TRABAJO ESPECIAL DE GRADO

Codificación de Video Escalable para Multimedia Streaming por Internet

Presentado ante la Ilustre
Universidad Central de Venezuela
por el Br. Ana K. De Abreu G.
para optar al Título de
Ingeniero Electricista.

Caracas, 2009

TRABAJO ESPECIAL DE GRADO

Codificación de Video Escalable para Multimedia Streaming por Internet

Prof Guía: Enrico Magli.
Tutor Académico: Prof. Zeldivar Bruzual.

Presentado ante la Ilustre
Universidad Central de Venezuela
por el Br. Ana K. De Abreu G.
para optar al Título de
Ingeniero Electricista.

Caracas, 2009

† A mis abuelos, Fe and João.

De Abreu Goes, Ana K.

CODIFICACIÓN DE VIDEO ESCALABLE PARA MULTIMEDIA STREAMING POR INTERNET

Prof. Guía: Enrico Magli. Tutor Académico: Prof. Zeldivar Bruzual. Tesis. Los Angeles, California. UCV. Facultad de Ingeniería. Escuela de Ingeniería Eléctrica. Opción: Comunicaciones. Institución: Politécnico di Torino y University of Southern California. 2009. 116 h.

Palabras claves: video, wireless, Scalable video coding (SVC), codificacion de canal, modulación digital, protección desigual de errores.

Resumen. Debido al rápido crecimiento de Internet y al incremento en la demanda de información multimedia, el video afluente (video streaming) sobre Internet ha generado gran interés en la industria y la academia. Dado que el Protocolo de Internet provee un servicio en tiempo real no fiable (también llamado del mejor esfuerzo, best-effort); sigue siendo una tarea difícil el diseño de sistemas de transmisión de video de alta calidad, que sean capaces de hacer frente al impredecible comportamiento de Internet y a las variaciones de las condiciones de la red. Este trabajo propone, como solución a este problema, un régimen de codificación y transmisión de video escalable, SVC (Scalable Video Coding) extensión de H.264/AVC, combinado con una protección desigual (UEP, Unequal error protection) del flujo de bits, utilizando para ello codificación de canal convolucional y esquemas de modulación digital, en lo que se llama *modulación codificada* (TCM, Trellis Coded Modulation). TCM es una técnica que combina la modulación y la codificación de canal para la transmisión de señales digitales a través de los canales de banda limitada. En este trabajo diferentes modelos de UEP son aplicados al flujo de bits escalables, donde los resultados de las simulaciones muestran que, con nuestro enfoque (TCM), la calidad del video es mucho mejor, para un valor fijo de velocidad de transmision y SNR.

Summary

The demand for Internet video streaming services has rapidly grown over the past few years. Since the Internet supports real-time services only in a best-effort manner, it remains to be a challenging task to design high-quality video streaming systems, which are able to cope with the Internet's unpredictable and varying network conditions. This work proposes an unequal error protection (UEP) scheme for scalable video transmission over an additive white Gaussian noise (AWGN) channel to increase the end-to-end video quality. Scalability of a video bit-stream allows for media bit rate as well as for device capabilities adaptation, which is a key feature if limitations in network resources (bandwidth variations and transmission errors) or supporting heterogeneous devices with a single bit stream, are considered. To verify the efficiency of the proposed strategy the end-to-end performance is evaluated through the Scalable Video Coding (SVC) standard, an extension of H.264/AVC, combined with unequal error protection using convolutional codes and modulation schemes (TCM) applied across the video layers. TCM or Trellis-coded modulation is a technique that combines modulation and channel coding for the transmission of digital signals over bandlimited channels. Several scalable video transmission of UEP schemes are compared, where simulation results reveal that, with our approach (TCM), the picture quality of a streamed video achieves a better error performance for the same data rate and SNR constraints.

Acknowledgements

After these months, I've got quite a list of people who contributed in some way to this thesis, for which I would like to express thanks.

My utmost gratitude goes to Prof. C.-C. Jay Kuo in the University of Southern California (USC) Viterbi School of Engineering, for allowing me to join his team, for his expertise, kindness, and most of all, for an enormous amount of faith in me and lots of wise words. I believe that one of the main gains of this months program was working with Prof. Kuo. Also, I am greatly indebted to Prof. Enrico Magli, who trusted me and gave me the opportunity to work with this incredible team.

A very special thanks goes to the PhD students of the USC for taking time out from their busy schedule to help me in many occasions. Specifically, I wish to thank: Tze-Ping Low for our philosophical debates, exchanges of knowledge and for the supervision that supervisors don't have time for; also, Tanaphol Thaipanich for specific help on the subjects of Scalable Video Coding (SVC) and Joyce Liang for some very special moments when the going was tough.

I have been extremely fortunate to have the support of very special friends to whom I am trully grateful.

I cannot end without thanking my family for the support they provided me through my entire life and in particular throughout this period of time.

I would like to dedicate this thesis to my grandparents. They have played an important role in the development of my identity and shaping the individual that I am today. Their unflinching courage and conviction will always inspire me.

Finally, I recognize that this research would not have been possible without the financial assistance of FIDETEL, in Caracas- Venezuela and the Politecnico di Torino Scholarship fund. I would like to express my gratitude to these agencies.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Significance of the Research

Multimedia communication is one of the major themes in today's information communication technologies. The rapid growth of interactive multimedia applications, such as mobile TV, video calling and video downloading has resulted in spectacular progress of wireless communications. With the great success of Internet video, wireless multimedia services are expected to be widely deployed in the near future. Namely, different types of wireless networks are converging into all-IP networks.

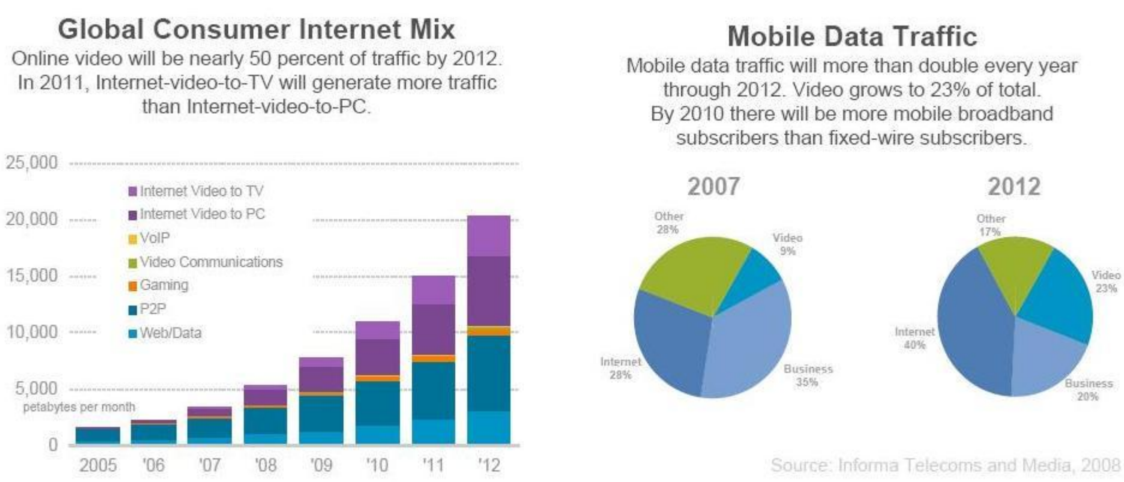


Figure 1.1. Cisco Visual Networking Index Charts.

According to Cisco System's recently updated Visual Networking Index forecast, the widespread adoption of video for entertainment and communications will drive sustained, rapid increases of global Internet-based network traffic, generating a projected 46 percent compound annual growth rate. Cisco also estimates that mobile data traffic will roughly double each year over the next four years (figure 1.1). For a more analytical look at the implications of the data presented above, please see the article entitled "Approaching the Zettabyte Era" [1].

To achieve this level of acceptability and proliferation of Internet video, there are many technical challenges that have to be addressed in the areas of video-coding and networking, specifically:

- Different types of networks inherently have different characteristics.
- There is dramatic heterogenity among end users, since terminals usually vary in display resolution and processing power capabilities according to their evolution state and category.
- Different applications have diverse QoS requirements in terms of data rates, delay bounds and packet loss probabilities.
- Varying network conditions.

To address the above challenges, scalability plays an important role. From a video-coding point of view, video scalability enables an application to adapt the streamed-video quality to changing network conditions, specifically bandwidth variation that is one of the primary characteristics of "best-effort" networks. From a networking point of view, scalability is needed to enable a large number of users to view any desired video stream, at anytime and from anywhere.

1.2 Research scope

Video transmission and storage systems using the Internet are typically based on RTP/IP for real-time services. Many pages could be filled with the properties and design considerations of RTP/IP, but literally hundreds of papers and books have been published on the topic [2]. For the purpose of this work, it is sufficient to say that most RTP/IP access networks are typically characterized by a wide range of connection qualities and receiving devices. The varying connection quality is resulting from adaptive resource sharing mechanisms of these networks, addressing the time varying data throughput requirements of a varying number of users. The variety of devices with different capabilities ranging from cell phones with small screens and restricted processing power to high-end PCs with high-definition displays results from the continuous evolution of these endpoints.

Layered compression and Layered transmission are attractive solutions to the problems posed by the characteristics of modern video transmission systems. That is, we must use a compression scheme that allows us to generate multiple levels of quality using multiple layers simultaneously with a network delivery model that allows us to selectively deliver subsets of layers to individual receivers.

The Scalable Extension of H.264/AVC (SVC) aims at achieving both high compression performance and adaptivity for video delivery over heterogeneous networks. SVC is based on H.264/AVC and provides a layered bit-stream with spatial, temporal and quality scalabilities. By packet discarding or truncation, a reduced spatial-temporal-quality resolution of a video bit-stream can be obtained with more graceful quality degradation as compared with the traditional non-scalable video codec.

Due to its real-time nature, video streaming typically has loss requirements. Packet loss is inevitable in the Internet and can produce undesirable effects at the decoder, including unpredictable behavior and unacceptable playback quality. However, the current best-effort Internet does not offer any quality of service (QoS) guarantees to streaming video over the Internet. In order to prevent such effects, it is necessary to design a robust transmission scheme to protect the quality of multimedia signals against transmission errors.

Due to the fact that unequal importance exist among the compresed scalable video layers, the type of transmission which achieves equal error protection is not optimal. Therefore, bit rate scalable media may be successfully combined with unequal error protection by employing forward error correction (FEC) codes, such as block codes and convolutional codes [3] where the more important bits are transmitted with more redundancy; however, this approach reduces compression gains. In bandlimited channels, hierarchical modulation schemes can be applied. The idea is to strongly protect the important part of the scalable media (base layer) and give less protection to the enhancement layer.

Such systems have been studied in great detail in many research publications [4, 5]; but must of these techniques have been limited by the non-availability of an efficient scalable video codec. Or only fixed modulation and coding have been considered in systems with the scalable extension of H.264/AVC (SVC).

Therefore the main objective of this research is to build a video transmission scheme that combines a scalable video coder with unequal error protection (UEP), based on previous work ([4], [5]) but implementing it with a different approach. To meet this objective the thesis addresses 3 separate goals:

- To record an understanding of the scalable video coding software development and, using this understanding to formulate an ideal software development environment.
- to investigate the challenges on video streaming.
- to explore and compare the transport prioritization mechanisms applied in previous works with our UEP approach, by applying it in two different video sequences.

In order to meet the above objectives, this thesis proposes a novel combined solution for "Joint Source-Channel Coding" (JSCC) for video transmission in error-prone environments. The scheme involves video source coding, using the scalable extension of H.264/AVC (SVC), and the unequal error protection (UEP) with coded modulation, a combination of Rate-Compatible Punctured Convolutional (RCPC) codes and digital modulation, for providing adaptive Forward Error Correction (FEC), in order to minimize the bit errors caused by channel impairments during video transmission, without increasing the channel bandwidth. We demonstrate that the proposed framework achieves significant improvement, and provides graceful degradation of reconstruction quality for increasing packet loss rate.

The Internet is often the target network for streaming video. It is certainly not the only network that could be used for streaming (for example, cellular and wireless networks.) But, the Internet is a difficult network for transporting real-time data; so, it is a perfect example of an heterogeneous, time-varying, network with no QoS control.

1.3 Reader's guide

The remainder of the document is structured as follows;

Chapter 2 presents the problem statement. A general background of the stated problem in the network system is explained followed by scalability and prioritization methods as a proposition for the practical problems.

Chapter 3 starts with a brief discussion of some of the properties of a video communication and streaming application, since these strongly influence the design of the system. Further, the chapter introduces some basic concepts, underlying video coding, and gives an overview of the latest scalable video coding, where each subsection talks about the basic capabilities and supported features of the coding standard. Also, several error control methods to combat the effect of losses in a video streaming system are briefly explained.

After the exposition of the principal features of our work, *chapter 4* talks about the different design considerations within the scalable video codec for the proposed hybrid temporal-spatial scalability. The chapter, also presents the design decisions and implementation details of the error control methods applied in this work.

Subsequently, *chapter 5* gives the performance analysis of scalable video coding combined with prioritization methods and shows the improvements of the unequal error protection (UEP) over the equal error protection (EEP) scheme. Special emphasis is given to highlight the improvements in using coded modulation combined with channel coding to achieve graceful degradation of the picture quality for a certain bit error rate value.

Chapter 6 concludes the thesis with a summary of the simulation results, the discussion about the potential future of the scalable extension of H.264 and future research directions. Finally, we provide a brief summary in Italian of our thesis in *chapter 7*.

Chapter 2

Problem Statement

2.1 The problem: Variable network conditions

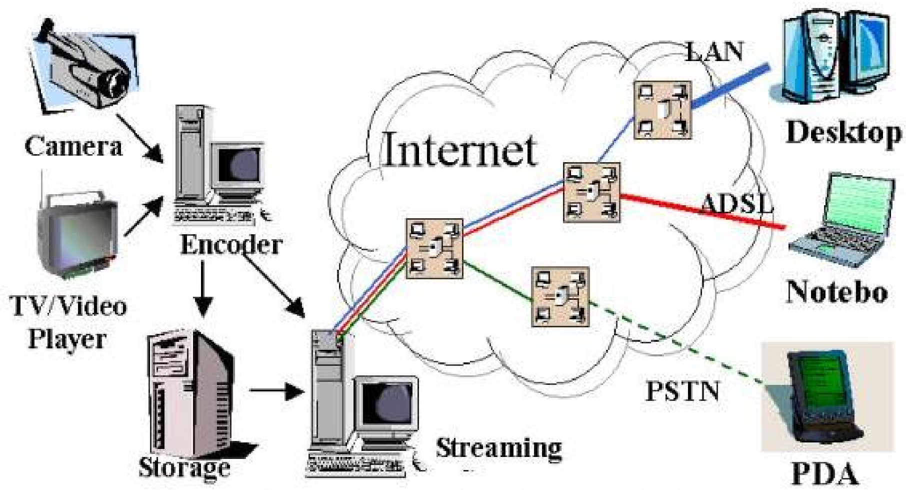


Figure 2.1. Real-time video over the internet.

The rapid growth of wireless communication and networking protocols over the web will ultimately bring video to our lives anytime, anywhere and on any device. Wireless video will enable remote classrooms, remote training facilities and remote hospitals anywhere in the world. Until this goal is achieved, wireless video delivery faces numerous challenges. The current Internet only provides "best-effort" service, and it does not guarantee or provide quality of services (QoS) for multimedia applications. Therefore, limited and unpredictably varying bit rates, heterogeneous networks and receivers, high error rates and random variations of wireless channels may adversely impact quality of content to deliver. These issues are briefly explained in the following.

1. **Finite bandwidth and limited bit rates**

   Bandwidth is the amount of data that can traverse through the network or a part of the network at any given time. Network bandwidth is a shared, limited resource and will vary with time. Networks may carry multiple video streams simultaneously or carry non-video data. Bandwidth variation is one of the primary characteristics of "best-effort" networks, and the Internet is a prime example of such networks.

   Rate control is important to multimedia streaming applications. If the sender transmits faster than the available bandwidth then congestion occurs, packets are lost, and there is a severe drop in video quality. If the sender transmits slower than the available bandwidth then the receiver produces suboptimal video quality. The goal to overcome the bandwidth problem is to estimate the available bandwidth and then match the transmitted video bit rate to the available bandwidth, thus efficient video compression is required.

2. **Heterogeneity**

   A heterogeneous network is a network whose parts (sub-networks) may have vastly unequal resources. For example, some parts of a heterogeneous network may have abundant bandwidth and excellent congestion control while other parts of the network are overloaded and congested by overuse or by a lack of phy-sical network resources. Different receivers on a heterogeneous network can experience different performance characteristics. When streaming video over a heterogeneous network, the video stream should be decodeable at optimal quality for users with a good network connection, and at useable quality for users with a poor connection.

In addition, video content is delivered to a variety of decoding devices with heterogeneous display and computational capabilities (see figure 2.2). In these heterogeneous environments, flexible adaptation of once-encoded content is desirable, at the same time enabling interoperability of encoder and decoder products from different manufacturers.

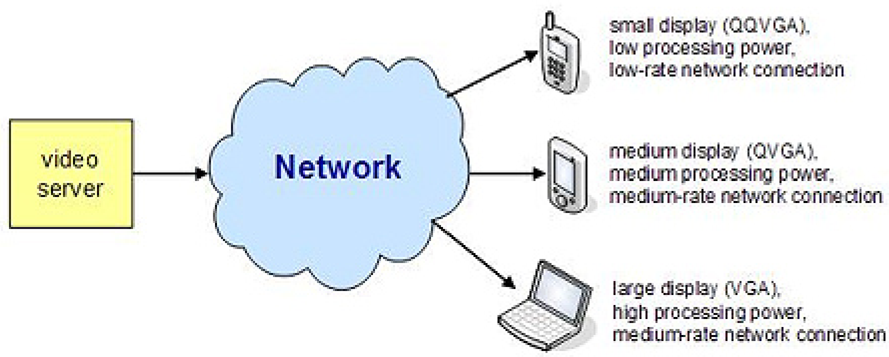


Figure 2.2. Example of video streaming with heterogeneous receiving devices and variable network conditions.

3. **High error rates**

Networks based on wireless technologies have much higher error rates than those based on more traditional technologies such as optical fiber, coaxial cable, or twisted pair wiring. Wireless signals share the same propagation medium with many competing signals, and as a result, there are many more opportunities for interference, that can result in bit errors.

During transmission, bit errors could render a whole packet useless. Packet losses can produce undesirable effects at the decoder, including unpredictable behavior of the decoder and an unacceptable playback quality. To combat the effect of losses, a video streaming system is designed with error control.

4. **Random fluctuations on time**

As the signal propagates through a wireless channel, it experiences random fluctuation in time if the object's transmitter surroundings are moving, due to changing reflections and attenuation. This implies that bandwidth, delay, loss, or other network characteristics can vary significantly over time, sometimes changing drastically in a matter of seconds which makes it difficult to dessign reliable system performance. When streaming over a network with time-variance, the video source should be able to adjust its parameters to changing network conditions (adaptability).

Next section focuses on these problems and discusses the approaches that have been taken for overcoming it.

2.2 Related work

When one compresses a video sequence the following parameters must be determined: frame size, frame rate, data rate, and de-compressed quality. One of the problems with many video compression methods is that these parameters are fixed at the encoding time and cannot be easily changed. For example, suppose one encodes a standard de-finition video sequence with MPEG-2 at 6 Mb/s and stores it on a network video server that will stream it over a network. At a later time if a receiver requests this video information but does not want it at 6 Mb/s, but at 4 Mb/s, the server would have to transcode the sequence to meet the new rate requirement, which is very computationally intense. Another way to address this problem is to store multiple copies of the compressed sequence at the video server that is then able to satisfy different users' needs; this is very expensive. Similarly a viewer may want the sequence at a different spatial resolution (i.e., different frame size) or at a different frame rate.

In this heterogeneous and dynamic environment it is obvious that one cannot expect that a single video bit rate stream can fulfill all requirements without limiting the user experience of most other receivers. For example, it is expected that a video stream is capable to adapt its media rate to the transmission conditions to provide at least acceptable quality at the receivers, but also explores the full benefits of available higher system or device resources.

The approach of layered video coding has already been included in different video coding standard in the past, like H.262-MPEG-2 Video [6], H.263 [7] and MPEG-4 Visual [8]. However, the scalable profiles of those standards have rarely been used. Reasons for that includes the inferior coding efficiency results and a large increase in decoder complexity compared to the non scalable profiles.

To combat channel errors, layered coding must be combined with transport prioritization so that the base layer is delivered with a higher degree of error protection. Different networks may implement transport prioritization using different means.

Prioritization can be realized with different error-control treatments to various layers: forward error correction or (rarely) retransmission. Retransmission is typically not suited for video traffic because, the latency involved in sending the retransmit request and reply is too great. In addition, the additional communication involved in retransmission is itself subject to error and loss. In such case, we can provide unequal protection by employing forward error correction (FEC) codes, such as block codes and convolutional codes [9].

An alternative approach to provide transport prioritization is to use the different bit error rates, caused by the characteristics of signal constellations in modulation [10].

Due to the limited bandwidth and the delay sensitivity of real-time video transmission, the optimal trade-off between source and channel coding under rate, delay or power constraints; has been studied extensively [11], [12]. Combined source / channel coding approaches in conjunction with scalable or layered source coding schemes can be used to minimize the overall distortion [13]; where unequal error protection was adopted and yielded unquestionable performance improvement over equal error protection.

On the other hand, layered transmission is very popular in image and video. Multi-carrier modulation (MCM) allows different bits to be transmitted at different subchannels while proper power allocation provides different error performance. A combined source-channel coding scheme using MCM to provide unequal error protectionis is developed in [14] where H. Zheng and K.J. Ray Liu propose a parallel transmission framework for reliable multimedia data transmission over spectrally shaped channels using multicarrier modulation.

These systems have been studied in great detail in many research publications and potentials of such technologies are well known, as it was mentioned above. However, most of these techniques have been limited by the non availability of an efficient scalable video codec. With SVC, which is an extension of H.264/AVC, in place the impact of such technologies will certainly grow over the next years.

For this purpose, this work attempts to provide some insight into a potential use case of SVC in wireless multimedia communication. More details are presented in the following.

2.3 Solution: Bit rate scalable media combined with prioritization methods

2.3.1 Layered Compression and Transmission

In this model, in order to make the most efficient use of wireless bandwidth and client resources, rather than distribute a single level of quality using a single network channel, the source distributes five video layers simultaneously across multiple network channels. In turn, each receiver individually tunes its reception rate by adjusting the number of layers that it receives. The net effect is that the signal is delivered to a heterogeneous set of receivers at different levels of quality using a heterogeneous set of rates. To fully realize this architecture, we must select a compression scheme that allows us to generate multiple levels of quality to selectively deliver subsets of layers to individual receivers.

Scalable Video Coding extensions of H.264/AVC

The SVC design (figure 2.3), which is an extension of H.264/AVC video coding standard, can be clasified as a layered video codec. SVC can combine certain layers in a flexible way in order to adapt to different bit rates, frame rates or spatial resolutions of the video content by removing parts of the bitstream. SVC-based layered video coding is suitable for different use-cases like:

- **Supporting heterogeneous devices with a single scalable bit stream**. With scalable video coding scheme, the source content has to be encoded only once, for the highest required resolution, resulting in a scalable bit stream from which representations with lower quality can be obtained by discarding selected data; thus, not even all streams have to be received by all terminals. For instance, the base layer of a scalable H.264/AVC stream to the low performance device only and guarantee that all layers of the stream reach the high performance terminal.

- **Adaptation to varying network conditions**. Network protocols cope with througput variations by adjusting the transmission rate. SVC explicitly provisions for removing packets from the bit stream, which implicity results in bit rate and by that, in presentation quality reduction of the video. For instance, When there is excess bandwidth, it will be efficiently used in a way that maximizes the perceptual quality.

Scalable video representation also provides inherent e-rror resilience and error concealment; that are essential for efficient transmission over error prone wireless networks.

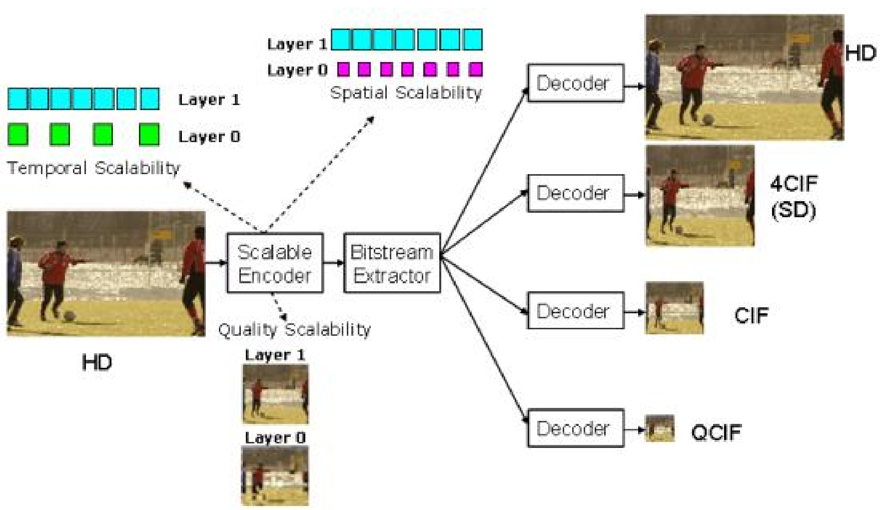


Figure 2.3. Scalable video coding (SVC).

Layered transmission

By combining the approach of layered compression with a layered transmission system, we can solve the heterogeneity problem. In this architecture,the scalable video coding produces a layered stream where each layer is transmitted on a different network channel. In turn, the network forwards only the number of layers that each physical link can support.

When bits are of non-uniform importance in contribution to distortion as in the case of multimedia data, scalable video coding approaches in conjunction with layered transmission can be used to minimize the overall distortion. This structure provides a very efficient, yet simple, level of abstraction between the encoding and the streaming process in a IP-based network. In this architecture, the scalable video coding produces a layered stream where each layer is transmitted on a different network channel. For instance, In an *unicast* streaming session the encoder only needs to know the range of bandwidth over it has to code the content and the streaming server has a total flexibility in sending any desired portion of any enhancement layer in parallel with the corresponding base layer; this means that a wideband channel is partitioned into a set of narrowband channels transmitting data in parallel subcarriers at a lower data rate. For *multicast* applications, combining layered source coding and layered transmission provides a flexible framework for the encoding, streaming and decoding process. Identical to the unicast case, the encoder compresses the content using any desired range of bandwidth. At time of transmission, the multicast server send the base and enhancement layers in different "multicast channels"; at the decoder side, the receiver can "subscribe" to the *base-layer channel* and to any number of *enhancement layer channels* that the receiver is capable of accessing, depending for example on the receiver access bandwidth.

2.3.2 Prioritization methods

A fundamental problem that afflicts video communication is losses. Losses can have a very destructive effect on the reconstructed video quality, and if the system is not designed to handle them, even a single bit error can have a catastrophic effect. The idea presented in this work is to strongly protect the important part of the scalable media (the base layer) in order to overcome worst-case error scenarios and give less importance to the enhancement layers in order to overcome the most tipical error situations. The prioritization methods used are:

- Unequal error protection(UEP).
- Hierarchical modulation schemes.
- Combined coded modulation.

These techniques adaptively change the level of modulations: Binary phase shift keying (BPSK), quadrature PSK(QPSK), 8-PSK, 16-quadrature amplitude modulation (16-QAM), and so on, as well as the amount of redundancy for an error correction code. A higher modulation (e.g., 16-QAM) with no error correction code can be used by enhancement layers resulting in graceful degradation of the perceptual quality according to the channel conditions and in higher bandwidth efficiency. A lower level of modulation (e.g., BPSK) with more redundancy for error correction is used by the base layer in bad channel conditions, but results in lower bandwidth efficiency.

Our solution emphasizes the adoption of *joint source/channel coding* from communications theory. Joint source/channel coding (JSCC) combines the design of compression and error-control coding to achieve better performance.

By considering how the pieces of a large design interact, better design decisions can be made for each individual component. Consequently, our video communications system is based on an interdisciplinary solution that combines scalable video coding and prioritization methods to achieve graceful degradation as the packet loss probability of an Internet connection increases.

Chapter 3

Video Representation and wireless transmission

This chapter provides a brief overview of the employed components and their relationship to our proposed solution, which is presented subsequently.

This chapter starts with a brief discussion of some properties of a video communication followed by basic concepts of video coding, continues with the principles of SVC (Scalable Video Coding), and concludes with a description of several error control methods designed specifically for multimedia transmission over wireless links.

3.1 Wireless Video Streaming

There exist a very diverse range of different video communication and streaming applications, which have very different operating conditions or properties. For example, video communication application may be for point-to- point communication or for multicast or broadcast communication, and video may be pre-encoded (stored) or may be encoded in real-time (e.g. interactive videophone or video conferencing). The video channels for communication may also be static or dynamic, may support a constant or variable bit rate transmission, and may support some form of Quality of Service (QoS) or may only provide best effort support. The specific properties of a video communication application strongly influence the design of the system. Therefore, we continue by briefly discussing some of these properties and their effects on video communication system design.

Point-to-point, multicast, and broadcast communications

Probably the most popular form of video communication is one-to-all communication or *broadcast* communication, where the most well known example is broadcast television. Broadcast is a very efficient form of communication for popular content as it is a method of transmitting radio, Internet or television signals (programs) to a number of recipients that belong to a large group, so the system must be designed to provide every intended recipient with the required signal. This is an important issue, since different recipients may experience different channel characteristics, and as a result the system is often designed for the worst-case channel. An example of this is digital television broadcast, where the source coding and channel coding were designed to provide adequate reception to receivers at the fringe of the required reception area, thereby sacrificing some quality to those receivers in areas with higher quality reception (e.g. in the center of the city). An important characteristic of broadcast communication is that, due to the large number of receivers involved, feedback from receiver to sender is generally infeasible – limiting the system's ability to adapt.

Another common form of communication is *point-to-point* or one-to-one communication, e.g. videophone and unicast video streaming over the Internet. In point-to-point communications, an important property is that a back channel could exists, so the receiver can provide feedback to the sender, which can then use it to adapt its processing.

Another form of communication with properties that lie between point-to-point and broadcast is *multicast*. Multicast is a one-to-many communication network technology for the delivery of information to a group of destinations simultaneously using the most efficient strategy to deliver the messages over each link of the network only once (it is not one-to-all as in broadcast). An example of multicast is IP-Multicast which is often employed for streaming media and Internet television applications. To communicate to multiple receivers, multicast is more efficient than multiple unicast connections (i.e., one dedicated connection to each client), and overall multicast provides many of the same advantages and disadvantages as broadcast.

Real-time encoding versus pre-encoded (stored) video

Video may be captured and encoded for real-time communication, or it may be pre-encoded and stored for later viewing. Interactive applications are one example of applications which require real-time encoding; e.g., videophone, video conferencing or interactive games. However real-time encoding may also be required in applications that are not interactive; e.g., the live broadcast of a sporting event.

In many applications video content is pre-encoded and stored for later viewing. The video may be stored locally or remotely. Examples of local storage include DVD and Video CD, and examples of remote storage include video-on-demand (VOD) and video streaming over the Internet (e.g., as provided by RealNetworks and Microsoft). Pre-encoded video has the advantage that it does not require a real-time encoding constraint. This can enable more efficient encoding such as the multi-pass encoding that is typically performed for DVD content. On the other hand, it provides limited flexibility as, for example, the pre-encoded video can not be significantly adapted to channels that support different bit rates or to clients that support different display capabilities than that used in the original encoding.

Static versus Dynamic Channels

Video communication system design varies significantly if the characteristics of the communication channel, such as bandwidth, delay, and loss, are static or dynamic (time-varying). Examples of static channels include ISDN (which provides a fixed bit rate and delay, and a very low loss rate) and video storage on a DVD. Examples of dynamic channels include communication over wireless channels or over the Internet. Video communication over a dynamic channel is much more difficult than over a static channel.

Constant-bit-rate (CBR) or Variable-bit-rate (VBR) Channel

Some channels support CBR, for example ISDN or DTV, and some channels support VBR, for example DVD storage and communication over shared packet networks. On the other hand, a video sequence typically has timevarying complexity. Therefore coding a video to achieve a constant visual quality requires a variable bit rate, and coding for a constant bit rate would produce time-varying quality. Clearly, it is very important to match the video bit rate to what the channel can support.

Quality of Service (QoS) Support

An important area of network research over the past two decades has been QoS support. QoS is the ability to provide different priority to different applications, users, or data flows, or to guarantee a certain level of performance to a data flow; e.g., guarantees on throughput, maximum loss rates or delay. Network QoS support can greatly facilitate video communication, as it can enable a number of capabilities including provisioning for video data, prioritizing delay-sensitive video data relative to other forms of data traffic, and also prioritize among the different forms of video data that must be communicated. The current Internet does not provide any QoS support, and it is often referred to as best-effort (BE), since the basic function is to provide simple network connectivity by best effort (without any guarantees) packet delivery . Different forms of network QoS that are under consideration for the Internet include Differentiated Services (DiffServ)[15] and Integrated Services (IntServ)[16].

3.2 Fundamentals of video coding

3.2.1 Data Compression

With the rapid advances in computers in the 1980s and 1990s came multimedia applications, where pictures and sound are combined in the same file. Such files tend to be large, which is why compressing them became a natural application.

Data compression is the encoding of a body of data D as a smaller body of data $\hat{D}$. There are two general types of compression: *lossless* and *lossy*

Lossless Compression

Lossless compression allows perfect recovery of the original digital data D given $\hat{D}$. This type of compression is required or highly desired by many applications, such as medical imaging, remote sensing and image archiving.

Lossy Compression

Lossy compression is compression in which some of the information from the original message sequence is lost. This means that the original sequence cannot be regenerated from the compressed sequence. Lossy compression is often used for applications such as images or videos where certain losses might be completely imperceptible to a human viewer (i.e., the loss of very high frequencies).

Video compression is normally lossy. Encoding a frame F_i in terms of its predecessor F_{i-1} introduces some distortions. As a result, encoding the next frame F_{i+1} in terms of (the already distorted) F_i increases the distortion. Even in lossless video compression, a frame may lose some bits. This may happen during transmission or after a long shelf stay.

3.2.2 Video Compression

Video compression is based on two principles. The first is the *spatial redundancy* that exists in each frame. The second is the fact that most of the time, a video frame is very similar to its immediate neighbors. This is called *temporal redundancy*.

A typical technique for video compression should therefore start by encoding the first frame using a still image compression method. It should then encode each successive frame by identifying the differences between the frame and its predecessor, and encoding these differences. If a frame is very different from its predecessor (as it happens with the first frame of a shot), it should be coded independently of any other frame. Therefore, in video a frame can be coded in three different ways:

- I-frame. An I or intra, is a frame that is coded independently of any other frame, using only spatial redundancies for prediction and coding.
- P-frame. Inter frame, labeled P (for predictive), is a frame coded based on the previously coded one. A P frame is coded by using temporal redundancies from the previous frame.
- B-frame. B frame refers to a bi-directionally predicted frame and requires information from previous and following I, P or B frames.

Figure 3.1 shows a sequence of such frames in the order in which they are ge-nerated by the encoder (and input by the decoder); and also the order in which the frames are output by the decoder and displayed.

Figure 3.2 depicts a typical encoder and decoder system. As seen at the encoder, the raw video is transformed by discrete cosine transform (DCT), quantized and coded by variable length coding (VLC).

The *discrete cosine transform* (DCT) is used on blocks that are, for instance eight pixels wide by eight pixels high, and it helps to separate the more perceptually significant information from the less perceptually significant information. Later steps in the compression algorithm, encode the low-frequency DCT coefficients with high precision, but use fewer or no bits to encode the high-frequency coefficients, thus

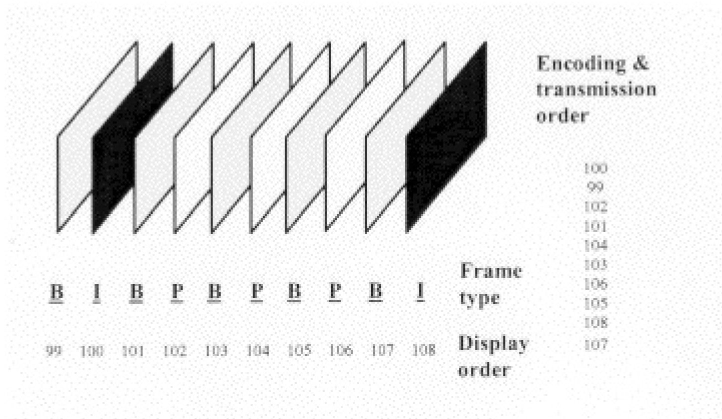


Figure 3.1. Coding and display order.

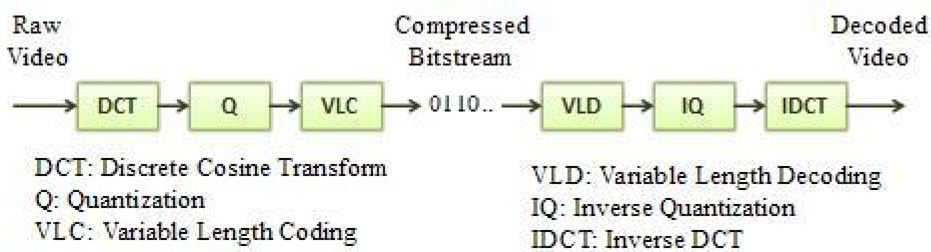


Figure 3.2. Typical video coding system.

discarding information that is less perceptually significant. This encoding of coefficients is accomplished in two steps: First, quantization is used to discard perceptually insignificant information; next, statistical methods are used to encode the remaining information using as few bits as possible.

Quantization rounds each DCT coefficient to the nearest of a number of predetermined values. After quantization, the next step in the compression process is to encode the quantized DCT coefficients in the digital bit stream using as few bits as possible, where many of this DCT coefficients have a value of zero (for the vast majority of high-frequency DCT coefficients). A technique called *run-length coding* takes advantage of this fact by grouping consecutive zero-valued coefficients (a “run”) and encoding the number of coefficients (the “length”) instead of encoding the individual zero-valued coefficients. This is accomplished by scanning the eight-by-eight-coefficient matrix in a diagonal zig-zag pattern.

Run-length coding is typically followed by *variable-length coding* (VLC). In *variable-length coding*, each possible value of an item of data is called a symbol. Commonly occurring symbols are represented using code words that contain only a few bits, while less common symbols are represented with longer code words; so that on average, VLC requires fewer bits to encode the entire image. Huffman coding is a variable-length coding scheme that optimizes the number of bits in each code word based on the frequency of occurrence of each symbol.

Then the compressed video stream is transmitted to the decoder through the network. At the decoder, the received compressed video stream is first decoded by variable length decoding (VLD), then inversely quantized (IQ), and inversely DCT (IDCT) transformed. The interested reader is referred to [17] for a more detailed discussion.

Using the techniques described above, still-image compression algorithms such as JPEG can achieve good image quality. Video compression algorithms, however, employ *motion estimation* and *compensation* to take advantage of the similarities between consecutive video frames. This allows video compression algorithms to achieve good video quality at high compression ratios.

Motion estimation attempts to find a region in a previously encoded frame (called a “reference frame”) that closely matches each macro block in the current frame. For each macro block, motion estimation results in a “motion vector”. The motion vector is comprised of the horizontal and vertical offsets from the location of the macro block in the current frame to the location in the reference frame of the selected region. The video encoder typically uses VLC to encode the motion vector in the video bit stream.

On the other hand, *motion compensation* uses the motion vectors encoded in the video bit stream to predict the pixels in each macro block. If the horizontal and vertical components of the motion vector are both integer values, then the predicted macro block is simply a copy of the region of the reference frame. If either component of the motion vector has a non-integer value, interpolation is used to estimate the image at non-integer pixel locations. Next, the prediction error is decoded and added to the predicted macro block in order to reconstruct the actual macro block pixels.

While *motion estimation* is the process of determining motion vectors that describe the transformation from one image to another, the concept of *motion compensation* contains the motion estimation between video frames. The combination of motion estimation and motion compensation is a key part of video compression, it is an efficient tool to reduce temporal redundancy between frames.

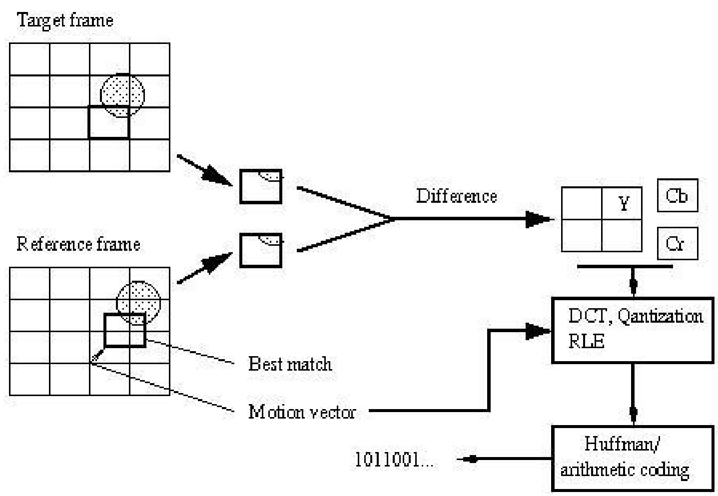


Figure 3.3. Motion estimation and compensation.

Error Metrics

Developers and implementers of lossy image compression methods need a standard metric to measure the quality of reconstructed images compared with the original ones. The better a reconstructed image resembles the original one, the bigger should be the value produced by this metric. Such a metric should also produce a dimensionless number, and that number should not be very sensitive to small variations in the reconstructed image. A common measure used for this purpose is the *peak signal to noise ratio* (PSNR). It is familiar to workers in the field, it is also simple to calculate, but it has only a limited, approximate relationship with the perceived errors noticed by the human visual system. This is why higher PSNR values imply closer resemblance between the reconstructed and the original images, but they do not provide a guarantee that viewers will like the reconstructed image.

PSNR most easily defined via the *mean squared error* (MSE). Denoting the pixels of the original image by P_i and the pixels of the reconstructed image by Q_i; the MSE is defined as:

$$MSE = 1/n \cdot \sum_{i=1}^{n} (P_i - Q_i)^2 \tag{3.1}$$

The PSNR is defined as:

$$PSNR = 10 \cdot log_{10}(\frac{MAX_i^2}{MSE}) = 20 \cdot log_{10}\frac{MAX_i}{\sqrt{MSE}} \tag{3.2}$$

In equation 3.2, MAX_i is the maximum possible pixel value of the image. When the pixels are represented using 8 bits per sample, this is 255. Typical PSNR values range between 20 and 40 dB. Acceptable values for wireless transmission quality loss are considered to be about 20 dB to 25 dB.

3.3 Scalable Video Coding

Scalability has already been presented in the video coding standards MPEG-2 Video, H.263, and MPEG-4 Visual in the form of scalable profiles. However, the provision of spatial and quality scalability in these standards comes along with a considerable growth in decoder complexity and a significant reduction in coding efficiency (i.e., bit rate increase for a given level a reconstruction quality) as compared to the corresponding non-scalable profiles. These drawbacks, which reduce the success of the scalable profiles of the former specifications, are addressed by the new SVC amendment of the H.264/AVC standard.

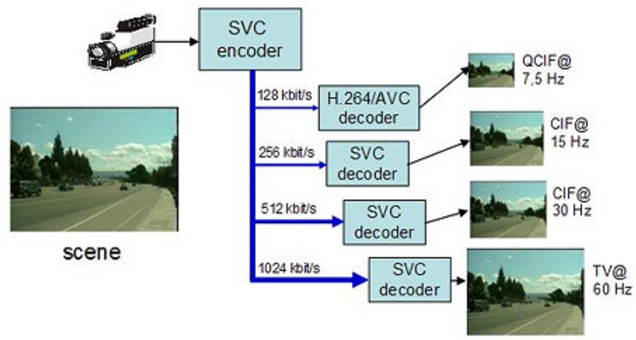


Figure 3.4. Scalable Video Coding (SVC).

A video bit stream is called scalable when parts of the stream can be removed in a way that the resulting sub-stream forms another valid bit stream for some target decoder, and the sub-stream represents the source content with a reconstruction quality that is less than that of the complete original bit stream but is high when considering the lower quantity of remaining data. Bit streams that do not provide this property are referred to as single-layer bit streams.

3.3.1 Basic concepts of scalable video coding

The basic idea of SVC is to extend the hybrid video coding approach in a way that a wide range of spatiotemporal and quality scalability is achieved. An SVC bit-stream consists of a base layer and one or several nested enhancement layers. The base layer is a plain H.264/MPEG4-AVC bit-stream, ensuring backward compatibility to existing receivers and enhancement layers bring additional information about quality, resolution or frame rate.

Temporal scalabitilty

Temporal scalability allows different picture rates (expressed in Hz); Typical frequencies in Europe are 50 Hz, 25 Hz or 12.5 Hz. It is based on the concept of B frames (bi-directionally predicted frames), where each B picture of a higher temporal enhancement level is encoded with a higher quantization parameter (QP), thus the fidelity per picture is decreasing with the decreasing importance in terms of the number of succeeding references by other pictures.

Temporal scalability can typically be used in video transmissions over mobile networks where bandwidth capacity can change very often, or if the target terminal has very low CPU capacities: in such cases, it is interesting to drop the enhancement layers and send only the base layer (which could, for example, contain only half the number of images per second).

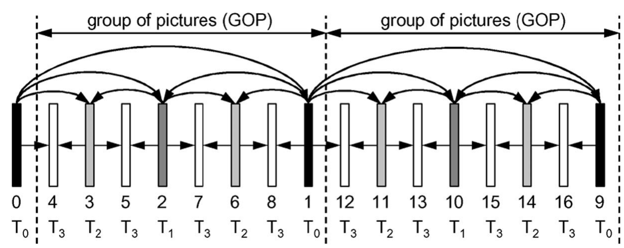


Figure 3.5. Temporal scalability.

Spatial scalabitilty

Spatial scalability describe cases in which subsets of the bit stream represent the source content with a reduced picture size. This implies that streams with different frame resolutions, such as QCIF (176 x 144 pixels), CIF (352 x 288), and 4CIF (704 x 576), are extractable from a single bit stream.

For supporting spatial scalable coding, SVC follows the conventional a-pproach of multi-layer coding, which is also used in H.262/MPEG-2 Video, H.263, and MPEG-4 Visual. In each spatial layer, motion-compensated prediction and intra prediction are employed as for single-layer coding. In addition to these basic coding tools of H.264/AVC, SVC provides so-called inter-layer prediction methods, which allow the usage of as much lower layer information as possible for improving rate-distortion efficiency of the enhancement layers (figure 3.6).

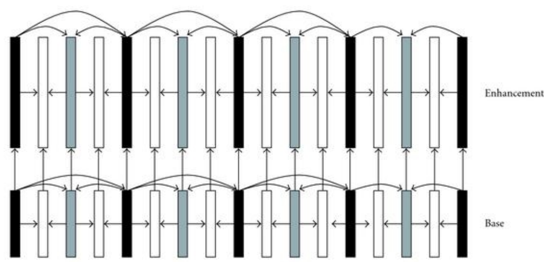


Figure 3.6. Two-layer spatial scalability intra- and interlayer prediction dependencies.

Spatial scalability can typically be used for transmission of the same video bitstream to PCs and portable devices, or to SD and HD television sets.

Quality scalabitilty

Quality scalability, also commonly referred to as fidelity or SNR scalability, allows the reconstruction of a given frame at different quality levels with corresponding different numbers of bits, while having the same spatial and temporal definitions. It is based on a progressive refinement (PR) approach, where the extension layers contain refinement quality information of the base layer in a progressive way. A further categorization is possible depending on how many different quality layers are required, with the most popular division into two classes, coarse granularity scalability (CGS) which only allows a few layers and fine granularity scalability (FGS) which should enable complete flexibility down to a single bit, or more practically single network packet.

Quality scalability can typically be used for: HD transmission to customers that are eligible for HD (full quality) and people not eligible for HD quality, but still equipped with HD screens (top enhancement layer is dropped); or for extra refinements when the bandwidth increases in mobile environments.

More rarely required scalability modes are region-of-interest (ROI) and object-based scalability, in which the sub-streams typically represent spatially contiguous regions of the original picture area. The different types of scalability can also be combined, so that a multitude of representations with different spatio-temporal resolutions and bit rates can be supported within a single scalable bit stream.

3.3.2 H.264/AVC towards an SVC standard

SVC was developed as an extension of H.264/AVC with all of its well-designed core coding tools being inherited (motion-compensated prediction, intra prediction, transform coding, entropy coding, deblocking filter), while only a few components have been added or modified. The key features of the scalable extension of H.264/AVC are:

- **Base layer compatibility with H.264/AVC**, that would enable the video to be decoded by those having an H.264/AVC currently-deployed decoder and no SVC decoder.
- Usage and extension of the **NAL unit** concept of H.264/AVC. The coded video data of H.264/AVC and its scalable extension is made up of Network Abstraction Layer (NAL) units, each of which is effectively a packet with a header of a few bytes (containing information about the payload) and a payload corresponding to the compressed information. A set of successive NAL units, sharing the same properties, forms a NAL access unit. The decoding of an access unit results in exactly one decoded picture. A set of consecutive access units with certain properties is referred to as a coded video sequence.
- **Hierarchical prediction structures**, employed in *temporal scalability* to generate the bit-streams. SVC extends the tools already provided by H.264/ AVC (hierarchical P or B slices coding), structuring the bitstream into a hierarchy of images, thus allowing the easy removal of the lower level(s) in the hierarchical description. Any picture can be marked as reference picture and used for motion-compensated prediction of following pictures independent of the corresponding slice coding types. These features allow the coding of picture sequences with arbitrary temporal dependencies. So called key pictures are

coded in regular intervals by using only previous key pictures as references. The pictures between two key pictures are hierarchically predicted. It is obvious that the sequence of key pictures represents the coarsest supported temporal resolution, which can be refined by adding pictures of following temporal prediction levels.

- **Inter-layer prediction methods** that enable the usage of lower layer information for rate-distortion optimization of the enhancement layers. In H.262/ MPEG-2 Video, H.263, and MPEG-4 Visual, the only supported inter-layer prediction methods employ the reconstructed samples of the lower layer signal. Although the reconstructed lower layer samples represent the complete lower layer information, they are not necessarily the most suitable data that can be used for inter-layer prediction. Usually, for sequences with slow motion and high spatial detail, the temporal prediction signal typically represents a better approximation of the original signal than the upsampled lower layer reconstruction. In order to improve the coding efficiency for spatial scalable coding, two additional inter-layer prediction concepts have been added in SVC: *prediction of macroblock modes and associated motion parameters* and *prediction of the residual signal* [18]. All inter-layer prediction tools can be chosen on a macroblock or sub-macroblock basis allowing an encoder to select the coding mode that gives the highest coding efficiency.

- **Progressive refinement.** For the encoding of quality enhancement representation layers, a new slice called Progressive Refinement (PR) has been introduced. For SNR scalability, coarse-grain scalability (CGS) and fine-granular scalability (FGS) are distinguished. In order to support FGS scalability, progressive refinement (PR) slices have been introduced in the scalable extension of H.264. Each PR slice represents a refinement of the residual signal that corresponds to a bisection of the quantization step size. NAL units can be truncated at any arbitrary point and the coding order of transform coefficient levels has been modified for the progressive refinement slices, so that the quality of the SNR base layer can be refined in a fine-granular way.

3.4 Error control for overcoming channel losses

Losses can have a very destructive effect on the reconstructed video quality, and if the system is not designed to handle losses, even a single bit error can have a catastrophic effect and during transmission could render a whole packet useless. Particularly, when transmitting a low bitrate video sequence over a wireless link, a coded picture typically fits in one packet; therefore, a transmission error will lead to a loss of a whole slice or frame. A number of algorithm have been proposed in the literature; for instance, [20] contains a good review of error control and other robust encoding techniques.

A video streaming system is designed with error control to combat the effect of losses. There are four rough classes of approaches for error control: (1) retransmissions, (2) forward error correction (FEC), (3) error concealment, and (4) error-resilient video coding. The first two classes of approaches can be thought of as channel coding approaches for error control, while the last two are source coding approaches for error control. A video streaming system is typically designed using a number of these different approaches. Forward error correction and error concealment techniques are discussed in the following subsections. In addition, joint design of the source coding and channel coding is very important and it is also discussed.

3.4.1 Forward error correction

The goal of FEC is to add specialized redundancy that can be used to recover from errors. FEC has the effect of increasing transmission overhead and hence reducing usable bandwidth for the payload data. The two main categories of FEC are block coding and convolutional coding. Block codes work on fixed-size blocks (packets) of bits or symbols of predetermined size. Convolutional codes work on bit or symbol streams of arbitrary length. Within this work, convolutional codes are employed; where in some cases, a performance very close to channel capacity is achievable.

Convolutional codes

With convolutional codes, the incoming bit stream is applied to a shift register, that consist of K (k-bit) stages and n linear algebraic function generators, as shown in figure 3.7. For each shift of the shift register, k new bits are inserted and n code bits are delivered, so the code rate is $R_c = k/n$. The power of a convolutional code is a function of its constraint length, K. Large constraint length codes tend to be more powerful. Unfortunately, with large constraint length comes greater decoder complexity.

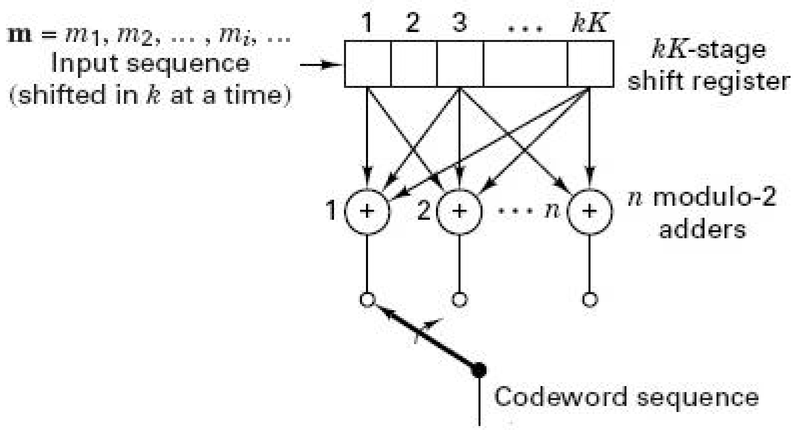


Figure 3.7. Convolutional encoder

Punctured convolutional codes

In some practical applications, there is a need to employ high-rate convolutional codes, e.g. rates of $(n-1)/n$. As you can observe in [3], the implementation of the decoder of a hight-rate code can be very complex. This complexity can be avoided by designing the high-rate code from a low-rate code, in which some of the coded bits are deleted from transmission. The deletion of selected coded bits at the output of a convolutional encoder is called *punctuting*. Thus, one can generate high-rate convolutional codes by puncturing rate $1/n$ codes with the result that the decoder maintains the low complexity, but reduces the free distance of the rate $1/n$ code by some amount that depends on the degree of puncturing. The low-rate encoder is called the *original* or *mother* code, while the high-rate encoder obtained after puncturing is called *punctured* encoder.

If a rate-$1/n$ original encoder is punctured by deleting some of the nP encoded bits corresponding to P information bits, then P is called the puncturing period. The puncturing pattern can be represented as an nxP matrix **P**, whose elements are 1's and 0's with a 1 indicating inclusion and a 0 indicating deletion. Thus, the code rate is determined by the period P and the number of bits deleted. If we delete N bits out of nP, the code rate is $P/(nP - N)$, where N may take any integer value in the range 0 to $(n - 1)P - 1$. Hence the achievable code rates are:

$$R_c = \frac{P}{P + M}, \qquad M = 1,2,..., (n-1) \cdot P \tag{3.3}$$

FEC provides a number of advantages and disadvantages. Compared to retransmissions, FEC does not require a back-channel and may provide lower delay since it does not depend on the round-trip-time of retransmits. Disadvantages of FEC include the overhead for FEC even when there are no losses, and possible latency associated with reconstruction of lost packets. Most importantly, FEC-based approaches are designed to overcome a predetermined amount of loss and they are quite effective if they are appropriately matched to the channel. If the losses are less than a threshold, then the transmitted data can be perfectly recovered from the received, lossy data. However, if the losses are greater than the threshold, then only a portion of the data can be recovered, and depending on the type of FEC used, the data may be completely lost. Unfortunately the loss characteristics for packet networks are often unknown and time varying. Therefore the FEC may be poorly matched to the channel, making it ineffective (too little FEC) or inefficient (too much FEC).

3.4.2 Error concealment

The severity of residual errors can be reduced if error concealment techniques are employed to hide visible distortion as much as possible . The basic goal of error concealment is to estimate the lost information or missing pixels in order to conceal the fact that an error has occurred. Error concealment is an extremely important component of any error robust video codec. The key observation is that video exhibits a significant amount of correlation along the spatial and temporal dimensions. This correlation is used to achieve video compression, and unexploited correlation can also be used to estimate the lost information. Therefore, the basic approach in error concealment is to exploit the correlation by performing some form of spatial and/or temporal interpolation (or extrapolation) to estimate the lost information from the correctly received data.

Three general approaches for error concealment are: (1) spatial interpolation, (2) temporal interpolation (freeze frame), and (3) motion-compensated temporal interpolation. It is important to select the appropriate error concealment technique to obtain a reasonably good visual quality. In *spatial interpolation* the implication is that a coefficient in a damaged block is likely to be close to the corresponding coefficients in spatially adjacent blocks. Therefore, the missing pixels are estimated by smoothly extrapolating the surrounding correctly received pixels. Correct recovery of the missing pixels is extremely difficult, however it provides significantly better concealment than assuming that the missing pixels have an amplitude of zero. In *temporal extrapolation* the missing pixels in a damaged macroblock are replaced by the pixels at the same spatial location in the previous correctly decoded frame (freeze frame). This approach is very effective when there is little motion, but visual artifacts can be produced in the presence of large motion. An improvement can be obtained by the *motion-compensated temporal extrapolation* where the missing block of pixels are replaced by the motion compensated block from the previous correctly decoded frame. The problem in this approach is that motion information may not be available in all circunstances. Possible approach include using a neighboring motion vector as an estimate of the missing motion vector, or compute a new motion vector by leveraging the correctly received pixels surrounding the missing pixels; but these can lead to large errors in reconstructed images.

In traditional concealment methods, the spatial and temporal correlations of frames are exploited, as shown in figure 3.8. However, the layered structure of the scalable extension of H.264/AVC means that it is possible to use other layers in the concealment of lost/erroneous frames in different layers of the scalable bit-stream to improve performance (Figure 3.9).

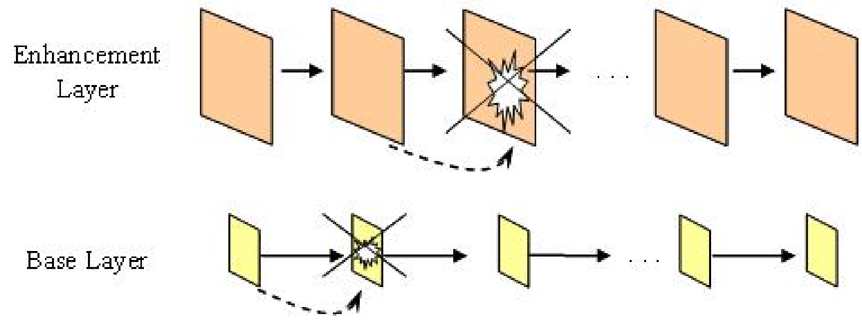


Figure 3.8. Traditional error concealment.

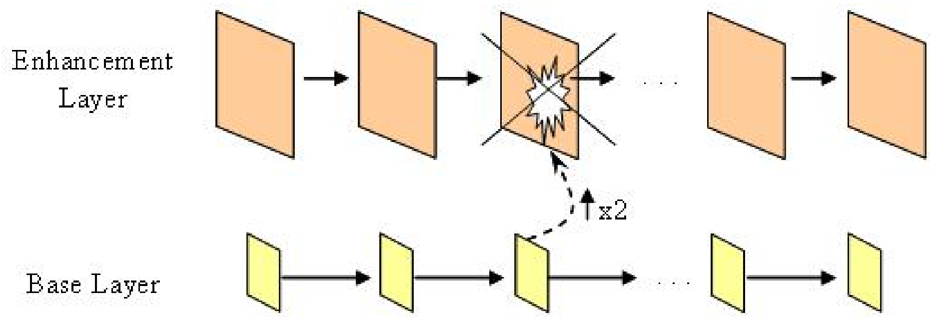


Figure 3.9. Error concealment in SVC.

3.4.3 Joint Source/Channel Coding

Data and video communication are fundamentally different. In data communication all data bits are equally important and must be reliably delivered, though timeliness of delivery may be of lesser importance. In contrast, for video communication some bits are more important than other bits, and often it is not necessary for all bits to be reliably delivered. On the other hand, timeliness of delivery is often critical for video communication. An example of coded video data with different importance include the different layers in a scalable coding; i.e., base layer is critically important and each of the enhancement layers is of successively lower importance. In general, *joint source and channel coding* is accomplished by designing the quantizer and entropy-coder at the source, and the design of FEC and modulation schemes at the channel coder for given channel error characteristics to minimize the effect of transmission errors . For example, for data communication all bits are of equal importance and FEC is designed to provide equal error protection for every bit. However, for video data it is desirable to have unequal error protection (UEP) and, instead of a common modulation scheme and error correction code for all bits, it is desirable to transmit some groups of information bits with more redundancy than others and use unequal (or prioritized) modulation strategies.

Chapter 4

Wireless communication system

We focus now on the major conceptual components in a wireless video communication system shown in figure 4.1, that describes the realization of our proposed solution to the problem statement. At the transmitter side, video packets are first generated by a video encoder, which performs compression by exploiting both temporal and spatial redundancies. At this stage compression efficiency of the video signal is most important, as the content is usually encoded with relatively high quality and independently of any actual channel characteristics. We note that the heterogeneity of client networks makes it difficult for the encoder to adaptively encode the video contents for a wide degree of different channel conditions; this is specially true for wireless clients. Subsequently, after passing through the network protocol stack, transport packets are generated and then transmitted over a wireless channel that is lossy in nature. Therefore, the video sequence must be encoded in an error resilient way that minimizes the effects of losses on the decoded video quality. The set of source coding parameters directly control video delivery quality; this includes prediction mode and quantization step size.

Wireless channels typically exhibit high variability in throughput, delay and packet loss. Providing acceptable video quality in such an environment is a demanding task for the video encoder and decoder, as well as the communication and networking infrastructure. Therefore, the power used to transmit each bit, the modulation mode, and the channel coding rate may be adapted based on source content and available

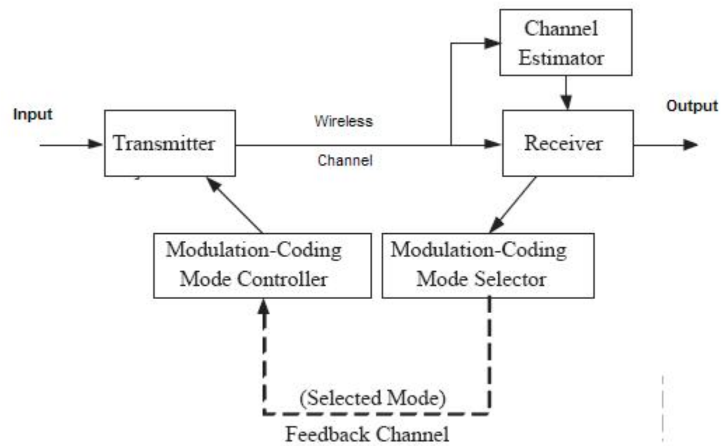


Figure 4.1. Wireless communication system

channel state information (CSI). Perfect channel state information (CSI) available at the receiver is an assumption adopted in this thesis.

At the receiver, the demodulated bit stream is processed by the channel decoder, which performs error detection and/or correction. Corrupt packets are usually discarded by the receiver, and are therefore considered lost. The video decoder employs concealment techniques to mitigate the effects of packet losses. Here, the goal is to achieve the best video delivery while using a minimum amount of transmission energy. Then, the video decoder decompresses video packets and displays the resulting video frames in real time.

4.1 Scalable video coding

As it was explained in the previous chapter, the Scalable Video Coding (SVC) design, which is an extension of the H.264/AVC video coding standard, enable the signal to be separated into multiples layers of different visual importance, this provides fast adaptation to bandwidth variations and complexity scalability properties that are essential for efficient transmission over error prone wireless networks. In this section the reference software used as well as the scalability features required by our application are explained.

4.1.1 SVC Reference Software (JSVM software)

The JSVM (Joint Scalable Video Model) [21] software is the reference software for the Scalable Video Coding (SVC) project of the Joint Video Team (JVT) of the ISO/IEC Moving Pictures Experts Group (MPEG) and the ITU-T Video Coding Experts Group (VCEG). It is written in C++ and is provided as source code. The JSVM software package contains a Software Manual, which provides information on using this software.The JSVM Software is still under development and changes frequently.

The usage of the JSVM software is illustrated by means of example for the scalable scenario. For the encoder and decoder part, there are two configuration files, in which parameters for the video encoding and decoding can be set. Generally, the configuration files present a collection of configuration parameters. Each parameter is specified in one line of the configuration file.

Let us consider a spatial scalable coding with two spatial resolutions, QCIF and CIF. Where, the base layer (layer 0) represents a QCIF sequence with a frame rate of 15 Hz and the enhancement layer, a CIF sequence with a frame rate of 30 Hz. Examples for the main and layer configuration files are depicted in figure 4.2, 4.3 and 4.4 respectively.

```
# JSVM Main Configuration File

OutputFile              test.264    # Bitstream file
FrameRate               30.0        # Maximum frame rate [Hz]
FramesToBeEncoded       150         # Number of frames (at input frame rate)
GOPSize                 16          # GOP Size (at maximum frame rate)
BaseLayerMode           2           # Base layer mode (0: AVC w larger DPB,
                                    #   1:AVC compatible, 2:AVC w subseq SEI)
SearchMode              4           # Search mode (0:BlockSearch, 4:FastSearch)
SearchRange             32          # Search range (Full Pel)
NumLayers               1           # Number of layers
LayerCfg                layer0.cfg  # Layer configuration file
```

Figure 4.2. Example main configuration file “main.cfg” for spatial scalable coding.

```
# JSVM Layer Configuration File

InputFile       BUS_176x144_15.yuv      # Input  file
ReconFile       rec_layer0.yuv          # Reconstructed file
SourceWidth          176                # Input  frame width
SourceHeight         144                # Input  frame height
FrameRateIn          15                 # Input  frame rate [Hz]
FrameRateOut         15                 # Output frame rate [Hz]
```

Figure 4.3. Example layer configuration file “layer0.cfg” for spatial scalable coding.

```
# JSVM Layer Configuration File

InputFile            BUS_CIF30.yuv  # Input  file
SourceWidth          352            # Input  frame width
SourceHeight         288            # Input  frame height
FrameRateIn          30             # Input  frame rate [Hz]
FrameRateOut         30             # Output frame rate [Hz]
InterLayerPred       2              # Inter-layer Pred. (0: no, 1: yes, 2:adap.)
```

Figure 4.4. Example layer configuration file "layer1.cfg" for spatial scalable coding.

The most important parameters that need to be specified in the main configuration file are the name for the bitstream *OutputFile*, the frame rate *FrameRate*, the number of frames to be encoded *FramesToBeEncoded*, the GOP size *GOPSize*, and the base layer mode *BaseLayerMode*. Furthermore, for the parameter *NumLayers*, at least one layer configuration file has to be specified via the parameter *LayerCfg*. The parameter *InterLayerPred* is set equals to 2, that specifies that the base layer (layer 0) is employed for inter-layer prediction. We additionally specified *SearchMode* and *SearchRange* to speed-up the encoder execution. For all other parameters of the main configuration file, default values are used.

It is assumed that the original sequence "BUS_CIF30.yuv" is given in CIF resolution with a frame rate of 30 Hz. In a first step, the resampling tool is used for generating a spatially and temporally downsampled sequence "BUS_QCIF15.yuv" in QCIF resolution with a frame rate of 15 Hz. Then the encoder is called with the main configuration file "main.cfg" and additional options to specify the quantization parameter. Figure **??** illustrates an example encoder call, also the corresponding encoder output is shown, which summarizes the supported spatio-temporal resolutions and bit-rates.

The generated scalable bit-stream contains 9 different representations, 4 of them are QCIF representation with frame rates of 1.875, 3.75, 7.5, and 15 Hz, and 5 of them are CIF representations with frame rates of 1.875, 3.75, 7.5, 15, and 30 Hz. All of the included QCIF representations are AVC compatible and can be decoded with a standard AVC decoder. Sub-streams for any of the 9 included representations can be extracted using the *BitStreamExtractorStatic*. The extracted substream is decoded using *H264AVCDecoderLibTestStatic* and the PSNR of the decoded video sequence and the bit-rate of the extracted sub-stream are measured using the *PSNR tool*.

```
>DownConvertStatic 352 288 BUS_CIF30.yuv 176 144 BUS_QCIF15.yuv 0 1

>H264AVCEncoderLibTestStatic -pf main.cfg -lqp 0 30 -lqp 1 32
...
SUMMARY:
                         SNR Level bitrate    Y-PSNR   U-PSNR   V-PSNR
                         --------- -------- -------- -------- --------
   176x144 @  1.8750      0.0000    71.0145  36.9930  41.3140  42.6465
   176x144 @  3.7500      0.0000   104.8926  36.2550  41.1325  42.4912
   176x144 @  7.5000      0.0000   143.5089  35.5507  40.9885  42.3923
   176x144 @ 15.0000      0.0000   187.6496  34.8944  40.9219  42.2986
   352x288 @  1.8750      0.0000   299.4675  37.3048  41.7122  43.4969
   352x288 @  3.7500      0.0000   412.7953  35.9055  41.3231  43.1962
   352x288 @  7.5000      0.0000   536.1900  34.8822  41.0904  43.0406
   352x288 @ 15.0000      0.0000   682.5856  34.1586  40.9484  42.9109
   352x288 @ 30.0000      0.0000   809.1168  33.5824  40.7982  42.7801
```

Figure 4.5. Example encoder call for spatial scalable coding.

4.1.2 Hybrid Temporal-Spatial Scalability

Within this work an hybrid Temporal-Spatial scalability is used for differentiation between video layers. Temporal scalability is an important tool for enhancing the motion smoothness of compressed video. Typically, a base-layer stream coded with a frame rate f_{BL} is enhanced by another layer consisting of video frames that do not coincide (temporally) with the base layer frames. Therefore, if the enhancement layer has a frame rate of f_{EL} then the total frame of both the base- and enhancement-layer streams is $f_{BL} + f_{EL}$. Based on the Spatial scalability structure described above, the frame rate of the transmitted video is "locked" to the frame rate of the base-layer regardless of the available bandwidth and corresponding transmission bitrate. At the same time, one of the design objectives of scalable video is to cover a relatively wide range of bandwidth variation over IP networks (e.g., 100 kbit/s to 1 Mbit/s). Consequently, it is quite desirable that the spatial scalability be completed with a temporal scalability tool. This, for example, can be used in response to users' preferences and/or real-time bandwidth variations at transmission time. For typical streaming applications, both of these elements are not known at the time of encoding the content.

4.2 Transport prioritization

The system resources available to a digital communication system designer are the signal power and the channel bandwidth. Given the channel noise characteristics, the design objective typically is to optimize the usage of these resources in maximizing the information throughput, while striving to meet certain performance criterion such as the probability of error at a given signal to noise ratio. If more bandwidth is available and the signal power is limited, the desired probability of error can be achieve by using channel coding. On the other hand, if the bandwidth is limited, the signal power has to be increased to meet the performance requirements. This point of view assumes that channel coding and modulation are treated as separate components of the system and are separately optimized.

A recent application of convolutional coding has been in communication systems that combine channel coding and modulation in one step to achieve better performance with no bandwidth expansion. This approach was proposed by Ungerboeck [27, 28, 29] and has been called *trellis-coded modulation* (TCM)

These concepts will be applied to combat channel errors, where layared coding is combined with prioritization methods (FEC, modulation, TCM) with an UEP approach, where the base layer is delivered with a higher degree of error protection.

4.2.1 Digital Modulation

Digital modulation involves mapping discrete symbols (which we assume to be a binary sequence) to a signal in a set or constellation. The signal sequence to be transmitted over the channel can be either memoryless or with memory.

The memoryless modulation scheme is used in this work. In this case, the binary sequence is parsed into subsequences each of length k, and each sequence is mapped into one of the $s_m(t)$, $1 \leq m \leq 2^k$, signals regardless of the previously transmitted signals. The demodulator recovers the m bits by making an independent M-ary nearest-neighbor decision on each signal received.

There are four major classes of digital modulation techniques used for transmission of digitally represented data:

- Amplitude-shift keying (ASK).
- Frequency-shift keying (FSK).
- Phase-shift keying (PSK).
- Quadrature amplitude modulation (QAM).

All of these, convey data by changing some aspect of a base signal, the carrier wave (usually a sinusoid), in response to a data signal. In this work we focus on PSK, mainly, and on QAM.

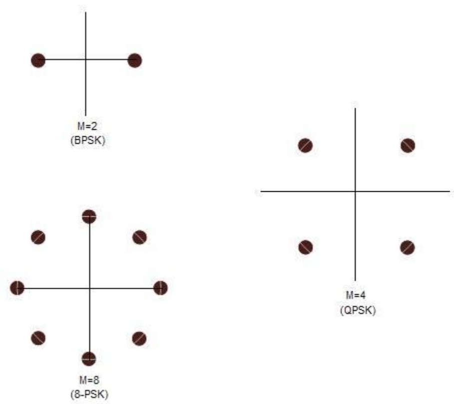


Figure 4.6. Signal space diagrams for BPSK, QPSK and 8-PSK.

In PSK the phase is changed to represent the data signal. The constellation points chosen are usually positioned with uniform angular spacing around a circle. This gives maximum phase-separation between adjacent points and thus the best immunity to corruption. They are positioned on a circle so that they can all be transmitted with the same energy. Signal space diagrams for BPSK (binary PSK, $M = 2$), QPSK (quaternary PSK, $M = 4$), and 8-PSK are shown in figure 4.6

In QAM, an inphase signal and a quadrature phase signal are amplitude modulated with a finite number of amplitudes, and summed. It can be seen as a two-channel system, each channel using ASK. The resulting signal is equivalent to a combination of PSK and ASK. Within this thesis only a rectangular 16-QAM is used in our simulations (figure 4.7).

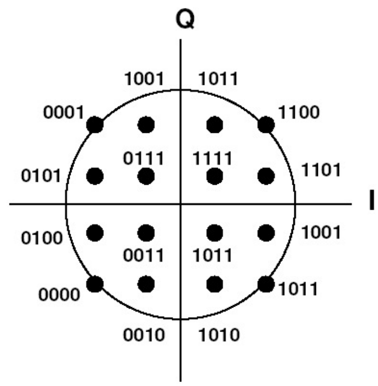


Figure 4.7. Signal space diagram for 16-QAM.

In the following section, the performance of PSK, when the channel corrupts the transmitted signal by the addition of white noise, is briefly described. In case of QAM, the performance analysis will be evaluated with the implementation of coded modulation (4.2.3).

Performance analysis

A signal set is characterized by the minimun distance between two signal points. In digital phase modulation, the minimun distance is given by [3]

$$d_{min} = 2\sqrt{\left(log_2 M \ x \ sin^2\left(\frac{\pi}{M}\right) \ E_b\right)} \tag{4.1}$$

If the signal power is kept constant, then selecting a larger signal set (e.g., 8-PSK instead of 4-PSK) results in a system that is more bandwidth efficient, since it can transmit at a higher bit rate in each hertz of bandwidth; but with a net reduction in the performance of the scheme (the probability of error increases), due to the reduction of the minimun distance 4.1.

For binary phase modulation, the error probability is:

$$P_b = Q\left(\sqrt{\frac{2E_b}{N_0}}\right) \tag{4.2}$$

where we have used the definition of the SNR per bit as

$$\frac{E_b}{N_0} = \frac{E}{N_0\, log_2 M} \tag{4.3}$$

where E indicates the signal energy and N_0 the double-sided power spectral density of the noise process.

When M = 4, we have in effect two binary phase-modulation signals in phase quadrature. Since there is no crosstalk or interference between the signals on the two quadrature carriers, the bit error probability is identical to that in 4.2. On the other hand, the symbol error probability for M =4 is determined by noting that

$$P_e = 1 - (1 - P_b)^2 = 2Q\left(\sqrt{\frac{2E_b}{N_0}}\right)\left[1 - \frac{1}{2}Q\left(\sqrt{\frac{2E_b}{N_0}}\right)\right] \tag{4.4}$$

For the case when $M > 4$ and the SNR is large, we can use an approximation to the symbol error probability P_e

$$P_e \approx 2Q\left(\sqrt{(2log_2 M)\, sin^2\left(\frac{\pi}{M^2}\right)\frac{E_b}{N_0}}\right) \tag{4.5}$$

Since the most probable errors due to noise result in the erroneous selection of an adjacent phase, most k-bit symbol errors contain only a single-bit error. Hence, the equivalent bit error probability for M-ary PSK is well approximated as

$$P_b \approx \frac{1}{k} P_e \tag{4.6}$$

4.2.2 Forward Error Correction: RCPC codes

If low signal-to-noise ratio limits the performance of the modulation system, the ability to correct errors can justify the rate loss caused by sending redundant check symbols.

Similarly, long delays in error-recovery procedures can be a good reason for trading transmission rate for forward error-correction capability. Our Forward error correction (FEC) scheme uses RCPC *(Rate-compatible Punctured Convolutional)* codes.

In the transmission of multimedia signals there is a need to transmit some groups of information bits with more redundancy than others (unequal error protection); therefore, instead of using separate codes to encode the different groups of bits, it is desirable to use a single code that has variable redundancy. This can be accomplished by RCPC codes, that are constructed from a low-rate code $1/n$ by periodic puncturing.

Two encoders of different rates are said to be rate-compatible if the redundant symbols generated by the higher rate encoder are also generated by the lower rate encoder. The lower rate encoder, of course, generates additional redundant symbols. This is the *rate-compatible criterion* introduced by Hagenauer (1988). Due to the rate-compatible criterion, the code rate of RCPC codes can be changed during transmission and thus unequal error protection is obtained [24, 25].

A family of RCPC encoders is described by the original $(n,1,m)$ encoder, and nxP puncturing matrices; where, the encoding rates given by the equation 3.3 are between $P/(P+1)$ and $1/n$.

Performance Analysis

Conventional encoders and decoders for error correction operate on binary, or more generally Q-ary, code symbols transmitted over a discrete channel. With a code of rate $k/n < 1$; where, $n-k$ redundant check symbols are appended to every k information symbols. Since the decoder receives only discrete code symbols, Hamming distance (the number of symbols in which two code sequences or blocks differ, regardless of how these symbols differ) is the appropriate measure of distance for decoding and hence for code design. A minimum Hamming distance d^H_{min}, also called “free Hamming distance” in the case of convolutional codes, guarantees that the decoder can correct at least [(d^H_{min} -1)/2] code symbol errors.

The bit error rate of convolutional codes may be found by extensive simulations. A good estimated value could be found by calculating error bounds. For rate $R = k/n$ convolutional codes, this bound can be generalized to apply also to punctured convolutional codes, and is given by [25]

$$P_b \leq \frac{1}{p} \sum_{d=d_f}^{\propto} c_d P_d \tag{4.7}$$

where c_d is the sum of bit errors (the information error weigth) for error events of distance d, and d_f is the free distance of the code. P_d is the pairwise error probability, which for coherent BPSK on an Additive White Gaussian Noise (AWGN) channel is given by [3, p. 488]

$$P_d = Q\left(\sqrt{2d\frac{E_b}{N_0}}\right) \tag{4.8}$$

where E_b denotes the received energy per information bit, $N_0/2$ the double-sided power spectral density of the noise process, and $Q(x) = \left(\sqrt{2\pi}\right)^{-1} \int_x^{\propto} e^{-z^2/2} dz$

In the appendix we present RCPC codes used in this work, taken by [26], with constraint length $K = 7$ using mother code rate $R = 1/4$ ODS (optimum distance spectrum) codes. In [26], P. Frenger, P. Orten and T. Ottoson study the influence of changing the puncturing period, p; and indicate that p=8 is a good choice. We have thus chosen to present RCPC codes only for $p = 8$.

4.2.3 Coded modulation: TCM

As it was mentioned before, moving from binary modulation to 2^k-ary modulation has the advantage that the number of bits per symbol is increased by a factor of k, thus increasing the spectral efficiency of the system. On the other hand, the required average energy of the signal increases or the distance between modulation symbols decreases. In practice, transmitted power is limited to a maximum value, this implies that the signal points becomes closer to each other.

Increasing the bandwidth of the signal is not a practical solution, since the channel bandwidth is expensive or limited. Decreasing the data rate is a possible solution, but places a limit in the number of services or applications offered. In addition, the increased delay may not be acceptable in a video transmission. In this sense, Massey [30] pointed out that communication system designs that treat the coding and modulation components of the system as being independent of each other do not achieve the best performance. Coded modulation is the joint design of error correcting codes and digital modulation schemes to increase the bandwidth efficiency of a digital communication system. Trellis-Coded Modulation (TCM) is a basic approach to design coded modulation systems.

In TCM the basic idea is to expand the constellation to obtain the redundancy needed for error correcting coding, and then to use coding to increase the minimun Euclidean distance between sequences of modulated signals. In the receiver, the noisy signals are decoded by a soft-decision maximum-likelihood sequence decoder. The term “trellis” is used because these schemes can be described by a state transition (trellis) diagram similar to the trellis diagrams of binary convolutional codes.

TCM combines rate $R = k/(k+1)$ binary convolutional codes with an M-ary signal constellation ($M = 2k+1$) in a way that achieves coding gain without increasing the transmission bandwidth. For example a rate 2/3 convolutional code can be combined with 8-PSK modulation by mapping the 3 encoded bits into one 8-PSK symbol. This combination has the same spectral efficiency (2 information bits/symbol) as an uncoded QPSK modulation. Here the redundant bits produced by coding are not used to transmit extra symbols, but instead they are used to expand the size of the signal constellation relative to an uncoded system. Hence coded modulation causes signal set expansion, not bandwidth expansion.

Set Partitioning

Set partitioning divides a signal set successively into smaller subsets with maximally increasing smallest intra-set distances. The concept of set partitioning is of central significance for TCM schemes. A $2^{m+1}-$th modulation signal set is partitioned in $m+1$ lebels. In each partition level, the signal set is divided into two subsets such that the intra-set distance d_i is maximized. Figure 4.8 shows it for a 8-PSK signal set.

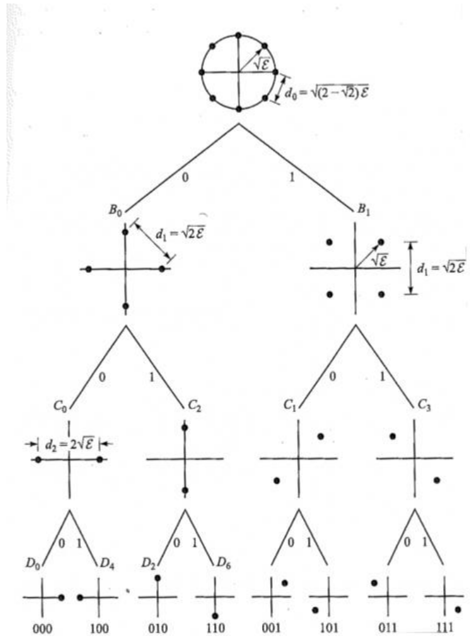


Figure 4.8. Set partitioning of an 8-PSK modulation [3]

The general structure of a TCM encoder is shown in figure 7.8. According to this figure, TCM signals are generated as follows: When m bits are to be transmitted per encoder/modulator operation, $\tilde{m} \leq m$ bits are expanded by a rate-$\tilde{m}/(\tilde{m}-1)$ binary convolutional encoder into $\tilde{m}+1$ coded bits. These bits are used to select one of $2^{\tilde{m}+1}$ subsets of a redundant 2^{m+1}-ary signal set. The remaining $m-\tilde{m}$ uncoded bits determine which of the $2^{m-\tilde{m}}$ signals in this subset is to be transmitted [29].

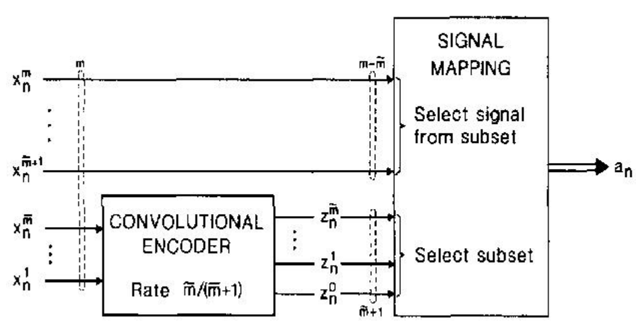


Figure 4.9. General encoder for trellis-coded modulation.[29]

Simple four-state TCM schemes can improve the robustness of digital transmission against additive noise by 3 dB, compared to conventional,uncoded modulation. With more complex TCM schemes, the coding gain can reach 6 dB or more. The best codes currently known for one-, two-,four-, and eight-dimensional signal sets are given in [29].

Performance analysis

In this work, and based on the studies made in [27], we consider the fact that for high SNR the event-error probability and bit-error probability become identical, because in this particular case the free Euclidean distance occurs for single-signal error events with only one wrong bit per event.

The event error probability is the probability that at any given time the decoder makes a wrong decision among the signals associated with parallel transitions, or starts to make a sequence of wrong decisions along some path diverging for more than one transition from the correct path [27, 28, 29]. If soft ML-decoding is applied, this probability, at high SNR, is well approximated by

$$P_e \simeq N_{free} \cdot Q[d_free/(2\sigma)] \tag{4.9}$$

where N_{free} denotes the (average) multiplicity of error events with distance d_{free}, and $Q(\cdot)$ is the Gaussian error probability function. The above approximate formula expresses the fact that at high signal to noise ratios probability of error events associated with a distance larger then d_{free} becomes negligible.

A table of optimum codes for TCM schemes used in our simulations, is shown in the Appendix B.

4.3 Channel model

For Internet video streaming the video frames at the sender are encoded, packetized and transmitted to the receiver where they are decoded and displayed. In the following we assume additive white Gaussian noise (AWGN) channel, that is a channel whose sole effect is addition of white Gaussian noise process to the transmitted signal. This channel is mathematically described by the relation:

$$r(t) = s_m(t) + n(t) \tag{4.10}$$

where $s_m(t)$ is the transmitted signal which, as we have seen in the previous section is one of the M possible signals; $n(t)$ is a sample waveform of a zero-mean white Gaussian noise process with power spectral density of $N_0/2$; and $r(t)$ is the received waveform. This channel model is shown in figure 4.10

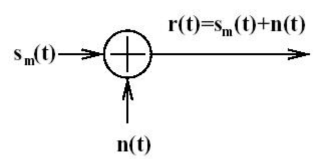


Figure 4.10. Model for received signal passed through an AWGN channel.

Although the AWGN channel model seems very limiting, its study is beneficial because noise is the major type of corruption introduced by channels; therefore, isolating it from other channel impairments and studying its effects results in better understanding of its effects on all communication systems.

4.4 Prioritization methods assignment

4.4.1 Packet loss rate

For Internet video streaming the video frames at the sender are encoded, packetized and transmitted to the receiver where they are decoded and displayed. Visual communication is simulated assuming an RTP/UDP/IP transport. As agreed at the 15th JVT meeting in Busan, only packet losses are considered and bit errors are not considered, as UDP would discard packets with bit errors. As is specified in [31] the simplest mode of packetization is considered, where a single NAL unit is transported in a single RTP packet.

The packet loss rate is given by

$$P_p = 1 - (1 - e)^N \tag{4.11}$$

where e is the bit error when a certain RCPC channel code rate and/or modulation scheme is applied and N is the size of packets in bits.

4.4.2 Channel code rate assignment

It is necessary to assign the channel code rate to each video layer according to its importance so that the limited bandwidth can be used efficiently and the total distortion can be minimized. Based on [32] we solved this problem by the following search:

1. Get the PSNR value for each video layer. We emulated the dropping of each layer, perform its error concealment and then compute PSNR between the concealed data and the error free data.
2. List the video layers in descending order of their PSNR values.
3. Try every combination of channel code rates that satisfies the ascending rule. Specifically, $S = \{r_k = (r_{k,1}, r_{k,2}, ... r_{k,n}) : r_{k,i} \leq r_{k,i}$ for $i < j\}$ where $r_{k,i}$ denotes a channel code rate for the layer i.

4. Select the set S that minimizes the total distortion or PSNR, subject to the constraint on the overall transmission rate

With the above procedure, we reduced the candidate set S. For instance, when the overall transmission rate is constrained to be smaller than twice the original source rate, and there are 5 video layers where each of this layers can have one of four channel rate (1, 1/2, 1/3 and 1/4) this results in 5 candidates vectors shown in table 4.1. On the contrary, if full search method is employed, we need to calculate the PSNR in $4^5 = 1024$ cases.

	Layer 1	Layer 2	Layer 3	Layer 4	Layer 5
1	1/2	1/2	1/2	1/2	1/2
2	1/3	1/2	1/2	1/2	1
3	1/3	1/3	1/2	1	1
4	1/4	1/2	1/2	1	1

Table 4.1. Example of the channel code rate assignment.

Although the transmission rate was constrained to be smaller than twice the original source rate, the best channel rate combination happens only when the total transmission rate is close to the rate constraint.

4.4.3 Modulation scheme assignment

As with channel code rates, it is necessary to assign a modulation scheme to each video layer according to its importance. The digital modulation methods described above can be compared in a number of ways. For example, one can compared them on the basis of the probability of error achieved at a specified SNR. However, such a comparison would not be very meaningful, unless it was made on the basis of some constraint, such a fixed data rate of transmission or, equivalently, on the basis of a fixed bandwidth.

To measure the bandwidth efficiency, the parameter r was defined in [3] as the ratio of bit rate of the signaling scheme to the bandwidth of it

$$r = \frac{R}{W} \qquad b/s/Hs \tag{4.12}$$

A system with a larger r is a more bandwidth-efficient system since it can transmit at a higher bit rate in each hertz of bandwidth. For the two dimensional signaling schemes used in this work (QAM and MPSK) $r = log_2 M$. As we have seen before in all these systems, the power efficiency decreases as M is increased. Therefore, the size of constalletion in these systems determines the tradeoff between power and bandwidth efficiency. These systems are appropriate where we have limited bandwidth and desire $r > 1$, and where there is sufficiently high SNR to support increases in M.

In our implementation, we assumed that all video layers at the output of our SVC encoder have the same bit-rate, and the assigned modulation scheme should satisfy the overall bandwidth efficiency constraint. For instance, if the overall bandwidth efficiency is constrained to be twice the original one with binary transmission and there are three choices of modulation schemes: BPSK, QPSK and 8-PSK , the procedure yields the only three candidates presented in table 4.2.

	Layer 1	Layer 2	Layer 3	Layer 4	Layer 5
1	QPSK	QPSK	QPSK	QPSK	QPSK
2	BPSK	QPSK	QPSK	QPSK	8-PSK
3	BPSK	BPSK	QPSK	8-PSK	8-PSK

Table 4.2. Example of the modulation scheme assignment.

Although the same bit rate for all video layers assumption was made in developing this algorithm, the bit rates are actually varying in our system. It is, however, worthwhile to point out that this constraint provides an acceptable performance (slight discrepancies in the total bandwidth employed between the candidates) at a low complexity.

Note that hierarchical modulation proposed in table 4.2 does not provide any error protection like channel coding. As a result, according to the diagram, high channel SNR is required for an acceptable BER. What is actually offered by hierarchical modulation is the unequal priority control to the different parts of the data. To achieve good quality of service in low channel SNR values, one may add some channel coding techniques to the current system.

4.5 Error concealment

Error concealment capabilities are included in the decoder of SVC, so the severity of artifacts resulting from transmission errors is minimized. JSVM software version 9.8, support the following methods:

1. Frame copy.

2. All macroblocks are assumed to be coded in direct mode.

In our simulations replacing a damaged frame with the previous one, produced adverse visual effects due to the fact of the motion characteristics of the videos used: Bus and Mobile (Complex and slow motion respectively), where Bus sequence had the worst performance (lower PSNR) with this method, because of the presence of large motion.

On the other hand significant improvement could be obtained by using *direct mode*. This method is effective in our case (layered coding) because the error concealment is made using the corresponding base layer MVs. Since the projected moving objects in the base and the enhancement layers are at the same direction, there is a high correlation between the base and the enhancement MVs. Another advantage of selecting this mode is that the base layer MVs are typically better protected than the enhancement layer ones and so the direct mode has a better immunity to channel errors. Therefore, direct mode was the error concealment applied in our simulations.

For additional details about the performance of error concealment methods, the interested reader is refered to [33] where *frame copy* method turns out to be the worst method, especially for cases where there is error prorogation.

The generic term “error concealment” will be used throughout this work to refer to the *direct mode* case.

Chapter 5

Simulations

In this chapter we simulate an Internet video transmission scheme which uses a two layer scalable video coder with standard compatible H.264 base layer coding and one enhancement layer, with the transmission scheme presented in chapter 4. An overview of the whole scheme is given in fig. 5.1.

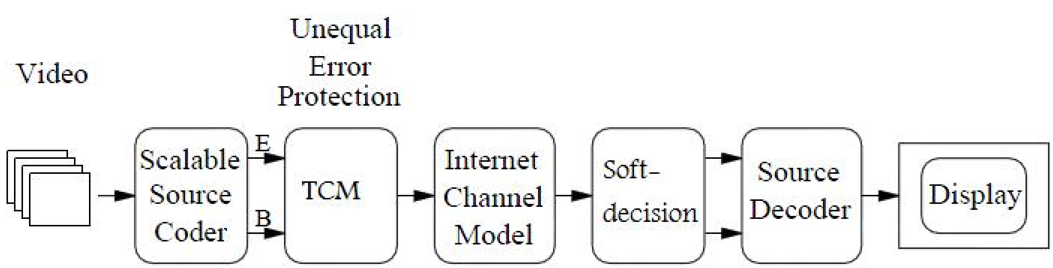


Figure 5.1. Outline of the video transmission scheme. *The scalable source coder: SVC encodes the input video in two layers, a base layer B carrying the most important and an enhancement layer E carrying the less important image information. Before the layers are transmitted they are packetized and protected against packet loss according to their importance by the TCM method described in chapter 4. Packet losses are simulated by the Internet channel model, presented in the same chapter. At the receiver, the noisy signals are decoded by a soft-decision maximum-likelihood sequence decoder. The source decoder reacts to missing packets with appropriate concealment techniques which are described in chapter 3.*

To demonstrate the efficacy of our UEP method and coded modulation, three experiments were performed. In the following, we will present simulation results based on different test conditions that show the influence of the selection of different channel code rates and/or modulation schemes.

5.1 Simulation environment

Simulations are presented using two sequences:

1. Bus, which is a fast-complex sequence, in CIF size with a frame rate of 30Hz. A total of 50 coded frames were used.

2. Mobile, that is a slow sequence, also in CIF size with a frame rate of 30Hz with a total of 50 coded frames used.

For all the test the SVC test model software, version JSVM 9.8 was used as the reference software, to produce hybrid temporal-spatial scalability; which accomplishes two of the design objectives of our work, that are to cover a relatively wide range of bandwidth variation over IP networks and to support heterogeneous devices with a single scalable bit stream.

The simulation parameters were set as follows:

- JSVM software was run in scalable mode for supporting spatial scalability, with two spatial resolutions, QCIF (176 x 144 pels) and CIF (352 x 288 pels).
- The QCIF layer was coded at a frame rate of 15 Hz, while the CIF layer was coded at a frame rate of 30 Hz. in order to obtain a temporal scalable representation.
- GOP size of 4. A GOP (group of pictures) consists of a key picture, which is generally coded as P picture, and several hierarchically coded B pictures that are located between the key pictures. The parameter GOPSize in the JSVM software must be equal to a power of 2. When specifying a GOP size of 4 pictures in the main configuration file, the effective GOP sizes that are used for the QCIF and CIF layer are 2 and 4, respectively. The hierarchical coding structure with 2 spatial layers employed, is illustrated in figure 5.2.

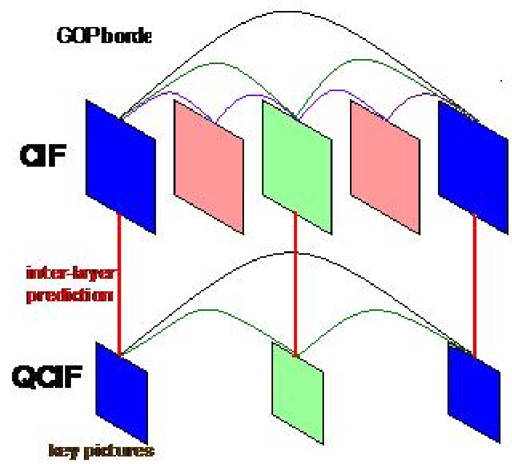
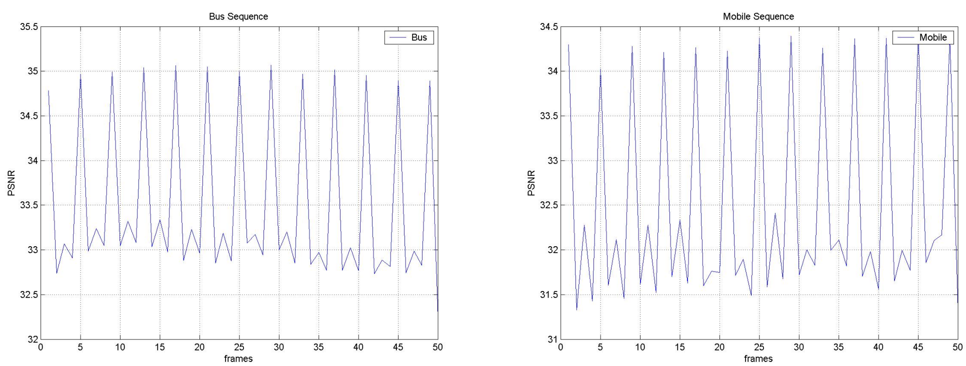


Figure 5.2. Hierarchical coding structure (GOP=4).

- The quantization parameter was set to 30 and 32 for the base and enhancement layer, respectively.
- Inter-layer prediction (*InterLayerPred*). For rate-distortion efficient coding, *InterLayerPred* should be set equal to 2 for any enhancement layer. *InterLayerPred* equal to 2 specifies that the base layer (layer 0) is employed for inter-layer prediction. The red arrows in figure 5.2 indicate the usage of inter-layer prediction. Inter-layer prediction can only be used inside an access unit, and thus between base and enhancement layer pictures at the same time instant. Since, the frame rate of the CIF enhancement layer is twice the frame rate of the QCIF base layer, the enhancement layer pictures of the highest temporal level are coded without inter-layer prediction. These pictures are only predicted using motion-compensated temporal prediction.
- AVC compatible bit-stream with additional sub-sequence SEI (Supplemental Enhancement Information) messages. SEI message provides statistical information, required for enabling the extraction of downsampled sub-streams without having to delve deeply into the syntax of a compressed bitstream.
- The generated scalable bit-stream contained 5 different representations, 2 of them were QCIF representation with frame rates of 7.5, and 15 Hz, and 3 of them were CIF representations with frame rates of 7.5, 15, and 30 Hz.
- Error concealment as discussed in the previous chapter.

For channel coding, RCPC with five rates, which are 8/10, 8/12, 8/16, 8/24 and 8/32, were employed. The modulation schemes used were: BPSK, QPSK, 8-PSK and 16-QAM, except where noted otherwise in the description of the experiments.

For the diagrams, a regular PSNR measurement was used. The reported PSNR is the arithmetic mean over the decoded luminance PSNR over all frames of the encoded sequence and since the locations of errors affect the quality of the reconstructed video significantly, each curve was obtained by averaging PSNRs over 20 different error patterns.



(a) Bus (b) Mobile

Figure 5.3. PSNR in 50 frames of video

Figure 7.4 shows the performance of the codec for two test sequences. For each sequence, PSNR values for decoding at the higher bitrate, corresponding to the higher resolution, are illustrated. In this case, the PSNR for each video was measured without the introduction of any transmission errors.

Since every 4th frame was coded in Intra mode (GOP=4), error propagation due to motion compensated prediction is limited, unless the Intra frame is also lost. Of course, error propagation can further be reduced by increasing the number of Intra macroblocks in the Inter coded frames.

5.2 Experiments performed

In classical digital communication systems, the functions of modulation and error-correction coding are separated. Modulators and demodulators convert an analog waveform channel into a discrete channel, whereas encoders and decoders correct errors that occur on the discrete channel. Recently, however, there has been an increasing interest in some types of combined modulation and coding schemes, called trellis-coded modulation (TCM), that achieve coding gain without using additional bandwidth. For this reason, we want to compare the following three groups of transmission modes with different methods of transport prioritization:

1. RCPC coded in BPSK mode.

2. Uncoded (without RCPC) M-ary PSK modes, where $M = 2^n$, n = 1, 2 and 3.

3. Trellis-coded modulation, using QPSK and rate-1/2 , 8-PSK and rate-2/3 and 16-QAM and rate-3/4.

This transmission modes are listed under tables 5.1, 5.4 and 5.7, respectively; where the modulation scheme, channel code rate and bandwidth efficiency are specified for each video layer encoded with SVC when equal error protection (EEP) and unequal error protection (UEP) modes are employed. It is clear that the UEP method provides a significant performance improvement as compared to the EEP method.

5.2.1 Transmission mode 1: RCPC coded in BPSK mode

In transmission mode 1, we investigated the performance of the channel rate allocation that consists in a traditional bandwidth expansive coding technique. Two different bit error rate (BER) environment, determined by two values of SNR: 6dB and 5dB, test 1 and test 2 respectively (table 5.2), were implemented. These values were found as described in section 4.2.2.

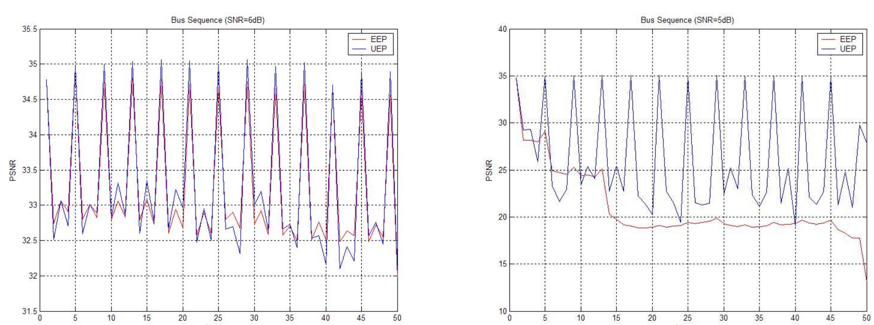
	Layer	1	2	3	4	5
EEP	Modulation	BPSK	BPSK	BPSK	BPSK	BPSK
	Code rate	1/2	1/2	1/2	1/2	1/2
	r(b/s/Hz)	0.5	0.5	0.5	0.5	0.5
UEP	Modulation	BPSK	BPSK	BPSK	BPSK	BPSK
	Code rate	1/3	1/3	2/3	4/5	4/5
	r(b/s/Hz)	0.33	0.33	0.66	0.8	0.8

Table 5.1. Transmission mode 1 with equal and unequal error protection (EEP-UEP).

	Modulation	BPSK	BPSK	BPSK	BPSK
	Rate	1/3	1/2	2/3	4/5
Test 1	BER	4.3732e-9	2.9257e-8	3.6587e-7	4.1020e-6
Test 2	BER	3.5532e-7	1.6250e-6	1.5122e-5	1.1398e-4

Table 5.2. Bit error rate in transmission mode 1 for SNR=6dB (Test 1) and SNR=5dB (Test 2)

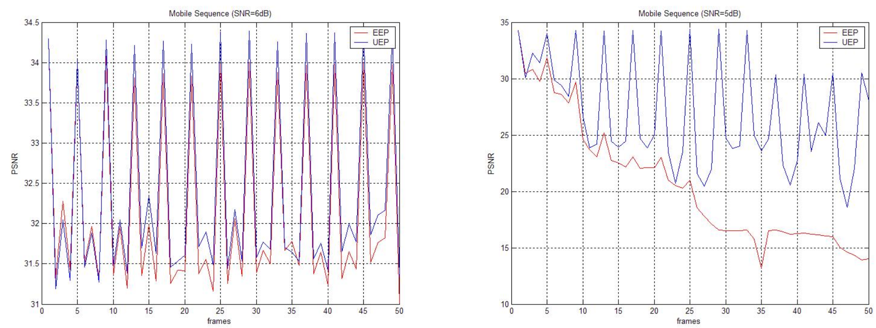
Figures 5.4 and 5.5 show the PSNR performance of unequal error protection (UEP) and the equal error protection (EEP) , which uses the same channel code rate (1/2) for all base and enhancement layers.



(a) Test 1 (b) Test 2

Figure 5.4. PSNR comparison for the “Bus” sequence, in 50 frames

As expected, at low bit error rates (in test 1), the performances of both EEP and UEP methods are close to each other since the convolutional code is capable of recovering lost packets in both cases. On the other hand, at higher bit error rates, the base layer in the EEP method is subject to packet losses that cannot be recovered by the convolutional code; therefore the PSNR degradation of enhancement layer increases



(a) Test 1 (b) Test 2

Figure 5.5. PSNR comparison for the “Mobile” sequence, in 50 frames

since the motion of base layer cannot be used of enhancement in the error concealment. Nevertheless, the UEP method can still recover these lost packets since the error protection level that is associated with the base layer is significantly higher than the corresponding one in the EEP method. This results in a better performance of the UEP method compared to the EEP method.

Comparing these four curves, in figures 5.4 and 5.5, we can see that for the slow motion sequence (Mobile), the gap between UEP and EEP is larger than that for the fast motion sequence (Bus). This was expected, since the main body of the compression in these experiments is a motion vector.

	Test 1		Test 2	
	EEP	UEP	EEP	UEP
Mobile	32.1546	32.3580	20.6160	26.8009
Bus	33.2388	33.2752	20.9459	26.0131

Table 5.3. PSNR, mean values (Transmission mode 1)

In 5.3, average PSNR performance comparison of these four schemes is given. Two standard video sequences Mobile and Bus, have been selected for testing. It is clear that the PSNR performance improvement is achieved from the UEP strategy.

5.2.2 Transmission mode 2: Uncoded in M-ary PSK modes

In transmission mode 2 a bandlimited channel that might use a nonbinary modulation signal set, in an attempt to increase the information throughput without increasing bandwidth, was employed. For such a case, the performance of convolutional coding techniques, like those described in the previous transmission mode, are disappointing. Therefore, in this simulations we investigated the performance of the unequal priority control, using uncoded hierarchical modulation to achieve a mean value of bandwidth efficiency of two (r=2 b/s/Hz). As a result, higher SNR is required for an acceptable BER. In this mode a SNR of 12dB was used.

Table 5.4 shows the proposed UEP (refer to section 4.4.3) and EEP, which uses the same modulation scheme (QPSK) for all base and enhancement layers.

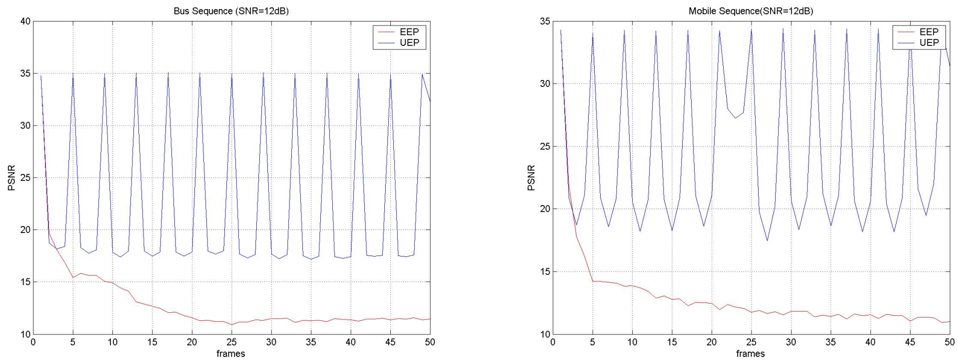
	Layer	1	2	3	4	5
EEP	Modulation	QPSK	QPSK	QPSK	QPSK	QPSK
	Code rate	1	1	1	1	1
	r(b/s/Hz)	2	2	2	2	2
UEP	Modulation	BPSK	BPSK	QPSK	8-PSK	8-PSK
	Code rate	1	1	1	1	1
	r(b/s/Hz)	1	1	2	3	3

Table 5.4. Transmission mode 2 with equal and unequal error protection (EEP-UEP).

Modulation	BPSK	QPSK	8-PSK
Rate	1	1	1
BER	9.0060e-9	3.4303e-5	1.0399e-2

Table 5.5. Bit error rate in transmission mode 2

Although the SNR considered was higher than the one implemented in the previous transmission mode, higher bit error rate was obtained; therefore, the average PSNR was poor (Table 5.6). However, in these conditions and according to the previous analysis, UEP outperformed clearly EEP method for both sequences (Bus and Mobile).



(a) Bus sequence (b) Mobile sequence

Figure 5.6. PSNR comparison for SNR=12dB

	PSNR	
	EEP	UEP
Mobile	13.0049	24.4337
Bus	12.9935	22.4832

Table 5.6. PSNR, mean values (Transmission mode 2)

5.2.3 Transmission mode 3: Trellis-coded modulation

In these simulations we will focus on transmission modes with convolutionally coded modulations for a bandlimited channels, using TCM to achieve coding gain expansion, where coding gain is defined as the reduction in required SNR for a given error probability, at the same bandwidth efficiency (r) per layer as in transmission mode 2 and at the same SNR value (12dB). On the basis of this constraint (r=2 b/s/Hz), we can compare transmission mode 2 and 3 on the basis of the probability of error and the average PSNR achieved at a specifed SNR. The schemes used for EEP, which uses rate-2/3 and 8-PSK modulation for all video layers, and UEP are presented in table 5.7.

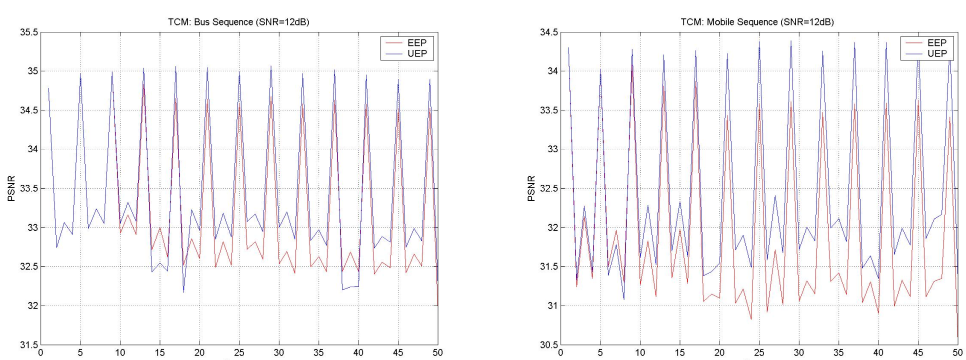
	Layer	1	2	3	4	5
EEP	Modulation	8-PSK	8-PSK	8-PSK	8-PSK	8-PSK
	Code rate	2/3	2/3	2/3	2/3	2/3
	r(b/s/Hz)	2	2	2	2	2
UEP	Modulation	QPSK	QPSK	8-PSK	16-QAM	16-QAM
	Code rate	1/2	1/2	2/3	4/5	4/5
	r(b/s/Hz)	1	1	2	3	3

Table 5.7. Transmission mode 3 with equal and unequal error protection (EEP-UEP).

Exact closed-form expressions for PER (packet error rate) and BER (bit error rate) are not available for TCM. We hence based our results on previous work of Ungerboeck (Appendix B) and these are shown in table 5.8. In case of rate-1/2 and QPSK scheme, and based on the information found in [3], where QPSK is seen as a combination of two BPSK in phase quadrature, because there is no crosstalk or interference between the signals on the two quadrature carriers; the same equations used in transmission mode 1 were applyed to find the BER values in this case.

Modulation	QPSK	8-PSK	16-QAM
Rate	1/2	2/3	3/4
BER	1.2371e-18	9.0060e-9	9.5220e-7

Table 5.8. Bit error rate in transmission mode 3



(a) Bus sequence

(b) Mobile sequence

Figure 5.7. PSNR comparison for SNR=12dB

	PSNR	
	EEP	UEP
Mobile	31.9023	32.4024
Bus	33.1985	33.3905

Table 5.9. PSNR, mean values (Transmission mode 3)

The desire to improve the poor coding results seen on bandlimited channels of transmission mode 2 was the motivation that led to trellis-coded modulation (TCM). Due to the FEC advantage, transmission mode 3 had better error performance than transmission mode 2 (TM2) for the same data rate and SNR, which results in higher average PSNR (table 5.9) or in higher spectral efficiency for both slow and fast motion video. This coding gain could be achieved without sacrificing data rate or expanding bandwidth, but at the expense of increasing decoder complexity.

Chapter 6

Conclusions

In this work, we described the potential use of Scalable Video Coding (SVC), which is an extension of H.264/AVC, in a transmission scheme for Internet video streaming. The main purpose was to employ Trellis-coded modulation (TCM) scheme to provide unequal error protection (UEP) for scalable delivery of video in heterogeneous environments, and demonstrate that communication system designs that treat the coding and modulation components of the system as being independent of each other do not achieve the best performance. In experimental results, based on hybrid temporal-spatial scalability, the efficiency of coded modulation scheme was measured in terms of video frames and average luma peak signal-to-noise ratio (PSNR).

From our results, it is evident that hierarchical modulation is preferred in cases where we need to consider big amounts of transmitting energy. On the other hand, when convolutional coding scheme was used to provide UEP, it helped to decrease the bit error rate significantly. This in turn, results in the transmission of signals of specified quality with a smaller transmit power. In other words, it leads to higher power efficiency, but lower bit rate. The general finding of this thesis was that compared with uncoded modulation, the same amount of information can be transmitted within the same bandwidth with coding gains of 3-6 dB when Trellis coded-modulation is used as a prioritization method with SVC encoder.

Although they benchmark theoretical performance, the assumptions that an additive white Gaussian noise (AWGN) is the channel of the video streaming, may not always hold true. One possible extension of this work is to analyze our unequal error protection scheme with fading channel, to achieve a good understanding of the key physical parameters and modeling issues of the wireless channel . Generalizations of layered transmission was considered in this thesis. A deep study of multicarrier wireless system shared by multiple users; namely, a multiaccess system (i.e., OFDM and OFDMA) could also be investigated.

Many systems that combines bit rate scalable media with FEC or hierarchical modulation have been studied before in great detail; but they have been limited by the absence of an efficient video codec. With SVC, the impact of such technologies will certainly grow over the next years. Rate-scalable media will be an important enabler to fully exploit the potentials of emerging delivery methods, improved receiver capabilities and new service offerings to satisfy increasing user expectations in a resource and cost efficient manner. SVC is a great solution for these kinds of services, in part because it enables backward interoperability with existing systems.

There is a huge potential for future work in this area. However, to have the full benefits of SVC encoders the three elements: encoder, server and player must be "SVC-aware"; therefore, it's going to be awhile to fully enjoy this technology. Luckily, companies involved in IP multimedia are embracing the standard, and it represents a viable opportunity for enhanced services in the future. For instance: Vidyo has licensed its H.264 SVC codec to Google for the Gmail Chat. RADVISION has incorporated the codec in its video development platform and testing tools. RADVISION and Samsung introduced a Video conferencing console that takes advantage of the scalable extension of H.264 and is aimed at a more general market threshold. Also, Apple, which recently added a YouTube channel to its Apple TV hardware, might have prompted YouTube to change its video formats, and start using the H.264 extension. Clearly, Adobe, Microsoft and Move are all aggressively pushing their proprietary alternatives. But with the ITU and ISO both behind H.264, and mobile emerging as the next great frontier for streaming, it's tough to bet against a standard.

Video communication over the internet has witnessed much progress in the past few years. Developments in algorithms and in compute, communication and network infrastructure technologies have continued to change the landscape of streaming media, each time simplifying some of the current challenges and spawning new applications and challenges. Therefore, we believe that video streaming will continue to be a compelling area for exploration, development, and deployment in the future.

Chapter 7

Riassunto

7.1 Introduzione

Comunicazione multimediale è uno dei temi più importanti nel settore delle tecnologie dell'informazione e delle comunicazioni di oggi. La rapida crescita delle applicazioni multimediali interattive, come la TV mobile, videochiamata e video download ha portato progressi spettacolari alle comunicazioni wireless. Con il grande successo del video su Internet, i servizi wireless multimediali dovrebbero essere ampiamente diffusi nel prossimo futuro. Cioè , diversi tipi di reti wireless stanno convergendo in reti all-IP.

Per raggiungere questo livello di accettabilità e proliferazione dei video su Internet, ci sono molte sfide tecniche che devono essere affrontate in materia di videocodifica e rete:

- Diversi tipi di reti hanno caratteristiche differenti.
- Eterogeneità drammatica tra gli utenti finali, dovuto a che i terminali di solito variano nella risoluzione del display e potenza, in base al loro stato di evoluzione e categoria.
- Diverse applicazioni hanno diversi requisiti di QoS, in termini di velocità di trasmissione di dati, limiti di ritardo e probabilità di perdita di pacchetti.

- Condizioni di rete variabile. Throughput di dati che variano nel tempo di un numero variabile di utenti.

"Layered compression" e " layered trasmission" sono interessanti soluzioni ai problemi delle caratteristiche dei moderni sistemi di trasmissione di video. Cioè, dobbiamo utilizzare uno schema di compressione che permetta generare più livelli di qualità, con un modello di consegna di rete che ci permetta fornire in modo selettivo, i sottoinsiemi di strati ai ricevitori individuali.

Inoltre, l'attuale Internet "best-effort" non offre una qualità di servizio (QoS), garanzia per lo streaming di video su Internet. Per questo motivo, la perdita di pacchetti è inevitabile in Internet e può produrre effetti indesiderati al decoder, come un comportamento imprevedibile e una qualità di riproduzione inaccettabile. Al fine di evitare tali effetti, è necessario progettare un sistema robusto di trasmissione per proteggere la qualità dei segnali multimediali contro gli errori di trasmissione.

Dovuto al fatto che l'importanza esistenti tra gli strati di video è disuguale, il tipo di trasmissione, che realizza una protezione pari degli errori non è ottimale. Pertanto, il bit rate scalabile può essere combinato con successo con una protezione disuguale degli errori, utilizzando Forward Error Correction (FEC) e modulazione gerarchica.

7.1.1 Obiettivi

Tali sistemi sono stati studiati in gran dettaglio in molte pubblicazioni di ricerca, ma la maggior parte di queste tecniche sono state limitate dalla mancata disponibilità di un scalabile video codec efficiente. O solo modulazione e codifica fissa è stata considerata nei sistemi con l'estensione scalabile di H.264/AVC (SVC).

Pertanto, l'obiettivo principale di questa ricerca è quello di costruire un sistema di trasmissione di video, che combine un codificatore di video scalabile con una protezione diseguale degli errori (UEP), con un approccio diverso ai lavori precedenti. Per raggiungere questo obiettivo la tesi indirizzi 3 reti separate:

- Registrare una comprensione del software per la codifica di video scalabile, usando questa conoscenza per formulare un ambiente ideale per lo sviluppo del software.
- Indagare le sfide in video streaming.
- Esplorare e confrontare i meccanismi di definizione delle priorità di trasporto, applicate in lavori precedenti, con il nostro approccio UEP.

Al fine di conseguire questi obiettivi, questa tesi propone una nuova soluzione combinata per "Joint Source-Channel Coding (JSCC)" per la trasmissione di video in ambienti soggetti a errori. Il modello utilizza l'estensione scalabile di H.264/AVC (SVC) per la codifica del video e una protezione diseguale degli errori (UEP) con modulazione codificata. Modulazione codificata è una combinazione di codici Rate-Compatible Punctured Convolutional (RCPC) e di modulazione digitale, per fornire Forward Error Correction (FEC) al fine di ridurre al minimo gli errori sul bit causati da problemi del canale durante la trasmissione del video, senza aumentare la larghezza di banda del canale. Noi vogliamo dimostrare che la soluzione proposta realizza un miglioramento significativoper una prescritta efficienza spettrale o throughput.

7.1.2 Struttura del riassunto

Il presente riassunto di tesi è strutturato nel seguente modo:

Nelle sezioni 7.2 e 7.3 cercheremo di capire quali fattori influenzano lo streaming video su Internet, quali sono i possibili problemi che si incontrano e quali le possibili soluzioni. La sezione 7.4 presenta il background del nostro lavoro. Innanzitutto, si parlerà di un generico sistema per la trasmissione video packet-based, considerando le caratteristiche, le limitazioni e le complessità di implementazione. Anche, in questa sezione, vengono presentati i concetti generali dello Scalable Video Coding (SVC). In seguito presentiamo i fondamenti del Trellis-coded modulation (TCM) e i metodi per il controllo di errore al trasmettitore e al decoder per ottenere una trasmissione più robusta agli errori grazie a tecniche come FEC e l'error-concealment. La sezione 7.5 espone la teoria che sta alla base del modello proposto e presenta le simulazioni che attestano la validità del modello. In ultimo si tracciano le conclusioni ed eventuali sviluppi del modello proposto e testato.

7.2 Problema

L'attuale Internet fornisce solo servizio "best-effort", e non fornisce alcuna garanzia o QoS (Quality of Service) per applicazioni multimediali. Pertanto; limitato, imprevedibilmente e variabile bit rate; reti e ricevitori eterogenee , elevati tassi di errore e variazioni casuali di canali wireless possono influire negativamente la qualità dei contenuti da consegnare. Questi problemi sono brevemente spiegati di seguito.

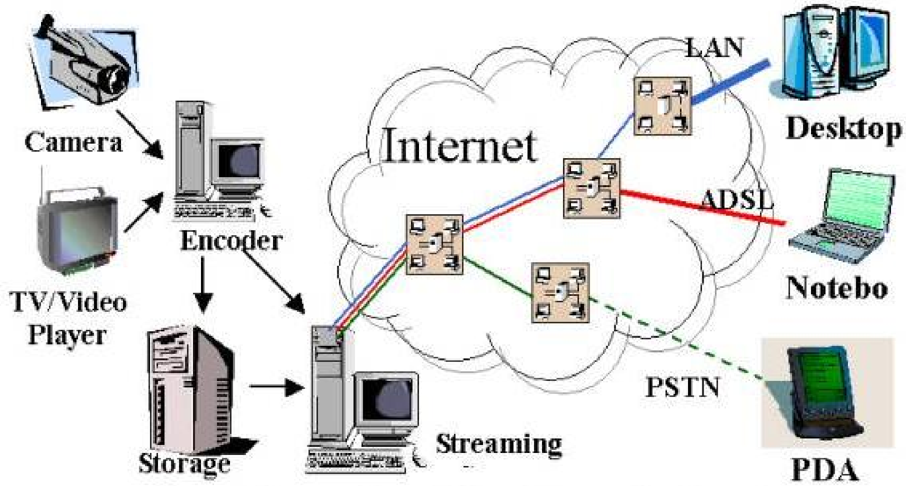


Figura 7.1. Video su internet.

1. **Larghezza di banda finita e bit rate limitato**

   La larghezza di banda è una risorsa condivisa, limitata che varia a seconda del tempo. La variazione della larghezza di banda è una delle caratteristiche primarie di reti "best-effort", e Internet è un ottimo esempio di tali reti.

   Il controllo del bit rate è importante per applicazioni di streaming multimediale. Se il mittente trasmette più veloce della larghezza di banda disponibile si verifica la congestione, i pacchetti vengono persi e si produce un grave calo della qualità del video. Se il mittente trasmette più lento rispetto alla larghezza di banda disponibile, il ricevitore produce una qualità di video sub-ottimali. L'obiettivo, per superare il problema della larghezza di banda, è di stimare la larghezza di banda disponibile e quindi corrispondere al bit rate trasmesso la larghezza di banda disponibile. In tal modo la compressione di video efficiente è necessaria.

2. **Eterogeneità**

   Una rete eterogenea è una rete le cui parti (sub-reti) possono avere una distribuzione disuguale delle risorse . Per esempio, alcune parti di una rete eterogenea possono avere larghezza di banda abbondanti e un controllo eccellente della congestione, mentre gli altri parti della rete sono sovraccarichati e congestionate da un uso eccessivo o da una mancanza di risorse di rete fisica. Differenti ricevitori su una rete eterogenea possono sperimentare caratteristiche di prestazioni diverse. Quando lo video è trasmesso su una rete eterogenea, questo deve essere decodeable a una qualità ottimale per gli utenti con una buona connessione di rete, e con una qualità utilizzabile per gli utenti con una connessione scadente.

   Inoltre, il video viene consegnato ad una varietà di dispositivi di decodificazione con capacità e display eterogenei. In questi ambienti eterogenei, adattamento flessibile del contenuto codificato è auspicabile, allo stesso tempo che consentono l'interoperabilità di encoder e decoder da produttori diversi.

3. **Alto tasso di errori**

   Reti basate su tecnologie wireless hanno un tasso di errore molto più elevati rispetto a quelli basati sulle tecnologie più tradizionali come la fibra ottica o cavo coassiale. I segnali wireless condividono lo stesso mezzo di propagazione con altri segnali, e di conseguenza, ci sono molte più opportunità per le interferenze che possono provocare errori di bit.

   Durante la trasmissione, errori sul bit potrebbero rendere tutto un pacchetto inutile. Perdite di pacchetti possono produrre effetti indesiderati al decoder, come un comportamento imprevedibile del decoder e una qualità inaccettabile di riproduzione. Per combattere l'effetto delle perdite, un video in streaming deve essere progettato con un sistema di controllo degli errori.

4. **Fluttuazioni casuali in tempo**

   Il segnale che si propaga attraverso un canale wireless, può esperimentare fluttuazioni casuali nel tempo. Ciò implica che, la larghezza di banda, ritardo, perdita, o altre caratteristiche della rete possono variare in modo significativo nel corso del tempo; ciò rende difficile un dessign affidabile delle prestazioni del sistema. Quando il video è trasmesso su una rete che varia nel tempo, la fonte dovrebbe essere in grado di adeguare i propri parametri alle condizioni variabili della rete (adattabilità).

7.3 Soluzione

7.3.1 Compressione e trasmissione a strati

In questo modello, al fine di rendere l'utilizzo della larghezza di banda wireless e le risorse del cliente più efficiente, piuttosto che distribuire un unico livello di qualità utilizzando un unico canale di rete, la fonte distribuisce cinque livelli di video contemporaneamente su canali paralleli. A sua volta, ogni ricevitore sintonizza il suo tasso di ricezione regolando il numero di strati che riceve. L'effetto netto è che il segnale è fornito a un gruppo eterogeneo di ricevitori a diversi livelli di qualità utilizzando un insieme eterogeneo di tassi. Per realizzare pienamente questa architettura, si deve selezionare uno schema di compressione che permetta generare diversi livelli di qualità per consegnare sottoinsiemi di strati a diversi ricevitori.

"Layered compression": Scalable Video Coding (SVC)

SVC, che è una estensione del standard di codifica di video H.264/AVC, può essere classificato come un video codec a livelli. SVC può combinare alcuni livelli in modo flessibile per adattarsi a diversi bitrate, frame rate o risoluzioni spaziali del contenuto del video, eliminando parti del bitstream. SVC è adatto per l'uso in diversi casi, come:

- **Supporto di dispositivi eterogenei con un unico bit stream scalabile.** Con lo schema di codifica scalabile, il video deve essere codificato una sola volta, per la risoluzione richiesta più alta , però con uno stream scalabile per cui rappresentazioni con qualità inferiore possono essere ottenute eliminando i livelli selezionati; quindi, nemmeno tutti gli streams devono essere ricevuti da tutti i terminali. Per esempio, lo strato di base di uno stream H.264/AVC scalabile deve essere ricevuto per il dispositivo a basso rendimento e garantire che tutti gli strati del video raggiungono il terminale ad alte prestazioni.

- **Adattamento alle diverse condizioni della rete.** I protocolli di rete fanno fronte alle variazioni di througput regolando la velocità di trasmissione. SVC esplicitamente permette la rimozione di pacchetti dal bit stream, che si traduce implicitamente in bit rate e con questo, riduzione della qualità di presentazione del video. Per esempio, quando cè un eccesso di banda, questo sarà utilizzato in modo efficiente in modo di massimizzare la qualità percettiva.

“Layered trasmission”

Combinando l’approccio di “layered compression” con un sistema di “ layered trasmission”, possiamo risolvere il problema dell’eterogeneità. In questa architettura, la codifica scalabile di video produce un flusso di livelli, dove ogni strato è trasmesso su un canale di rete differente. Dove, la rete trasmette solo il numero di strati che ogni collegamento fisico è in grado di supportare.

7.3.2 Metodi di definizione delle priorità

Un problema fondamentale in comunicazione di video sono le perdite. Le perdite possono avere un effetto molto distruttivo sulla qualità del video ricostruito , e se il sistema non è progettato per gestirgli, un singolo errore di bit può avere un effetto catastrofico. L’idea presentata in questo lavoro è proteggere fortemente la parte più importante del video scalabile (“base layer”) al fine di superare il caso peggiore di errore e dare meno importanza agli “enhancement layers”, al fine di superare le situazioni più tipiche di errore. I metodi utilizzati sono:

- UEP (Unequal error protection).
- Modulazione gerarchica.
- TCM (Trellis-coded modulation).

Queste tecniche cambiano il livello di modulazione: Binary phase shift keying (BPSK), quadrature PSK(QPSK), 8-PSK, 16-quadrature amplitude modulation (QAM), e così via; così come la quantità di ridondanza per un codice di correzione di errori. Una modulazione più elevata (ad esempio, 16-QAM), senza nessun codice di correzione di errore, può essere utilizzata da “enhancement layer” con conseguente degrado della qualità percettiva in base alle condizioni del canale e in una maggiore efficienza della larghezza di banda. Un livello inferiore di modulazione (ad esempio, BPSK), con più ridondanza per la correzione di errori, è utilizzato dal “base layer” in male condizioni di canale, questo si traduce in minore efficienza della larghezza di banda.

La nostra soluzione enfatizza l’adozione di *joint source/channel coding* (JSCC) o di una “progettazione congiunta della codifica di sorgente e codifica di canale” dalla teoria della comunicazione. JSCC unisce il design di compressione e codifica per il controllo di errori, per ottenere migliori prestazioni. Il nostro sistema di comunicazione video si basa su una soluzione interdisciplinare che combina la codifica di video scalabile e metodi di definizione delle priorità, per raggiungere una buona prestazione quando la probabilità di perdita di pacchetti aumenta in una connessione di Internet.

7.4 Background

7.4.1 Generalità

Il sistema di video mostrato nella figura 7.2, descrive la realizzazione di nostra soluzione, per il problema proposto. Nel trasmettitore, i pacchetti di video vengono prima generati da un video encoder, che effettua la compressione sfruttando la ridondanza temporale e spaziale. In questa fase l'efficienza di compressione del segnale di video è più importante, in quanto il contenuto è di solito codificato con una qualità relativamente elevata e indipendentemente da caratteristiche del canale. L'eterogeneità delle reti dei clienti, rende difficile per l'encoder codificare il video adattivamente per una certa gamma di condizioni di canale, questo è specialmente vero per i clienti wireless. Per questo motivo in questa tesi SVC è l'encoder utilizzato. L'estensione Scalable di H.264/AVC (SVC) ha elevati prestazioni di compressione e adattività per la distribuzione di video su reti eterogenee. SVC è basato su H.264/AVC e fornisce scalabilità spaziale, temporale e di qualità.

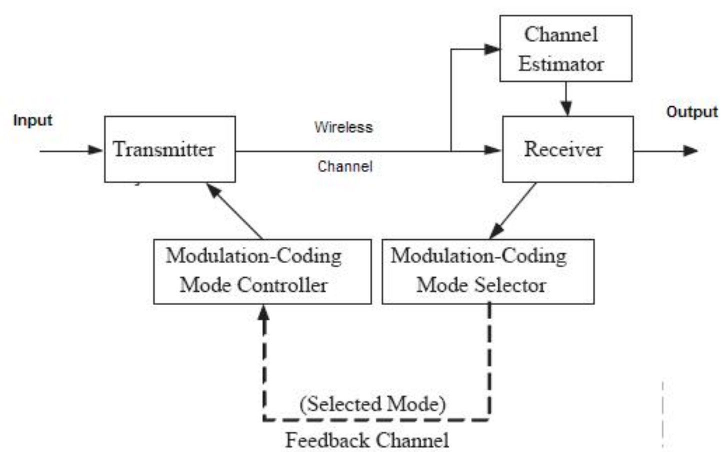


Figura 7.2. Sistema di video in comunicazione wireless.

Successivamente, i pacchetti sono generati e poi trasmessi su un canale wireless. I canali wireless in genere mostrano una gran variabilità nel throughput, ritardo e perdita di pacchetti. Fornire una qualità di video accettabile, in tale ambiente è un compito impegnativo per l'encoder e decoder video, così come per le infrastrutture di rete. Pertanto, la potenza utilizzata per trasmettere ogni bit, la modalità di modulazione e il tasso della codifica di canale, può essere adattata in base al contenuto di origine e le informazioni sullo stato dei canali disponibili (CSI). Perfetto CSI, disponibile al ricevitore è un presupposto adottato in questa tesi. Anche, per simplicità, si assume un canale con rumore AWGN (rumore bianco additivo gaussiano), che è un canale il cui unico effetto è l'aggiunta di rumore bianco gaussiano al segnale trasmesso.

Al ricevitore, il flusso demodulato di bit viene trasformato dal decodificatore di canale, che esegue il rilevamento e/o correzione di errore. I pacchetti corrotti vengono di solito eliminati dal ricevitore, e sono quindi considerate persi. Il decoder di video impiega tecniche di occultamento (error-concealment) per mitigare gli effetti delle perdite di pacchetti. Qui, l'obiettivo è quello di raggiungere una migliore consegna di video mentre si utilizza una quantità minima di energia di trasmissione. Poi, il decoder di video decomprime i pacchetti e mostra i fotogrammi di video in tempo reale.

In questa sezione i componenti più importanti del sistema di video utilizzati in questo lavoro, sono brevemente descritte.

7.4.2 Scalable video coding (SVC)

Un video bit-stream si chiama scalabile, quando le parti del flusso possono essere rimossi in modo che il sub-stream risultante forme un altro flusso di bit valido per alcun decoder, questo sub-stream rappresenta il contenuto di origine con una qualità di ricostruzione minore da quella del flusso di bit originale , ma è elevato se si considera la piccola quantità di dati rimanenti. Flussi di bit che non forniscono questa proprietà sono indicati come flussi di bit single-layer.

Scalabilità è già stato presentato nel video standard di codifica MPEG-2 video, H.263 e MPEG-4 Visual. Tuttavia, la disposizione di scalabilità spaziale e di qualità di questi standard, viene con un notevole incremento della complessità del decoder e con una riduzione significativa dell'efficienza di codifica rispetto al modo non scalabile. Questi inconvenienti, che hanno ridotto il successo dei profili scalabili, sono affrontati con il nuovo standard di H.264/AVC: SVC (Scalable video coding).

SVC bit-stream è costituito da un "base layer" e uno o più "enhancement layers”. Il “base layer” è un H.264/MPEG4-AVC bit-stream, che garantisce la compatibilità con ricevitori già esistenti, e gli “enhancement layers” portano ulteriori informazioni su qualità, risoluzione o frame rate.

Tipi di scalabilità

1. **Scalabilità temporale.**

   Scalabilità temporale permette diversi frame rates (espressa in Hz). Consiste nella possibilità di ricevere solo una frame ogni N ed essere comunque in grado di decodificare correttamente la sequenza video.

2. **Scalabilità spaziale.**

   Scalabilità spaziale descrive i casi in cui sottoinsiemi del flusso di bit rappresentano il contenuto di origine con una foto di dimensioni ridotte. L’obiettivo è quello di permettere all’utente di ricavare dal flusso codificato una sequenza video alla stessa risoluzione dell’originale, a una risoluzione più bassa oppure a una risoluzione maggiore.

3. **Scalabilità in qualità (SNR).**

   Scalabilità in qualità, anche comunemente indicato come scalabilità SNR, permette la ricostruzione di un frame determinato a livelli di differenti qualità con diversi numeri di bit, pur avendo la stessa definizione temporale e spaziale.

Software di Riferimento: JSVM

JSVM (Joint Scalable Video Model) software, è il software di riferimento per lo Scalable Video Coding (SVC), progetto del “Joint Video Team” (JVT) della norma ISO / IEC “Moving Pictures Experts Group” (MPEG) e ITU-T “Video Coding Experts Group” (VCEG). È scritto in C++ ed è fornito come codice sorgente. JSVM contiene un manuale, che fornisce informazione sull’utilizzo di questo software. The JSVM software è ancora in fase di sviluppo e di frequenti modifiche. In questa tesi, è stato utilizzato il software di versione JSVM 9.8.

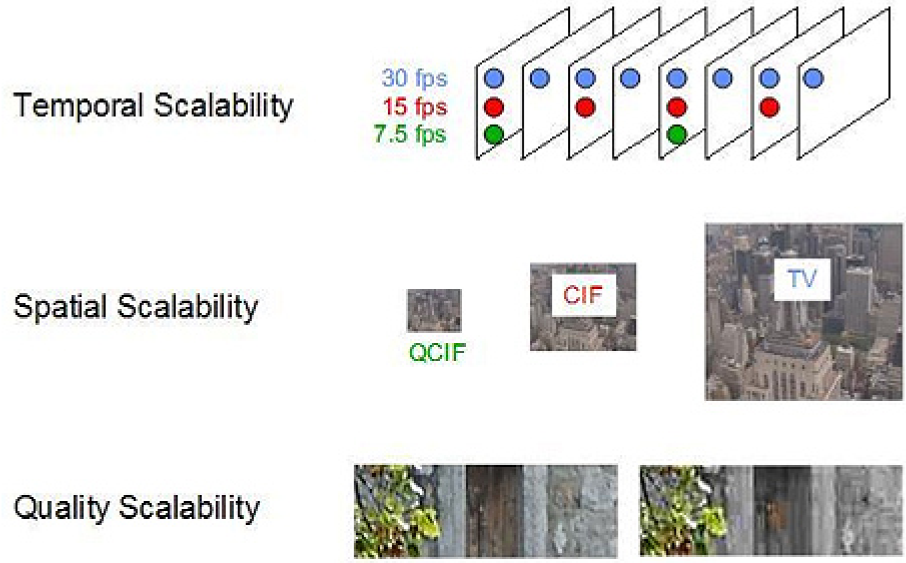


Figura 7.3. Tipi di scalabilità in codifica di video: SVC.

7.4.3 Modulazione digitale

La modulazione digitale consiste nella mappatura dei simboli discreti (che supponiamo di essere una sequenza binaria), ad un segnale in un insieme o costellazione. La sequenza del segnale da trasmettere attraverso il canale può essere sia di memoria o con memoria.

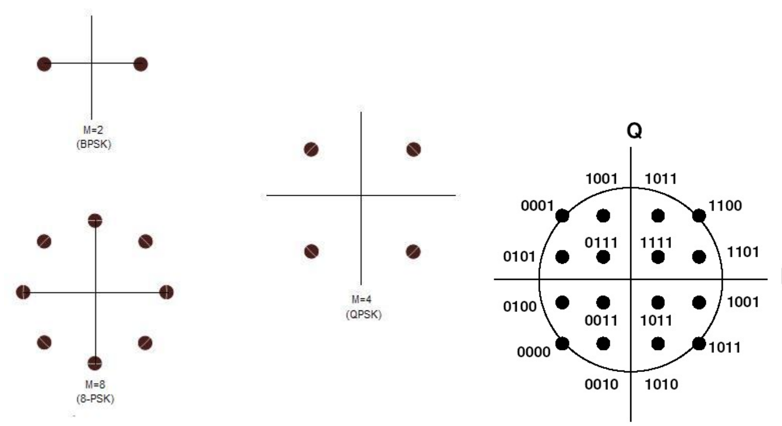
Lo schema di modulazione senza memoria viene utilizzato in questo lavoro. In questo caso, la sequenza binaria viene analizzata in sottosequenze di lunghezza k, dove ogni sequenza è mappata in uno dei $s_m(t)$, $1 \leq m \leq 2^k$ segnali, indipendentemente dalla precedente segnale trasmesso. Il demodulatore recupera i m bits facendo una decisione indipendente su ogni segnale ricevuto ("M-ary nearest-neighbor).

Ci sono quattro classi principali di tecniche di modulazione digitale utilizzata per la trasmissione digitale dei dati:

- Amplitude-shift keying (ASK).
- Frequency-shift keying (FSK).
- Phase-shift keying (PSK).

- Quadrature amplitude modulation (QAM).

Tutti questi, trasmettono dati cambiando qualche aspetto del segnale di base, la portatora d'onda (solitamente una sinusoide), in risposta ad un segnale di dati. In questo lavoro ci concentriamo su PSK, soprattutto, e QAM.



(a) BPSK, QPSK and 8-PSK (b) 16-QAM

Figura 7.4. Signal space diagrams for and 16-QAM.

7.4.4 Controllo degli errori per superare le perdite di canale

Un sistema di video streaming è stato progettato con controllo di errori per combattere l'effetto delle perdite. Ci sono quattro classi di approcci per il controllo di errore: (1) ritrasmissioni, (2) forward error correction (FEC), (3) error concealment, and (4) error-resilient di codifica di video. Le prime due classi di approcci possono essere pensati come approcci di codifica di canale per il controllo di errore, mentre gli ultimi due approcci sono codifica di fonte per il controllo degli errori. Un sistema di video di streaming è tipicamente progettato utilizzando un certo numero di questi diversi approcci. Forward error correction ed error concealment sono discussi nei paragrafi seguenti. Inoltre, la progettazione congiunta della codifica di sorgente e codifica di canale è molto importante ed è anche discusso in questa sezione.

Forward error correction (FEC): Codici convoluzionali e codici RCPC

L'obiettivo del FEC è quello di aggiungere la ridondanza necessaria che può essere usata per correggere gli errori. FEC ha l'effetto di aumentare l'overhead di trasmissione e quindi di ridurre la larghezza di banda utilizzabile per i dati utile. Le due principali categorie di FEC sono codifica a blocco e codifica convoluzionale. All'interno di questo lavoro, i codici convoluzionali sono impiegati.

Con i codici convoluzionali, il flusso di bit in ingresso viene applicato a un shift register, che consiste in K (k-bit), fasi e n generatori lineari, come è mostrato nella figura 7.5. Per ogni turno del shift register, k bits vengono inserite e n bits di codice sono consegnati, in modo che il tasso di codice è di $R_c = k/n$. La potenza di un codice convoluzionale è una funzione della sua "constraint length", K. Codici di constraint length grande tendono ad essere più potente. Purtroppo, con constraint length di grande dimensioni viene maggiore complessità del decoder.

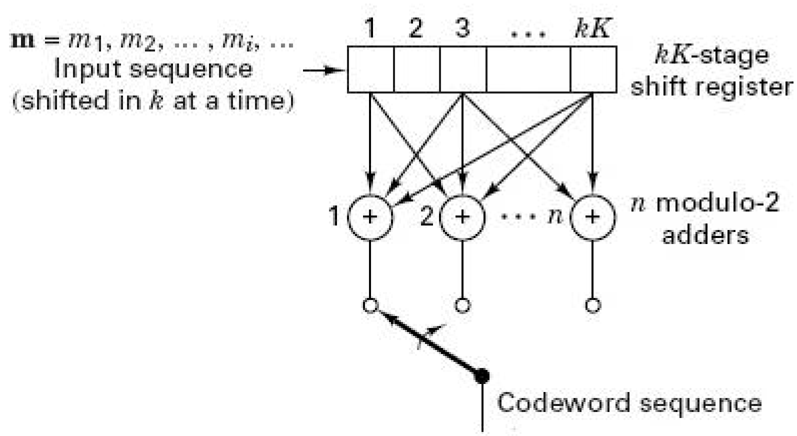


Figura 7.5. Codificatore convoluzionale.

Il nostro Forward Error Correction (FEC) regime utilizza codici RCPC *(Rate-compatibile Punctured convoluzionali)*. Nella trasmissione di segnali multimediali cè la necessità di trasmettere alcuni gruppi di bit di informazioni con maggiore ridondanza rispetto ad altri (UEP), pertanto, invece di utilizzare codici distinti per codificare i diversi gruppi di bit, è opportuno utilizzare un unico codice con ridondanza variabile. Ciò può essere eseguito dai codici RCPC, che sono costruiti da un codice a basso tasso $1/n$, con "puncturing".

"Puncturing" è la cancellazione dei bit codificati selezionati in uscita di un codificatore convoluzionale. Così, uno è in grado di generare codici convoluzionali di alto tasso "puncturing" i codici con un tasso di $1/n$, con il risultato che il decoder mantiene la complessità bassa, ma riduce la distanza libera del codice di rate $1/n$ una certa quantità che dipende dal grado di "puncturing".

Error-concealment

La gravità degli errori può essere ridotta se tecniche di error-concealment, o di occultamento di errore, sono impiegate per nascondere, quanto più possibile, la distorsione visibile. L'obiettivo di fondo è di stimare la informazione persa o pixels mancanti, al fine di nascondere il fatto che un errore si è verificato. Error-concealment è un componente estremamente importante di qualsiasi video codec. L'osservazione fondamentale è che il video mostra una notevole quantità di correlazione spaziale e temporale. Questa correlazione è utilizzata per realizzare la compressione di video, e la correlazione non sfruttata può anche essere utilizzata per stimare la perdita di informazione. Pertanto, l'approccio di base in error-concealment è quello di sfruttare la correlazione eseguendo una qualche forma di interpolazione (o estrapolazione) temporale o spaziale per stimare i dati persi dai dati correttamente ricevuti.

Nei metodi di occultamento tradizionali, le correlazioni spaziali e temporali di fotogrammi vengono sfruttati, come mostrato nella figura 7.6. Tuttavia, la struttura stratificata della estensione scalabile di H.264/AVC significa che è possibile utilizzare altri livelli per l'occultamento di fotogrammi smarriti oppure errati, in diversi strati del bit-stream scalabile per migliorare le prestazioni (figura 7.7).

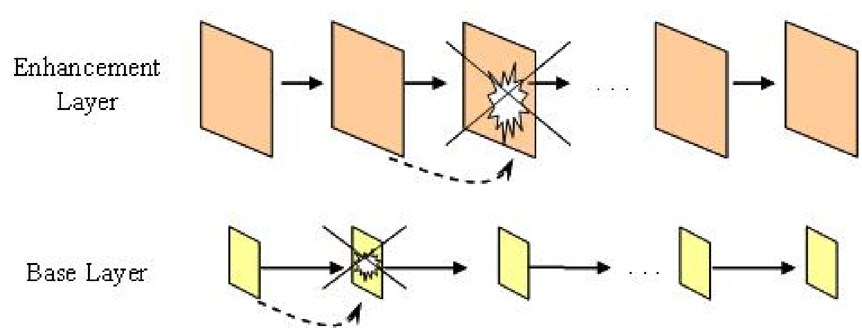


Figura 7.6. Traditional error concealment.

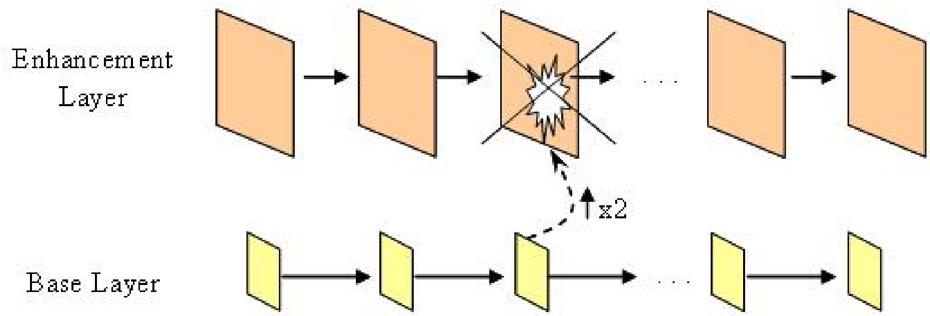


Figura 7.7. Error concealment in SVC.

Tecniche di error-concealment sono inclusi nel decoder di SVC. Versione del software JSVM 9.8, supportano i seguenti metodi:

1. Copia del frame precedente.
2. Si assumono che tutti i macroblocchi siano codificati in modo diretto.

In nostre simulazioni, la sostituzione di un frame danneggiato con quello precedente ha prodotto effetti visivi negativi. D'altra parte, un significativo miglioramento è stato ottenuto utilizzando il *modo diretto*. Questo metodo è efficace in nostro caso ("Scalable Video Coding"), perchè l'error-concealment è fatto utilizzando i MVs (motion vectors) corrispondenti al livello di base ("base layer"). Dal momento che gli oggetti in movimento proiettate nei base ed "enhancement layers" sono nella stessa direzione, cè una forte correlazione tra loro MVs. Un altro vantaggio di scegliere questa modalità è che i MVs del "base layer" sono in genere meglio protetti da quelli degli "enhancement layers" e quindi il *modo diretto* ha una migliore immunità agli errori del canale. Pertanto, il *modo diretto* è stato l'error-concealment applicato in nostre simulazioni.

Progettazione congiunta della codifica di sorgente e codifica di canale

In generale, la progettazione congiunta della codifica di sorgente e codifica di canale, si ottiene con la progettazione del quantizzatore della fonte, e la progettazione di FEC e gli schemi di modulazione al codificatore di canale, per le varie caratteristiche di errore del canale, per ridurre al minimo l'effetto di errori di trasmissione. Ad esempio, per la comunicazione dei dati tutti i bit sono di pari importanza, pertanto FEC deve essere progettato per fornire una protezione di errore uguale per ogni bit. Tuttavia, per il video, che è auspicabile avere una protezione di errore diseguale (UEP), invece di avere un regime comune di modulazione e lo stesso codice di correzione degli errori per tutti i bit, è opportuno trasmettere alcuni gruppi di bit di informazioni con maggiore ridondanza di altre e utilizzare differenti strategie di modulazione.

7.4.5 Trasporto con priorità

Le risorse disponibili dei sistemi di comunicazione tema digitale per un sono la potenza del segnale e la larghezza di banda del canale. Considerate le caratteristiche di rumore del canale, l'obiettivo di progettazione è in genere di ottimizzare l'utilizzo di queste risorse nel massimizzare il throughput di informazione, mentre si sforzano di soddisfare certo criterio di rendimento, come la probabilità di errore a un dato SNR. Se maggiore larghezza di banda è disponibile e la potenza del segnale è limitata, la probabilità di errore desiderato può essere conseguito mediante codifica di canale. D'altra parte, se la larghezza di banda è limitata, la potenza del segnale può essere aumentata per soddisfare i requisiti di prestazione. Questo punto di vista presuppone che la codifica di canale e modulazione sono trattati come elementi distinti del sistema e sono ottimizzati separatamente.

Una recente applicazione di codifica convoluzionale è stata utilizzata nei sistemi di comunicazione. Questa tecnica unisce la codifica di canale e modulazione per raggiungere prestazioni migliori, senza l'espansione della larghezza di banda. Questo approccio è stato proposto da Ungerboeck [27, 28, 29] ed è stato chiamato *Trellis-coded Modulation* (TCM).

Questi concetti saranno applicati per combattere gli errori di canale, in cui la codifica layared è combinata con metodi di priorità (TCM) con un approccio UEP, dove lo strato di base è fornito con un grado maggiore di protezione dagli errori.

Trellis-coded modulation (TCM)

La modulazione codificata è la progettazione congiunta di codici per correzione di errori e gli schemi di modulazione digitale, per aumentare l'efficienza della larghezza di banda di un sistema di comunicazione digitale. Trellis-Coded Modulation (TCM) è un approccio per progettare questi sistemi.

In TCM l'idea di base è quella di ampliare la costellazione per ottenere la ridondanza necessaria per la codifica di correzione di errori, e quindi di utilizzare il codice per aumentare la minima distanza euclidea tra le sequenze di segnali mo-dulati. Nel ricevitore, i segnali rumorosi vengono decodificati da un "soft-decision maximum-likelihood" decoder. Il termine "trellis" viene utilizzato in quanto questi sistemi possono essere descritti da un schema di transizione di stato (trellis), simile a quello dei diagrammi a trellis dei codici convoluzionali binari.

TCM combina un codice convoluzionale, di rate $R = k/(k+1)$, con una M-ary costellazione del segnale ($M = 2k+1$) in modo che raggiunga un guadagno di codifica senza aumentare la larghezza di banda di trasmissione. Ad esempio, un rate di 2/3 del codice convoluzionale può essere combinato con la modulazione 8-PSK, dove i 3 bit sono codificati in un 8-PSK simbolo. Questa combinazione ha la stessa efficienza spettrale (2 bit di informazione / simbolo) di una modulazione QPSK non codificata. Qui i bit ridondanti, prodotti dalla codifica non sono utilizzati per trasmettere i simboli extra, ma sono utilizzati per espandere le dimensioni della costellazione del segnale rispetto a un sistema non codificato. Quindi la modulazione codificata causa una espansione del segnale, non l'espansione della larghezza di banda.

La struttura generale di un codificatore TCM è mostrata nella figura 7.8. In base a questa figura, i segnali TCM sono generati nel modo seguente: quando i m bits devono essere trasmesse per encoder / modulatore, $\tilde{m} \leq m$ bits vengono espanse da un codificatore convoluzione binario di rate $\tilde{m}/(\tilde{m}-1)$ in $\tilde{m}+1$ bits. Questi bits vengono utilizzati per selezionare uno dei $2^{\tilde{m}+1}$ sottoinsiemi di un ridondante 2^{m+1}-ary set di segnale. I restanti $m - \tilde{m}$ bits non codificati determinano quale delle $2^{m-\tilde{m}}$ segnali in questo sottoinsieme è stato trasmesso.

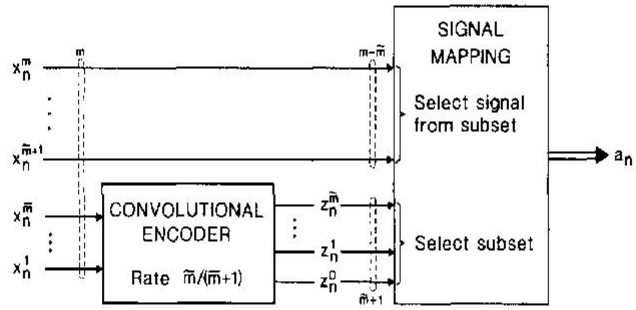


Figura 7.8. Codificatore generale del “trellis-coded modulation” (TCM).

7.5 Simulazioni

In questo capitolo si simula un sistema di trasmissione di video su Internet che utilizza un codificatore di video scalabile di due livelli, un “base layer”, compatibile con lo standard di codifica H.264, e un‘ “enhencement layer”. Si utilizza il sistema di trasmissione presentato nel capitolo 4. Una panoramica dell’intero sistema è dato in fig 7.9.

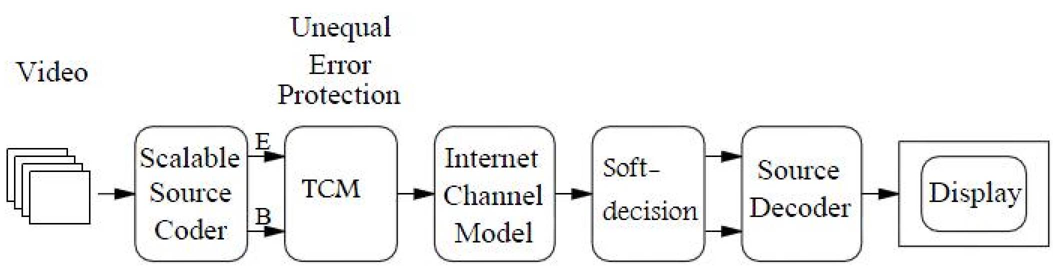


Figura 7.9. Schema del sistema di trasmissione video proposto.

7.5.1 Implementazione

Per dimostrare l’efficacia del metodo UEP e la modulazione codificata (TCM), tre esperimenti sono stati effettuati. Le simulazioni in questo riassunto sono presentati con solo una sequenza:

- Mobile, che è una sequenza lenta, in formato CIF con un frame rate di 30 Hz, con un totale di 50 frames codificati.

I parametri principali di simulazione sono stati fissati come segue:

- JSVM software è stato eseguito in modalità scalabile per supportare la scalabilità spaziale con due risoluzioni, QCIF e CIF.
- Lo strato QCIF è stato codificato in un frame rate di 15 Hz, mentre lo strato CIF è stato codificato in un frame rate di 30 Hz., al fine di ottenere una rappresentazione temporale scalabile.
- Dimensione del GOP (Group of picture) di 4. La struttura gerarchica di codifica con 2 strati spaziali impiegati è illustrata nella figura 7.10.

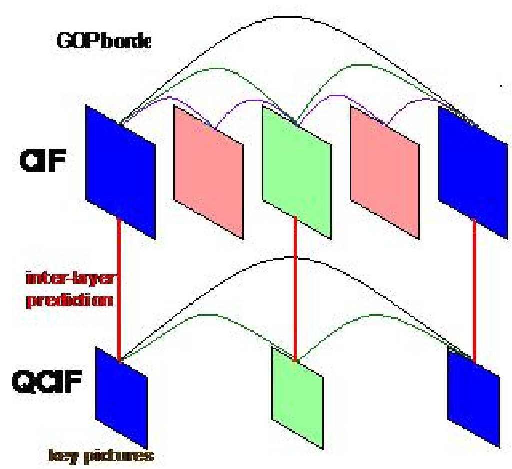


Figura 7.10. Struttura gerarchica di codifica (GOP=4).

- Lo scalabile bit-stream generato, conteneva 5 diverse rappresentazioni, 2 di loro erano QCIF con un frame rate di 7.5 e 15 Hz, e 3 di loro erano CIF con un frame rate di 7.5, 15 , e 30 Hz.
- *Modo diretto* come error concealment.
- Per la codifica dei canali, RCPC con cinque rates differenti, che sono 8/10, 8/12, 8/16, 8/24 e 8/32, sono stati impiegati. I sistemi di modulazione utilizzati sono

stati: BPSK, QPSK, 8-PSK e 16-QAM, eccetto dove sia indicato diversamente nella descrizione degli esperimenti.

- Per gli schemi, una misurazione regolare del PSNR è stata utilizzata. Però, come le posizioni degli errori incide sulla qualità del video ricostruito in modo significativo, ogni curva è stata ottenuta facendo la media dei PSNRs oltre più di 20 modelli diversi di errori.

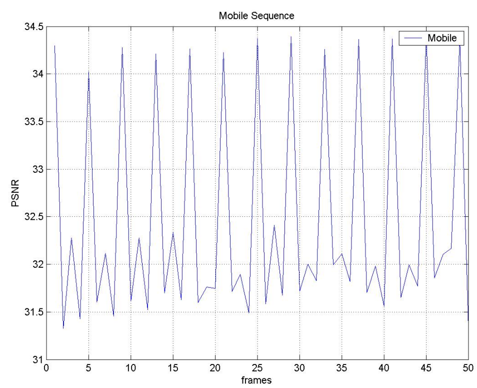


Figura 7.11. PSNR in 50 frames di video

La figura 7.11 mostra le prestazioni del codificatore utilizzato per la sequenza di test: Mobile (le prestazioni per la sequenza “Bus” è illustrata nel capitolo 5, figura 7.4). Per ogni sequenza, i valori di PSNR per la decodifica al bitrate più alto, corrispondente a la maggiore risoluzione, sono presentati. In questo caso, il PSNR del video è stato misurato senza l’introduzione di eventuali errori di trasmissione.

Nel seguito, presenteremo i risultati di alcuni simulazioni (i piu significativi), che mostrano l’influenza della selezione dei tassi di codice di canale e/o schemi di modulazione. Per maggiore informazione sulle condizioni specifiche di ogni esperimento presentato nella seguente sezione, referirsi ai capitoli 4 e 5 della tesi. Anche, nel capitolo 5, si fa una comparazione tra la sequenza “Bus”, una sequenza veloce, con la sequenza “Mobile”, e altre condizioni di errore sono presentate.

7.5.2 Esperimenti

Abbiamo preso in considerazione i seguenti tre gruppi di modalità di trasmissione:

1. RCPC codificati in modalità BPSK.

2. Non codificati con modulazione M-ary PSK , dove $M = 2^n$, n = 1, 2 e 3.

3. Trellis-coded modulation (TCM), con QPSK e rate-1/2 , 8-PSK con rate-2/3 e 16-QAM con rate-3/4.

Queste modalità di trasmissione sono elencati in dettaglio sotto i tavoli 7.1, 7.2 e 7.3, rispettivamente, dove il sistema di modulazione, code rate del canale e l'-efficienza della larghezza di banda sono specificati per ogni strato di video codificato con SVC quando le modalità “equal error protection” (EEP) e “unequal error protection” (UEP) sono impiegati. È chiaro che il metodo UEP fornisce un significativo miglioramento delle prestazioni rispetto al metodo EEP.

	Layer	1	2	3	4	5
	Modulation	BPSK	BPSK	BPSK	BPSK	BPSK
EEP	Code rate	1/2	1/2	1/2	1/2	1/2
	r(b/s/Hz)	0.5	0.5	0.5	0.5	0.5
	Modulation	BPSK	BPSK	BPSK	BPSK	BPSK
UEP	Code rate	1/3	1/3	2/3	4/5	4/5
	r(b/s/Hz)	0.33	0.33	0.66	0.8	0.8

Tabella 7.1. Modalità di trasmissione 3 con protezione uguali e disuguali dagli errori (EEP-UEP).

	Layer	1	2	3	4	5
	Modulation	QPSK	QPSK	QPSK	QPSK	QPSK
EEP	Code rate	1	1	1	1	1
	r(b/s/Hz)	2	2	2	2	2
	Modulation	BPSK	BPSK	QPSK	8-PSK	8-PSK
UEP	Code rate	1	1	1	1	1
	r(b/s/Hz)	1	1	2	3	3

Tabella 7.2. Modalità di trasmissione 2 con protezione uguali e disuguali dagli errori (EEP-UEP).

	Layer	1	2	3	4	5
EEP	Modulation	8-PSK	8-PSK	8-PSK	8-PSK	8-PSK
	Code rate	2/3	2/3	2/3	2/3	2/3
	r(b/s/Hz)	2	2	2	2	2
UEP	Modulation	QPSK	QPSK	8-PSK	16-QAM	16-QAM
	Code rate	1/2	1/2	2/3	4/5	4/5
	r(b/s/Hz)	1	1	2	3	3

Tabella 7.3. Modalità di trasmissione 3 con protezione uguali e disuguali dagli errori (EEP-UEP).

Nella prima modalitá di trasmissione abbiamo studiato le prestazioni del sistema quando viene utilizzata la codifica del canale, che consiste in una tecnica tradizionale di codifica espansiva di banda. In questo riassunto le prestazioni per un solo caso di errore di bit (BER), determinato da un valore di SNR: 5dB, sono presentati.

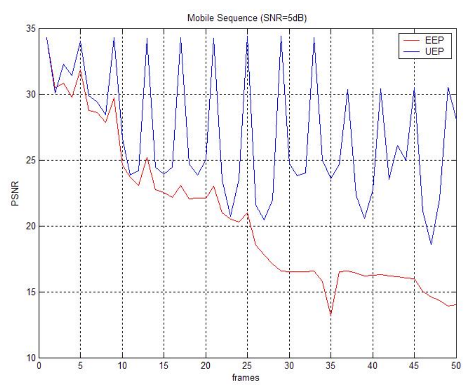
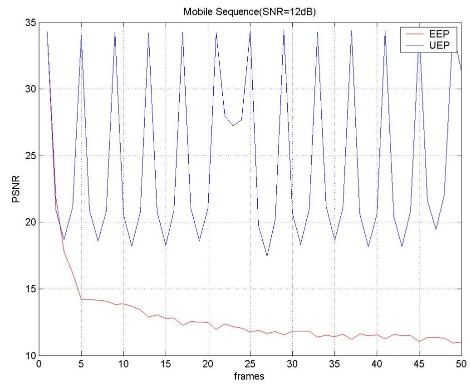


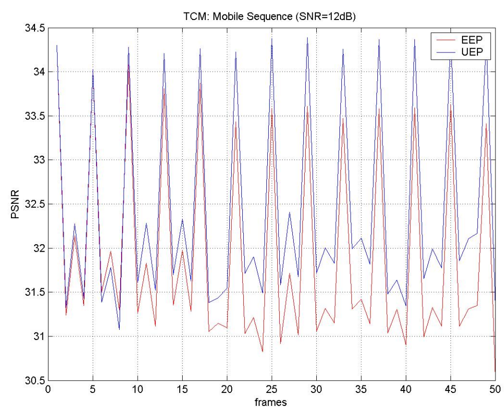
Figura 7.12. Modalità di trasmissione 1.

La figura 7.12 mostra le prestazioni del PSNR per i casi di UEP e EEP, che utilizza lo stesso tasso di codice di canale (1/2) per tutti i base ed "enhancement layers". In questo caso, dove cè un elevato tasso di errore sul bit, il "base layer" nel metodo EEP è soggetto a perdite di pacchetti che non possono essere recuperati dal codice convoluzionale, quindi il degrado del PSNR del "enhancement layer" aumenta in quanto i MVs del "base layer" non possono essere utilizzati per l'error-concealment. Tuttavia, il metodo UEP può recuperare questi pacchetti persi in quanto il livello di protezione di errore che è associato con il "base layer" è significativamente superiore a quello corrispondente al metodo EEP. Questo si traduce in una migliore performance del metodo UEP rispetto al metodo EEP.

Nelle modalità di trasmissione 2 e 3 (figura 7.13), un canale limitato in banda, che utilizza una modulazione del segnale, è stato impiegato nel tentativo di aumentare il throughput di informazione senza aumentare la larghezza di banda.



(a) Modalità di trasmissione 2.



(b) Modalità di trasmissione 3.

Figura 7.13. Modalità di trasmissione 2 e 3

Nella modalità di trasmissione 2, modulazione gerarchica non codificata è stata impiegata. Di conseguenza, più alto SNR è necessario per uno BER accettabile. In questo modo un SNR di 12dB è stato utilizzato.

Anche se il SNR è superiore a quello implementato nella modalità di trasmissione precedente, un tasso di errore sul bit più elevato è stato ottenuto, pertanto, il valore medio di PSNR è stato molto basso (tabella 5.6). Tuttavia, in queste condizioni e in base alle analisi precedenti, UEP ha superato chiaramente il metodo EEP per questa sequenza (Mobile).

Nella modalità di trasmissione 3, ci siamo concentrati sul modo di trasmissione di canali con una banda limitata, che utilizzano TCM per ottenere un guadagno di codifica, presso la stessa efficienza della larghezza di banda (r) per strato utilizzato nella modalità di trasmissione 2, e allo stesso valore di SNR (12 dB). Dove, guadagno di codifica è definito come la riduzione del SNR per una certa probabilità di errore. Sulla base di questo vincolo (r= 2 b/s/Hz), della probabilità di errore e del PSNR medio ottenuto per un SNR specificato, si possono confrontare le modalità di trasmissione 2 e 3.

Il desiderio di migliorare i risultati poveri di codifica, visto sui canali limitati in banda nel modo di trasmissione 2, è stata la motivazione che ha portato a TCM (trellis-coded modulation). A causa del vantaggio del FEC, la modalità di trasmissione 3 ha una migliore performance di errore rispetto al modo di trasmissione 2, per lo stesso tasso di dati e SNR, che si traduce in un valore maggiore di PSNR medio o in una maggiore efficienza spettrale. Questo guadagno di codifica è stato realizzato senza sacrificare la velocità di trasmissione di dati o ampliare la larghezza di banda, ma a scapito della crescente complessità del decoder.

7.6 Conclusioni

In questo lavoro, abbiamo descritto l'uso potenziale di Scalable Video Coding (SVC), estensione di H.264/AVC, in un sistema di trasmissione per il video streaming su Internet. Lo scopo principale era quello di impiegare TCM (Trellis-coded modulation) per fornire una protezione di errore diseguale (UEP) per la consegna di video scalabile in ambienti eterogenei, e dimostrare che la progettazione dei sistemi di comunicazione che trattano la codifica e modulazione del sistema indipendente di ogni altro non ottengono le migliori prestazioni. Nei risultati sperimentali, sulla base di scalabilità ibrida temporale-spaziale, l'efficienza del sistema di modulazione codificata è stata misurata in termini di PSNR per fotogramma di video.

Dai nostri risultati, è evidente che la modulazione gerarchica è preferibile nei casi in cui è necessario considerare la trasmissione con gran quantità di energia. D'altra parte, quando lo schema di codifica convoluzionale è stato utilizzato per fornire UEP, ha contribuito a diminuire il tasso di errore sui bits in modo significativo. Questo a sua volta, provoca la trasmissione di segnali di una determinata qualità con una minore potenza di trasmissione. In altre parole, conduce ad una efficienza di potenza superiore, ma bit rate inferiore. La constatazione generale di questa tesi è che rispetto alla modulazione non codificata, la stessa quantità di informazione può essere trasmessa all'interno della stessa larghezza di banda con guadagno di codifica di 3-6 dB, quando TCM è utilizzato come metodo di prioritizzazione con encoder SVC.

Molti sistemi che combinano scalabile bit rate dei media con FEC o modulazione gerarchica sono stati studiati prima in grande dettaglio, ma sono stati limitati dalla mancanza di un efficiente video codec. Con SVC, l'impatto di queste tecnologie crescerà nei prossimi anni. Scalabile bit rate sarà in grado di sfruttare appieno le potenzialità dei metodi di consegna emergenti, le capacità migliorata del ricevitore e nuove offerte di servizi, per soddisfare le crescenti aspettative degli utenti in un modo efficiente in quanto a risorsa e costo. SVC è una soluzione ideale per questi tipi di servizi, in parte perche esso permette l'interoperabilità con i sistemi esistenti.

Appendice A

RCPC codes

Many researchers have published list of convolutional encoders. In this appendix, for size limitations, the RCPC codes used in this work are given in table A.1. However, references in which encoders of various rates, lenghts and types can be found in [34, 35].

Pos.	R	d_f	(a_d,$d = d_f$,$d_f + 1$,...$d_f + 9$) [c_d,$d = d_f$,$d_f + 1$,...$d_f + 9$]
-	8/32	20	(16, 0, 48, 0, 56, 0, 120, 0, 200, 0) [24, 0, 136, 0, 256, 0, 0, 528, 0, 1040, 0]
(1,6)	8/24	14	(5,18,28,47,64,74,136,237,396,753) [8,48,91,200,338,409,789,1397,2559,5346]
(2,5)	8/16	8	(3,12,43,92,180,431,1015,2530,6112,14357) [4, 40, 196, 493, 1141, 3044, 8044, 22226, 59149, 152186]
(2,4)	8/12	6	(26, 82, 287, 1097, 4079, 15286, 57851, 218486, 821520, 3089271) [128, 512, 2272, 10717, 46140, 196244, 834321, 3493131, 14465684, 59272939]
(1,7)	8/10	4	(9, 55, 363, 2505, 16402, 106468, 701263, 4601195, 30186318, 198108997) [48, 460, 3884, 34919, 279497, 2132709, 16200442, 120276696, 880843487, 6383725051]

Tabella A.1. RCPC codes, $K = 7$, mother code (117, 127, 155, 171)

The puncturing is given relatively the mother code rate and the additional puncturing position (column “Pos.”) is given by the row and column number. Also given in the tables are the free distance and 10 terms of the spectrum (a_d and c_d, $d = d_f$,$d_f + 1$,...,$D_f + 9$), where a_d is the number of error paths of distance d and can be use to upper bound the firs-error event probability [3]. In calculating union bounds on the probability of bit error infinetely many terms need to be included to achieve a true value; but, by including many terms (At least 10) one will obtain a very good estimate of the true bit error rate. The generator polynomial of the mother codes is given in octal form and K denotes the constraint length.

Appendice B

TCM

Table B.1 ,which values are taken from [27, 29], provides a summary of coding gains achievable with rate-2/3 trellis-coded 8-PSK modulation and rate-3/4 trellis coded 16-QAM modulation that were used in our simulations. Also, the values of the parameters used in the calculation of the event error probability in equation 4.9 are given.

No. of states		**Parity-check coeficients**						
2^ν	m	h^2	h^1	h^0	d^2_{free}/Δ_1	$\frac{G_{8PSK}}{4PSK}[dB]$	$\frac{G_{16QAM}}{8PSK}[dB]$	N_{free}
4	1	-	2	5	4	3.01	-	1
512	2	510	346	1001	8	-	5.81	4

$\Delta_1(8 - PSK) = \sqrt{2}$
$\Delta_1(16 - QAM) = 2/\sqrt{5}$

Tabella B.1. Codes for 8-PSK and 16-QAM modulation

Parity-check coefficients are specified in octal form

The results in this table clearly illustrates the significant coding gains that are achievable with relatively simple trellis codes, with simple rates of $m\ /(m\ +1) =$ 1/2 or 2/3 .

Bibliografia

[1] Cisco Systems. *Approaching the Zettabyte Era*, USA, White papers, June 2008.

[2] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson. *RTP: A transport protocol for real-time applications*, RFC 1889, Jan. 1996.

[3] G. Proakis. *Digital Communications*, 4th edition, New York, McGraw Hill, 2008.

[4] U. Horn, K. W. Stuhlmuller and B. Girod. *Robust internet video transmission based on scalable coding and unequal error protection*, Image Commun., Special Issue on real-time video over the internet, Vol. 15, No.(1-2), pp. 77-94, Sep. 1999.

[5] H. Radha, M. van der Schaar and Y. Chen. *The MPEG-4 fine-grained scalable video coding method for multimedia streaming over IP*, IEEE Trans. Multimedia, Vol. 3, No. 1, pp. 53-68, March 2001.

[6] ITU-T and ISO/IEC JTC1. *Generic Coding of Moving Pictures and Associated Audio Information Part 2: Video*, ITU-T Recommendation H.262, ISO/IEC 13818-2 (MPEG-2), November 1994.

[7] ITU-T. *Video Coding for Low Bit Rate communication*, ITU-T Rec. H.263, Version 1: Nov. 1995, Version 2: Jan. 1998, Version 3: Nov. 2000.

[8] ISO/IEC JTC 1. *Coding of audio-visual objects Part 2: Visual*, ISO/IEC 14492-2 (MPEG-4 Visual), Version 1: Apr. 1999, Version 2:Feb. 2000, Version 3: May 2004.

[9] M. Gallant, F. Kossentini. *Rate-distortion optimized layered coding with unequal error protection for robust internet video*, IEEE Trans. Circuits Syst. Video Tech. 11, pp. 357372, 2001.

[10] L.-F. Wei. *Coded modulation with unequal error protection*, IEEE Trans. Commun. 41, pp 14391449, 1993.

[11] A. Ortega, K. Ramchandran. *Rate-distortion methods for image and video compression*, IEEE Signal Process. Magazine 15, pp. 2350, 1998.

[12] B. Hochwald, K. Zeger. *Tradeoff between source and channel coding*, IEEE Trans. Inform. Theory 43, pp. 14121424, 1997.

[13] G. Cheung and A. Zakhor. *Joint source/channel coding of scalable video over noisy channel*, IEEE Int. Conf. Image Processing (ICIP),vol.3, pp. 767-770, 1996.

[14] H. Zheng and K. J. Ray Liu, *Robust and video transmission over spectrally shaped channels using multicarrier modulation*, IEEE Trans. Multimedia, vol. 1, no. 1, pp. 88-103, March 1999.

[15] D. Grossman. *New terminology and classification for DiffServ*, RFC3260, April 2002.

[16] J. Wroclawski.*The use of RSVP with IETF Integrated Services*, RFC2210, September 1997.

[17] D. Salomon.*Data Compression, the complete reference*, 4th edition, New York ,Springer-Verlag , 1997.

[18] H. Schwarz, D. Marpe, and T. Wiegand. *Overview of the scalable video coding extension of the H.264/AVC standard*, IEEE Transactions on Circuits and Systems for Video Technology, vol. 17, no. 9, pp. 11031120, 2007.

[19] Fraunhofer. *Scalable extension of H.264/AVC,* http://ip.hhi.de/imagecom_G1/savce/index.htm.

[20] Y. Wang and Q. F. Zhu. *Error control and concealment for video communication: A review*, Proc. of the IEEE, vol. 86, no. 5, pp. 974-977, May 1998.

[21] *Joint Scalable Video Model (JSVM 2)*, http://ip.hhi.de/imagecom_G1/savce/index.htm.

[22] M. van der Schaar, S. Krishnamachari, S. Choi and X. Xu. *Adaptive cross-layer protection strategies for robust scalable video transmission over 802.11 WLANs,* IEEE J. Selected Areas in Commun., vol. 21, no. 10, pp. 1752-1763, December 2003.

[23] Y. Wang, S. Wenger, and M. Hannuksela. *System and transport interface of H264/AVC*, Nokia Research Center, pp. 165, Finland, 2006.

[24] J. Hagenauer. *The performance of rate-compatible punctured convolutional codes for digital mobile radio,*IEEE Trans. Comm., vol. 38, no. 7, pp. 966-980, July 1990.

[25] J. Hagenauer. *Rate-compatible punctured convolutional codes (RCPC codes) and their applications*, vol. 36, no. 4, pp. 389- 400, April 1988.

[26] P. Frenger, P. Orten and T. Ottoson. *Rate-compatible convolutional codes for multirate DS-CDMA systems* IEEE Transactions on Communications, vol. 47, no. 6, pp. 828836, June 1999.

[27] G. Ungerboeck. *Channel Coding with Multilevel/Phase Signals*, IEEE Trans. Inform. Theory, vol. 28, no. 1, pp. 5567, 1982.

[28] G. Ungerboeck. *Trellis-coded modulation with redundant signal sets* Part I: Introduction. IEEE Comm. Mag., vol. 25, no. 2, pp. 5-11 ,1987.

[29] G. Ungerboeck. *Trellis-coded modulation with redundant signal sets* Part II: State of the art. IEEE Comm. Mag., vol. 25, no. 2, pp. 12-21, 1987.

[30] J. L. Massey. *Coding and modulation in digital communications*, In Proc. IEEE Intl. Zurich Sem. Dig. Comm., pp. 248-250, Switzerland, 1974.

[31] Y.-K. Wang, S. Wenger and M.M. Hannuksela. *Common conditions for SVC error resilience testing*, JVT-P206, Jul. 2005.

[32] W. Kung, C. Kim, J. Kuo. *Packet video transmission over wireless channels with adaptive channel rate allocation* J. Vis. Commun. Image R. 16, pp. 475498, 2005

[33] C. Ying, X. Kai, Z. Feng, P. Purvin and B. Jill. *Frame Loss Error Concealment for SVC,* Journal of Zhejiang University, vol. 7, no. 5, pp. 677-683, Apr. 2006.

[34] Y. Yasuda, K. Kashiki, Y. Hirata. *High Rate Punctured Convolutional Codes for Soft Decision Viterbi Decoding,* IEEE Trans. on Com., vol. 32, pp. 315-319, March 1984

[35] P. K. Frenger, P. Orten, T.Ottosson, andA. B. Svensson. *Rate-compatible convolutional codes for multirate DS-CDMA systems,* IEEE Trans. Commun., vol. 47, no. 6, pp. 828836, June 1999.