [69]. It is also recognized that this nature either causes inefficient resource allocation or cell loss at multiplexing in the network, and that it can be reduced by smoothing [77]. Most smoothing schemes mentioned in literature are based on cell spacing: enforcing a minimum number of empty time slots between two consecutive cells. This is equivalent to the transmission of cells in a burst at peak cell rate as shown in Figure 4.2b, while using a peak rate smaller than the maximum rate expected for the chosen layer in MPEG. As a result, the extra cells are transmitted in the following burst. In [78] such a cell spacing algorithm is proposed as a policing function in the network instead of pick-up mechanisms such as the GCRA. Knightly and Rossaro [79] show that when using such a mechanism before transmission, fewer network resources will have to be allocated than when using bufferless packetizing. The shaping algorithm in [80] aims at a deterministic output, but does not take the delay into account. Although cell spacing reduces the peak cell rate considerably, the peak rate is still determined by the most detailed scene so that no smoothing occurs in less detailed scenes. The effect of cell spacing in the *car* sequence using a peak bit rate of 4.5 Mbit/s is shown in Figure 4.6a. The burstiness according to (4.11) is reduced from 99.0 to 37.5.

4.2.2 The Δ-smoothing Rule

The fear of large delays caused by smoothing is countered by Ott et al [81], who propose the Δ-smoothing rule algorithm, taking into account a delay constraint. This algorithm computes a series of upper bounds on the cell rates based on the idea that if the transmission rate is too high, the buffer will go empty and the link will become idle, which increases the burstiness. Similarly, a series of lower bounds on the cell rates is computed; this is based on a delay constraint, which is violated if the transmission rate is too low. The series are computed using predictions of future cell rates. It is shown that the upper bounds decrease when more future rates are predicted, and the lower bounds increase. The transmission rate is chosen at the

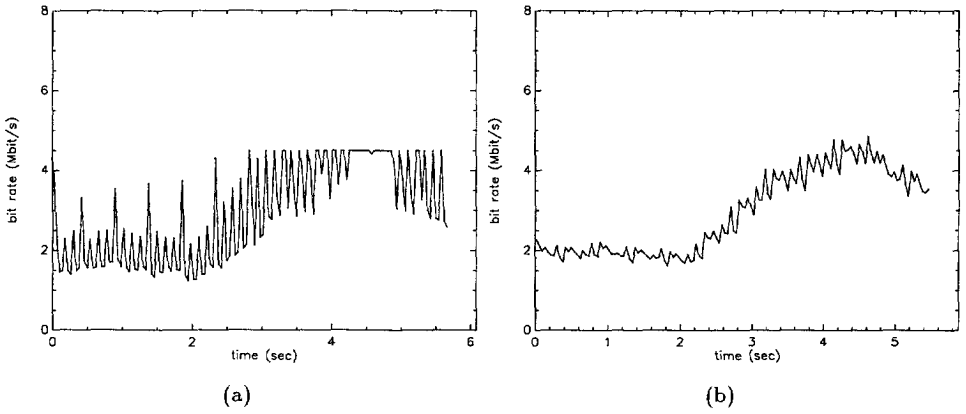


FIGURE 4.6: EFFECT OF (A) CELL SPACING AND (B) THE Δ-SMOOTHING RULE ON CELL RATE FLUCTUATIONS. IN (B) IT IS ASSUMED THAT THE CELL RATES OF 5 FUTURE PICTURES ARE KNOWN IN ADVANCE.

crosspoint of the two series. When no crosspoint occurs, the mean between the upper and lower bound is taken. Although it was argued that this is the exception more than the rule, it occurred regularly in our experiments using the predictions of up to 20 future cell rates. Further, the experiments presented in [81] are based on the picture cell rates of a teleconference sequence. Our data is much more bursty, so that it is more difficult to make the predictions. Therefore, Figure 4.6b shows the results for a maximum delay of 3 picture periods using perfect predictions, assuming that future cell rates are known in advance. Consequently, the results are an upper bound of the possible results for this algorithm. The burstiness measure yields $\beta = 3.8$ for the results in the figure.

4.2.3 Spreading

Considering a maximum delay, the simple algorithm in [82] spreads the cells generated in a picture uniformly over the delay interval. The total number of cells $Y(k)$ to be transmitted in the next picture interval of size T_{pic} with cell rate:

$$y(k) = \frac{Y(k)}{T_{pic}} \tag{4.12}$$

is derived from the generated number of cells $X(k-i)$:

$$Y(k) = \frac{1}{N} \sum_{i=0}^{N-1} X(k-i) \tag{4.13}$$

for a maximum delay of N picture periods. Because of the periodic structure of pictures in a GOP, the results depend on the relation between the delay and the GOP structure. The experimental results in Figure 4.7a show that the best results are achieved if the delay is a multiple of the distance between consecutive P pictures, in our case 3 picture periods. Figure 4.7b shows the effects of spreading over 3 picture periods.

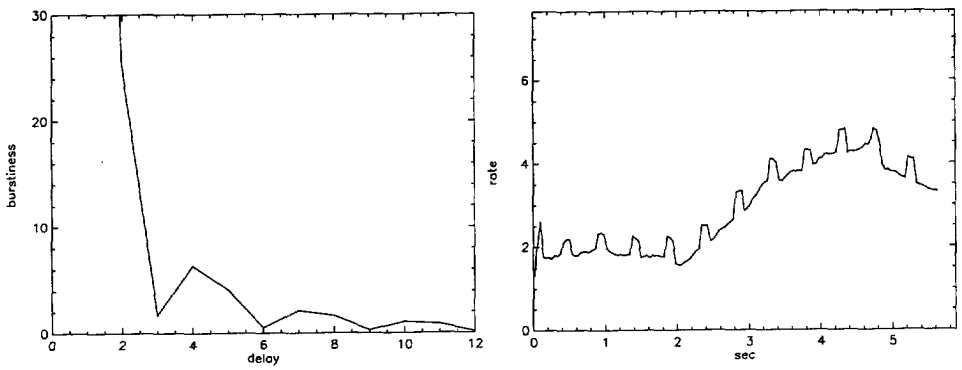


FIGURE 4.7: (A) BURSTINESS AS A FUNCTION OF DELAY WHEN THE CELLS ARE SPREAD OVER THE DELAY PERIOD. (B) EFFECT OF SPREADING FOR A DELAY OF 3 PICTURE PERIODS.

When we consider the delay in spreading over the delay interval, it is obvious that in most cases a much smaller delay is achieved than the maximum, since the buffer mostly holds no more than the cells from 1 or 2 pictures. Therefore, a much longer spreading interval can be used than the maximum delay. Ideally, the spreading interval is chosen to correspond with a GOP interval [82]. This yields for the output cell rate:

$$Y_{GOP}(k) = \frac{1}{N_{GOP}} \sum_{i=0}^{N_{GOP}-1} X(k-i). \tag{4.14}$$

In this case, however, an extra flushing of the buffer is required in the case where the delay constraint is not satisfied by the spreading function [82]. To determine the number of cells that need to be flushed, consider that the cells $Y(k)$ produced by picture k together with the contents of the buffer when these become available need to be flushed within the maximum delay period. If the number of cells in the smoothing buffer is indicated by $B(k)$, the buffer fullness changes at picture intervals according to:

$$B(k) = B(k-1) + X(k-1) - Y(k-1). \tag{4.15}$$

From the delay constraint, the minimum number of cells to be flushed can be calculated according to:

$$MIN(k) = \frac{B(k) + X(k)}{N}, \tag{4.16}$$

where N is the delay interval, expressed in the number of picture periods. To account for the constraints of the previous pictures, a maximum of these minimum numbers of cells is taken:

$$MAX(k) = max(MIN(i)) \qquad i = k - N + 1, .., k. \tag{4.17}$$

Eventually, the number of cells $Y(k)$ is determined by:

$$Y(k) = max(Y_{GOP}(k), MAX(k)). \tag{4.18}$$

Figure 4.8a shows that the output cell rate when using spreading is smooth, unless the delay constraint becomes active when the cell rate increases. The burstiness $\beta = 0.95$ in this figure.

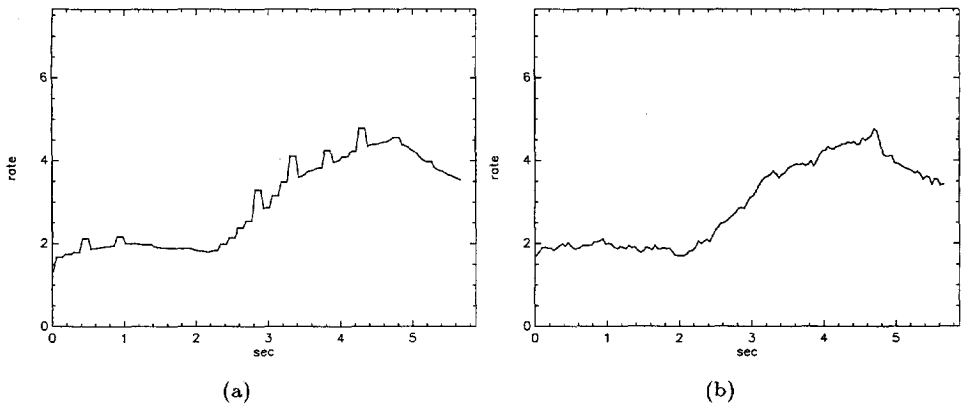


FIGURE 4.8: (A) EFFECT OF SPREADING OVER THE GOP LENGTH WITH A DELAY CONSTRAINT OF 3 PICTURE INTERVALS. (B) EFFECT OF USING THE GOP STRUCTURE IN SPREADING.

4.2.4 Predict the Mean Rate in a GOP

The spreading function of the previous section has a drawback in that the output cell rate responds slowly to the changes in the input cell rate. This causes the delay constraint to be active if the cell rate increases because of increasing activity in the

encoded video. We propose to prevent this by using a prediction of the future cell rates. A good prediction is achieved by using only the cell rates in the previously encoded pictures of each picture type [67]. If a GOP consists of N_I I pictures, N_P P pictures, and N_B B pictures, and X_I, X_P, and X_B are the number of cells produced in the last I, P, and B pictures, the mean number of cells per picture in the current GOP can be predicted by:

$$Y_{GOP}(k) = \frac{N_I X_I + N_P X_P + N_B X_B}{N_I + N_P + N_B}. \tag{4.19}$$

Using this calculation of the output number of cells yields a faster response to the input number of cells, which is shown in Figure 4.8b. The burstiness $\beta = 0.75$ in this figure.

4.2.5 Use Scene Changes

The prediction of $Y_{GOP}(k)$ in (4.19) can be used when the GOP structure is fixed throughout the whole sequence. The reference codec, however, starts a new GOP at every scene change, leaving a short GOP to end the previous scene. To calculate a smooth output rate in this situation, we should consider that smoothing mainly distributes the extra cells I pictures produce, compared to P and B pictures over the rest of the GOP. Therefore, at the beginning of each GOP the buffer should be empty or nearly empty. However, when a new I picture (new GOP) is started before the last GOP is finished, the buffer still contains cells. To overcome this problem, an algorithm is applied which empties the buffer entirely before the next GOP. At each point where the cells from a new picture become available, the number of cells in the buffer plus the number of cells to be expected in the rest of the current GOP will have to be spread over the remaining number of picture periods in the GOP. The number of cells in the buffer right after the cells from a new picture become available equals $B(k) + X(k)$. The number of cells to be expected in the rest of the GOP, considering that only the first picture in a GOP is an I picture, equals $N_P^r X_P + N_B^r X_B$, where N_P^r and N_B^r are the remaining number of P and B pictures in the current GOP. The remaining number of picture periods equals $1 + N_P^r + N_B^r$, since the current picture period is also available. Hence, the output cell rate is calculated right after the cells from a picture become available according to:

$$Y_{GOP}(k) = \frac{B(k) + X(k) + N_P^r X_P + N_B^r X_B}{1 + N_P^r + N_B^r}. \tag{4.20}$$

With this algorithm, the cells that are in the buffer at a scene change are smoothed over the next GOP, preventing a sudden peak.

The algorithm is based on the assumption that the relation between the number of cells generated in I, P, and B pictures is stationary during a scene. At scene changes,

however, this relation also changes, and X_P and X_B in (4.20) should be replaced by an estimation of the number of cells to be produced in future P and B pictures of the new scene. This estimation is based on the number of cells produced in the first (I) picture and an estimation of the new ratio between the number of cells in I, P, and B pictures. As an estimation, we can choose either the most frequently occurring ratio, the previous ratio, or the worst possible ratio, which is the ratio that may lead to the highest output rate after a scene change. The latter may be chosen to prevent an action of the delay constraint fall-back mode, which may occur if the output cell rate after a scene change appears to be too low. In other situations, however, taking the worst ratio may generate an unnecessary peak in the output rate. Since using the previous IBP ratio also does not lead to a good prediction, the mean IBP ratio is used as the new ratio here.

4.2.6 Results and Discussion

To evaluate the performances of the different smoothing algorithms, several video sequences of 24 seconds each were encoded with the reference coder described in Chapter 3. Each source was smoothed with a maximum delay of 3 picture periods using the four algorithms described in the previous sections: the Δ-smoothing rule with a horizon of 5 picture intervals and perfect prediction ("Δ rule"), spreading over the GOP interval ("Spreading"), prediction of the mean cell rate in a GOP ("Pred. GOP"), and incorporating scene changes ("Scene ch."). The results with respect to the burstiness measure β are shown in Table 4.1.

TABLE 4.1: SMOOTHING RESULTS FOR DIFFERENT ALGORITHMS AND SEQUENCES.

	β in sequence				
algorithm	soccer	news	tennis	dune	road
No Sm.	19.65	13.66	34.93	5.56	10.23
Spacing	17.01	4.87	19.29	2.77	2.39
Δ rule	1.77	0.81	2.61	0.52	0.99
Spreading	0.58	0.48	1.12	0.17	0.44
Pred. GOP	0.40	0.55	0.93	0.17	0.38
Scene ch.	0.54	0.56	1.13	0.38	0.60

Table 4.1 shows that predicting the mean rate in a GOP yields the lowest burstiness and that incorporating the scene changes does not provide better smoothing results. Moreover, even the spreading algorithm outperforms the incorporation of scene changes. The cause of this is that at scene changes a quick adaptation of the

output cell rate to the new scene is attempted. Consequently, a sudden change in the output cell rate is introduced, which is reflected in the burstiness measure. The spreading and GOP prediction methods, however, smooth at scene changes as well, which explains the difference in performance.

The burstiness measure is not the only criterion that should be judged. The behaviour at scene changes is of importance to the description of the traffic. For the network, it may be important to identify a change in the cell rate of a source caused by a scene change. A smooth transition between the cell rates of different scenes in this case is not desired. Therefore, the results for β', an adaptation of β in which the fluctuations at scene changes are not taken into account, are shown in Table 4.2. Then, incorporating scene changes does provide better smoothing in most of the sequences.

TABLE 4.2: SMOOTHING RESULTS WHEN THE TRANSITIONS AT SCENE CHANGES ARE NOT CONSIDERED.

	β' in sequence				
algorithm	soccer	news	tennis	dune	road
Spreading	0.50	0.42	1.10	0.16	0.40
Pred. GOP	0.26	0.44	0.91	0.15	0.33
Scene ch.	0.25	0.36	0.54	0.29	0.33

Comparing the different video sequences with each other, an indication of the amount of fluctuations can be seen. Obviously, the burstiness differs for different types of video. It is remarkable, however, that the burstiness in the *news* sequence is not as low as one would expect from a news reader. This can be explained by the fact that, in that sequence, the amount of detail is very high compared to the activity. Therefore, the number of bits generated in I pictures is higher than three times the mean bit rate, which results in a high burstiness in non-smoothed bit stream. Further, it causes an action of the delay constraint in the smoothing algorithms, increasing the burstiness in the smoothed bit streams as well.

Comparing the different algorithms with each other, they all try to reduce the burstiness while respecting the delay constraint. It seems that the Δ smoothing rule incorporates both objectives in its design, while the others only aim at perfect smoothing. By using the delay constraint, however, the same is achieved with much less effort. Because of the delay constraint, however, suddenly too many cells are transmitted by the spreading and prediction algorithms, causing an underflow of the buffer later

on and thus an increase in burstiness. This is prevented in the algorithm that incorporates scene changes, since the fullness of the buffer is considered in the algorithm. This algorithm is used for smoothing in the rest of this thesis.

4.3 Traffic Parameter Control

The parameters in the network contract of each source are negotiated at connection set-up. The choice for these parameters is based on the data expected to be transmitted. Since a policing function will be implemented in the network to check the validity of the traffic to the contract, a control function will have to be implemented in the network interface as well, to prevent policing actions. Although the network will probably use the GCRA algorithm described in Section 4.1.3, or other algorithms implemented on the cell level, the adaptation to the network proposed in this chapter indicates that the control actions at the user's site can be implemented at burst, and thus picture level. For both the peak and the mean or sustainable cell rate, a maximum cell rate for the following picture interval can be deduced from the policing algorithm. For the burstiness, it is not yet clear what a policing function would look like. Nevertheless, it is easy to see that if no change in traffic intensity is allowed because of a burstiness constraint, the maximum cell rate for the following picture interval should be equal to the cell rate in the previous picture interval. In the following, we assume that a maximum output cell rate $y_{max}(k)$, and thus a maximum number of cells $Y_{max}(k)$ for the following picture interval, is determined from the traffic parameters.

In the smoothing algorithm discussed in the previous section, the number of cells to be transmitted was calculated, after which the cells were evenly spaced over the following picture period. From the maximum output cell rate and the smoothing algorithm, the maximum number of cells to be generated by the encoder for a certain picture can be calculated, which is described in Section 4.3.1. The techniques to be used by the encoder to adhere to this maximum are described in Section 4.3.2.

4.3.1 The Maximum Number of Cells

Since the algorithm to smooth the VBR traffic and the maximum number of cells are known, the maximum number of cells to be generated in each picture can be calculated. When the delay constraint is not considered, (4.20) describes the relation between $Y(k)$ and $X(k)$. If we rewrite it to extract $X_{max}(k)$ from $Y_{max}(k)$, we get:

$$X^{max}(k) = (1 + N_P^r + N_B^r)Y_{max}(k) - B(k) - N_P^r X_P - N_B^r X_B, \qquad (4.21)$$

which is the maximum number of cells to be generated if the picture to be encoded is an I picture. If it is a P picture, however, we should consider that in the computation

of the output rate the number of cells in the last P picture X_P is derived from $X(k)$, i.e. $X_P = X_P^{max}(k)$. Further, the number of remaining P pictures before encoding is one more than that number after encoding. Hence, the equation for the maximum number of cells for P pictures becomes:

$$X_P^{max}(k) = \frac{(N_P^r + N_B^r)Y_{max}(k) - B(k) - N_B^r X_B}{N_P^r}. \tag{4.22}$$

Equivalently, for B pictures, the maximum number of cells is calculated by:

$$X_B^{max}(k) = \frac{(N_P^r + N_B^r)Y_{max}(k) - B(k) - N_P^r X_P}{N_B^r}. \tag{4.23}$$

A different situation arises at scene changes. Although (4.21) still holds, the parameters X_P and X_B taken into the calculation of the output rate depend on the encoded rate $X(k)$ of the first (I) picture of the scene and on the newly chosen ratio between I, P, and B pictures:

$$X_{sc}^{max}(k) = \frac{(1 + N_P^r + N_B^r)Y_{max}(k) - B(k)}{1 + N_P^r r_{IP} + N_B^r r_{IB}}, \tag{4.24}$$

where r_{IP} and r_{IB} are the newly assumed cell rate ratios between I and P pictures and I and B pictures in the new scene, respectively.

When the maximum number of cells is calculated from the delay constraint, we get:

$$X^{max}(k) = NY^{max}(k) - B(k). \tag{4.25}$$

We should consider this maximum if it is smaller than the maximum resulting from the smoothing algorithm. Thus, for I pictures, if

$$NY^{max}(k) - B(k) < (1 + N_P^r + N_B^r)Y_{max}(k) - B(k) - N_P^r X_P - N_B^r X_B. \tag{4.26}$$

With $N = 3$, this reduces to:

$$N_P^r X_P + N_B^r X_B < (N_P^r + N_B^r - 2)Y^{max}. \tag{4.27}$$

Hence, the maximum number of cells for an I picture is determined by the delay constraint if the expected number of cells in the rest of the GOP is smaller than this maximum times the remaining number of pictures in the GOP minus two. But in such a case, there is no danger of an excess of the maximum number of cells. For the maximum number of cells in P and B pictures, similar considerations can be made. In conclusion, the maximum number of cells to be generated in a picture is never determined by the delay constraint.

4.3.2 Control Algorithm

From CBR coding, two possible techniques to control the output bit (and thus the cell) rate of a video coder are known. First, feedbackward techniques adjust the control parameter of the codec according to cell rates generated in the past. Second, feedforward techniques use a priori knowledge about the relation between the control parameter and the output bit rate in order to choose the control parameter which leads to the desired output bit rate.

In our situation, the control parameter Q is fixed and should only be adjusted if the number of cells to be generated were to otherwise become too high. Since a feedbackward control cannot guarantee a certain (maximum) output bit rate, a feedforward control would be more appropriate in our situation. However, the collection of accurate knowledge about the relation between the control parameter and the bit rate requires a lot of computational power and is not attractive in a real-time encoder. Further, since the maximum number of cells for a picture is never determined by the delay constraint, an excess of this maximum can be stored in the smoothing buffer. Therefore, a feedbackward system is used here.

In CBR coders, the use of a feed-back algorithm to control the bit rate is very common. There, the fullness of the buffer determines the quantization scaling factor Q in the encoder. When the buffer is almost full, future quantization will become coarser, so that fewer bits are generated. When the buffer is almost empty, future quantization will become finer, so that the quality will be improved, and more bits will be generated. To adopt this regulation in a maximum rate control, some adaptations are necessary.

The fullness of the smoothing buffer itself cannot be used to control the maximum number of cells for a picture in a VBR coder because of the variable output rate. Instead, the fullness of a virtual buffer as shown in Figure 4.9 is used to perform this task. The output rate of this buffer is the maximum bit rate as described in the previous section, which is fixed during the encoding of a picture. The input rate is the rate with which the bits are generated by the VBR coder. At the beginning of a picture, the fullness of the virtual buffer is set at the critical level for the chosen value of the quantizer scaling factor Q, which is 4. Only when the fullness of the virtual buffer comes above that value, is Q increased to generate fewer bits.

When the fullness of the virtual buffer is higher than the critical value at the end of the encoded picture, more bits are generated than the allowed maximum. This would lead to the calculation of an output cell rate that is higher than the maximum cell rate. By ignoring this calculating and transmitting at the maximum cell rate, the extra generated cells are stored in the smoothing buffer. Then, the virtual buffer will not be reset to the critical value, so that the quality in the following pictures will be reduced to compensate for the higher bit rate in the current picture.

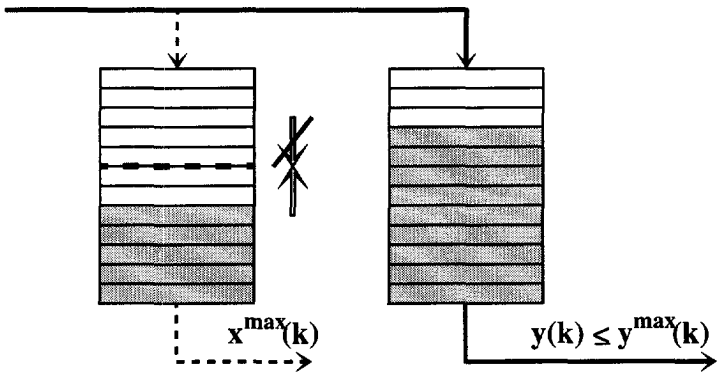


FIGURE 4.9: ILLUSTRATION OF THE VIRTUAL AND SMOOTHING BUFFER.

4.4 Conclusions

The characteristics of the traffic generated by a VBR video coder depend on the adaptation to the network. In addition to the peak and mean bit rate of the source, a parameter specifying the burstiness is needed in order to specify the fluctuations of the bit rate in time. In this chapter, a technique is proposed to smooth the traffic generated by our reference VBR codec in order to reduce the burstiness. This technique predicts the mean bit rate in a GOP and incorporates scene changes. By doing this, the output traffic is highly predictable, except for the change in traffic intensity at scene changes.

In practical situations, the traffic parameters are negotiated in the network contract and therefore need to be monitored by the network and controlled by the encoder. The smoothing algorithm allows for an easy implementation of the traffic parameter control at the encoder. If a good choice of parameters is made, a control action by the encoder is very rarely necessary. Nevertheless, the consequences of a control action should be as limited as possible. By adjusting the control parameter Q, the overall quality of the video is reduced, but this is much better than the loss of information, which would occur in the case of a policing action in the network.

The reference VBR coder is extended with a simple control algorithm. However, if we assume that the traffic parameters are properly chosen, control actions are very rare. Therefore, in the remainder of this thesis, no control actions are considered and the results are therefore independent of the parameter choices.

Chapter 5

Multiplexing of VBR Video

The strength of VBR video lies in the fact that the overall compression performance is better than in CBR sources: the mean bit rate in VBR is lower than the CBR rate. Most transmission and storage media, however, use a physical medium on which a constant information stream needs to be transmitted. Therefore, the VBR stream(s) will have to be converted to a CBR stream using stuffing and/or multiplexing.

When a single VBR stream is converted to a CBR stream by stuffing dummy bits into it, no advantage is taken of the low mean bit rate. When the aggregate bit rate of a number of VBR sources is considered, however, the standard deviation from the mean bit rate decreases with an increasing number of sources, according to the weak law of large numbers. This can be exploited by transmitting the different bit streams over a CBR channel using multiplexing and stuffing, while the channel rate is significantly lower than the sum of the peak bit rates of the individual streams. By doing this, there is always a probability that the sum of instantaneous bit rates is higher than the channel rate, which would result in loss of information. Consequently, multiplexing yields a trade-off between information loss and efficient use of the channel.

Multiplexing is defined as the transmission of multiple bit streams on a single channel while the individual sources can still be identified. An extra requirement for the multiplexing of VBR sources is that the bandwidth can be shared dynamically. To accomplish this, a framework will have to be agreed upon in which the multiplexing takes place. Section 5.1 discusses the possibilities and shows that ATM may serve as a general multiplexing framework for all kinds of media.

The performance of multiplexing can be expressed in terms of efficient use of the channel at a certain probability of information loss. It depends on the characteristics

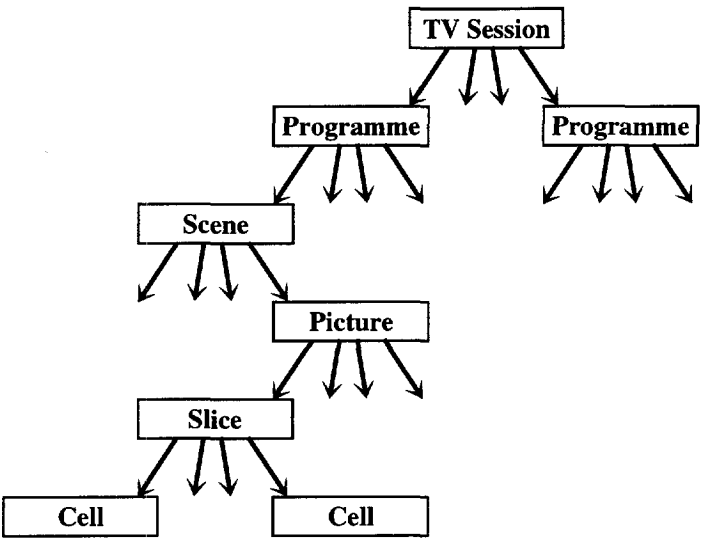


FIGURE 5.1: HIERARCHICAL MODEL OF THE BIT RATE CHARACTERISTICS IN A TV SESSION.

of the input bit streams. If these characteristics are modelled properly, the multiplexing performance can be evaluated analytically. Otherwise, the performance will have to be determined by simulation.

Although, in ATM, the peak and mean cell rate of a VBR source, together with a parameter describing the burstiness, are provided in the contract with the network, they do not describe the traffic characteristics accurately enough to predict the multiplexing performance. More complicated models are needed to describe voice, data, and video sources [83, 84]. For VBR video, the characteristics of the traffic depend on the video itself, on the encoding technique used, and on the adaptation to the network. Therefore, several levels to describe the traffic can be identified. In [85] the most detailed hierarchical model is described, which is shown in Figure 5.1.

The actual multiplexing takes place at the cell level. The behaviour is determined by the cell spacing in the network interface. Nevertheless, delay and delay jitter caused by switching and multiplexing in the network may affect it. Consequently, the characteristics of the traffic at the last node of the network may differ from the characteristics at the source. Therefore, separate models for this level are of interest and are discussed in Section 5.2.

Above the cell level, the slice and picture level are included in the hierarchical model to reflect the different number of bits that are generated in different slices of a picture and in consecutive pictures, which may differ in type (I, P, and B pictures in MPEG).

The smoothing algorithm in the network interface, however, reduces the effects of these encoder-specific characteristics, which indicates that the traffic characteristics are mainly determined by the properties of the input video. Section 5.3 discusses the extent to which this is true.

The properties of the input video are incorporated by the picture, scene, programme and TV session levels in Figure 5.1. This division is chosen to reflect not only the occurrences at scene changes, where abrupt changes in activity occur, but also the difference in activity between, for instance, video clips and talk shows, or between sports and news channels. Models at these levels are called video models and are the subject of discussion in Section 5.4.

Accurate models of VBR traffic are mostly too complicated to be used in an analytic evaluation of multiplexing performance. Therefore, Section 5.5 uses a simple model to analyze the effects of the network adaptation, proposed in Chapter 4, to the multiplexing performance. Finally, Section 5.6 evaluates the performance of multiplexing experimentally.

5.1 Multiplexing Framework

As discussed in Chapters 2 and 4, multiplexing is performed in the network layer of media that support VBR, but it can alternatively be performed in the transport layer for media that do not support VBR, as is shown in Figure 5.2. In this case, a framework in which the multiplexing takes place will have to be agreed upon by the transmitting and receiving terminals. In this framework, the input is a number of VBR bit streams and the output a CBR bit stream which is offered to the physical layer [1].

There are several ways to multiplex multiple bit streams onto the same physical channel, but when the bandwidth needs to be dynamically shared among the sources, time division multiplexing is inevitable. The data of each source will have to be split into variable or fixed-sized units with source identification in the header of each unit, so that units of different sources can be concatenated to form a single stream. How the data is split depends on the application. In MPEG-1, for instance, a system part is standardized to multiplex an audio and a video elementary stream into a system (also called programme) stream [28], in which the multiplexing units may be of variable and relatively great length, to limit the amount of overhead. In practical situations such as CD-i and video-CD, however, small fixed size packets are usually used because of physical limitations.

[1] Note that the network layer is mostly empty in CBR media and the data-link layer is empty in all networks that support real-time data transfer. Therefore, they are omitted here.

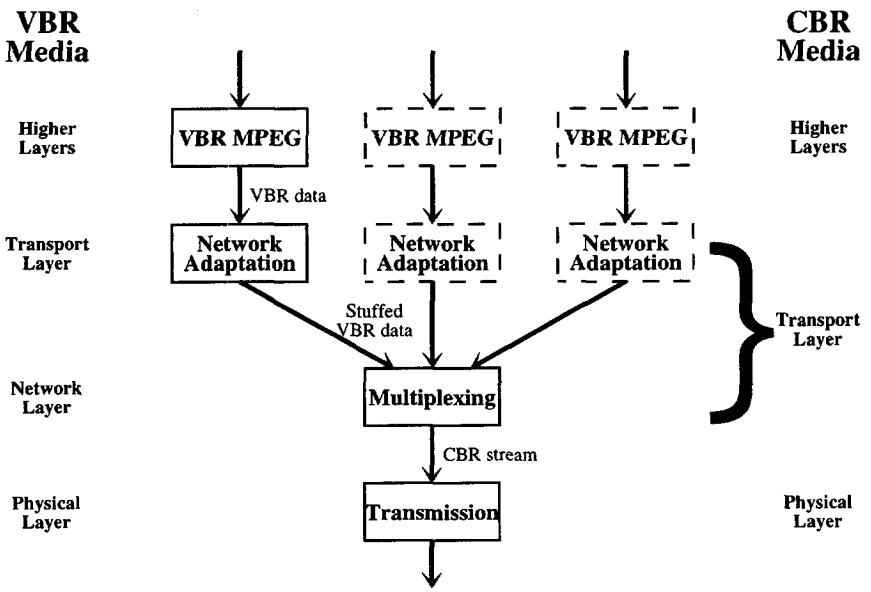


FIGURE 5.2: MULTIPLEXING CAN BE PERFORMED IN THE TRANSPORT LAYER FOR CBR MEDIA.

If multiplexing is performed in the transmitting terminal, it is possible to prevent an overflow of the output channel by an additional rate control [86]. It is even better to apply a joint bit rate control on all sources to be encoded. Such a control distributes the available bits of a CBR channel to the different sources, based on their activity [87, 88]. Here, we do not use these possibilities since we assume that the different sources originate from different encoders which have no knowledge about the other sources with which they are multiplexed. This is a more general approach which permits more applications such as demultiplexing and remultiplexing of sources either in the network or at a terminal.

Considering the above, we choose to apply multiplexing by splitting the individual bit streams into ATM cells and by simulating an ATM switch. By doing this, a general framework is created in which it is possible to provide all kinds of functionalities as they exist in ATM, such as a higher priority for important data. Before multiplexing, an adaptation of each source to an ATM network is applied as described in Chapter 4.

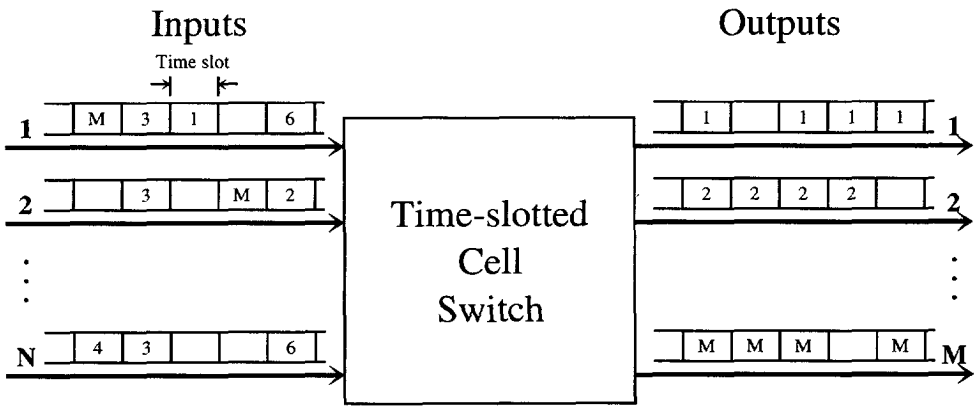


FIGURE 5.3: SCHEMATIC DIAGRAM OF AN ATM CELL SWITCH.

Multiplexing in an ATM Switch

In ATM, multiplexing is performed in every network node. Consequently, every node consists of a cell switch, which conceptually is a box with N inputs and M outputs as shown in Figure 5.3. At each time slot, cells may arrive at any input, addressed to any output, the routing information in the header of each packet indicating which output must be chosen. In order to route a cell from an input to an output, a path within the switch will have to be established. When this path is not available, the cell is buffered or, when the buffer is full, the cell is lost.

Methods to buffer the cells within an ATM node can be divided into different classes, dependent on the location of the buffers. We confine ourselves to output queuing because of its simplicity and good performance results compared with input queuing [89]. In this technology, cells are buffered at the output when more than one cell is routed to this output. Each input may contain cells from different sources, which may be directed to different outputs. Hence, multiplexing in an ATM node can be divided into three stages: demultiplexing at the input links, routing of each cell to the correct output link and the actual multiplexing of the different sources at each output link. When the switch is a non-blocking network, that is, if there are enough switch points to provide simultaneous, independent paths between arbitrary pairs of inputs and outputs, cell losses only occur in the last stage, when the number of cells routed to an output link exceeds the capacity of this link and the buffer is not large enough to store the cells in excess. Hence, only this last stage is considered in the remainder of this chapter.

5.2 Modelling at Cell Level

At the lowest level in the hierarchy, traffic is described according to the arrivals of cells. Different analytical models are proposed to describe the probabilities of cell occurrences and to calculate the multiplexing performance from these models using queuing theory [90].

5.2.1 Bernoulli and Poisson Processes

When we consider that in ATM switches the cells arrive only at discrete moments, the time slots, a simple characterization of the traffic is that of assuming a constant probability p of cell arrival in each time slot. Consequently, the probability of no cell arrival equals $1-p$. This kind of stochastic process is called the *Bernoulli* process and its corresponding inter-arrival time process has a geometric probability mass function (PMF):

$$P(A_i = k) = (1-p)^k p, \quad m = \frac{1-p}{p}, \quad \sigma^2 = \frac{1-p}{p^2}, \tag{5.1}$$

with $P(x = k)$ the probability to observe the arrival of a cell after k empty time slots, m is the mean inter-arrival time and σ^2 its variance. When we consider the counting process $P(X_n = i)$, which is the probability to observe i arrivals within n time slots, we obtain the binomial variable with probability mass function:

$$P(X_n = i) = \left(\begin{array}{c} n \\ i \end{array} \right) p^i (1-p)^{n-i}. \tag{5.2}$$

The counting process for a memoryless source also represents the batch arrival process, i.e. the probability that i cells arrive on n input links. The mean and variance of this process are given by:

$$m = np, \quad \sigma^2 = np(1-p). \tag{5.3}$$

If the number of time slots (or input links) n increases, while the mean np remains constant due to a lower probability p, the Bernoulli process converges to a Poisson process with variable $\lambda = np$. This process is characterized by the probability to observe i cell arrivals in the interval Δt:

$$P(X_{\Delta t} = i) = e^{\lambda \Delta t} \frac{(\lambda \Delta t)^i}{i!} \tag{5.4}$$

The Poisson process assumes independent inter-arrival times with a negative exponential PMF. An important property of the Poisson process is that a superposition

of independent Poisson processes is another Poisson process. The rate λ of the resulting Poisson process is equal to the sum of the rates of the individual Poisson processes:

$$\lambda = \sum_{i=1}^{n} \lambda(i). \tag{5.5}$$

Because of this property, the multiplexing performance can be easily determined when the sources are modelled by independent Poisson (or Bernoulli) processes. In VBR compressed video, however, consecutive arrivals are not independent, because of the correlation in the video material. Therefore, more complicated models are required to describe the dependency in cell arrivals.

5.2.2 Renewal and Point Processes

An important class of arrival processes arises when we consider the case where the cell arrivals are not independent, but the inter-arrivals still are. For instance, the probability of observing a cell immediately after the previous one is smaller than the probability of observing a cell a few time slots later. Thus, the probability to observe the arrival of a cell depends only on the previous arrival:

$$P(A_i = k | A_{i-1} = k - l) = f(l). \tag{5.6}$$

This kind of process is a generalization of the Poisson process and is called a *renewal* process. Renewal processes are characterized by the PMF of the inter-arrival time, which may have any general shape. They are more flexible in describing cell traffic, sufficiently to reflect real-time voice traffic accurately enough to predict its multiplexing performance. For VBR video, however, a further generalisation is needed.

If also the constraint of independent inter-arrivals is dropped, a generalization of the renewal processes can be made. Members of this general family of arrival processes are called *point* processes. There are many different classes of point processes, two of which are described in [69, 91], to reflect the behaviour of VBR video sources. The first is the Doubly Stochastic Point Process (DSPP), which is defined as a Poisson process where the rate λ is a realization of a continuous-time stochastic process $\lambda(t)$. The case where the rate of arrival $\lambda(t)$ is derived from a continuous-time irreducible Markov chain of m states is called the Markov Modulated Point Process $MMPP(m)$. The second class to be considered is the compound Poisson process, which is composed of a cascade of two arrival rates λ_1 and λ_2, where λ_1 defines the arrival of bursts and λ_2 the arrival of cells within a burst.

The superposition of a number of renewal and point processes is considered in [69, 91]. In general, the resulting process is an intricate point process with correlated inter-arrivals. Nevertheless, under the assumptions that the number of superposed sources is sufficiently large, and that the individual processes contribute

less frequently to the superposed process with an increasing number of sources, the resulting process is a Poisson process. These assumptions, however, lead to the observation that the multiplexing performance can only be analyzed over relatively short time intervals.

The two classes of point processes mentioned above are able to reflect the cell arrival processes of VBR video sources better than a simple Bernoulli or Poisson process. The parameter estimation of both processes, however, depends very much on the implementation of the encoder if bufferless packetizing in the transport layer is applied. If a network adaptation, as proposed in Chapter 4, is applied, the traffic at cell level can be modelled more easily by a fixed inter-arrival time within consecutive picture boundaries.

5.2.3 On-Off Model

Another model that is often used and easily analyzed in queuing theory is the on-off source model. It is especially used for modelling voice traffic in packet networks [83] because of its capability to capture bursty traffic. A number of adaptations of the model are proposed in literature to model VBR video traffic as well.

The on-off model is a two-state Markov model, and thus characterized by the state transition probabilities μ and ν and the activity level A of the on state, as is shown in Figure 5.4. When the source is on, cells are generated with a constant inter-arrival time $T = \frac{1}{A}$. The probability that the source is on is given by

$$P_{on} = \frac{\mu}{\mu + \nu}, \tag{5.7}$$

and the probability that a source is off by

$$P_{off} = 1 - P_{on} = \frac{\nu}{\mu + \nu}. \tag{5.8}$$

The transition rates μ and ν determine the dynamic behaviour of the source. This is reflected in the autocorrelation function, which is given by

$$R_{xx}(t) = e^{-(\mu+\nu)t}. \tag{5.9}$$

The superposition of on-off sources is studied extensively for the multiplexing of voice and data sources (see for instance [92]). The superposition of M identical and independent on-off sources can be modelled by a multi-minisource model, which is an (M+1)-state Markov Chain describing the number of sources which are currently active. Hence, this is a simple one-dimensional birth-death process as shown in Figure 5.5.

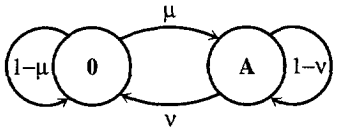


FIGURE 5.4: THE ON-OFF MODEL.

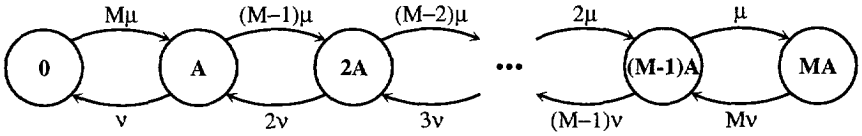


FIGURE 5.5: THE MULTI-MINISOURCE MODEL.

On-off sources assume that cells are transmitted at peak cell rate in a burst. Consequently, they only describe the behaviour at burst level. For VBR video, this kind of modelling is only appropriate if the transmission of cells in a picture or part of a picture is at peak cell rate. Otherwise, other models will have to be used at cell level.

5.2.4 Conclusions

The behaviour of VBR video traffic at cell level depends on the adaptation of the video source to the network. Complex point processes are needed to describe the traffic if bufferless packetizing is applied, and an on-off source can be used to describe the influence of cell spacing. If a uniform distribution of cells within picture boundaries, as proposed in Chapter 4 is applied, a simple point process with fixed inter-arrival times within picture boundaries is adequate. Then, the inter-arrival time is determined by modelling at higher levels in the hierarchy. Only if the cell traffic further on in the network is affected by jitter, may an adaptation to this model be needed. The worst case in that situation is when the traffic within picture boundaries is characterized by a Bernoulli process.

5.3 The Influence of Smoothing

The behaviour at levels higher in the hierarchy than the cell level depends on the part of a picture that is currently being encoded. Most encoders process the data

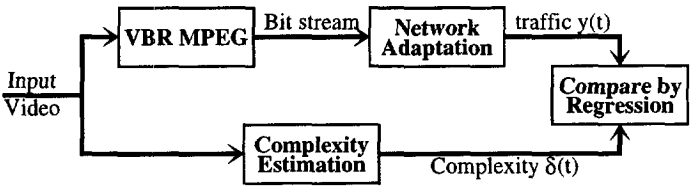


FIGURE 5.6: VERIFICATION OF INFLUENCE OF INPUT VIDEO ON CELL RATE CHARACTERISTICS.

block-wise, scanning a picture from left to right. One horizontal scan is often called a stripe or, in MPEG terminology, a slice. The number of cells to be produced in a slice depends on the location of the slice in the picture. Because of the correlation between corresponding slices in consecutive pictures, the bit, and thus cell rate, generated per slice has a periodic nature. Further, the bit rate also depends on the encoding technique used. In MPEG, the three different encoding techniques (I, P, and B pictures) are used periodically, determined by the GOP structure, yielding another periodic nature for the bit rates per picture.

There have been several attempts in literature [93] to model the periodic nature of bit rates per slice and picture. For instance, the slice periodicity can be modelled by a discrete non-stationary periodic process (DNSPP) or a cyclo-stationary process [85]. The picture periodicity in MPEG is accounted for by separately modelling the behaviour for I, P, and B pictures [94], and cycling through these models based on the GOP structure [95]. Another approach is to include the GOP pattern after generating a bit rate profile according to the activity in the scene [96].

In Chapter 4, a network adaptation for VBR MPEG video was proposed in which smoothing is applied to remove the periodic nature of slice and picture cell rates. Consequently, when such an adaptation is used, there is no need to model the behaviour at these levels. Instead, it is expected that the resulting traffic depends only on the complexity of the input video and on the quality (which depends on the amount of data reduction). Since the quality of VBR compressed video should be constant, the traffic is expected to be linearly dependent on the complexity of the video. To verify this, we will use regression as shown in Figure 5.6. The traffic resulting from VBR MPEG encoding and network adaptation is compared by regression with a complexity measure of the input video.

5.3.1 Determination of Complexity

We use a complexity measure of input video to describe how redundant the information in the pictures is. Since compression algorithms try to remove this redundancy,

such a measure is assumed to reflect the compressibility of the video. Where compression algorithms use complicated techniques to optimize compression, the complexity measure should be as simple as possible in order to obtain a quick insight into the compressibility of video.

The complexity of a signal can be reflected by the auto-correlation function (ACF) of the signal [3], which quantifies the closeness of two samples $x(n)$ and $x(n+k)$ as a function of their separation n in time or space:

$$R_{xx}(k) = \frac{1}{N} \sum_{n=0}^{N-k-1} x(n)x(n+k). \tag{5.10}$$

Mostly, the normalized ACF is used by dividing (5.10) by the variance of the signal. This function decays quickly in highly detailed, and slowly in less detailed scenes. In our situation, a single measure is needed to reflect the correlation in both time and space. Therefore, a temporal differential picture $\Delta p(x,y,t)$ is obtained first:

$$\Delta p(x,y,t) = p(x,y,t) - p(x,y,t-1). \tag{5.11}$$

Here $p(x,y,t)$ is the luminance value at position (x,y) in the picture at time t. Then, the complexity is measured looking at adjacent pixels only. Since a linear regression is used with the VBR traffic, we want to avoid a quadratic term as in (5.10). Therefore, horizontal and vertical complexity are defined as:

$$\delta_h(t) = E[|\Delta p(x,y,t) - \Delta p(x-1,y,t)|] \tag{5.12}$$

$$\delta_v(t) = E[|\Delta p(x,y,t) - \Delta p(x,y-1,t)|]. \tag{5.13}$$

From these values we define the complexity parameter $\delta(t)$ as the mean between the horizontal and vertical components:

$$\delta(t) = \frac{\delta_h(t) + \delta_v(t)}{2}. \tag{5.14}$$

In Figure 5.7a, $\delta(t)$ is shown for each picture in a video sequence containing scene changes and a diversity of motion scenes (24 seconds of the video clip *dune*).

5.3.2 Regression

Different implementations of the network adaptation and different GOP structures can be evaluated by regression of the complexity $\delta(t)$ against the traffic $y(t)$. As an example, the scatter diagram of $\delta(t)$ and the bit rate $y(t)$ generated with an IPPP... GOP structure without smoothing is shown in Figure 5.7b. The outliers in this diagram result from the behaviour at scene changes, where $\delta(t)$ is relatively

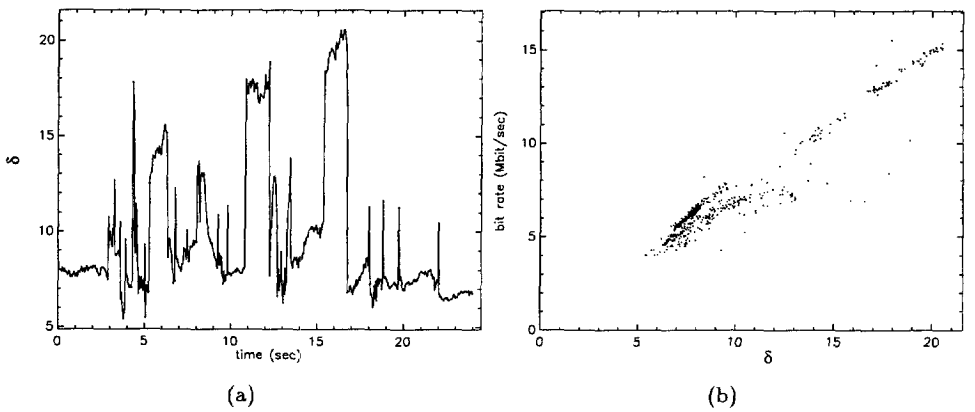


FIGURE 5.7: (A)PARAMETER $\delta(t)$ AND (B) SCATTER DIAGRAM OF BIT RATE $y(t)$ ON $\delta(t)$ FOR THE SEQUENCE *dune*.

high, although the bit rate results from an IPPP... GOP structure also shows a peak there. With other GOP structures and with smoothing, similar outliers occur at scene changes.

Using regression with the least-squares estimation method, the parameters α and β in $y(t) = \alpha\delta(t) + \beta$ can be found by

$$\alpha = \frac{\sum_{t=1}^{N}(\delta(t) - \bar{\delta})(y(t) - \bar{y})}{\sum_{t=1}^{N}(\delta(t) - \bar{\delta})^2}, \tag{5.15}$$

and

$$\beta = \bar{y} - \alpha\bar{\delta}, \tag{5.16}$$

where $\bar{y}$ and $\bar{\delta}$ are the mean values of $y(t)$ and $\delta(t)$ respectively. As a measure for the dependency between $y(t)$ and $\delta(t)$, the correlation coefficient r can be used:

$$r = \frac{\sum_{t=1}^{N}(\delta(t) - \bar{\delta})(y(t) - \bar{y})}{\sqrt{\sum_{t=1}^{N}(\delta(t) - \bar{\delta})^2 \sum_{t=1}^{N}(y(t) - \bar{y})^2}}. \tag{5.17}$$

Regression is performed on a long video sequence resulting from the concatenation of the sequences *dune*, *soccer*, *tennis*, *road*, and *news*, representing a video

clip, two sports events, a cartoon, and a news report, respectively. The cell rates are derived from encoding with three different GOP structures (IPPP..., IBP and IBBPBBPBBPBBP, respectively) and two quality levels (quantization scaling factor $Q = 4$ and $Q = 10$, with weighting matrices and local adaptive coding according to the reference codec). Two different network adaptations are considered: one with and one without smoothing. Smoothing is denoted by "sm". The regression results are shown in Table 5.1.

GOP	Q	sm	α	β	r
IPPP	4	no	0.70	0.36	0.96
IBP	4	no	0.69	0.83	0.84
IBP	4	yes	0.71	0.65	0.97
I4P8B	4	no	0.79	-0.68	0.86
I4P8B	4	yes	0.83	-1.05	0.96
I4P8B	10	no	0.32	-0.79	0.79
I4P8B	10	yes	0.33	-0.84	0.92

TABLE 5.1: REGRESSION RESULTS FOR DIFFERENT GOP STRUCTURES.

The correlation coefficient r for each entry in Table 5.1 is larger than 0.79, while a correlation higher than 0.5 indicates a linear relation between two quantities. Consequently, a linear relation is justified for both smoothed and non-smoothed traffic. For a strict linear dependency, however, a correlation coefficient of 1 should be found. Although this is not the case, the values for smoothed cell rates are larger than 0.95 for $Q = 4$. This implies that the cell rates after smoothing indeed reflect the complexity in the video, independent of the GOP structure used for encoding. For $Q = 10$, the correlation is somewhat smaller because of a coarser quantization, and thus there is more reduction of the information present in the video scene. Nevertheless, we conclude that after smoothing, the cell rates of a VBR MPEG source depend merely on the input video and the quality. Consequently, to model VBR video traffic, a model for the input video is needed.

5.4 Modelling the Video

Video signals originate from one or more cameras, depending on the type of service the video supplies. For video-phone and video-conference, for instance, a single camera is used, while the number of cameras used for a TV channel is considerable. The switching between different cameras results in a division of the video signal into

different scenes. Therefore, to describe a video source, both the occurrences within a scene and between different scenes should be modelled.

5.4.1 Intra-scene Modelling

Within a scene, consecutive pictures are very similar, and thus the amount of detail slowly changes. To model this behaviour, several different models are proposed in literature. Among them, the Auto-Regressive (AR) model and the multi-minisource model are the most popular ones.

Auto-Regressive Model

Auto-regressive models define the next random variable in the sequence as an explicit function of values from the past. Linear auto-regressive models have the following form [84, 73, 77]:

$$X(t) = \sum_{i=1}^{m} a_i X(t-i) + be(t). \tag{5.18}$$

There are several similar ways to derive a bit rate sequence $y(t)$ from an AR process. Here, we assume that $e(n)$ is a Gaussian-distributed random number with variance 1 and mean 0. The bit rate sequence is then derived according to [73]:

$$y(t) = E[y(t)] + X(t). \tag{5.19}$$

The order m of the auto-regressive model can be chosen according to the goodness of fit of the model to the measured data. It is shown in [77] that an AR(2) model is representative of the data from a video-conference scene. To reduce complexity, however, most papers use an AR(1) model to reflect the data [84, 73].

Although AR models are known to reflect the behaviour within video scenes very well, it is extremely difficult to obtain the multiplexing performance analytically with these models. Therefore, they are mostly used in simulation studies only [97].

Multi-minisource Model

Since the fluctuations in complexity and thus bit rate within a scene are minimal, a multi-mini source, where jumps occur only between states that are one level apart, can be used to reflect these fluctuations [97]. Then, the possible bit rate is quantized into finite discrete levels. Evidently, the superposition of a number of identical and independent multi-mini source processes is again a multi-mini source process. When a model is used for a single source, the switching between the states would occur at the picture boundaries. When the aggregate traffic of a number of VBR video sources is modelled, the switching is assumed to be sampled at random Poisson times.

Although this model is not very accurate for modelling single sources, it can be used to predict the multiplexing performance. Unfortunately, it is rather difficult to obtain the appropriate parameters for the model if sources of different types or quality levels need to be multiplexed, which makes this model less useful.

5.4.2 Inter-scene Modelling

In most video applications, the video data consists of multiple scenes with varying contents. When we assume that each scene is associated with a process, a scene change is a transition between processes. When the intra-scene variations are modelled by an AR or multi-minisource process, then a scene change can be modelled by changing the parameters of that model. In [98] this is done for an AR process, where the number of states depends on the nature of the video scenes.

Another approach to modelling video consisting of multiple scenes is to extend the multi-minisource model to support more activity classes. In [99], two types of on-off sources represent two activity classes. In [100], this approach is generalized by using a superposition of heterogeneous on-off sources, hence allowing more activities. A more general approach is by allowing state transitions between all states of a Markov chain, each state representing an activity class [77]. Such a model can even be used to model the periodicity of the bit rates at picture level in MPEG [101]. A drawback of using a Markov chain with transitions between all states is that it has many parameters and no apparent connection between the parameters and some easily measured statistics of the data [77].

One phenomenon of the bit rates at scene level requires extra attention. In some video coders, inter-coding is applied in each picture, so that a peak occurs at the first picture of each scene. When an extensive Markov chain is used as a model, these peaks may be incorporated by that model. When one or more AR processes are used, an extra Markov process can be added to generate these peaks [102]. In our reference encoder, a new I picture is applied at the first picture of a scene, and the smoothing algorithm aims at reducing all peaks.

For a simple evaluation of multiplexing performance, it is argued that the intra-scene variations are small compared to those resulting from a scene change and can thus be ignored [73]. Hence, the traffic characteristics are defined by the scene length and scene activity distribution only, which are mostly assumed to be independent [103]. Different distributions have been proposed to model both events [73, 103, 85].

The distribution of scene lengths and scene activity is expected to vary with the type of programme. Talk shows, for instance, have much less active scenes than sports programmes. Within the scene length PDF found in [85], two lobes are found, indicating two existing fashions of producing TV programmes: short scenes in movies and long scenes taken in a studio. In addition, where the distribution of scenes

depends on the type of programme being broadcasted, the types of programmes may depend on the broadcasting channel. Music and sports channels are expected to produce much more high activity programmes than news channels.

5.4.3 Conclusions

A video model consists of a model for the duration and activity of the scenes, possibly extended with a model for the intra-scene variations in complexity. The parameters of such a model depend on the type of video to be considered. Therefore, a relevant subset of programmes need to be considered for the derivation of these parameters. Then, it is expected that the multiplexing performance can be predicted by the models. However, to verify this, first an analysis of the behaviour at cell level will have to be made.

5.5 Multiplexing Analysis

The characteristics of VBR video traffic depend on the adaptation to the network. With cell spacing and smoothing in the transport layer of the OSI model, the characteristics reflect the complexity of the input video. Therefore, the multiplexing performance is expected to depend on the input video only. To verify this, this section analyzes the multiplexing at cell level. It describes the influence of cell spacing in terms of an adaptation of the Bernoulli process.

In a Bernoulli process, the probability of observing a cell in an arbitrary time slot equals p. The batch arrival process, the probability of k arrivals in one time slot of n incoming links, equals the binomial distribution as expressed in (5.2). With the output queuing model, cell loss in an ATM switch occurs when the number of cells routed to an output link is larger than the available buffer space. Hence, when B is the size of the buffer and B_c the number of cells in the buffer, the probability of cell loss equals

$$P_{CL} = \sum_{i=0}^{B} P(B_c = i)P(X_n > B - i). \tag{5.20}$$

The overall probability of cell loss can be calculated as the steady-state solution of this equation.

If cell spacing is applied in the network interface, the Bernoulli process can be adapted by allowing no cell arrivals when the inter-arrival time is smaller than min (the cell rate of the input link divided by the peak cell rate) time slots:

$$P(A_i = t|A_{i-1} = T) = \begin{cases} 0, t - T < min \\ q, t - T \geq min \end{cases} \tag{5.21}$$

where A_i is the time slot in which cell i arrives. q can be calculated from p and min to have the same arrival rate:

$$q = \frac{p}{(1 - min)p + 1}. \tag{5.22}$$

The batch arrival process in the case of cell spacing depends on the batch arrivals in the previous time slots: n in (5.2) must be substituted by the number of sources that did not have a cell arrival in the last min time slots. For instance, when min is equal to the number of input links and the buffer size, a buffer content of 3 cells means that at least 3 input links cannot produce a cell in the next time slot. Then, only $min - 3$ cells may arrive, which can be handled by the buffer. This is true for all buffer contents in this case, and thus, buffer overflow will not occur when min is larger than or equal to the number of input links. Note that this means that the sum of peak rates is less than the output rate.

When min is decreased, and the mean bit rates are kept constant (hence, q increases), the burstiness increases. We can see from our model that the probability of cell loss also increases. When min is one, the sources reflect a Bernoulli process again.

If smoothing as described in Chapter 3 is applied, the cells in a burst are uniformly distributed over the duration of the burst, in our case a picture period. Then, we can adapt the above model assuming $q = 1$ and min is determined by the calculated cell rate in the burst. At each picture boundary a new min will be calculated. Consequently, if the buffer size is larger than or equal to the number of input links, the probability of cell loss within the picture boundaries is completely determined by the calculated cell rates, and thus by the input video.

5.6 Experiments

For an accurate evaluation of the multiplexing performance in practical situations, either an accurate model for video material of different types or a large amount of encoded video would be needed. At present, neither are available and therefore, only a rough indication can be obtained via experiments with encoded video. Here, a number of different video sequences are encoded and the resulting bit rates are used in a simulation model to obtain insight into the achievable multiplexing gain.

In this chapter, we have discussed the effect that the cell rates after smoothing reflect the complexity of the input video, and that the multiplexing performance when uniform cell spacing is applied within a picture period is determined by the calculated cell rates only. Here, the numerical effects on the multiplexing performance of both techniques are shown by experiment.

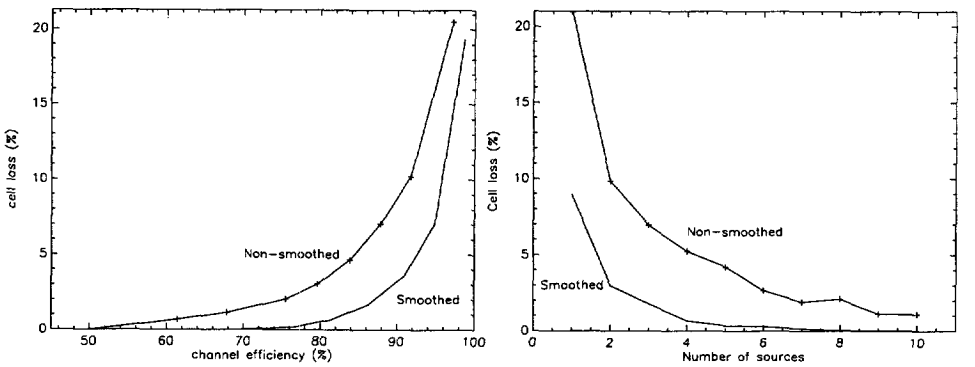
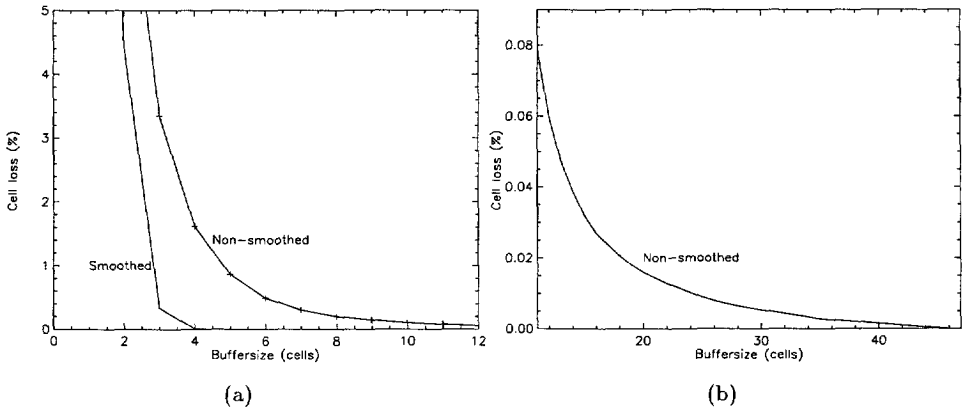


FIGURE 5.8: (A) CELL LOSS PERCENTAGE VERSUS CHANNEL EFFICIENCY FOR THE MULTIPLEXING OF 5 SOURCES. (B) CELL LOSS PERCENTAGE AS A FUNCTION OF THE NUMBER OF MULTIPLEXED SOURCES, THE CHANNEL EFFICIENCY IS KEPT CONSTANT AT 80 %.

The multiplexing performance can be expressed in terms of cell loss probability at a certain network efficiency and buffer size. At higher network efficiencies, the cell loss probability will be higher. Further, the performance is expected to improve when more sources are multiplexed. Figure 5.8 shows both effects for non-smoothed and smoothed sources, while in both cases the cells are uniformly distributed over a picture period. Figure 5.8a shows the cell loss percentages for the multiplexing of 5 sources as a function of the network efficiency, and Figure 5.8b shows the cell loss percentages as a function of the number of multiplexed sources for a network efficiency of 80 %. The buffer sizes in both experiments are equal to the number of sources. From both experiments, the advantage of smoothed sources over non-smoothed sources is clear.

The smoothing technique used in the experiments predicts the mean bit rate in a GOP and incorporates scene changes. If other smoothing techniques are used, the results are expected to differ according to the burstiness of the smoothed traffic. Table 5.2 shows the cell loss percentages for a channel efficiency of 80 % if 5 sources with different smoothing techniques are multiplexed. In this table the mean burstiness of the sources according to the burstiness parameter β proposed in Chapter 4 is also listed. We conclude that with the burstiness parameter indeed an evaluation of the multiplexing performance can be made. The improvement of incorporating scene changes in the multiplexing performance, however, is remarkable since the burstiness increases due to the behaviour at scene changes. The intra-scene burstiness β' indicates that within a scene the burstiness is lower. A possible explanation



(a) (b)

FIGURE 5.9: CELL LOSS PERCENTAGE AS A FUNCTION OF BUFFER SIZE. THE RESULTS FOR MULTIPLEXING 5 SOURCES AT A NETWORK EFFICIENCY OF 70 % ARE SHOWN. (A) PERFORMANCE FOR SMALL BUFFER SIZES, UNIFORM AND BERNOULLI DISTRIBUTION. (B) PERFORMANCE FOR LARGER BUFFER SIZES, BERNOULLI DISTRIBUTION.

for the (slightly) improved multiplexing performance is that although the transition at scene changes is less smooth, this yields a smaller probability of an action from the delay constraint, which would result in higher cell rates.

TABLE 5.2: CELL LOSS PERCENTAGES FOR DIFFERENT SMOOTHING ALGORITHMS.

Smoothing Technique	Burstiness β	Intra-scene β'	Cell loss Percentage
No smoothing	16.81	16.32	3.03
Spacing	8.25	8.11	1.77
Δ rule	1.34	1.26	0.72
Spreading	0.56	0.52	0.66
Pred. GOP	0.49	0.42	0.67
Scene Ch.	0.64	0.35	0.65

Evidently, the use of larger buffers in an ATM switch decreases the cell loss probability. If the cells within a picture period are uniformly distributed and the buffer size is larger than or equal to the number of input links, however, the multiplexing performance depends on the cell rates only. Hence, the effect of increasing the buffer size is expected to be limited. If the cells are not uniformly distributed, for instance

due to delay jitter introduced earlier in the network, the buffer size has indeed influence on the cell loss probability. Figure 5.9 shows the cell loss percentages for multiplexing 5 smoothed sources, while a Bernoulli distribution is enforced within a picture period. This is done by calculating the probability of a cell occurrence from the (smoothed) picture cell rate. The results show that a buffer size of 46 cells is needed for zero cell losses in this experiment, while a buffer size of 5 cells would be sufficient if the cells were uniformly distributed over a picture period.

5.7 Conclusions

From the multiplexing analysis and experiments, it is clear that smoothing of the traffic generated by our reference VBR codec has a positive effect on the multiplexing performance. Indeed, the smoothing technique incorporating scene changes as chosen in Chapter 4 appears to outperform the other smoothing techniques mentioned there, which is remarkable because of the higher burstiness caused by the behaviour at scene changes. The differences in multiplexing performance for the different smoothing techniques, however, are small as long as the burstiness is significantly reduced (i.e. $\beta < 1.5$).

The multiplexing performance depends on the characteristics of the traffic. For the smoothed traffic from our reference VBR coder, these characteristics are merely determined by the input video. Therefore, to evaluate the multiplexing performance, a model for the video is needed. From our experiments, only a rough indication can be given. These experiments imply that an effective use of the network capacity ($>$ 60 % efficiency) with an acceptable cell loss probability can be achieved when about 4 or more smoothed VBR sources are multiplexed. This number of sources can be used in the transport layer for networks (or storage media) that support CBR bit streams only. For multiplexing in the network, more sources are usually considered, so that the efficiency can be even higher.

Chapter 6

Cell Loss Concealment

In all data transmission and storage applications, there is a probability that some of the data will be lost. As was shown in Chapter 2, there are two general causes for these losses. First, the occurrence of noise on the transmission or storage medium may result in single-bit errors. The number of such errors can be controlled, for instance by using proper repeater spacing in the physical layer and error control codes in the transport layer of the OSI model. The second cause of errors is by an overload of the capacity in packet switched networks, resulting in the loss of complete packets. When statistical multiplexing of VBR sources is applied, a certain amount of lost packets (or cells in ATM terminology) cannot be avoided. Although retransmission protocols in the transport layer may compensate for these losses, they would result in delays that are too long for real-time services. Hence, the receiver has to apply concealment techniques to reduce the effects of cell loss.

The impact of a lost cell is very different in the different applications. In general, the more the data is compressed, the more the damage a lost cell will cause. The type of damage depends on the compression techniques applied. Section 6.1 describes the effects of cell loss in the reference VBR MPEG codec when no concealment is applied.

To reduce the visual effects of cell loss, conventional concealment techniques try to make use of the spatial and temporal correlation in video sequences. This is done by copying the lost picture parts from neighbouring parts in the same picture or from corresponding parts in preceding pictures. Section 6.2 gives a review of these techniques.

A better prediction of the lost data can be made when some of the data of the damaged picture region is still present. We propose to use layered coding in order to split the data into several streams which are transmitted separately. When some

data of one stream is lost, the data from the other stream can be used to reconstruct the damaged region. Layered coding is described in Section 6.3.

Cell loss concealment is necessary to reduce the visual impairment of network failure. The success of concealment very much depends on the region where data is lost and on the environment of that region. While it is difficult to capture quality in a quantitative measure, it is even more difficult in the case of visual impairment since a loss of detail will have to be compared with erroneous detail. Therefore, this chapter shows the effects of the different concealment techniques in a practical situation, where a region is impaired at the edge of a moving object, where it is most difficult to conceal.

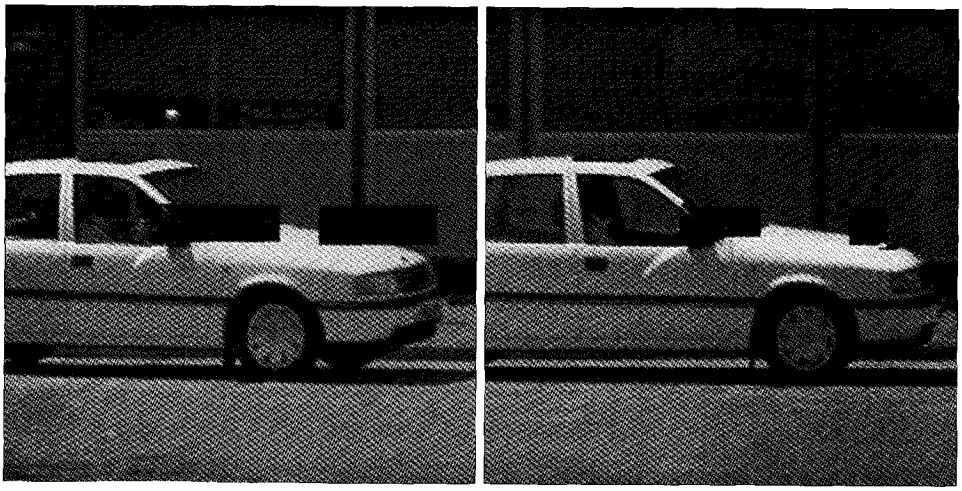
6.1 The Impact of Lost Cells in MPEG

The impact of a lost cell on a decoded picture depends on the size of the picture region the cell contained data of. Because of the use of VLC codes in MPEG, however, the data in the cell is not the only data that is corrupted as a result of cell loss. To be able to correctly decode VLC codes, the decoder has to know where the code word boundaries are, i.e. it needs to be synchronized with the code words in the bit stream. When a part of the bit stream is missing, this is impossible and synchronisation will be lost. In MPEG, resynchronisation occurs by means of start codes at the beginning of each slice. Hence, the impact of a lost cell concerns the rest of the slice currently being decoded and is therefore larger at the beginning than at the end of a slice.

If we assume that no error concealment is applied in the decoder, the lost picture parts are black in the decoded pictures (green when chrominance values are also zero). In Figure 6.1, the effect of a lost cell in an I picture of the sequence *car* is illustrated. One slice corresponds with a horizontal stripe of macroblocks.

Apart from the placement of the lost data, another point of interest is the type of picture to which the lost data belongs. Since we assume no concealment, the impact of a lost cell in a picture is independent of the type of the picture. The effect of loss in a reference picture, however, extends to the pictures which are predicted from that reference. Hence, the effects of loss in a B picture are only visible in that picture, while the effects of a cell loss in an I or B picture extend until a new GOP (or I picture) begins. Figure 6.2a illustrates the effects of the lost cell in Figure 6.1 on the following P picture, and Figure 6.2b on the last B picture of the GOP. We see that the region of the picture where data may be corrupted extends due to motion compensation, but that the regions where the loss is visible is reduced because of intra-coded macroblocks in P and B pictures.

FIGURE 6.1: IMPACT OF LOST CELL ON DECODED I PICTURE IN THE SEQUENCE *car*.



(a) (b)

FIGURE 6.2: IMPACT OF LOST CELL ON (A) FOLLOWING P PICTURE AND (B) LAST B PICTURE IN THE GOP.

An obvious method to reduce the impact of a lost cell is by allowing shorter slices at the encoder. Then, resynchronization occurs faster and less data will be lost. By allowing shorter slices, the number of slices and thus slice headers increases. Each slice header consists of a synchronization word of 32 bytes plus the slice position (8 bits) and the quantizer scaling factor (5 bits). Further, in a new slice, no references to the previous slice can be made. Hence, no predictive coding for motion vectors, macroblock positions and DC coefficient values can be applied. This yields extra costs for the first macroblock in a slice. Consequently, the possibility to achieve an early resynchronization is at the cost of a smaller compression efficiency.

Because of the relative small impact of a lost cell in B pictures, a loss of efficiency can be prevented by allowing very long slices in B pictures and short slices in I and P pictures [104]. Nevertheless, there is still a need to use some kind of concealment technique.

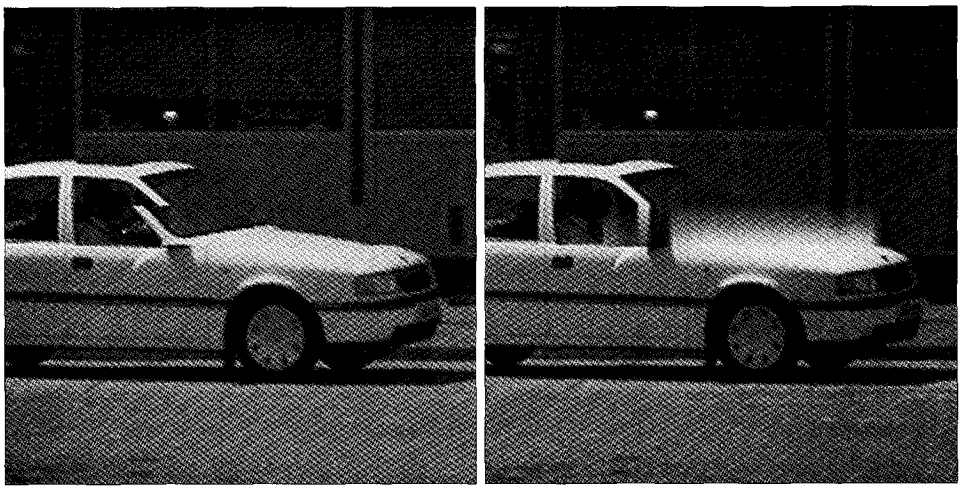
6.2 Conventional Concealment Techniques

In order to conceal the data that is lost in a picture, it can be replaced by data that resembles it. Therefore, a prediction of the lost data will have to be made at the decoder side. There are two ways to predict the contents of a lost picture part. First, because of the spatial correlation, the data can be predicted from neighbouring regions in the picture. Second, the temporal correlation can be exploited by predicting the picture contents from previously decoded pictures. Here, these techniques will be treated separately in the following sections.

6.2.1 Temporal Replacement

The assumption that the regions of a picture for which the data is lost are set to black is very unrealistic. In both software and hardware decoders, the picture memories are overwritten with new data while decoding. Hence, when the data is not overwritten due to lost cells, the old data remains in the memory. Since the correlation between consecutive pictures is very high, the old data may resemble the new data so that a very simple temporal concealment is achieved without actively pursuing this. Figure 6.3a shows the result of this concealment for the losses shown in Figure 6.1.

In stationary parts of a picture, the loss of data is concealed very well by the data of earlier transmitted pictures. In areas of moving objects, however, the errors remain noticeable and very annoying, because of details that are stationary while they ought to be moving. Therefore, a motion compensated prediction would be desired, but the motion vectors of the lost macroblocks are also lost. However, they can be



(a) (b)

FIGURE 6.3: RESULTS OF (A) TEMPORAL AND (B) SPATIAL CONCEALMENT IN THE SEQUENCE *car*.

predicted from the motion vectors of neighbouring blocks. Then, the lost areas can be replaced by blocks of the previous picture shifted by this reconstructed motion vector [105]. The picture quality obtained by using this technique improves in large moving areas, but it drops if the neighbouring blocks have different motion vectors, which happens, for instance, at the edges of moving objects.

6.2.2 Spatial Interpolation

In the absence of a previously decoded picture that is sufficiently correlated with the current one, the corrupted region of a picture can be predicted by adjacent regions in the same picture. In a DCT coding scheme like MPEG, a simple implementation is to copy a block (or macroblock) from the row above or below the corrupted block. The correlation between the blocks, however, does not hold for the details in a block. Therefore, only the coefficients with low spatial frequency should be used. Further, instead of copying, better results can be obtained when these coefficients are predicted by using the coefficients from neighbouring blocks plus a smoothing constraint [106].

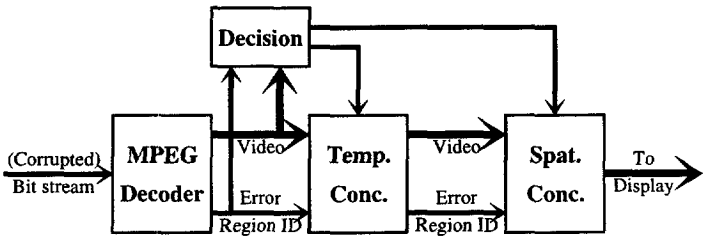


FIGURE 6.4: BLOCK DIAGRAM OF ADAPTIVE CONCEALMENT.

Instead of predicting the DCT coefficients, an alternative approach is to predict the corrupted pixels directly from the pixels of neighbouring blocks. If a change of intensity is present between the blocks above and below the corrupted macroblock, good results are obtained by interpolation. Then, the pixels of a corrupted macroblock are replaced by a linear interpolation between the row of pixels immediately above and below the corrupted macroblock [107]. Hence, a smooth transition between the intensities of the neighbouring blocks is achieved. The result for the losses of Figure 6.1 is shown in Figure 6.3b. The method can be extended by a directional interpolation, to interpolate the sharp edges in the neighbouring blocks [108].

All the concealment schemes described above work better when not only the neighbours above and below, but also the left and right neighbours of the lost macroblock are present. Because of the lost synchronization, this can only be achieved when block interleaving is applied [106], which means that the macroblocks of two adjacent slices are shuffled so that two consecutively transmitted macroblocks belong to different slices. The problem with this technique is that it is not implemented within the MPEG standard. Slice interleaving [109] only prevents adjacent slices from being lost, but it can be implemented as a post-encoding procedure, making it compatible with MPEG.

6.2.3 Adaptive Concealment

Where temporal replacement may fail in scenes with large and irregular motion, spatial interpolation fails to predict the details within a lost block. Hence, an adaptive combination is proposed in [110], which makes a choice between the two techniques based on the local activity within a picture. The overall concealment procedure consists of two stages and is shown in Figure 6.4. First, the corrupted bit stream is decoded and the error regions are identified by a standard MPEG decoder. Based on easily obtained measures from neighbouring macroblocks, the decision is taken as to which concealment technique is used in the error region. Finally, the chosen concealment technique is applied.

6.3 Layered Coding

Since there is always a probability of cell loss in ATM and because of the effects a lost cell may have on the quality of real-time services, ATM allows distinguishing vital data from less critical data by means of the priority bit in the cell header. Further, since different virtual channels (VCs) can have different qualities of service, several VCs can be used for a single service, therin distinguishing several classes of data. Each of these classes defines its own quality of service (QoS) to be specified in the network contract. The techniques to divide the data for a single service into different classes to be transmitted separately are covered by the term *layered coding.*

Layered coding is especially attractive for VBR video for two reasons. First, some parts of the transmitted bit stream are more important than others in terms of their contribution to the visual quality of the reconstructed pictures. For instance, low frequency components are more essential to the visual quality than high frequency components. Second, because of the large impact of a lost cell, a high QoS would be required for the whole cell stream. With VBR, this can only be guaranteed if the network allocates a CBR channel at the peak bit rate of the VBR source. Such an allocation, however, would be relatively expensive.

In addition to cell loss concealment, multi-resolution is another reason to use layered coding. Since this is expected to be an important requirement for future broadcast services [111], it is attractive to combine it with error concealment in a layered coding scheme. It requires a single compression system to facilitate decoding of the video at various levels in terms of spatial or temporal resolution. Multiple levels of temporal resolution can be simply obtained by transmitting the different picture types of MPEG in different layers. Hence, in the following, the term multi-resolution coding is used for the support of multiple spatial resolution levels. For this purpose, the information needed to decode the signal at the lowest resolution level is used as a prediction for the higher levels. Thus, the signal can be compressed better than with *simulcast*, where each level is encoded separately. This approach is used in both subband coding [112] and MPEG-like schemes [111].

Although the techniques used in layered coding for error concealment and multi-resolution seem similar, the requirements with respect to the bit rates and quality may differ. Section 6.3.1 discusses the requirements to the division of the bit rates in both cases. Then, Section 6.3.2 describes the layering techniques that can be used in MPEG and how the choice for a technique influences both coding efficiency and quality. Section 6.3.3 evaluates the possible layering techniques from an error-concealment point of view.

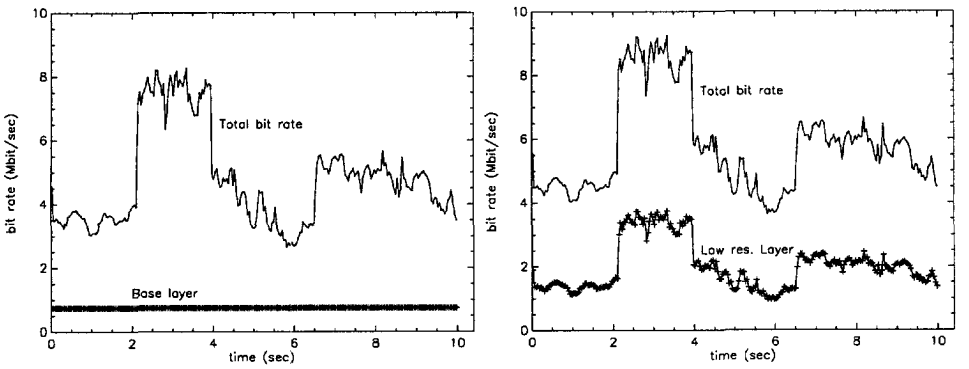


FIGURE 6.5: EXAMPLE OF LOWER LAYER AND TOTAL BIT RATES GENERATED BY A TWO-LAYER CODING SCHEME FOR (A) ERROR CONCEALMENT AND (B) MULTI-RESOLUTION.

6.3.1 Bit Rate Division

Layered coding means that the data needed to reconstruct the video signal is divided into two or more layers. An important issue in layered coding is how the bit stream is divided. This division depends on the requirements imposed on the lower layer(s), depends on the reason for using the layering.

For error concealment, the most important requirement is that if occasionally some of the less important data is lost due to cell losses in the "inexpensive" but unreliable VC, an acceptable reconstruction which is based on information in the "expensive" but reliable VC can still be made at the decoder. As the base layer is never decoded in an entirely stand-alone fashion, the additional bandwidth necessitated by using the concept of layered coding to achieve occasional error concealment must be negligible. Further, since the base layer is transmitted with a high QoS, the costs for this layer are high. The bit rate in this layer should thus be as low as possible and constant, so that resources can be allocated in advance. An example of the bit rates of such a two-layer error concealment scheme is shown in Figure 6.5a.

In multi-resolution coding, the lower layer is decoded stand-alone by some users. The resulting quality should conform to the quality specified for the corresponding resolution level. Hence, when a constant quality is also required for the lower levels, VBR in the lower layer(s) is inevitable. Further, the requirements on the base layer quality may yield a higher overall bandwidth, which is not a problem as long as compression is optimized and the overall bit rate is lower than when simulcast is applied. An example of the bit rates is shown in Figure 6.5b.

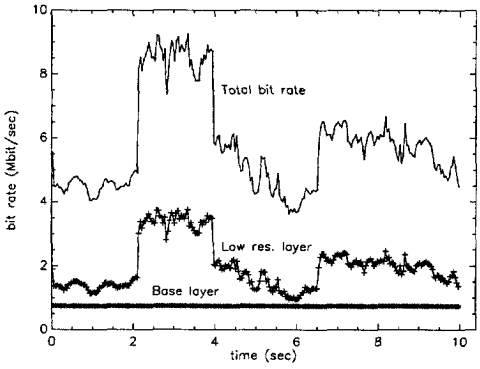


FIGURE 6.6: THREE-LAYER SOLUTION FOR AN ERROR CONCEALMENT MULTI-RESOLUTION CODER.

When both multi-resolution and error concealment are required in a layered scheme, a combination of techniques should be applied. There are three ways to do this. First, it is possible to use the base layer in an error-concealment layered coder as the low resolution layer for multi-resolution coding. Then, a low quality for this layer must be accepted because of the constraints on the base layer imposed by the error-concealment requirement. The second scenario is to use the normal multi-resolution coder. When data from the higher layer(s) is lost, error concealment can be applied by using the information from the lower layer(s) only. However, if the lowest layer in multi-resolution coding has either a variable or a relatively high bit rate, it is too expensive to protect this layer adequately enough for error concealment. A third option is to split the lowest layer of the multi-resolution coder into a CBR base layer for error concealment, and a VBR low resolution layer as is shown in Figure 6.6. This yields an extra inefficiency with respect to a 2 layer scheme because of the extra data needed to split the low resolution layer. The choice for one of the scenarios depends on how efficiently layering can be applied, and on how this efficiency is affected by the demand for multi-resolution. This, in turn, depends on the techniques to be applied for layering, as is shown in the following section.

6.3.2 Layering Techniques

Although the objective of early work on layered coding was error concealment, most of the proposed schemes were based on a stand-alone coder at the base layer. A twin coding scheme such as proposed in [113] can be simply adjusted to support multi-resolution coding by adding up- and downsampling, as shown in Figure 6.7. Further,

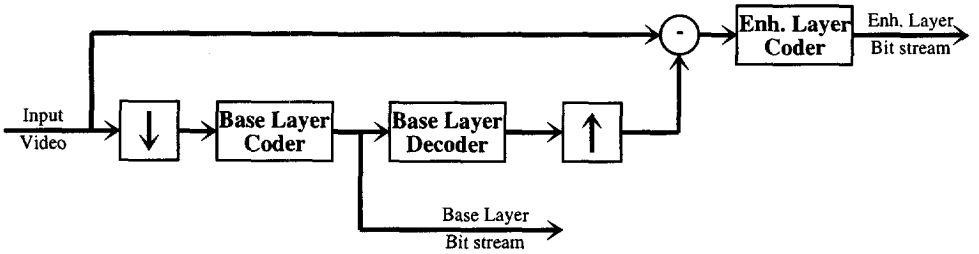


FIGURE 6.7: BASIC COMPATIBLE TWIN CODING SCHEME.

when a standard encoding algorithm is used as the base layer encoder, compatibility with this standard is obtained, which is a desired feature in developing new standards [111]. This is the reason why this kind of layered coding is adopted in MPEG-2 as one of the layering capabilities, called *spatial scalability*.

When the information transmitted in the enhancement layer is analysed, it appears that the DCT coefficients to be transmitted are residual quantization distortions of the base layer [114]. Further, since the base layer operates as a stand-alone coder, the information of the enhancement layer is not used in the prediction loop of the encoder. Therefore, the efficiency of the enhancement layer coder can be improved by using this information. Consequently, a general approach to a layered coding scheme is given in Figure 6.8, where the coefficients resulting from quantizer Ql are transmitted in the base layer and the quantization errors are requantized by Qh before transmission in the enhancement layer. In this scheme, it is possible to quantize only a subset of DCT coefficients, for instance the 4x4 lower coefficients, by Ql. Then, the rest of the coefficients are transmitted in the enhancement layer only. Multi-resolution can be achieved by applying a smaller, in this case 4x4, inverse DCT on the subset in a low resolution decoder.

When switch (a) in Figure 6.8 is closed, the motion compensation is performed on both layers, yielding a higher efficiency. In this case, however, a decoder cannot make the same prediction as the encoder if only the base layer is decoded. When (a) is open, only the base layer is used for motion compensation, so that a correct prediction can be made in a decoder using the base layer only. Instead of closing switch (a), a separate prediction loop in the enhancement layer is proposed in [115], thus improving the efficiency. Since our objective is to provide error concealment by means of efficient layering, we consider only the case with switch (a) closed.

Although an error concealment, and thus efficient layered coding scheme is discussed here, also the performance of stand-alone decoding of the base layer is considered. This is relevant to the performance when large bursts of data are lost. Further, it gives an indication of the overhead needed for a multi-resolution scheme.

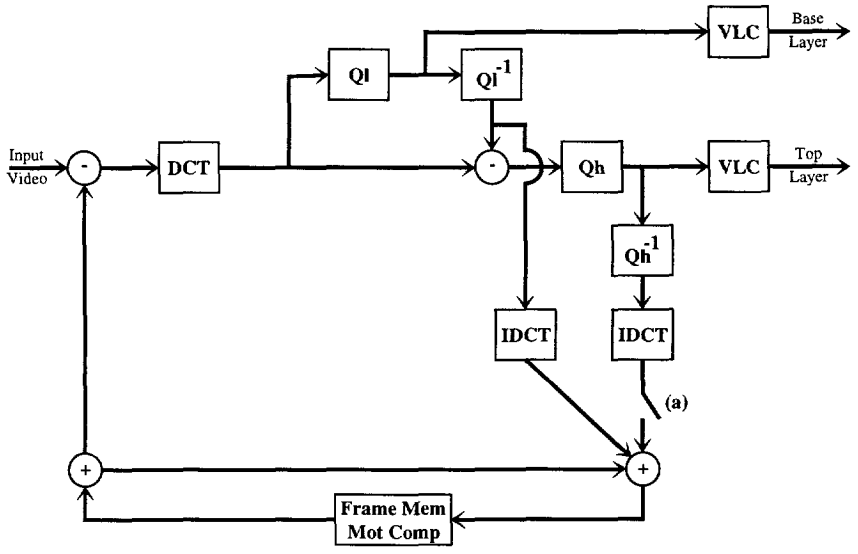


FIGURE 6.8: A GENERAL LAYERED CODING SCHEME WITH LAYERING IN THE TRANSFORM DOMAIN.

Efficiency

The efficiency of a layered coding scheme can be defined as the bit rate generated by a single-layer VBR coder divided by the total bit rate of the layered VBR coder. Alternatively, the inefficiency is the amount of overhead bits needed for layering compared with the single-layer bit rate. In our general layered coding scheme, we assume that the enhancement layer transmits only the refinement of the DCT coefficients. Side information such as macroblock types and motion vectors are transmitted in the base layer. Hence, the efficiency depends only on the encoding of the DCT coefficients in both layers.

In MPEG, the DCT coefficients are encoded using run-length coding. This means that the occurrences of zeros are exploited best. Hence, the efficiency is optimized when each DCT coefficient is zero in either the base or the enhancement layer or both. This can be achieved by taking the same quantizer scaling factor in Ql as in Qh. Then, layering is achieved by sending some of the quantized coefficients in the base layer and the rest in the enhancement layer. When a CBR bit stream needs to be transmitted in the base layer, the control algorithm controls only the number of coefficients to be transmitted in the base layer, not the quantizer scaling factor [116].

The efficiency of the layered scheme with equal quantizer scaling factors depends only on the amount of overhead needed to transmit the coefficients in two layers instead of one. This overhead depends on how the coefficients are divided over the two layers. Here, we distinguish two methods for this division.

The first method to split the bit stream into two layers is by taking a subset of DCT coefficients for which the data is transmitted in the base layer. If this subset is chosen according to the zigzag scan, the run-length codes of a single-layer MPEG coder can be simply divided over the two layers. In this case, the decoder reads the base layer bit stream up to the end-of-block marker, then it switches to the enhancement layer to read the rest of the coefficients until another end-of-block marker is read. Consequently, the overhead consists of the sequence, GOP, picture and slice headers, which are needed for synchronisation, and the end-of-block markers. An advantage of this method is that the syntax of the base layer is compatible with the syntax of the single-layer coder if at least one run-length code is transmitted per block. If a smaller IDCT is to be used in a low resolution decoder, a different scanning technique is required, so that the coefficients needed for the smaller IDCT are scanned first. This leads to an increase in the overhead required for layering and the loss of compatibility.

The second method to split the bit stream is by transmitting a fixed number of run-length codes per block in the base layer. In this case, the decoder does not need an end of block marker in the base layer, since it switches to the enhancement layer after reading the specified number of run-length codes. Therefore, the overhead is reduced to the synchronization headers. This kind of layered coding, where an extra code determining the number of run-length codes for the base layer is transmitted in the slice header, is also standardized in MPEG-2 where it is called *data partitioning* [117]. A drawback of this technique is that no knowledge is present about the layer in which the information of a certain coefficient is transmitted. Further, the syntax of the base layer is not compatible with the syntax of the single-layer coder.

Stand-alone Base Layer: Drift

The quality of the pictures obtained by decoding the base layer depends only on the amount of data that is transmitted in the base layer. As an indication, Figure 6.9a shows the result for a decoded I picture in the sequence *car* if 3 run-length codes are transmitted in the base layer. Figure 6.9b shows the result if the information for the lowest 6 coefficients (plus an end-of-block marker) are transmitted in the base layer. Both techniques require approximately 40 % of the bit rate to be assigned to the base layer in the sequence *car*.

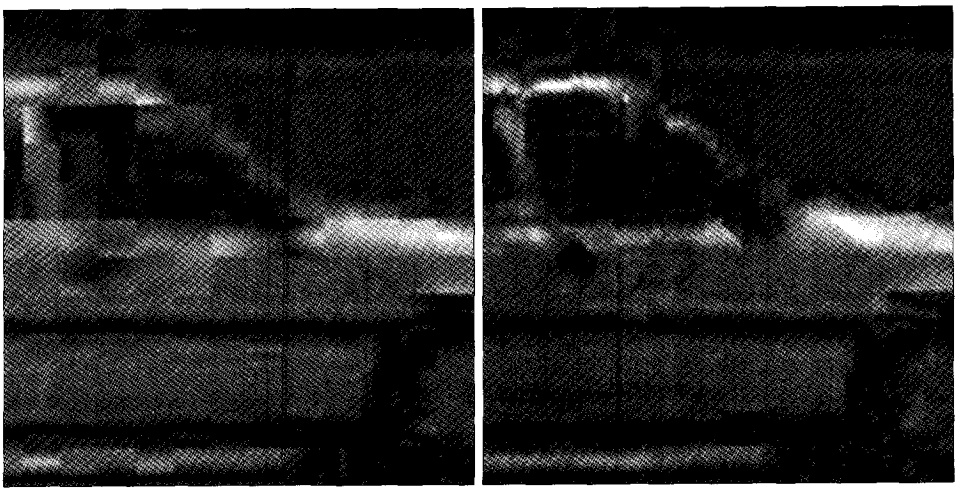
For P and B pictures, motion compensation is applied, but the prediction in the encoder is based on the information sent in both layers, and can, therefore, not be identical to the prediction in a decoder that uses only the base layer for decoding. The mismatch between the two predictions increases with the number of pictures

(a) (b)

FIGURE 6.9: QUALITY IN I PICTURE IF ONLY THE BASE LAYER IS DECODED (A) THREE RUN-LENGTH CODES ARE TRANSMITTED IN THE BASE LAYER, (B) THE INFORMATION FOR SIX COEFFICIENTS ARE TRANSMITTED IN THE BASE LAYER.

encoded since the last intra-coded picture. This phenomenon is called *drift*. The prevention of drift is the major cause of the inefficiency of multi-resolution coders.

To analyse the amount of drift, we first consider a prediction loop without motion compensation. In this case, the prediction loop can be performed in the DCT domain and each DCT coefficient is predicted by a coefficient quantized with Qh. Hence, the cause of the drift is the difference between Ql and Qh for each coefficient. This drift can be removed by taking the same quantizer scaling factor for Qh and Ql, as used in the efficient layered coding scheme discussed earlier. If this is done, only the coefficients for which not all information is transmitted in the base layer of the reference picture are affected by drift. Therefore, if the information for a fixed number of coefficients is transmitted in the base layer, no drift occurs and the result for P and B pictures is similar to the result for I pictures as shown in Figure 6.9b. However, if a fixed number of run-length codes are transmitted in the base layer, drift occurs if a certain coefficient of an I picture is transmitted in the base layer, while its prediction error in the following P pictures is not transmitted in the base layer. The visual effect of this kind of drift is shown in Figure 6.10a for the sequence *car*.

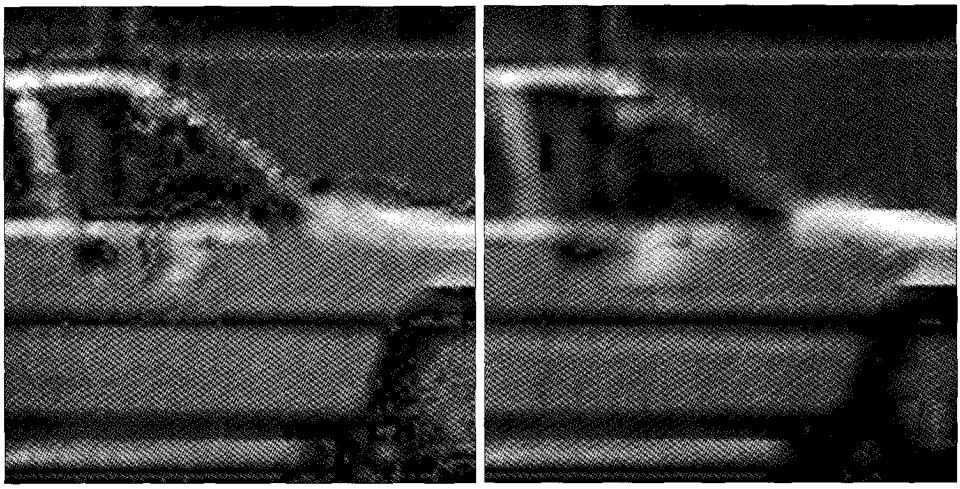


(a) (b)

FIGURE 6.10: QUALITY IN 14TH P PICTURE IF ONLY THE BASE LAYER IS DECODED IN WHICH THREE RUN-LENGTH CODES ARE TRANSMITTED (A) WITHOUT AND (B) WITH MOTION COMPENSATION.

If motion compensation is applied in the prediction loop, we should consider the DCT transformation of a shifted block. Since the DCT is block based, the coefficients do not represent pure frequencies and discontinuities occur at the DCT block boundaries. Consequently, the high frequency DCT coefficients of a shifted block depend on all coefficients of the referenced picture [118]. This causes an extra drift component, which is shown in Figure 6.10b for the case where three run-length codes are transmitted in the base layer, and in Figure 6.11a if the information for six coefficients is transmitted in the base layer. In the latter case, however, the drift in the higher coefficients can be removed by removal of these coefficients at the expense of an extra DCT and inverse DCT operation in the decoder. The result is shown in Figure 6.11b. Since the low frequency DCT coefficients of a shifted block also depend on the higher frequency DCT coefficients in the reference picture, which are not available in the base layer decoder, a small amount of drift is still present in Figure 6.11b. A detailed analysis of the drift effect is found in [118]. It is clear that it depends on the amount of motion in the scene and on the scene contents.

It is possible to take a smaller IDCT of the low frequency coefficients to achieve a lower spatial resolution for multi-resolution coding [119]. Hence, when the coefficients needed for the smaller IDCT are transmitted in the base layer, a multi-spatial resolution scheme is obtained. To do this, however, the zigzag scan of the coefficients



(a) (b)

FIGURE 6.11: QUALITY IN 14TH P PICTURE IF ONLY THE BASE LAYER IS DECODED, IN WHICH THE INFORMATION OF SIX COEFFICIENTS IS TRANSMITTED (A) WITHOUT AND (B) WITH REMOVAL OF THE HIGH FREQUENCY COEFFICIENTS.

will have to be replaced by a scanning that first transmits the coefficients needed for the smaller IDCT.

If a smaller IDCT is used in a low resolution decoder, an extra cause of drift occurs since the motion vectors at low resolution are more fractional [118]. This drift can be avoided by decoding at full resolution and downscaling afterwards.

Conclusions

Although efficiency and the prevention of drift are contradictary demands for a layered coding scheme, it is possible to apply an efficient layering by transmitting the information for a fixed number of coefficients in the base layer so that the drift in these coefficients is limited. Further, the drift in the higher coefficients can be removed by removal of these coefficients with an extra DCT and IDCT operation.

By transmitting the information of a fixed number of coefficients in the base layer, the base layer bit stream has a variable bit rate. To obtain a constant bit rate, the number of coefficients can be used to control the bit stream. If a proper control algorithm is applied, the number of coefficients remains nearly constant within a scene. Consequently, for both CBR and VBR base layers, the efficient layered coding scheme is applicable.

The different solutions for layered coding presented here are summarized in Table 6.1, they compare the inefficiencies in a qualitative fashion. Obviously, the quantitative inefficiency depends on the encoded picture sequence and the quality at which it is encoded. However, the values given in the table give a rough indication based on a wide variety of sequences and quality levels.

Scheme	Objective	Inefficiency	Drift
Split	Concealment	<1%	Medium
Split+EOB	Compatibility	5%	Low
small IDCT	Multi-Resolution	10%	High
Twin	Comp.+Multi-Res.	20%	No

TABLE 6.1: LOSS IN EFFICIENCY FOR DIFFERENT LAYERING TECHNIQUES.

From Table 6.1, we see that using data partitioning to split the bit stream into two layers ("Split") is very efficient, but the costs of reaching compatibility and less drift in a stand-alone base layer decoder by adding EOB markers ("Split+EOB") are limited. The possibility for low spatial resolution decoding by means of a smaller IDCT ("small IDCT") yields a higher degree of inefficiency because of the need for another scanning scenario than the zigzag scan. Since the drift is also a problem in this scheme, we will not consider this solution for layered coding. As a reference, the inefficiency of a twin codec ("Twin") is also shown.

6.3.3 Evaluation

In Section 6.3.1, we saw that the base layer for error-concealment layered coding should have a constant bit rate containing a small percentage of the overall bit rate. The percentage of the total bit rate for the base layer depends on the number of run-length codes or the number of coefficients for which information is transmitted there. This dependency is shown in Figure 6.12 for the sequence *car*. For less detailed sequences, the curves are even steeper, since most information will be concentrated in the lower coefficients.

Clearly, for error-concealment purposes, only a few run-length codes or coefficients can be transmitted in the base layer. Therefore, the performance in a cell loss environment for two cases are shown in Figure 6.13. It is assumed here that cell losses do not occur in the base layer since this layer is sufficiently protected. Figure 6.13a shows the effect of a lost cell in an I picture if only the DC coefficient is transmitted in the base layer. Figure 6.13b shows the effect if two additional run-length codes are transmitted in the base layer. The transmission of information for a fixed number of

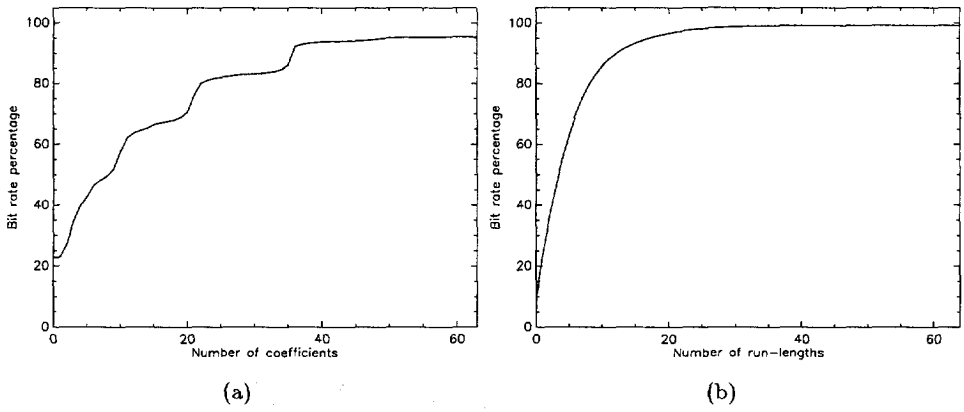


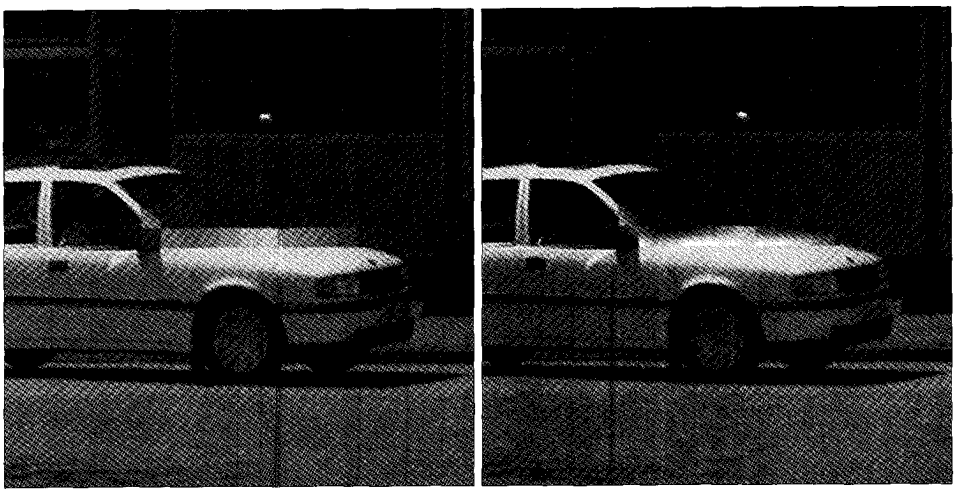
FIGURE 6.12: PERCENTAGE OF THE BASE LAYER BIT RATE AS A FUNCTION OF (A) THE NUMBER OF TRANSMITTED COEFFICIENTS AND (B) THE NUMBER OF TRANSMITTED RUN LENGTHS IN THE BASE LAYER.

coefficients in the base layer is less efficient and yields visually comparable results. Further, since drift is not relevant in an error-concealment environment and the bit rates in the base layer are not high enough for an acceptable quality in a stand-alone base layer decoder, only the transmission of a fixed number of run-length codes is considered here.

From Figure 6.13, it can be seen that the transmission of at least two run-length codes in the base layer is needed to provide better concealment results than the methods proposed in Section 6.2. For P and B pictures, however, the correct motion vector is always transmitted in the base layer. Figure 6.14a shows the effect of a lost cell in an I picture on the following P picture and Figure 6.14b shows the effect of a lost cell in the P picture itself. From these figures, we conclude that no coefficients need to be transmitted for error concealment in P and B pictures. For I pictures, however, at least 2 run-length codes together with the DC coefficient are needed. For the sequence *car*, this resulted in the transmission of 15 % of the bit rate in the base layer, which is acceptable in an error-concealment layered coding scheme.

6.4 Conclusions

Layered coding provides the means to conceal cell loss better than by using conventional techniques. Therefore, the traffic generated by our reference coder is split into two layers using data partitioning. The base layer transmits only motion vectors for P and B pictures and the DC coefficient plus an additional number of run-length

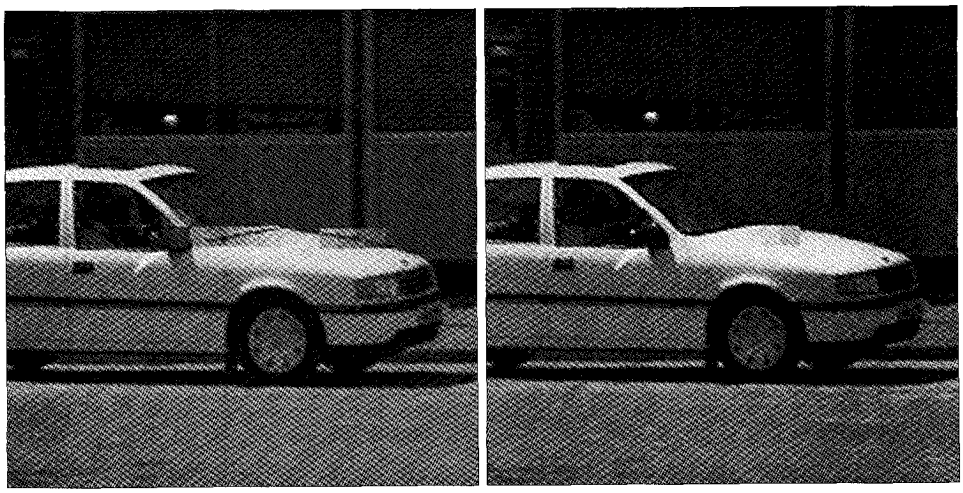


(a) (b)

FIGURE 6.13: EFFECT OF AN ERROR IN THE ENHANCEMENT LAYER FOR I PICTURES DEPENDENT ON THE NUMBER OF COEFFICIENTS IN THE BASE LAYER. (A) ONLY THE DC COEFFICIENT IS TRANSMITTED IN THE BASE LAYER, (B) THE DC COEFFICIENT PLUS TWO RUN-LENGTH CODES ARE TRANSMITTED IN THE BASE LAYER.

codes for I pictures. This number is determined by a control algorithm which ensures that the bit stream of the base layer has a constant bit rate. The characteristics of the total bit stream of the two layers remain unchanged, apart from a small offset.

Despite of the concealment achieved with layered coding, a loss of picture quality is inevitable in the case of a lost cell. Since VBR compression should offer a constant quality of reconstructed video, cell loss is an undesirable event. However, the extent to which the visual quality is corrupted by occasional impairments is not known. Subjective tests on impaired (and concealed) video will have to be performed in order to determine the extent to which the impairments should be concealed and how often an impairment is tolerable.



(a) (b)

FIGURE 6.14: (A) EFFECT OF LOST CELL IN I PICTURE ON FOLLOWING P PICTURE IF ONLY THE DC COEFFICIENT IS TRANSMITTED IN THE BASE LAYER. (B) EFFECT OF LOST CELL IN P PICTURE IF ONLY MOTION VECTORS ARE TRANSMITTED IN THE BASE LAYER.

Chapter 7

Discussion

In this thesis, we have developed a robust reference VBR video codec with proper network adaptation. By generating a variable bit rate, the overall compression ratio is better than with CBR coding. It is, however, difficult to take advantage of the low mean bit rate of this kind of VBR source at transmission and in storage. This chapter discusses the complexity and applicability of the proposed techniques in practical situations. It compares our reference VBR codec with CBR codecs and with other techniques proposed for VBR coding. Further, it contains suggestions for future development in both network management and VBR video.

Although the concept of VBR video compression differs from the concept of CBR video compression, for both techniques the performance depends on the complexity of the techniques applied. To compare VBR with CBR sources, a distinction should be made between the techniques used for compression, and the adaptation to the network. The simplest way to apply VBR compression is by using open-loop VBR coding, which is as complex as the simplest CBR technique. More advanced VBR algorithms aim at a constant quality and are more complex because of the non-linear relation between distortion and quality. However, since more advanced CBR coders aim at optimizing quality, they are coping with the same problem. Therefore, the complexity of VBR compression in general can be compared with the complexity of CBR compression. In practical situations, the traffic parameters of VBR sources need to be controlled. Such a control can be implemented by a simple feedback loop, as in the simplest CBR techniques. If smoothing is applied as proposed in this thesis, the number of bits generated in the last picture of each type will have to be stored, a calculation of the output bit rate will have to be made, and buffering will have to be applied. In advanced CBR coders, however, bit allocation between the different picture types is also based on the number of bits generated in the last picture of each type. Further, buffering is also needed and the size of the buffer is

approximately as large as the buffer needed for smoothing VBR sources. For both CBR and VBR coders, a larger buffer is necessary and a larger delay is introduced if more B pictures are applied in a GOP. In conclusion, the difference in complexity between VBR and CBR coders is negligible.

One of the most straightforward applications of VBR video coding is for storage on randomly accessible devices. The low mean bit rate of VBR sources is reflected in the amount of storage space a source needs for recording. If the storage device is shared with other services to be recorded, as, for instance, in a computer hard-disk, our reference VBR coder is particularly suitable since it aims at reducing the bit rate as much as possible, so that the capacity used is as small as possible. If the storage device is used for one particular source only, as, for instance, in the first DVD applications where movies are recorded on a single DVD, it is better to use two-pass VBR coding. This technique aims at optimizing the quality of the whole movie, with the total capacity as a constraint. In fact, this technique uses bit allocation and is therefore more related to CBR coding than to VBR coding as defined in Chapter 3. If our reference VBR codec is used, no guarantee can be given in advance about the capacity needed to store a movie.

Another application of VBR video is for transmission on a network that supports sources with a varying bit rate, i.e. packet-switched networks. Of these kinds of networks, we distinguish two types: the particular data-communication networks like the Internet, which offer no real-time services, and the hybrid networks like ATM, which support both data-communication and real-time services.

Video transmission over data-communication networks like the Internet is possible in two ways. First, off-line transmission treats compressed video as a file to be transmitted. At the receiver the file is stored before playback. Our reference codec is suitable for this kind of video compression, as stated above. Second, real-time transmission over the Internet is possible, but a high probability of packet loss will have to be accounted for. The smoothing in the network adaptation may reduce these losses, but no guarantee can be given about the quality of service obtained. The layering applied in the codec reduces the impact of lost packets as long as they occur in the enhancement layer. However, since no bandwidth can be allocated for the base layer, losses occur there as well, and these have a much higher impact. Consequently, data-communication networks like Internet are not suitable for real-time video transmission, which holds for our reference VBR codec as well as for all codecs, although the performance for our reference VBR codec is expected to be less bad.

Packet-switched networks that do support real-time services, like ATM, use statistical multiplexing and are therefore particularly suitable for our reference VBR codec. Because of statistical multiplexing, however, there is always a probability that the

offered traffic is higher than the network capacity, which leads to a loss of information. Although the layering in our reference VBR coder reduces the impact of these losses, network management, including resource allocation and monitoring the characteristics of the traffic for each source, is needed to reduce the probability of information loss.

Resource allocation is based on the traffic parameters defined in the network contract for each source. Currently proposed allocation algorithms (Call Admission Control (CAC) algorithms in ATM) are based on the peak and mean (or rather: sustainable) cell rates of the sources only. This thesis has shown that a burstiness parameter is very important in the description of the traffic from a VBR video source and in the achievable multiplexing performance for these sources. Therefore, resource allocation algorithms will have to take the burstiness into account and policing algorithms will have to be developed to control this parameter. One possible implementation of such a burstiness policing function is to allow only a limited number of decreases in cell inter-arrival times (increases in cell rate) per time unit, which is managed by, for instance, a leaky bucket.

Many researchers are sceptic about VBR compressed video and statistical multiplexing in ATM because of the non-zero probability of cell loss caused by traffic overload. With proper models for the VBR traffic, however, these probabilities can be calculated accurately enough to dimension the network so that a cell loss probability is achieved which is as low as the probability of errors in a conventional channel. An additional possibility not yet considered in ATM is the use of a feedback channel. If there is danger of congestion in the network because of an increase in network load, the network may use this channel to temporarily prohibit users from transmitting at a higher rate than they are currently doing. Such a prohibition is similar to CBR coding, and it is therefore a temporary limitation of the VBR concept.

For transmission and storage media that make use of fixed bandwidth channels, the use of VBR coding seems inappropriate. However, in an environment in which the video will be transmitted and stored in several different media, it can be advantageous to use VBR. Especially if smoothing is applied, the peak bit rate of a VBR source is in the range of the bit rate of a comparable CBR source. Therefore, the allocation of a CBR channel for a VBR source is no worse than applying CBR coding in the first place. In a later stage, advantage can be taken of the low mean bit rate of the VBR source if it is stored on a randomly accessible device, transmitted over ATM, or multiplexed with other sources on a CBR channel.

If multiple VBR sources are multiplexed on a CBR channel, there are two possible scenarios. In the first, the multiplexer is at the same location as the video coders. Then, it may influence the bit rate generated by the different sources. In the extreme case, this leads to multi-programme video compression, which is more related to CBR than to VBR coding, since bit allocation among the different sources is applied. For

a small number of sources, this kind of coding outperforms VBR coding, but if the video is transmitted or stored on other media later, it is better to use our reference VBR coder with network adaptation. In the second scenario, the multiplexer can have no influence on the video coders, since they are located elsewhere. In that case, our reference VBR codec with network adaptation needs to be applied. Then, an efficient use of the channel (> 60 %) can be achieved when at least foour sources are multiplexed.

If statistical multiplexing of VBR sources is applied, there is always a probability of information loss. Because of the layering, the impact of these losses is reduced. Further, if these losses are prevented by control algorithms or feedback channels, the quality of the reconstructed video is affected as well. Research will have to be applied in order to determine how often both (concealed) errors and control actions are tolerable in a video service.

Bibliography

[1] A. Tanenbaum, *Computer Networks.* Prentice Hall, 2nd ed., 1989.

[2] H. Nyquist, "Certain topics in telegraph transmission theory," *AIEE Transactions*, vol. 47, pp. 617–644, 1928.

[3] N. Jayant and P. Noll, *Digital Coding of Waveforms.* Englewood Cliffs, NJ: Prentice Hall, 1984.

[4] T. Berger, *Rate Distortion Theory: a mathematical basis for data compression.* Prentice Hall, 1971.

[5] A. Jain, "Image data compression: a review," in *Proceedings of the IEEE*, pp. 1349–1389, Mar. 1981.

[6] K. Ramchandran, A. Ortega, and M. Vetterli, "Bit allocation for dependent quantization with applications to MPEG video codec," in *International Conference on Acoustics, Speech and Signal Processing ICASSP'93*, Mar. 1993.

[7] G. Keesman, I. Shah, and R. Klein-Gunnewiek, "Bit-rate control for MPEG encoders," *Signal Processing: Image Communication*, vol. 6, pp. 545–560, Feb. 1995.

[8] J. Darragh and R. Baker, "Fixed distortion subband coding of images for packet-switched networks," *IEEE Journal on Selected Areas in Communications*, vol. 7, pp. 789–800, June 1989.

[9] M. Miyahara, K. Kotani, and V. Algazi, "Objective picture quality scale (PQS) for image coding," *SID digest*, pp. 859–862, 1992.

[10] S. Westen, R. Lagendijk, and J. Biemond, "Perceptual image quality based on a multiple channel HVS model," in *International Conference on Acoustics, Speech and Signal Processing ICASSP'95*, vol. 4, pp. 2351–2354, May 1995.

[11] C. Shannon, "The mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423 and 623–656, 1948.

[12] W. Stallings, *Data and Computer Communications.* Macmillan Publishing Company, third ed., 1991.

[13] K. Kovarik and P. Maveddat, "Multi-rate ISDN," *IEEE Communications Magazine*, vol. 32, pp. 48–54, Apr. 1994.

[14] R. Frederick, "Experiences with real-time software video compression," in *Sixth International Workshop on Packet Video*, pp. F1.1–F1.4, VISICOM, Sept. 1994.

[15] H. Jinzenji, S. Azegami, and T. Tajiri, "Audio and video communication system for Internet: "a server oriented communication system"," in *Sixth International Workshop on Packet Video*, pp. F2.1–F2.4, VISICOM, Sept. 1994.

[16] C. Aras, J. Kurose, D. Reeves, and H. Schulzrinne, "Real-time communication in packet-switched networks," *Proceedings of the IEEE*, vol. 82, pp. 122–139, Jan. 1994.

[17] M. de Prycker, *Asynchronous Transfer Mode, solution for Broadband ISDN.* New York: Ellis Horwood, 1991.

[18] H. Händel, N. Huber, and S. Schröder, *ATM Networks, Concepts, Protocols, Applications.* Reading, Massachusetts: Addison Wesley, second ed., 1994.

[19] A. Ortega, M. Garrett, and M. Vetterli, "Toward joint optimization of VBR video coding and packet network traffic control," in *5th International Workshop on Packet Video*, VISICOM, Mar. 1993.

[20] K. Sadashige, "Transition to digital recording: An emerging trend influencing all analog signal recording applications," *SMPTE Journal*, pp. 1073–1078, Nov. 1987.

[21] P. de With, *Data Compression Techniques for Digital Video Recording.* PhD thesis, Delft University of Technology, 1992.

[22] E. Frimout, *Fast Playback of Helical-scan Recorded MPEG Video.* PhD thesis, Delft University of Technology, 1995.

[23] P. Westerink, *Subband Coding of Images.* PhD thesis, Delft University of Technology, 1989.

[24] C.-H. Hsieh and J.-S. Shue, "Frame adaptive finite-state vector quantization for image sequence coding," *Signal Processing: Image Communication*, vol. 7, pp. 13–26, Mar. 1995.

[25] L.Chiariglione, "The development of an integrated audiovisual coding standard: MPEG," *Proceedings of the IEEE*, vol. 83, pp. 151–157, Feb. 1995.

[26] ISO/IEC 11172-2, IS, *Coding of moving pictures and associated audio for digital storage media up to about 1.5 Mbit/s - Part 2: Video.*

[27] ISO/IEC 11172-3, IS, *Coding of moving pictures and associated audio for digital storage media up to about 1.5 Mbit/s - Part 3: Audio.*

[28] ISO/IEC 11172-1, IS, *Coding of moving pictures and associated audio for digital storage media up to about 1.5 Mbit/s - Part 1: System.*

[29] B. Haskell, F. Mounts, and J. Candy, "Interframe coding of videophone pictures," *Proceedings of the IEEE*, vol. 60, pp. 792–800, July 1972.

[30] B. Girod, "The efficiency of motion-compensating prediction for hybrid coding of video sequences," *IEEE Journal on Selected Areas in Communications*, vol. 5, pp. 1140–1154, Aug. 1987.

[31] H. Mussmann, P. Hirsch, and H. Grallert, "Advances in picture coding," in *Proceedings IEEE*, vol. 73, pp. 523–547, Apr. 1985.

[32] M. Bierling, "Displacement estimation by hierarchical block matching," in *Visual Communications and Image Processing '88*, SPIE, pp. 942–953, Nov. 1988.

[33] N. Ahmed, T. Natarjan, and K. Rao, "Discrete cosine transform," *IEEE Transactions on Computers*, vol. 23, pp. 90–93, Jan. 1974.

[34] K. Ramchandran and M. Vetterli, "Rate-distortion optimal fast thresholding with complete JPEG/MPEG decoder compatibility," *IEEE Transactions on Image Processing*, vol. 3, pp. 700–704, Sept. 1994.

[35] P. Pancha and M. El Zarki, "MPEG coding for variable bit rate video transmission," *IEEE Communications Magazine*, vol. 32, pp. 54–66, May 1994.

[36] J. van Ewijk, "Bitrate control voor een MPEG-1 encoder," afstudeerverslag (in dutch), Information Theory Group, Dept of Electrical Engineering, Delft University of Technology, 1996.

[37] P. Barten, "Evaluation of subjective image quality with the square-root integral method," *Journal of the Optical Society of America A*, vol. 7, pp. 2024–2031, Oct. 1990.

[38] A. van Meeteren and J. Vos, "Resolution and contrast sensitivity at low luminances," *Vision Research*, vol. 12, pp. 825–833, May 1972.

[39] C. Carlson, "Sine-wave threshold contrast sensitivity function: dependence on display size," *RCA Review*, vol. 43, pp. 675–683, Dec. 1982.

[40] L. Olzak and J. Thomas, "Seeing spatial patterns," in *Handbook of Perception and Human Performance* (K. Boff, L. Kaufman, and J. Thomas, eds.), ch. 7, Wiley, New York, 1986.

[41] A. Vassilev, "Contrast sensitivity near borders: significance of test stimulus form, size and duration," *Vision Research*, vol. 13, pp. 719–730, 1973.

[42] S. Daly, "The visible differences predictor an algorithm for the assessment of image fidelity," in *Digital Images and Human Vision* (A. Watson, ed.), ch. 7, pp. 179–208, MIT Press, Cambridge, Massachusetts, 1993.

[43] T. Carney, S. Klein, and Q. Hu, "Visual masking near spatiotemporal edges," in *Human Vision and Electronic Imaging* (B.E.Rogowitz and J. Allebach, eds.), vol. 2657, pp. 393–402, SPIE, 1996.

[44] N. Nill, "A visual model weighted cosine transform for image compression and quality assessment," *IEEE Transactions on Communications*, vol. COM-33, pp. 551–557, June 1985.

[45] S. Klein, A. Silverstein, and T. Carney, "Relevance of human vision to JPEG-DCT compression," in *Human Vision, Visual Processing, and Digital Display III* (B.E.Rogowitz, ed.), vol. 1666, pp. 200–215, SPIE, 1992.

[46] H. Peterson, H. Peng, J. Morgan, and A. Watson, "Quantization of color image components in the DCT domain," in *Human Vision, Visual Processing, and Digital Display II* (B.E.Rogowitz, M. Brill, and J. Allebach, eds.), vol. 1453, pp. 210–222, SPIE, 1991.

[47] H. Peterson, "DCT basis functions visibility in RGB space," in *Society for Information Display Digest of Technical Papers* (J. Morreale, ed.), Society for Information Display, 1992.

[48] A. Ahumada and H. Peterson, "Luminance-model-based DCT quantization for color image compression," in *Human Vision, Visual Processing, and Digital Display III* (B.E.Rogowitz, ed.), vol. 1666, pp. 365–374, SPIE, 1992.

[49] H. Peterson, A. Ahumada, and A. Watson, "An improved detection model for DCT coefficient quantization," in *Human Vision, Visual Processing, and Digital Display IV* (B.E.Rogowitz and J. Allebach, eds.), vol. 1913, SPIE, 1993.

[50] J. Katto, K. Onda, and Y. Yasuda, "Variable bit-rate coding based on human visual system," *Signal Processing: Image Communication*, vol. 3, pp. 321–331, Sept. 1991.

[51] A. Basso, I. Dalgic, F. Tobagi, and C. van den Branden Lambrecht, "Study of MPEG-2 coding performance based on a perceptual quality metric," in *PCS'96 International Picture Coding Symposium*, vol. 1, pp. 263–268, Mar. 1996.

[52] J. Saghri, P. Cheatham, and A. Habibi, "Image quality measure based on a human visual system model," *Optical Engineering*, vol. 28, pp. 813–818, July 1989.

[53] C. van den Branden Lambrecht, "A working spatio-temporal model of the human visual system for image restoration and quality assessment applications," in *International Conference on Acoustics, Speech and Signal Processing ICASSP'96*, 1996.

[54] E. Peli, "Contrast in complex images," *Journal of the Optical Society of America*, vol. 7, pp. 2032–2040, 1990.

[55] N. Jayant, J. Johnston, and R. Safranek, "Signal compression based on models of human perception," *Proceedings of the IEEE*, vol. 81, pp. 1385–1422, Oct. 1993.

[56] A. Watson, "DCT quantization matrices visually optimized for individual images," in *Human Vision, Visual Processing, and Digital Display IV* (B.E.Rogowitz and J. Allebach, eds.), vol. 1913-14, SPIE, 1993.

[57] A. Webster, C. Jones, M. Pinson, S. Voran, and S. Wolf, "An objective video quality assessment system based on human perception," in *Human Vision, Visual Processing, and Digital Display*, vol. 1913, pp. 15–26, SPIE, 1993.

[58] F.-H. Lin and R. Mersereau, "A constant subjective quality MPEG encoder," in *International Conference on Acoustics, Speech and Signal Processing ICASSP'95*, pp. 2177–2180, 1995.

[59] I. Dalgic and F. Tobagi, "Constant quality video encoding," in *Proceedings International Conference on Communications ICC'95*, IEEE, June 1995.

[60] I. Dalgic and F. Tobagi, "A constant quality MPEG-1 video encoding scheme and its traffic characterization," in *PCS'96 International Picture Coding Symposium*, pp. 105–110, Mar. 1996.

[61] W. Pennebaker and J. Mitchell, *JPEG Still Image Data Compression Standard*. Van Nostrand Reinhold, New York, 1993.

[62] B. Girod, "The information theoretical significance of spatial and temporal masking in video signals," in *Human Vision, Visual Processing, and Digital Display*, vol. 1077 of *SPIE*, pp. 178–187, 1989.

[63] R. Safranek and J. Johnston, "A perceptually tuned sub-band image coder with image dependent quantization and post-quantization data compression," in *International Conference on Acoustics, Speech and Signal Processing ICASSP '89*, pp. 1945–1948, 1989.

[64] J. Pandel, "Variable bit-rate image sequence coding with adaptive quantization," *Signal Processing: Image Communication*, vol. 3, pp. 123–128, June 1991.

[65] ITU, Geneva, *CCIR Recommendation 500-1: Method for the Subjective Assessment of the Quality of Television Pictures*, 1978.

[66] J. Stuller and A. Netravali, "Transform domain motion estimation," *Bell System Technical Journal*, vol. 58, pp. 1673–1703, Sept. 1979.

[67] P. van der Meer, J. Biemond, and R. Lagendijk, "Modeling and multiplexing of VBR video without network constraints," in *Sixth International Workshop on Packet Video, Program and Proceedings*, pp. D8.1–D8.4, VISICOM, Sept. 1994.

[68] I. Sethi and N. Patel, "A statistical approach to scene change detection," in *Storage and Retrieval for Image and Video Databases III*, vol. 2420 of *SPIE*, pp. 329–339, 1995.

[69] J.-P. Leduc, *Digital Moving Pictures - Coding and Transmission on ATM Networks*, vol. 3 of *Advances in Image Communication*. Elsevier, 1994. Series editor: J. Biemond.

[70] F. Schoute, "Mixed traffic patterns and traffic capacity in ISDN," in *Innovative Services or Innovative Technology?* (J. Arnbak, ed.), pp. 97–105, Elsevier Science Publishers B.V., 1989.

[71] J. Hui, "Resource allocation for broadband networks," *IEEE Journal on Selected Areas in Communications*, vol. 6, pp. 1598–1608, Dec. 1988.

[72] The ATM Forum, *ATM User-Network Interface Specification*. Prentice Hall, 1993.

[73] N. Ohta, *Packet Video - Modeling and Signal Processing*. Artech House, Norwood, 1994.

[74] J. Solé-Pareta and J. Domingo-Pascual, "Burstiness characterization of ATM cell streams," *Computer Networks and ISDN Systems*, vol. 26, pp. 1351–1363, Aug. 1994.

[75] S. Jung and J. Meditch, "Design of a burstiness measure for variable bit rate video," in *Proceedings of Singapore International Conference on Networks SICON'95*, pp. 483–487, IEEE, July 1995.

[76] R. Onvural, *Asynchronous Transfer Mode: Performance Issues*. Boston: Artech House, 1994.

[77] D. Heyman, A. Tabatabai, and T. Lakshman, "Statistical analysis and simulation study of video teleconference traffic in ATM networks," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 2, pp. 49–59, Mar. 1992.

[78] F. Guillemin, P. Boyer, A. Dupuis, and L. Romoeuf, "Peak rate enforcement in ATM networks," in *Proceedings IEEE INFOCOM '92 (Florence, Italy)*, pp. 753–758, May 1992.

[79] E. Knightly and P. Rossaro, "Effects of smoothing on end-to-end performance guarantees for VBR video," in *Proceedings of the 1995 International Symposium on Multimedia Comunications and Video Coding*, May 1995.

[80] R. Lehr and J. Mark, "On traffic shaping in ATM networks," in *Proceeding of Singapore ICCS'94*, Nov. 1994.

[81] T. Ott, T. Lakshman, and A. Tabatabai, "A scheme for smoothing delay-sensitive traffic offered to ATM networks," in *Proceedings IEEE INFOCOM '92 (Florence, Italy)*, pp. 776–785, May 1992.

[82] K. Joseph and D. Reininger, "Source traffic smoothing for VBR video encoders," in *Sixth International Workshop on Packet Video, Program and Proceedings*, pp. G1.1–G1.4, VISICOM, Sept. 1994.

[83] J. Cosmas, G. Petit, R. Lehnert, C. Blondia, K. Kontovassilis, O. Casals, and T. Theimer, "A review of voice, data and video traffic models for ATM," *European Transactions on Telecommunications and Related Technologies*, vol. 5, pp. 139–154, Apr. 1994.

[84] V. Frost and B. Melamed, "Traffic modeling for telecommunications networks," *IEEE Communications Magazine*, pp. 70–81, Mar. 1994.

[85] J.-P. Leduc and P. Delogne, "Statistics for variable bit-rate digital television sources," *Signal Processing: Image Communication*, vol. 8, pp. 443–464, 1996.

[86] B. Haskell and A. Reibman, "Multiplexing of variable rate encoded streams," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 4, pp. 417–424, Aug. 1994.

[87] G. Keesman and D. Elias, "Analysis of joint bit-rate control in multi-program image coding," in *Visual Communications and Image Processing '94*, vol. 2308 of *SPIE*, pp. 1906–1917, 1994.

[88] G. Keesman, *Multi-program video data compression*. PhD thesis, Delft University of Technology, 1995.

[89] M. Karol, M. Hluchyj, and S. Morgan, "Input versus output queueing on a space-division packet switch," *IEEE Transactions on Communications*, vol. COM-35, pp. 1347–1356, Dec. 1987.

[90] L. Kleinrock, *Queueing Systems*, vol. 1 and 2. Wiley, New York, 1975.

[91] J.-P. Leduc, "Multiplexing digital television sources on ATM networks," *Signal Processing: Image Communication*, vol. 6, pp. 435–462, 1994.

[92] K. Sriram and W. Whitt, "Characterizing superposition arrival processes in packet multiplexers for voice and data," *IEEE Journal on Selected Areas in Communications*, vol. 4, pp. 833–846, Sept. 1986.

[93] M. Izquierdo and D. Reeves, "Statistical characterization of MPEG VBR-encoded video at the slice layer," in *Proceedings of the Conference on Multimedia Computing and Networking*, SPIE, 1995.

[94] D. Heyman, A. Tabatabai, and T. Lakshman, "Statistical analysis of MPEG2-coded VBR video traffic," in *Sixth International Workshop on Packet Video, Program and Proceedings*, pp. B2.1–B2.5, VISICOM, Sept. 1994.

[95] J. Enssle, "Modelling and statistical multiplexing of VBR MPEG compressed video in ATM networks," in *4th Open Workshop on High Speed Networks*, pp. 59–67, Sept. 1994.

[96] P. Bocheck and S.-F. Chang, "A content based approach to VBR video source modeling," in *Proceedings of the 6th International Workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDV96)*, 1996.

[97] B. Maglaris, D. Anastassiou, P. Sen, G. Karlsson, and J. Robbins, "Performance models of statistical multiplexing in packet video communications," *IEEE Transactions on Communications*, vol. 36, pp. 834–843, July 1988.

[98] F. Yegenoglu, B. Jabbari, and Y.-Q. Zhang, "Motion-classified autoregressive modeling of variable bit rate video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 3, pp. 42–52, Feb. 1993.

[99] P. Sen, B. Maglaris, N.-E. Rikli, and D. Anastassiou, "Models for packet switching of variable bit rate video sources," *IEEE Journal on Selected Areas in Communications*, vol. 7, pp. 865–869, June 1989.

[100] G. Awater, *Modeling, Analysis and Synthesis of an ATM switching Element.* PhD thesis, Delft University of Technology, 1994.

[101] P. Pancha and M. El Zarki, "Bandwidth-allocation schemes for variable-bit-rate MPEG sources in ATM networks," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 3, pp. 190–198, June 1993.

[102] G. Ramamurthy and B. Sengupta, "Modeling and analysis of a variable bit rate video multiplexer," in *INFOCOM '92*, pp. 0817–0827, 1992.

[103] M. Frater, J. Arnold, and P. Tan, "A new statistical model for traffic generated by VBR coders for television on the broadband ISDN," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 4, pp. 521–526, Dec. 1994.

[104] I. Richardson and M. Riley, "Improving the error tolerance of MPEG video by varying slice size," *Signal Processing*, vol. 46, pp. 369–372, Oct. 1995.

[105] S. Aign and K. Fazel, "Error detection and concealment measures in MPEG-2 video decoder," in *Signal Processing of HDTV, VI. Proceedings of the International Workshop on HDTV '94*, pp. 169–181, 1994.

[106] Q.-F. Zhu, Y. Wang, and L. Shaw, "Coding and cell-loss recovery in DCT-based packet video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 3, pp. 248–258, June 1993.

[107] M. Ghanbari and V. Seferidis, "Cell-loss concealment in ATM video codecs," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 3, pp. 238–247, June 1993.

[108] W. Zeng and B. Liu, "Geometric-structure-based directional filtering for error concealment in image/video transmission," in *Wireless Data Transmission at Information Systems/Photonics East '95*, vol. 2601 of *SPIE*, pp. 145–156, Oct. 1995.

[109] W. Luo and M. El Zarki, "Analysis of error concealment schemes for MPEG-2 video transmission over ATM based networks," in *Visual Communications and Image Processing '95*, vol. 2501 of *SPIE*, pp. 1358–1368, May 1995.

[110] H. Sun and J. Zdepski, "Adaptive error concealment algorithm for MPEG compressed video," in *Visual Communications and Image Processing '92*, vol. 1818 of *SPIE*, pp. 814–824, Nov. 1992.

[111] R. ter Horst, A. Koster, and K. Rijkse, "MUPCOS: A multi-purpose coding scheme," *Signal Processing: Image Communication*, vol. 5, pp. 75–89, Feb. 1993.

[112] F. Bosveld, R. Lagendijk, and J. Biemond, "Compatible video compression using subband and motion compensation techniques," in *International Workshop on HDTV '93 (Ottawa, Canada)*, Oct. 1993.

[113] M. Ghanbari, "Two-layer coding of video signals for VBR networks," *IEEE Journal on Selected Areas in Communications*, vol. 7, pp. 771–781, June 1989.

[114] M. Ghanbari, "An adapted H.261 two-layer video codec for ATM networks," *IEEE Transactions on Communications*, vol. 40, pp. 1481–1490, Sept. 1992.

[115] M. Ghanbari and V. Seferidis, "Efficient two-layer video coding techniques," in *Proc. 5th International Workshop on Packet Video (Berlin, Germany)*, Mar. 1993.

[116] P. van der Meer, J. Biemond, and R. Lagendijk, "An efficient layered coding scheme with scalable capabilities," in *Proceedings of the International Picture Coding Symposium PCS'94*, pp. 194–197, Sept. 1994.

[117] A. Eleftheriadis and D. Anastassiou, "Optimal data partitioning of MPEG-2 coded video," in *IEEE International Conference on Image Processing ICIP'94*, pp. 273–277, Nov. 1994.

[118] R. Mokry and D. Anastassiou, "Minimal error drift in frequency scalability for motion-compensated DCT coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 4, pp. 392–406, Aug. 1994.

[119] M. Civanlar and A. Puri, "Scalable video coding in frequency domain," in *Visual Communications and Image Processing '92*, vol. 1818 of *SPIE*, pp. 1124–1134, Nov. 1992.

Samenvatting

Gedurende de laatste decennia is er een grote variëteit in informatie-diensten ontstaan. In het verleden gebruikte elk van deze diensten een voor die dienst veelal specifiek medium voor transmissie of opslag, zoals het telefonie netwerk en het kabel TV netwerk. Echter, vanwege de geleidelijke verandering van analoge naar digitale representatie van de informatie worden bestaande, analoge media gebruikt voor de transmissie en opslag van nieuwe, digitale diensten. Verder ontstaan er nieuwe digitale media waarop verschillende digitale diensten verzonden en opgeslagen kunnen worden. De situatie ontstaat dat de beschikbare bandbreedte zo efficiënt mogelijk verdeeld moet worden tussen verschillende diensten.

De robuustheid van gedigitaliseerde informatie heeft als nadeel dat veel meer bandbreedte nodig is voor de transmissie en opslag hiervan in vergelijking met de originele analoge signalen. Met name in het geval van digitale videosignalen kan door middel van compressie dit effect tenietgedaan worden. Met compressie hebben digitale videosignalen juist minder bandbreedte nodig dan analoge videosignalen.

Op het gebied van videocompressie voor transmissie en opslag is er een vraag-en-aanbod-interactie tussen een tweetal onderzoeksgebieden, te weten de telecommunicatie en de digitale signaalbewerking: Aan de ene kant levert de telecommunicatiewereld de bandbreedte die aan de andere kant door de digitale signaalbewerkingswereld zo efficiënt mogelijk gebruikt wordt. Een betere verdeling van de bandbreedte zou echter bereikt kunnen worden indien de beide werelden wat meer zouden samenwerken.

Vanwege de specifieke eigenschappen van traditionele telecommunicatiekanalen zijn de meeste videocompressiealgoritmen ontworpen om een digitale bitstroom met een constante bandbreedte (ofwel constante bit capaciteit, in het Engels Constant Bit Rate, CBR) af te leveren. Vanwege de variërende hoeveelheid activiteit in de video zal de kwaliteit van het gecomprimeerde videosignaal dan echter ook varieren. Omdat een menselijke waarnemer de kwaliteit zal beoordelen aan de hand van de kwalitatief slechtste delen, comprimeren CBR algoritmen de video nooit optimaal. Door een variabele bit capaciteit (Variable Bit Rate, VBR) te ondersteunen,

wordt optimale compressie met een constante kwaliteit mogelijk. Dit proefschrift beschrijft diverse aspecten van VBR gecomprimeerde video in relatie tot verschillende transmissie- en opslagmedia. Het laat zien hoe VBR video bitstromen geschikt gemaakt kunnen worden voor een CBR telecommunicatiekanaal, zodat voordeel behaald kan worden uit de hoge compressie van VBR video.

Het concept van VBR videocompressie verschilt van dat van CBR videocompressie. Bij CBR compressie wordt de kwaliteit geoptimaliseerd door de beschikbare bits zo goed mogelijk tussen de verschillende beelden en delen van beelden te verdelen. Bij VBR compressie wordt de compressie geoptimaliseerd, waarbij ervoor gezorgd dient te worden dat de door de compressie geïntroduceerde vervorming onder normale kijkomstandigheden niet waargenomen wordt. Om dit te doen dient de niet-lineaire relatie tussen vervorming en kwaliteit bestudeerd te worden. Elke videocompressietechniek kan parameters zo instellen, dat de vervorming de zichtbaarheidsdrempel niet overschrijdt. Ook in MPEG kunnen de kwantisatie-schaalfactor en weegmatrices met dit doel bepaald worden. De compressie kan echter verder verbeterd worden door gebruik te maken van het effect dat vervorming vaak gemaskeerd wordt door de beeldinhoud. Het is daardoor mogelijk om zo'n 20 % meer compressie te verkrijgen zonder dat de extra vervorming waargenomen wordt.

De lage gemiddeld benodigde bandbreedte van VBR gecomprimeerde video is voordelig indien de video opgenomen wordt op een willekeurig toegankelijk medium (Random Accessible Memory, RAM) zoals een optische schijf (CD, DVD). Als de opslagapparatuur echter niet willekeurig toegankelijk is, of indien de video verzonden dient te worden, is er een conversie van de VBR bitstroom naar een CBR bitstroom nodig. Dit kan bereikt worden door opvulling van de bitstroom met loze bits, of door meerdere VBR bitstromen te multiplexen met aanvullende opvulling om één enkele CBR bitstroom te vormen. In het laatste geval kan voordeel behaald worden uit de lage gemiddelde bandbreedte van de VBR bitstromen indien de bandbreedte van de resulterende CBR bitstroom lager is dan de som van de piek bandbreedtes van de individuele VBR bitstromen. In dat geval bestaat er echter ook de kans dat op een bepaald moment de capaciteit van de resulterende CBR bitstroom niet toereikend is om alle informatie uit de VBR bitstromen te verzenden, zodat een verlies van informatie ontstaat. Vandaar dat er altijd een afweging gemaakt dient te worden tussen de bezettingsgraad van het CBR kanaal en de kans op informatieverlies.

Het multiplexen van verschillende VBR bronnen op een CBR kanaal kan uitgevoerd worden door het netwerk zelf of, indien het beoogde netwerk dit niet ondersteunt, door een multiplexer in de koppeling met het netwerk. In beide gevallen dient de multiplexer voor iedere bron bandbreedte te reserveren zodat een zekere maximale kans op informatieverlies kan worden gegarandeerd terwijl een hoge bezettingsgraad van het kanaal wordt bereikt. Deze gereserveerde bandbreedte zal hoger zijn dan de te verwachten gemiddelde bandbreedte van de VBR bron, maar lager dan de piekbandbreedte. Om dit te doen dient de bron het te genereren verkeer dus te

beschrijven aan de hand van een aantal parameters, zoals de te verwachten piek en gemiddelde bandbreedte. De multiplexer heeft dan ook de mogelijkheid om te controleren of het verkeer ook daadwerkelijk aan deze parameters voldoet. Ook de compressor zal rekening moeten houden met de karakteristieken van het gegenereerde verkeer. Met behulp van parameter controlerende functies zal hij de kwaliteit moeten aanpassen indien te veel bits gegenereerd worden. Een goede beschrijving van het verkeer is dus van essentieel belang om teveel kwaliteitsverlies te voorkomen en een hoge bezettingsgraad van het kanaal te bereiken.

Om het verkeer van een VBR videobron nauwkeurig te beschrijven, is er behalve de piek en gemiddelde bandbreedte ook een parameter nodig die de fluctuaties in de bitstroom beschrijft. Het blijkt dat de grootte van deze zogenaamde *burstiness*-parameter in hoge mate bepaald wordt door de gebruikte algoritmen in de koppeling tussen de videocompressieapparatuur en het netwerk. Door in deze koppeling de bitstroom te effenen kan de burstiness gereduceerd worden, zodat het VBR verkeer aanzienlijk beter voorspeld kan worden.

Het resultaat van multiplexen kan beschreven worden aan de hand van de kans op het verlies van informatie bij een bepaalde bezettingsgraad. Het blijkt dat deze kans afhangt van de karakteristieken van het verkeer. Om deze karakteristieken te voorspellen zijn vele complexe modellen van VBR video bitstromen in de literatuur te vinden. Dit proefschrift laat echter zien dat het VBR verkeer, wanneer het geëffend is, het gedrag van de activiteit in de video weerspiegelt. Om het resultaat van het multiplexen van geëffende VBR videobronnen, dat sterk verbeterd is ten opzichte van niet geëffende bronnen, te voorspellen, dienen dus modellen van de inhoud van de te comprimeren videobronnen beschikbaar te zijn.

Indien VBR bronnen verzonden of opgeslagen worden met behulp van statistische multiplexing, zal er altijd een kans op verlies van informatie aanwezig zijn. Een verhulling van deze verliezen is daarom noodzakelijk zodat het kwaliteitsverlies beperkt blijft. Dit is mogelijk door gelaagde compressie toe te passen, waarbij de data gegenereerd door een VBR videobron gesplitst. De meest vitale informatie wordt dan verzonden in een apart CBR kanaal met lage bandbreedte. Hierdoor wordt een goede verhulling van informatieverlies bereikt ten koste van een zeer kleine inefficientie ($< 1\%$).

De technieken die in dit proefschrift beschreven staan hebben geleid tot de beschrijving van een gelaagde VBR MPEG referentie-compressor met parameter controlerende functies en een effeningsalgoritme in de koppeling met het transmissie- of opslagmedium. Het door deze encoder gegenereerde verkeer heeft een lage burstiness en een lage gemiddelde bandbreedte, terwijl de piekbandbreedte in de orde ligt van de bandbreedte van een CBR bron die een vergelijkbare kwaliteit videosignaal genereert. Voordeel van de lage gemiddelde bandbreedte kan behaald worden op een willekeurig toegankelijk opslag medium of door verschillende bronnen te multiplexen

voor verzending of opslag op een CBR medium. Een hoge bezettingsgraad van het CBR kanaal kan bereikt worden vanwege de lage burstiness en de daardoor ontstane voorspelbaarheid van het verkeer.

Acknowledgements

At the end of a period it is a pleasure to be able to thank everybody who supported me and therefore contributed, direct or indirect, to this thesis.

First of all, I would like to thank my promotor and co-promotor, Jan Biemond and Inald Lagendijk, for offering me a job at the Information Theory Group, somewhat less than five years ago. They are also the ones that had largest influence on the contents of this thesis, which is another reason to mention them first. My work was performed within the DART project, and therefore the members of this project are mentioned next. Of these people, Pepijn Sitter was the one who had to make my ideas work in the demonstrator, for which I thank him. Emmanuel Frimout was responsible for writing most parts of the MPEG software, and it was a pleasure to cooperate with him to improve it and make it compatible with public domain MPEG. Further, the german partners in the project from Deutsche Thomson Brandt, Mr. Bachnick and Mr. Streckenbach, get a special mention for the friendly cooperation and hospitality when I was in Hannover. I also thank the rest of the project members for the fruitful and pleasant meetings.

At the Information Theory Group, I want to mention my roommates Stefan Westen and Ruggero Franich first. Stefan helped me with Chapter 3, Ruggero's contribution is found in my current knowledge of computers and software. Another (ex-)roommate, Richard Kleihorst, influenced my choice in text and drawing processors (LaTeXand programming in PostScript). Peter van Roosmalen helped me with the recording and encoding of large video sequences. I thank them and the rest of the HDSP-group for all interesting discussions and the numerous lunches at 11.45 AM. Further, I would like to thank Annett Bosch and Hatin Tekis for all their administrative activities and the rest of the Information Theory Group for the pleasant atmosphere.

Concerning my social life, there are many who contributed to my well-being. I cannot mention them all, but I must mention "De Vodjes", with whom I experienced most of my adventures. Further, I thank my family for their non-demanding and supporting attitude, providing the basics for everything. At home, I thank my cat Chicky for

the company, especially in the lonely days, and Sabine, who removed those days and who supported me in many, many ways despite my stubborn personality.

Curriculum Vitae

Patrick Johannes van der Meer was born in Roelofarendsveen, the Netherlands on February 16, 1968. In 1986 he obtained his VWO diploma (secondary school, VWO stands for preparation for scientific education) from the Bonaventura college in Leiden. He acquired his degree in Electrical Engineering (Ir., which is similar to the M.Sc. degree) from Delft University of Technology in 1992. The work which led to his graduation was carried out at the Dr. Neher Laboratory of PTT Research in Leidschendam and was supervised by the Information Theory Group at the Department of Electrical Engineering. The title of his thesis was "3DTV: Source coding of a stereoscopic video signal".

After his graduation he joined the Information Theory Group to work in the RACE-DART project, sponsored by the European Union and concerning the development of a digital data recording terminal based on a helical scan video recorder. He researched the interconnection between future B-ISDN networks, based on ATM technology, and the recorder and was responsible for one of the demonstrators of the project, showing the recording of VBR compressed video on a CBR recorder. Eventually, this work has led to his writing of this Ph.D. thesis.

Since November 1996 he has been working at the Development and Strategy Group of the Technology and Research Department of N.V. CASEMA, the largest cable company in the Netherlands. There, his knowledge of networking and video compression contributes to the migration of CASEMA from a cable television company to a full service provider.