Antonio Ortega and Masoud Khansari. “Video Transmission over Wireless Links: State of the Art and Future Challenges”
Handbook of Emerging Communications Technologies: The Next Decade.
Ed. Saba Zamir
Boca Raton: CRC Press LLC, 2000

3 Video Transmission over Wireless Links: State of the Art and Future Challenges

Antonio Ortega and Masoud Khansari

CONTENTS

In this chapter we give an overview of potential applications of wireless video and discuss the challenges that will have to be overcome for these systems to become reality. We focus on video coding issues and outline how the requirements of (i) low bandwidth and low power consumption and (ii) robustness to variable channel conditions are being addressed in state of the art video compression research. We provide an overview of these areas, emphasizing in particular the potential benefits of designing video coding algorithms that are aware of the channel transmission conditions and can adjust their performance in real time to increase the overall robustness of the system.

3.1 INTRODUCTION

The increased popularity of mobile phones among consumers has made the wireless communications industry one of the fastest growing industries ever. At the beginning, the industry concentrated on providing users with telephone access from their cars. The technology, however, was at its early stages, as was evident by the need to use bulky handset radios, which were mounted onto the cars. This, in comparison with the current small handsets, shows the significant progress that has been made in both the system and the radio transmission technologies.

As wireless access becomes more commonplace, it will be necessary to support services other than voice and data. In particular, image and video communications over wireless links are likely to grow in importance over the coming years. This chapter addresses wireless video, which we believe to be one of the most promising emerging technologies for future mobile communications. Our goal is to report on the state of the art in this technology, especially from the perspective of video compression, and to highlight some of the remaining challenges that will have to be addressed to make wireless video a reality.

3.1.1 POTENTIAL APPLICATIONS OF WIRELESS VIDEO

While there are numerous applications in which wireless delivery of video is likely to play a role, we distinguish between two major scenarios, namely, those where interactive and noninteractive communication takes place. Each scenario has different implications when it comes to system-level trade-offs.

In an interactive environment the goal will be to support two-way video or to be able to respond in a very fast manner to user commands (for example, fast forward in a one-way playback application). In this case, there will be strict constraints on delay, which will have to be kept low so that interactivity can be preserved. For example, a good rule of thumb is that system delays of the order of 100ms are completely transparent to the users. Although longer delays may be tolerable, they can have a negative impact on the perceived quality of the service. In addition to delay constraints, interactive applications involving two-way video also require that the portable device provide video encoding in addition to decoding, which will place severe demands on power consumption (not only for the purpose of signal processing, but also for transmission, display, capture, etc.)

In a noninteractive application we can assume one-way video being delivered without as strict a constraint on the delay. This could be the case if we considered the local distribution of video signals (for example displaying a particular video feed to several receivers in a household or office). Here, we are likely to have access to higher bandwidth and to deal with more predictable channel characteristics. This chapter will concentrate more on the low rate, high channel-variability case, as it presents the most challenges for efficient video transmission.

To further illustrate the significant differences between these two types of situations, let us consider two concrete examples of wireless video transmission.

Local video distribution

The success of the cordless phone in the consumer market has demonstrated the usefulness of providing mobility even within a limited area such as a home. It is thus likely that this type of short-range mobility will be supported for other types of applications, in addition to traditional telephony. For example, given that personal computers (PCs) have found their way into many households within the United States and Europe, it is likely that they will be used more frequently to control household appliances. PCs are thus envisioned to expand their role as central intelligence of the household, to control appliances within the home, and to provide connectivity to the outside world. Based on this vision, one has to provide radio connectivity between the PC and the many networked appliances. Providing such connectivity, one such application is to send video signals representing the screen from the PC to a light portable monitor. The user can then access and control the PC remotely (e.g., to check e-mail). Another application is to use the DVD player of the PC (or the video-on-demand feed received over the network) and use a TV monitor to display the video signal. These applications will likely require a high-bandwidth channel from the PC to the receiver and a lower bandwidth one from mobile to PC. The relatively limited mobility and range make it possible to have a more controlled transmission environment, thus higher bandwidth and video quality may be possible.

Car videophone

As an example of an interactive environment consider the provision of video conferencing services to mobile receivers as a direct extension of the existing wireless telephony. Here one can assume that low bandwidth and the effects of mobility (time-varying channels) are going to be the most significant factors. Unless the system has to be fully portable (so that it can be used outside of the car), power will not be a major issue. In recent years, there has been a significant increase in the amount of data that is delivered to moving vehicles; for example, in addition to telephony, some cars are equipped to be linked to geopositioning systems (GPS). These kinds of services are likely to grow, and one can foresee maps, traffic information, video news updates, and even two-way video being proposed as extra features for high-end vehicles in coming years. The most significant characteristics of these applications are the low bandwidth available, the potentially extreme variations in channel conditions, and the delay sensitivity of the system, in the case where interactivity is required.

3.1.2 Present and Future Wireless Communications Systems

While our main interest here is the support of video services, we start by briefly discussing the evolution of wireless communication systems in recent years. The major trend to be observed is the move towards all digital service, with data services being supported in addition to traditional voice service.

With the first generation of wireless communications systems, service providers have developed a new infrastructure parallel to the traditional phone network (Public System Telephone Network or PSTN).[1] This infrastructure uses the cellular concept to provide capacity (via frequency reuse) and the necessary range to cover a wide geographical area. In first generation systems, radio transmission is analog and the necessary signaling protocols have been developed to provide seamless connectivity despite mobility.[2] Parallel to this activity, cordless phone standards such as CT2 and DECT were developed to provide wireless connectivity to the telephone network within home and office environments.[3] In comparison with the cellular system, these systems are of lower complexity and use smaller transmission power because of the limited required range.

The more advanced second generation wireless systems use digital transmission to improve the overall capacity of the system, and also to provide security and other added functionalities. Different incompatible systems have been developed and implemented in Japan, Europe, and North America for the radio transmission.[4–8] Sophisticated speech compression methods such as Code-Excited Linear Prediction (CELP) are used to compress digital speech while almost achieving the toll-quality speech of PSTN.[9] Also, new higher frequency bands have been allocated and auctioned by the Federal Communications Commission (FCC) for the rollout of the new Personal Communication System (PCS). The PCS uses the existing telephony signaling infrastructure to provide nationwide coverage and uses smaller transmission power, which translates into smaller handsets and longer battery life.

In the meantime, the emergence of the Internet with data applications such as e-mail or web-browsing and the popularity of lap-tops have resulted in increased interest in providing wireless connectivity for data networks. Different wireless LAN protocols have been proposed, and the IEEE 802.11 working group has defined a new MAC protocol with three different physical layer possibilities.[10] At the same time, in Europe a new protocol known as High Performance Radio Local Area Network (HIPERLAN) has been proposed.[11] These proposals tend to use the public Industrial Scientific Medical (IMS) bands and can provide an aggregated bandwidth of up to 2 Mbps. They, however, support only a limited mobility and coverage.

The current cellular system is primarily targeted at the transmission of speech, and even though it can support extensive user mobility and coverage, it provides only a limited bandwidth (around 10 Kbps).[12] This is clearly inadequate for many multimedia applications. Therefore, a new initiative, known as third-generation cellular systems (IMT-2000), has been started; it emphasizes providing multimedia services and applications. The proposed systems are based mostly on Code Division Multiple Access (CDMA) and are able to support applications with a variety of rate requirements. Two main candidates are CDMA-2000 (proposed by Qualcomm and supported by North American standard groups) and Wideband CDMA (WCDMA) (proposed jointly by Japan and Europe).[13–18] The main improvements over second-generation systems are more capacity, greater coverage, and high degree of service flexibility. Third-generation systems also provide a unified approach to both macro- and microcellular systems by introducing a hierarchical cell organization. Third-generation systems provide enough bandwidth and flexibility (e.g., possibility of adaptive rate transmission) to bring multimedia information transmission (specifically video) closer to reality.[18]

3.1.3 Video over Wireless: Algorithm Design Challenges

While significant progress is being made in developing a digital transmission infrastructure, there is by no means a guarantee that advanced real-time services, such as video transmission, will be widely deployed in the near future. This is in part because of the demanding requirements placed on video transmission to achieve efficiency over the challenging wireless channels. These requirements can be derived by considering the characteristics of typical transmission environments, namely low bandwidth, low power, and time-varying behavior.[19] In this chapter we address these issues by describing first the video coding algorithms then discussing how channel characteristics need to be taken into account in the video transmission design.

First, the low bandwidth of the transmission link calls for low or very low rate compression algorithms. We outline some of the progress made in this area in recent years, in particular through algorithms like MPEG-4[20] or H.263.[21] In addition, it is obvious that the devices to be used for video display and capture have to be particularly efficient because they have to be portable. This places constraints on the weight and power consumption of these devices thereby calling for compression algorithms that are optimized for power consumption, memory, etc.

Given the variable nature of the channels we consider, it is necessary to consider video compression algorithms that are scalable, that can compress the same video input at various rates (consequently with different decoded output qualities). In other words, if the channel can provide different qualities of service (QoS) the video encoder should be able to likewise provide different rates and video qualities. We present some of the approaches that have been proposed for scalable video and indicate how these can be incorporated into practical video transmission environment. An alternative approach to deal with channel variability is to let the source coder adjust its rate to match the expected channel behavior (i.e., transmit fewer bits in instances of poor channel quality). We also discuss these rate control approaches.

Finally, even if efficient channel coding is used, the variations in channel conditions due to roaming will result, in general, in losses of information. Thus, the video applications will have to provide sufficient built-in robustness to ensure that the quality of the decoded video is not overly affected by the channel unreliability.

3.1.4 Outline of the Chapter

This chapter is organized as follows. Section 3.2 provides a brief introduction to the typical architecture of the wireless access network and describes the channel behavior to be expected under such transmission conditions. Section 3.3 introduces the basic components in currently used video compression environments and discusses the delay constraints imposed by the transmission environment and how these are translated into conditions on the video encoding process. It also briefly describes how some of the requirements of efficient wireless video (very low rate and power consumption) are met in state of the art systems. Section 3.4 motivates that error correction techniques are not sufficient to provide the required robustness and that, in fact, video coding algorithms have to be especially designed to support transmission over a wireless link. For each of the techniques that can be used to increase the video coding robustness (e.g., redundancy, packetization, etc.) we describe how specific video techniques can be found to improve the performance over purely channel-coding approaches.

3.2 WIRELESS ACCESS NETWORK ARCHITECTURE AND CHANNEL MODELS

We start by providing a quick overview of the channel characteristics that are most important for video transmission. First we describe a typical wireless access network architecture, then the methods used to achieve robustness in mobile environments. Finally, we describe some of the models that are used to characterize overall transmission performance in terms of packet losses.

3.2.1 ARCHITECTURE

In a typical cellular system a mobile station (MS) can communicate only with the nearest base station (BS). Each BS will be connected to a network that allows it to communicate with other BS,[1–2] so that communication between MSs and between a MS and the fixed telephone network is possible. Obviously, power is much more constrained at the MS than at the BS, and a BS has much more processing power than a MS. Also, since every connection within a given cell is established through the corresponding BS, each BS has knowledge of the status of all the connections within its cell. As a result, there is a significant amount of network intelligence at the BS which does not exist at the MS.

Therefore the two links or connections in a wireless access network (the one from a base station to a mobile station, the downlink, and the one in the reverse direction, the uplink) have very different characteristics. For example, for the downlink channel, the transmitter has an almost unlimited amount of processing power, whereas the conservation of power at the receiver is of utmost importance. Consequently, for this link the transmitter can be significantly more complex than the receiver. The situation is reversed for the uplink channel; the transmitter would tend to be simpler than the receiver.

Note that the same will be true when designing a video compression algorithm targeted for transmission over a wireless link. For example, consider the case when a video signal is transmitted to many mobile stations simultaneously. In this scenario, one can tradeoff a more complex video compression encoder for a simple decoder (see Meng et al.[22] for examples of this approach). Alternatively, video transmission from a MS to another user may be power-limited to such an extent that a low complexity (and low compression performance) algorithm may be preferred, even if bandwidth efficiency is sacrificed.

It is also worth noting that there is another popular architecture, the so-called *Ad hoc* network, in which mobile stations can communicate with each other without the need of a base station. In this network all the communication links are symmetric and there is no difference between the transmitter and the receiver. Transmitting video signals over this network favors the use of symmetric video coding algorithms in which both the encoder and the decoder are of about the same complexity. *Ad hoc* networks are more likely to be used in a local configuration to provide communication between a number of users in a reduced geographical area.

3.2.2 Channel Characteristics and Error Correction

Let us consider now the characteristics of typical wireless links. We first discuss the various impairments that affect the physical channel behavior, then we discuss how these can be addressed using various channel-coding techniques.

3.2.2.1 Physical Channel Characteristics

The first type of channel impairment is due to the loss of signal as the distance between the transmitter and the receiver increases or as shadowing occurs. Clearly this loss of signal will depend heavily on the surrounding geographical environment, with significant differences in behavior between, for instance, urban and rural environments. These channels are also subject to the effect of various interferences, which produce an ambient noise usually modeled as an additive white Gaussian noise.

A major contributor to the degradation of the received signal is what is known as multipath fading. This occurs when different duplicates of the same transmitted signal reach the receiver, and each version of the signal has a different phase and signal level. This situation is common in any transmission environment where the signal is reflected off buildings and other objects. Reflections can result in changes in phase and in attenuation of the signal, such that signals arriving through different paths may combine to generate a destructive addition at the receiver. If this is the case, the result can be a considerable drop in the signal-to-noise ratio (SNR). As the SNR fluctuates, the bit error rate will vary as well, and if the fluctuations are sufficiently severe the connection itself can be dropped. Typically the magnitude of the received signal under fading conditions is modeled using the Rayleigh distribution, and its phase is assumed to follow a uniform distribution. Other models, based on two or four paths, are also used in practice. Since our concern here is not the behavior of the physical link but the performance of the link after channel coding, suffice it to say that fading conditions are characterized by the fact that the distributions of low SNR periods is "bursty" in nature. Thus, rather than observing randomly distributed errors, we may be seeing that during fading periods hardly any information gets across correctly. For fading channels, measuring the average performance (i.e., the error rate averaged over the duration of a long transmission period) may not be as meaningful as characterizing the worst case performance; for example, if severe fading occurs does the connection get lost? Given that fading is determined by location, speed of the user, etc., there is a great degree of variability in channel conditions.

Typically, the fading experienced by each frequency component in the transmitted signal is statistically independent. However, in situations where the mobile station is not stationary this is no longer true, due to Doppler frequency shift. Because of the Doppler shift, the frequency of the received signal is shifted from that of the transmitted signal based on a random cosine distribution. The effect of Doppler shift becomes more pronounced when higher frequency bands (e.g., tens of Ghz) are used or the mobile stations are moving at a faster speed.

Radio transceivers use a combination of channel coding, different methods of diversity, channel equalization, and power control to combat the channel impairments. However, wireless channels are not only a hostile environment because of multipath and Doppler effects, but in many cases they are also a shared environment where many users compete for the same available bandwidth. While users can cooperate to avoid having a negative impact on each other's transmission quality, there is no guarantee that conflicts can be avoided when, for example, two users try to use the same transmission resources. Because one user's signal can have destructive effects on the other users, the challenge is therefore not to improve the performance of a specific user or connection but to simultaneously improve the overall performance of the system for all the users and the links.

3.2.2.2 Channel Coding for Wireless Channels

There are two main types of error control techniques that can be used in wireless environments, specifically open-loop and closed-loop techniques. We describe first open-loop systems, where the channel encoder provides redundancy which can be used by the decoder to correct potential channel errors.

The most popular approach to achieve robustness in wireless channels is to introduce diversity. Diversity is achieved when multiple coded versions of the same information arrive at the receiver, with each replica degraded by different *independent* noise. Spatial, frequency, and time diversity are the main modalities, and in each the underlying assumption is that while multipath fading may degrade the SNR, its effect is "local."

For example, in a broadband signal only a few frequencies will be attenuated, given that multipath fading is typically frequency selective. Thus, by spreading the information over a broad range of frequencies it is possible to achieve robustness. This frequency diversity approach is the one taken in CDMA spread spectrum techniques and explains their wide popularity. Alternative methods using frequency multiplexing are also used in practice.

The fading period (as the user moves) will not last a long period of time. Then if one can time interleave the data produced by the channel coder for transmission, bits transmitted at different times are likely to be subject to different noise. Thus, while the error patterns affecting the transmitted stream will be bursty, after time de-interleaving, the error patterns observed by the channel decoder will look more "random." This time diversity approach is also popular. In fact, time interleaving combined with channel coding is useful for nonfading channels, as demonstrated by the recent success of "turbo" codes.[23]

Finally, while the user remains stationary, fading is also spatially localized; that is, small changes in the position of the antenna can result in large increases and decreases of SNR. Thus spatial diversity approaches operate by providing multiple antennas at either transmitter or receiver, either to ensure that at least one of the antennas receives a strong signal, or to combine the information received over all the antennas to achieve better overall SNR performance.

With any of these diversity scenarios, the aggregated received noise becomes an Additive White Gaussian Noise (AWGN), and the performance of the system asymptotically approaches that of the AWGN channel.

Closed-loop approaches are such that the channel coder is allowed to adjust to the channel conditions. These approaches are more complicated to implement because the transmitter must estimate the state of the channel in order to adjust its transmission parameters. Channel state estimation is achieved with the help of the receiver, which can send information back to the transmitter about observed channel conditions. For example, if information has been corrupted due to excessive channel errors, the receiver can request retransmission. Conversely, if the receiver detects that few channel errors occur (so the channel conditions are good), it can request the transmitter to lower its transmit power or to reduce the amount of redundancy in the bitstream. Examples of closed-loop power control can be found in the design of the uplink of the CDMA-based IS-95 systems.[6] Closed-loop recovery methods based on Automatic Repeat reQuest (ARQ) have also been proposed[24] where the receiver can request retransmission of the erroneously received data. This method is especially efficient for channels with burst errors because it requires only the redundancy needed to provide error correction during fade periods. In contrast, open-loop approaches operate with basically the same level of redundancy under all channel conditions. This means that during periods of good channel behavior (e.g., when the user is close to a BS) the amount of channel protection is excessive.

Closed-loop approaches are better suited for data applications with no stringent delay requirement. While it may appear a priori questionable that such closed-loop methods be applicable to video (given the delay-sensitive nature of the video data), we will see that in fact closed-loop methods can also prove useful for video.

3.2.3 Channel Modeling

While channel coding aims at supporting error-free transmission, this becomes a nearly impossible goal as users roam about. Thus, given that channel errors are unavoidable, we provide now an overview of models that characterize the effective channel rate variations that the end-user application observes. These models aim at characterizing the effective channel performance after any applicable channel decoding has been used. Thus, in typical video transmission environments video data will be packetized and error detection techniques will be used to determine if a particular packet has been corrupted, even after error correction has been applied. If the video encoder is to adjust its coding parameters when channel variations occur, the parameter of interest will then be the probability that packets of video data will be lost. Therefore the models we describe provide statistical characterization of the probability of packets' being corrupted.

As indicated above, fading will tend to produce localized changes in SNR so that overall packet loss will also tend to be bursty. Consequently, if a packet is lost, subsequent packets are likely to be lost too. Various models have been proposed that attempt to capture the burstiness of errors in wireless channels. Of particular interest are Markov models[24–29] that are characterized by a series of states, each having associated different channel behavior. Transitions between states occur with given probabilities and are used to capture the variability of the channel. Figure 3.1 shows an example of a two-state Markov model,[25,28] where we assume that the channel can be in only one of two states (*good* and *bad* channel behavior). The modeling process consists of finding estimates for the transition probabilities. Models such as these can be used in conjunction with rate control techniques to let the source coder adjust to channel variations, as will be discussed further in Section 3.4.5.

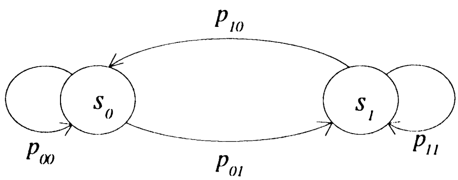

Figure 3.1 Two state Markov channel model

3.3 VIDEO COMPRESSION FOR WIRELESS COMMUNICATIONS

While a complete introduction to state of the art video coding techniques goes far beyond the scope of this chapter, we present here a brief description of the main principles underpinning video compression design. We then consider the specific changes required to adapt these algorithms to the demanding wireless environment, and in particular how low rate, low power, and low delay can be supported. In Section 3.4 we revisit some of these design issues but will consider them only from the perspective of adding robustness to the transmission.

3.3.1 Typical Video Compression Algorithms

Most practical video compression algorithms achieve high compression ratios with good quality decoded images by exploiting the various types of redundancy (spatial, temporal, and statistical) existing in video sequences. The main trade-off in video coding is that between the fidelity of the approximation to the original images (i.e., the distortion introduced by coding) and the number of bits required to represent the images (i.e., the rate). Compression performance can then be evaluated by considering the rate required to achieve a specific quality level. What is important is how redundancy is exploited to achieve good compression performance.

First, in most natural images neighboring image pixels exhibit significant similarity. For example, in typical scenes the images contain many flat regions (i.e., where low frequencies predominate) or regions with regular patterns or textures (i.e., where a specific range of frequencies contains most of the information). This spatial redundancy can be exploited so that only a few bits are needed to represent these regions. Most practical systems employ a transform, such as the discrete cosine transform (DCT), which provides estimates of the frequency contents of each image region (e.g., each block of 8×8 pixels, in the DCT case). Given that most of the energy is concentrated in a few coefficients, only a few bits will be needed to provide a good approximation to the original image block.

Second, consecutive frames in a video sequence will also exhibit similarity, especially in video scenes with low motion. This temporal redundancy is exploited through motion estimation, a technique that breaks video frames into nonoverlapping blocks and then searches the previous frame for blocks that provide the most similarity to each block in the current frame. For example, this technique predicts a block belonging to an object in the current frame from a block in the same object in the previous frame.

Finally, when data has been quantized we are left with a number of symbols to be transmitted, and these symbols are drawn from a discrete set, i.e., they can take a finite number values after quantization. Given that these symbols, e.g., the quantized DCT coefficients, are not equally likely, further compression gains can be achieved by taking advantage of this statistical redundancy. Gains are achieved through *entropy coding*: the most frequently observed symbols are represented by shorter codewords, while the least frequently used are represented by longer codes. Variable length coding techniques are very useful in terms of compression but, as will be seen later, are risky when used in a lossy environment. A single bit error can trigger a loss of synchronization in the bitstream and lead to loss of most of the information in the sequence.

Our purpose here is not the basics of image and video coding. Detailed discussions of these topics can be found in numerous textbooks.[30,31] Instead, our emphasis is that compression techniques that exploit temporal or spatial redundancy will result in different rates for different frames at a given quality level. In fact, the degree of redundancy, therefore the resulting rate for a given distortion, can fluctuate widely from scene to scene. For example, scenes with high motion content will require more bits than more stationary ones.

Figure 3.2 provides a diagram of the basic building blocks of an MPEG coder,[31] which include a DCT transform, block-based motion estimation, and quantization. It should be noted that the selection of quantization step size decided by the encoder determines the rate-distortion trade-off, and it allows the encoder to adjust the number of bits/frame to suit various constraints, such as those imposed by the channel. Note also that this video coder operates essentially as a closed-loop predictor, that is, the encoder and the decoder both use previously quantized frames to generate predictions for the current frame (the decoder uses this predicted value to reconstruct the frame once it receives the coded residue from the encoder). This is a particular concern when operating over a lossy channel: if, after the occurrence of an undetected channel error, the encoder and decoder start to operate with different frames in the prediction loop, the result can be a complete loss of synchronization between the two, with potentially disastrous results for the video quality at the receiver.

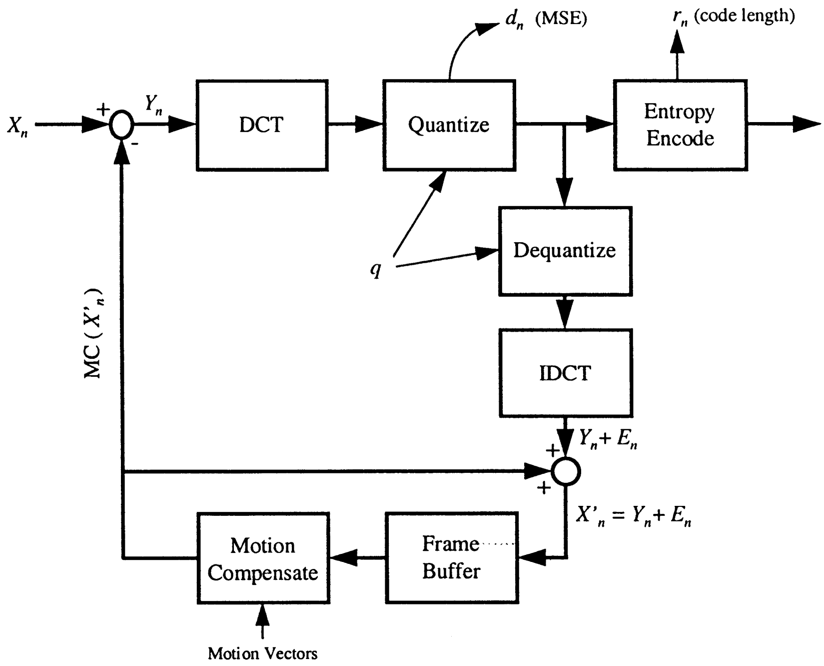

Figure 3.2 Block diagram of a typical MPEG coder. The quantization parameter can be adjusted to make the rate comply with the channel constraints. X_m represents the current video frame, and $MC(X^1_m)$ is the motion-compensated prediction frame, based on the previously transmitted frame.

3.3.2 Delay Constraints in Video Communications

Let us now consider a typical real-time transmission as illustrated in Figure 3.3. As just described, video frames require a variable bit rate, thus it is necessary to have buffers at encoder and decoder to smooth the bit rate variations. Assuming the video input and output devices capture and display frames at a constant rate, and no frames are dropped during transmission, the end-to-end delay in the system will remain constant.[32]

Let us call ΔT the end-to-end delay: a frame coded at time t has to be decoded at time $t + \Delta T$. That imposes a constraint on the rate that can be used for each frame (it has to be low enough that transmission can be guaranteed within the delay). For example, let a coding unit be coded at time t and assume that it will have to be available at the decoder at time $t + \Delta T$. If each coding unit lasts t_u seconds, then the end-to-end delay can be expressed as $\Delta N = \Delta T/t_u$ in coding units. For example, if a

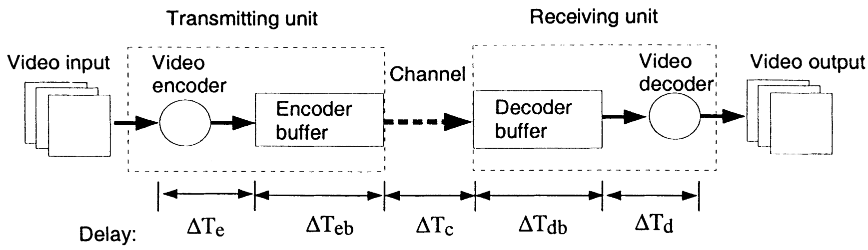

Figure 3.3 Delay components in a communication system

video encoder compresses 30 frames/sec and the system operates with an end-to-end delay of $\Delta T = 2$ sec, then the decoder will wait 2 sec to decompress and display the first frame (assuming no channel transmission delay) and at any given time there will be $\Delta N = 2/(1/30) = 60$ video frames in the system (stored in the encoder or decoder buffers or being transmitted). The video encoder will have to ensure that the rate selection for each frame is such that no frames arrive too late at the decoder.

Consider the case when transmission takes place over a constant bit rate (CBR) channel. Of the delay components of Figure 3.3 only ΔT_{eb} and ΔT_{ed}, which is the time spent in encoder and decoder buffer, respectively, will now be variable. Consider, for example, ΔT_{eb}. This delay will be at most B_{max}/C, where B_{max} is the physical buffer size at the encoder and C is the channel rate in bits/sec. It is clear that B_{max} has to be smaller than $\Delta T \cdot C$; otherwise we could store in the buffer frames which will then experience too much delay.

If we consider the transmission of a variable rate sequence we will either (1) have to use very large buffers (and correspondingly long end-to-end delays), or (2) have to adjust the source rate, thus the delivered quality, to make it possible to use a smaller buffer (shorter delay) without losing any data. The delays required for such a sequence would be exceedingly long; therefore, in practical applications it is necessary to perform rate control to adjust the coding parameters and meet the delay constraints.

As shown in Figure 3.2, it is possible to adjust the video rate (and the quality) by modifying the quantization stepsizes used for each frame. A natural way to approach these problems is to consider the Rate-Distortion trade-offs in the allocation. R-D techniques for rate control have been described in the literature.[33–36] See Ortega and Ramchandran[37] for a more detailed treatment of these techniques.

Note that even in cases where transmission is performed over a variable bit rate (VBR) channel, or where the sequence is preencoded and stored (e.g., in a Digital Versatile Disk, DVD), it is also necessary to perform rate allocation. For example, to store a full-length movie in a DVD it may be necessary to preanalyze the movie, then allocate appropriate target rates to the various parts of the movie. In this way, allocating more bits to the more challenging scenes and fewer to the easier ones will result in a globally uniform quality. R-D based approaches for rate control for VBR channels have also been studied.[38–41] Approaches more suitable for storage in disk-based video servers are considered in Miao and Ortega.[42]

In summary, video coders produce a variable number of bits/frame and are subject to strict delay constraints. Thus, any rate adjustment at the video encoder to accommodate changing channel conditions will have to take into account the delay constraints to avoid having data arrive at the decoder too late to be decoded.

The overall delay in the system ΔT is an important parameter. For noninteractive applications its sole significance is the latency it introduces (i.e., decoding does not start until ΔT seconds after data was first received at the decoder). For interactive applications ΔT is even more critical, since large values of end-to-end delay degrade the perceived interactivity. On the other hand, since channel variations are bursty, a longer delay allows increased robustness to changes (if the delay is long enough the transmitter can wait until the fading conditions have subsided to transmit data without errors). Thus, the selection of an end-to-end delay for a particular application will have to balance conflicting demands for high robustness, low latency, and high interactivity.

3.3.3 Coding for Low Rate Channels

The basic principles outlined in Section 3.3.1 still apply to video coding at low rates, i.e., the main sources of compression gains come from exploiting temporal, spatial, and statistical redundancies.

The ITU-T standards for video conferencing and video phone (H.261[43–44] and its successor H.263[21]) have traditionally been representative of the state of the art for low bit rate video, targeting bit rate ranges in the order of 64-384kbps. While these standards initially were meant for a fairly narrow range of applications (e.g., they assumed "head and shoulders" sequences as their expected source material), technology has improved enough that the last generation of low bit rate coders in the ITU-T family, the H.263 series,[21] can handle more challenging sequences with good results. Still, to achieve good quality at very low transmission rates (64kbps and below) one has to start with video source material captured at a reduced frame rate and with limited frame sizes. Typically, frame sizes of 176×144 pixels (QCIF) are typical of these applications, with frame rates as low as ten frames/sec or less.

Most of the gains achieved in going from H.261 to H.263 can be attributed to more efficient motion estimation, rather than to specific optimizations of the systems for low rates. In the mid nineties, as the process to define the ISO MPEG-4 standard[20] got underway, there was substantial interest in defining very low bit rate coding algorithms that would go beyond the transform coding paradigm and introduce so called second-generation coding techniques,[45] such as segmentation, or even more advanced object-based approaches.[46–47] However, the investigation of these techniques seems to have led to the conclusion that second-generation image-coding approaches are not sufficiently mature yet, and their complexity may still place them beyond the reach of affordable and reliable implementations with today's technology. As a result the emerging MPEG-4 standard[20] has not abandoned the transform coding techniques. However, one of its main novelties has been to provide added functionality for multimedia applications, notably the possibility of choosing as the basic coding unit video objects that have arbitrary shapes, instead of being limited to rectangular frames.

In summary, recent technology has contributed to increased quality for low bit rate video, along with, in the case of MPEG-4, some interesting additional functionality. However, these techniques address generic low bit rate applications (e.g., including video streaming over the Internet). As will be seen in Section 3.4, additional features are required to provide robustness in an error-prone environment such as a wireless link. As will be seen, many useful robustness tools have been proposed within the H.263 Annexes.[21,48]

3.3.4 Power Constraints and System Level Trade-offs

As anybody who has used a cell phone can easily attest, power efficiency and the related issue of weight are two of the most important considerations in a useful wireless communications system. There is an excellent overview of recent work on these issues.[49] Obviously there are many components that affect the power consumption of a wireless communications device, including the RF circuitry, the communication-specific processing, and of course the image/video compression and corresponding display, if video support is included. See the *IEEE Personal Communications Magazine*[49] for a detailed analysis of some of these issues.

It is also clear that, compared to the required power for display or transmission, video processing power may not always rank among the top power-consumption components in a portable video communicator. Still, video processing will be a significant component in power consumption budget. Recent research has demonstrated that the power consumption can be dramatically reduced by intelligent trade-offs in the image coding/decoding process. For example, Meng et al.[22] and references show reductions of two orders of magnitude in power consumption with respect to a JPEG[50] image decoder. The coding algorithm used is designed such that decoding is particularly simple, with the DCT replaced by simple (i.e., integer coefficient) subband filtering. In addition, coding efficiency is based on a pyramid vector quantization (PVQ) scheme which has further advantages in terms of complexity, as decoding can be implemented based on a look-up table. The basic lesson from this and related work is that it is possible to implement algorithms with low power in mind, but even larger gains can be achieved if one is willing to *design* the compression algorithms with power efficiency, and not just coding efficiency, in mind.

There is an increased awareness in the coding community that implementation issues should be at the forefront of the coding algorithm design, rather than being addressed as an afterthought. For example, in the ongoing JPEG2000[51] standardization process, which is set to define the next generation of wavelet-based image coding techniques, special care has been given to taking into account both the complexity and the memory usage of competing algorithms.[52]

Since in cellular and most other wireless access networks the communication among mobile stations is done through a base station, there are always two different links to consider: downlink (forward link) and uplink (reserve link). If the video application is interactive and the same video codec is used in both directions, then the performance of the overall system is determined by the worst-case link. Note that the decision of which link has poorer performance is dependent on the channel recovery method being used. In one-way applications or when it is possible to use different video encoders for each direction, by using asymmetric video codec, one can improve the performance. For example, for the downlink we use a more complex video encoder to simplify the operation of the decoder. The situation is reversed for the uplink where one would like to use a simple encoder, but since there is more processing power at the base station the constraint on the complexity of the decoder is more relaxed.

As in Meng et al.[22] it may be a good idea to consider very simple decoders to minimize the power consumption at the receiver, since the MS receivers are clearly much more power limited than the BS. For the same reasons, it may be easier to have higher power (thus rate) in transmitting from the BS to the MS. In general, therefore, power is the fundamental source of asymmetry in the overall system design and this may lead to interesting system-level trade-offs. Power efficiency can thus be addressed at three levels: (1) efficient implementation of existing systems, (2) algorithm design driven by power efficiency, and (3) intelligent system-level trade-offs. The latter two levels have the most potential for significant power reductions but are also the least studied and understood.

3.4 ACHIEVING ROBUSTNESS THROUGH SOURCE AND CHANNEL CODING

The goal of this section is to motivate that video coding techniques are required to support robust communication over a wireless channel and to complement channel coding approaches such as those described in Section 3.2. While in other transmission environments it is possible to effectively decouple source and channel coding, time-varying channels such as a wireless link make it more difficult to guarantee error-free transmission. If errors do occur then video coding techniques are needed in order to achieve graceful degradation, i.e., to provide lower, but acceptable, quality video as the channel conditions worsen.

3.4.1 The Separation Principle and the Need for Joint Source Channel Coding

In his seminal work on transmission of information over a noisy channel, Shannon[53] showed that the two fundamental functions of source coding (compressing a source so that it requires the minimum number of bits to represent it, given a fidelity criterion) and channel coding (adding redundancy to a bitstream to protect it against channel errors) can be done separately without loss of optimality. This is known as the *separation principle* and has been a fundamental guideline in the design of communication systems. Based on this principle, a source encoder tries to remove as much redundancy as possible from the source, irrespective of the channel and the channel encoder being used. At the same time, the channel encoder adds enough redundancy to provide reliable, i.e. error-free, transmission of the information over the noisy channel, and again this can be done irrespective of what the information being transmitted is. Shannon showed that if the entropy of the transmitted information is smaller than the channel capacity, then reliable communication is possible and can be achieved through separate source and channel coders. If source entropy exceeds channel capacity, then reliable communication is impossible. An implicit assumption in developing the separation principle is that there is no delay constraint on the operation of both the source and the channel encoder/decoders. Also, it is assumed that the channel is ergodic (hence stationary).

It is fair to say that in a wide variety of applications, in particular point-to-point transmission over channels that exhibit mostly time-invariant behavior (e.g., phone lines), these assumptions are valid and provide a perfect framework to design communication systems. However, both broadcast channels[54] and time-varying channels, such as those considered in this chapter, no longer fit into the framework assumed by the separation principle.

In a system design based on the separation principle the channel encoder takes the output of the source encoder (bits) and adds redundancy so that all bits are equally protected. In fact, not all bits produced by the encoder have equal importance. For example, when transform coding is used to compresss images, coefficients representing low frequencies have more importance in terms of perceptual quality. In general, for each type of bits produced by a source coder, it is possible to measure the effect errors would have on the reconstructed signal, and these can then be used to provide unequal levels of protection to the different types of data.[55,56] Obviously, unequal error protection is needed only if channel errors cannot be avoided; otherwise one would be better off by simply providing a high level of protection to all the information so that it is all received error-free.

The separation principle guided us in the design of a system that would be virtually error free and this would be true in general, assuming the statistical characteristics of the channel do not change. However, typical radio channels in a wireless access network, as discussed in Section 3.2, are characterized by their time-variability. Thus a standard design based on separation would have to assume a specific model for the channel: if conditions were worse than predicted, the service would be interrupted. If the channel conditions are worse than modeled only a small fraction of the time, this might be acceptable. However, to ensure that a service interruption happens only rarely, the system may have to be designed under very pessimistic assumptions. This in turn means that during periods of favorable channel conditions the system may be operating at a higher level of redundancy than would be required by the channel at that time, resulting in inefficient use of resources.

In summary, for both time-varying and broadcast channels, trade-offs tend to be more complex than for the more traditional point-to-point channel with time-invariant characteristics. This leads to the development of *joint source channel-coding* techniques, which aim at alleviating the effect of time variability by enabling quality of service that can be adjusted to match different types of channel behavior.

Although there is a more detailed discussion of joint source channel coding,[57] we outline here some of the basic principles. In a joint source source-channel coder design, the two objectives are (1) to design source coders that clearly separate source information into different classes according to their importance as far as the decoded image quality is concerned (these are called multiresolution, scalable, or layered codecs), and (2) to provide channel coding mechanisms that allow unequal error protection, so important information can be better protected than less important information. Note that we provide an overview of a broad range of techniques which we put under the umbrella of joint source channel coding. Some are truly joint in that the source and channel coding techniques are designed together. In other cases the designs are carried out independently, but channel coding is made aware of the relative importance of the information that is being carried. In fact, the latter set of techniques, which do not involve a joint design, are by far the most widely used in practice.

3.4.2 Packetization and Synchronization

A simple way of achieving some degree of protection, without necessarily introducing modifications in the source coder, is to packetize the data and provide mechanisms that limit error propagation. While these are not strictly speaking joint source channel coding techniques (since the channel coder is not modified) they do show how having knowledge of data can help reduce the effect of errors. In fact, most of the current activities on error-resilient source coding have focused on methods to prevent error propagation or to facilitate resynchronization after errors have occurred.

Packetization consists of breaking up the video data into segments and transmitting each of these segments or packets, with individual error detection being performed on each of them to determine whether the information, or payload, of the packet was corrupted. If the packet was corrupted the corresponding information can be discarded by the decoder. A priori, the packetization process could take place without any knowledge of the contents of the bit stream. However, careful packetization is essential to improve the system's performance. For example, if packet boundaries always coincide with codeword boundaries it is easier to prevent the propagation of errors from a corrupted packet to the next.

More generally it is desirable to design compression algorithms with reduced error propagation, such that a particular error affects only local (spatial or temporal) video data. This can be achieved not only through packetization, but also by introducing explicit synchronization markers which are unlikely to be corrupted (for example, to be corrupted a synchronization marker formed by a fixed pattern of 16 bits would require the occurrence of many consecutive bit errors.)

There are two major reasons for the existence of error propagation in compressed video bitstreams: the use of variable length codes for entropy coding and the reliance on predictive techniques to increase the coding performance.

Entropy coders used in practice generate *variable length* codes, i.e., the number of bits transmitted can be different for each symbol. Unique decoding of their output is possible because they satisfy the prefix condition; no valid codeword is a prefix of another valid codeword. When channel errors occur and a value of even one single bit is altered, it may no longer be possible to correctly decode not only the current symbol but, in some cases, the complete sequence of symbols following the location of the error. This may happen if the current symbol is incorrectly decoded with a different length from that of the original transmitted symbol. When this happens, the decoder may not be able to determine the output correctly until decoded symbols fall again at the correct codeword boundaries.

Resilience can be increased by introducing synchronization markers, which guarantee that the error will not propagate beyond the marker. However, explicit synchronization has a cost since the markers themselves do not convey any useful video information. In some cases it is possible to replace explicit synchronization markers by introducing constraints in the bit stream, such that groups of codewords have a total fixed length, even if individual codewords themselves have variable length. See the reference section[58–61] for examples of synchronization techniques used in image coding.

Another approach for robust entropy coding consists of using fixed-length codewords. For example, it is possible to group several symbols so that together they are assigned a fixed-length representation. This approach is used in the design of the scalar vector quantizers,[62] for example. It has the drawback of added complexity, as the encoder has to assign a single (fixed length) code to a block of inputs.

Finally, even variable-length entropy codes can be designed with inherent error resilience properties. For example, reversible variable length codes (RVLC)[63] are such that it is possible to decode them both in the forward and the backward directions (i.e., starting from the beginning of the sequence and working to the end, or vice versa). When, due to a bit error, synchronization is lost while decoding in the forward direction, one can start decoding in the backward direction. Using this method, it is possible to isolate the erroneously received codeword and to limit the error propagation. In some instances it is even possible to correct some of the errors.

The second source of error propagation is the utilization of predictive techniques, in particular motion-compensated prediction as used in most state of the art video coders, including both the ITU-T H.263 coders and the ISO MPEG family. As discussed earlier, for a given block within the current frame, the motion estimation algorithm finds the best match in the previous frame. The difference between the block and its best match is what is known as residual signal. The information transmitted over the channel consists of the residual signal and the corresponding motion vector field. Clearly, since prediction is being used, any erroneously received motion vector and residual information can affect not only the current but many subsequent reconstructed frames: the current (and erroneously reconstructed) frame is used to predict future frames and because the predictors in encoder and decoder will be different, so will the reconstructed images. This situation will last until the next intracoded frame (a frame coded without motion estimation) is transmitted. Increased robustness can be achieved by forcing intraframe coding to be used at regular intervals or by refreshing all the blocks in a frame following some pseudo-random pattern.[64]

Other approaches are possible if a closed-loop system is used. For example, in H.263+ coders[21] it is possible to use arbitrary frames as references in the prediction loop. We can then use as a reference frame not the previous frame (which may not have been correctly received) but rather the *the latest acknowledged frame* — the latest frame known to have been received correctly by the decoder. In this way, we guarantee that encoder and decoder will use the same information in their prediction loops. Other approaches use the feedback channel to increase the speed at which the system can be resynchronized after an error by forcing a refresh of regions that are known to have been corrupted.[65] The decoder can request the refresh through the back channel after detecting that errors have occurred.

3.4.3 Redundancy vs Robustness Trade-off

As described in Section 3.3.1 video compression algorithms operate by removing redundancy from the original source. But, considering the robustness issues discussed above as an example, the price for robustness is a reduction in the coding efficiency. This is a fundamental trade-off in designing error-resilient source coding: resilience can be achieved at the cost of increased redundancy, thus lower coding efficiency. This section considers approaches that explicitly avoid removing all the redundancy in order to increase the robustness.

Illustrating this trade-off is a very simplistic example where an image is transmitted and a few random bit errors occur. Assume first that the image is transmitted without compression. Then these bit errors (1) may be only perceptually noticeable if they affect the most significant bits of a particular pixel and (2) will never affect more than one pixel at a time, since each pixel is coded independently (no prediction or variable-length coding is used). Conversely, if the image has been efficiently compressed, variable-length codes are likely to be used; therefore a single bit error can cause complete loss of synchronism and render the decoded image useless. However it is not clear that it is better to transmit a very large (uncompressed) image to make the system more error resilient.

Of course typical trade-offs are not quite as drastic, and one can achieve reasonable compression performance while being robust. For example, it is possible to compress the data so that even though individual codewords have variable length, groups of codewords have constant length. In this way, full advantage of the statistical redundancy of the data is not taken, but some additional robustness is achieved. Another example is a system where motion estimation can be turned off to avoid errors propagating through several frames, i.e., the temporal redundancy is not fully removed.

It is important to note that redundancy is useful not just to limit the error propagation but also to allow better estimation of the information that was lost due to errors. For example, if not all spatial redundancy has been removed one can interpolate lost information from the received information in the local neighborhood. Similarly, as Hagenauer[66] has shown, if there is residual statistical redundancy in the transmitted codewords this can be used to aid in decoding. For example, if transmitted signals are not equally likely, it is possible to take this into account in weighting the likelihood of the missing information.

Thus, we can in general state that an image/video compression algorithm which has been optimized to squeeze out all redundancy will tend to be less robust than one where redundancy (statistical or temporal redundancy as in the two examples above) is carefully preserved to prevent error propagation or enable easier reconstruction.

3.4.3.1 Diversity and Multiple Description Coding

A particularly interesting approach for robust coding is to explicitly design a source encoder that can preserve some redundancy. Note that this is contrary to the traditional communication system design where redundancy is introduced by the channel coder only and is independent of the type of data transmitted.

Redundant source encoders can introduce structured correlation and redundancy at the signal level, before the entropy coding is applied. One such approach is Multiple Description Coding (MDC), where the same signal is encoded using two (or possibly more) encoders. An appropriate amount of structured correlation is constructed between these streams. Each stream is then transmitted along a separate path, or, in the case of wireless transmission, using a different path in the diversity structure. For example, each version of the signal could be transmitted on a different frequency, or at different times, or from a different antenna. Thus, assuming that not all the "channels" (frequencies, times, etc.) fade at the same time, the system is conceptually sending information over several virtual links, each subject to independent loss probabilities. Note that in a standard diversity system the redundancy is introduced only through the channel coding, while in a MDC approach redundancy is built into the source as well.

At the receiver, it is possible to reconstruct the original signal using either of the streams. Moreover, because of the inherent correlation between the two transmitted streams, one can estimate to some extent the lost information and improve the reconstructed signal. This is in contrast with the layered coding approaches to be discussed next, where no reconstruction is possible at the absence of the base layer (containing the most important signal information). When both streams are received without error, a better replica of the transmitted signal can be reproduced. In comparison with repetitive transmission of the same signal, MDC provides better performance both when there is no information loss and when one of the streams is received erroneously. MDC is particularly attractive in conjunction with transmission diversity.

Interest in MDC has been ongoing with progress being made since the early eighties in both the theory,[67,68] including the design of optimal quantizers[69–70] and its applications to speech coding under various transmission environments.[71–72] Some of the recent developments have included the development of Internet-based applications which use MDC to provide robustness[73–74] and the extension of MDC to transform coding[75–76] and wavelet coding.[77–78] The recent burst of activity in this area indicates that it may be one of the most promising approaches for robust image/video transmission.

3.4.4 Unequal Error Protection and Scalable Coding

So far, we have assumed that all robustness is achieved exclusively through source coding techniques, under the assumption that channel errors could happen, and that errors would occur at random and affect with equal likelihood the whole bitstream.

However, in practical source encoders each bit of the compressed bit sequence contains different amounts of information, and the effect of a loss caused by erroneously receiving a bit will be different in each case. It is therefore natural to try to protect different portions of a bit stream differently, so that important information is protected using stronger channel codes. This approach, known as Unequal Error Protection (UEP), assumes that channel coders can adequately provide different levels of protection.

3.4.4.1 Channel Coding Techniques Providing UEP

UEP can be achieved in many different ways in a channel coding environment. One basic principle of UEP codes is that each channel codeword carries several types of information, with different levels of protection for each. For example, assume two levels of protection are provided; *coarse* data is the information that is heavily protected, and *detail* is the information that is less protected. UEP can be achieved by *embedding* the coarse information in the fine information.

This can be best illustrated by the concept of clouds and satellites as introduced by Cover.[54] The fine information is transmitted by selecting one among all the possible channel codewords. However, these codewords are also grouped into *clouds*, i.e., sets of codewords that are relatively close to each other while being somewhat more distant from codewords in other clouds. Thus a selection of a cloud conveys information that is better protected than the information conveyed by the satellites within the cloud.

When one of the satellite points is received the information about which cloud was transmitted can be decoded; this can be done even if the error is large, since confusing different points within the cloud does not affect the cloud information. The least protected information is carried by the specific codeword within the cloud, which is itself more susceptible to errors. Examples of codes designed with this property can be found in Kasami et al.[79] and Lin, Lin, and Lin.[80] This approach can also be used with the selection of modulation points[81] rather than with specific codes. Generally in a given modulation scheme it is possible to control the degree of protection provided by each modulation point by selecting the distances between the modulation points.

Once one allows the selection of modulation points in arbitrary positions it is possible to combine the design of source coders and channel modulation schemes. This truly joint source channel coding design then no longer has the goal to minimize distortion for a given rate, but rather to minimize overall (received) distortion for a given average transmission power budget.[82–84]

Probably the most popular example of UEP codes are the Rate-Compatible Punctured Convolutional (RCPC) codes introduced by Hagenauer.[85–87] RCPCs are essentially convolutional codes that can be *punctured* at predetermined positions (i.e., certain output bits are not transmitted). In this way the transmitter can control the rate. If the puncturing is increased fewer bits are transmitted, thus a lower degree of protection is offered; conversely, if less puncturing is performed then more redundancy is present and more protection will be offered.

3.4.4.2 Scalable Video Coding Techniques

While any compression algorithm will tend to produce bits with unequal importance, it may be desirable to design algorithms where the encoder has control of the relative importance of the information. This will lead us to the design of scalable or multi-resolution algorithms.

In layered coding, a source signal is separated into a number of streams (or layers). This separation is usually based on the importance of the information of each layer in the reconstruction of the signal. Therefore, different layers contain information with different levels of significance. The most important layer, called base layer, contains information without which the reconstruction of the signal would not be possible. By adding more layers to this base layer, a better replica of the original signal can be reconstructed. The base layer corresponds to high priority information whereas the enhancement layer is of lower priority and its loss does not have catastrophic consequences.

Different methods can be used to produce multiple layers from a single source signal. One such method is to split the bit stream generated by the source encoder into multiple streams. For example, for the motion-compensated based source encoders (such as MPEG and H.263), the motion vectors are of more significance for the reconstruction than is the residual information. Therefore, one can construct a base layer containing bits corresponding to these motion vectors and other control information, and another stream, the enhancement layer, can then contain the bits corresponding to the residual information. A different splitting strategy than the above can be used.

The MPEG-2 standard[31] provides several methods to achieve scalability: temporal, SNR, and spatial scalability. In a temporal scalability framework, the base layer may consist of video data encoded at a reduced frame rate, for example 10 frames/sec instead of the original 30 frames/sec of the source material. The enhancement layer provides the remaining frames. SNR scalability consists of transmitting the low frequency DCT coefficients in the base layer (as they tend to convey most of the energy in the sequence) and the higher frequency coefficients in the enhancement layer. Finally spatial scalability consists of coding as the base layer a sequence that is produced by reducing the size of the frames in the original sequence. The enhancement layer contains the difference between the the base layer frames interpolated to the larger size and the original frames.

Other examples of scalable coding techniques include wavelet or subband coding,[88–89] where the resulting image subbands can be, if needed, separately coded and transmitted. An alternative method which offers additional flexibility is the so-called Pyramid coding.[90–91] In this framework a coarse version of the signal (generated, for example, through spatial decimation) is compressed as the base layer, then the difference between the decoded base layer interpolated to the full resolution and the original image is coded as the enhancement layer.

In summary, with a scalable coder and efficient UEP channel codes it is possible to match important source information to high levels of protection. For example in wavelet coders such as those described by Shapiro[92] and Said and Pearlman,[93] the bits produced by the encoder are generated in order of importance, but if a single bit is lost all remaining bits in the file are useless. Thus, as discussed in Sherwood and Zeger,[56] one can use an RCPC as a channel code with the maximum redundancy (and protection) available at the beginning of the file, and with little or no protection available towards the last bits in the file.

3.4.5 Adaptation to Channel Conditions and Rate Control

So far we have described approaches that are mostly suited for open-loop error control environments, i.e., where no feedback from the receiver/decoder is provided to the source encoder/transmitter. We now show how feedback information can be used to further improve the performance. The basic principle as illustrated by Figure 3.4 is simple: we should transmit less information during periods of poor channel performance. This is achieved by feeding back information about the channel state to the rate control algorithm at the encoder. The channel state information can be, as in Figure 3.4, simply the result of the latest acknowledgment in an ARQ scheme: a negative acknowledgment indicates poor channel behavior.

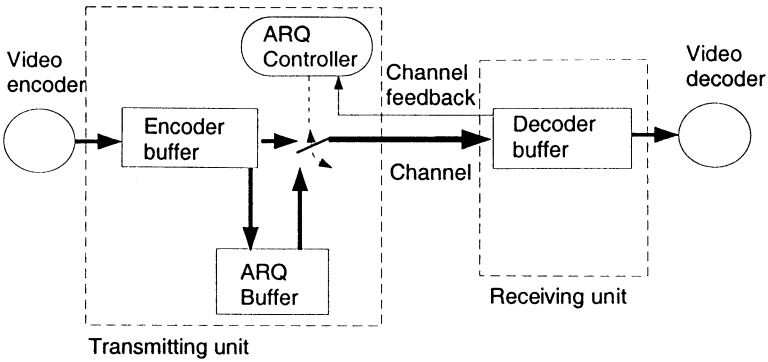

Figure 3.4 Diagram of buffers in the system

Let us formalize this by considering once again the delay constraints affecting video transmission that were disussed in Section 3.3.2. In both CBR and VBR transmission cases data is stored in buffers at encoder and decoder. Assume a variable channel rate of $C(i)$ during the i-th coding unit interval, then, the encoder buffer state at time i is

$$B(i) = \max B(i-1) + r_{ix(i)} - C(i), 0$$

with $B(0) = 0$ being the initial state of the buffer, and $r_{ix(i)}$ is the rate required by frame i when quantizer $x(i)$ is used.

Let us consider now what constraints need to be applied on the encoder buffer state (controlling the encoder buffer suffices to guarantee that the delay constraints are met[32,38]). First, the buffer state $B(i)$ cannot grow indefinitely because the physical buffer memory will be limited. If B_{max} is the physical memory available then we need to guarantee that $B(i) \leq B_{max}$ at all times. In addition, in order for the delay constraints not to be violated we need to guarantee that the data corresponding to coding unit i is transmitted before $t_i + \Delta T$, that is, transmission has to be completed during the next ΔN coding unit intervals. Intuitively, in order for this constraint to be met, all we need to ensure is that the future channel rates, over the next ΔN units, are sufficient to transmit all the data in the buffer.

Let us define the effective buffer size, $B_{eff}(i)$, as the sum of future channel rates over the next ΔN intervals:

$$B_{eff}(i) = \sum_{k=i+1}^{i+\Delta N} C(k),$$

Then it is easy to see, as demonstrated in Hsu et al.[38] and Reibman and Haskell,[32] that correct transmission is guaranteed if

$$B(i) \le B_{eff}(i), \forall i.$$

As an example, consider the case where $C(i) = \bar{C} = R_T / N$ is constant. Then if the system operates with an end-to-end delay ΔN the buffer can store no more than $\Delta N \cdot \bar{C}$ bits at time i. For a detailed analysis of the relationship between buffering and delay constraints, see Reibman and Haskell[32] and Hsu et al.[38]

We call this the *effective* size because it defines a constraint imposed regardless of the physical buffer size. In general the applicable constraint will be imposed by the smallest of $B_{eff}(i)$ and B_{max}. Assuming that sufficient physical buffer storage is available, that B_{max} is always larger than $B_{eff}(i)$, our problem becomes, to find the optimal set of quantizers $x(i)$ for each i such that the buffer occupancy

$$B(i) = \max(B(i-1) + r_{ix(i)} - C(i), 0),$$

is such that

$$B(i) \le B_{eff}(i)$$

and some metric $f(d_{1x(1)}, d_{2x(2)}, \ldots, d_{Nx(N)})$is minimized. The metric to be used could be, for example, the mean squared error.

It is interesting that the constraints depend on the channel rates. When the channel rates can be chosen by the user (e.g., transmission over a network) this leads to interesting questions on the best combination of source and channel rates given constraints on the channel rates that can be utilized.[38–40]

In the wireless environment we are considering we have a variable channel rate because packets of video information can be randomly lost. However, we have no control over the behavior of the summer and consequently on the channel rates. Thus, deterministically ensuring that transmission will be successful requires us to know $B_{eff}(i)$ at all times i. Because the behavior of the channel is not deterministic and $B_{eff}(i)$ is a function of the *future* channel rates, we will never be able to know $B_{eff}(i)$ exactly. Instead what we can do is use some of the models described in Section 3.2.3 in order to *predict* $B_{eff}(i)$. This approach is described in more detail elsewhere[41,94–96] and it can be shown to perform better than approaches where there is no feedback about the channel state. In general, if a closed-loop system is used we can take advantage of the available channel information to adapt the information transmitted by the encoder and improve overall performance. Other examples of this approach can be found in Hafez and Rajugopal[97] and Liu and Zarki.[98]

In summary, given the constraints imposed by the delay and our observation of the channel state, we can optimize the selection of source coding rates. Other approaches may allow us to modify the level of redundancy or the power used for transmission. The principle remains the same: in a closed-loop error control scheme it is possible to adjust the parameters (both source and channel coding) to best match the observed state of the channel.

3.5 CONCLUSIONS

In this chapter we have introduced some of the major issues in the design of video coding algorithms that are suitable for wireless communications. The requirements for these algorithms to become reality go beyond achieving reasonable quality at low rate and with low power consumption. In fact one of the key requirements for these systems to become reality is to provide a robustness to errors that, as discussed in Section 3.4, is best achieved when the video encoder is aware of the characteristics of the underlying channel. This section has also illustrated a number of techniques to increase the robustness, ranging from simple packetization to sophisticated rate adaptation algorithms based on channel state feedback. We expect much of the research involved in making wireless video a reality to concentrate on some of these issues.

REFERENCES

1. C. Lee, *Mobile Cellular Telecommunications*. McGraw-Hill Inc., New York, 1995.
2. R. Steele (Ed), *Mobile Radio Communications*. Pentech, London, 1992.
3. W. Tuttlebee (Ed), *Cordless Telecommunication in Europe*. Springer-Verlag, 1990.
4. *GSM Recommendation, European Telecommunications Standardization Institute, Sophia Antipolis, France,* 1988.
5. *Dual-Mode Subscriber Equipment –Network Equipment Compatibility Specification*. Telecommunications Industry Association, 1989.
6. *Mobile-station-base-station compatibility standard for dual-mode wideband spread spectrum cellular system*. Telecommunications Industry Association, July 1993.
7. *Recommended minimum performance standards for dual-mode wideband spread spectrum mobile stations*. Telecommunications Industry Association, Jan. 1994.
8. *Mobile Station-Base Station Compatibility Standard for Dual-Mode Wideband Spread Spectrum Systems*. Telecommunications Industry Association, Washington, D.C.
9. *Coding of Speech at 8 Kbit/s using conjugate-structure algebraic code-excited linear prediction* (*CS-CELP*). ITU draft recommendation G.729, Feb. 1996.
10. *Wireless LAN Medium Access Control* (*MAC*) *and Physical Layer* (*PHY*) *specification*, 1997.
11. *Radio Equipment and Systems (RES); High Performance Radio Local Area Network (HIPERLAN), Type 1, functional specifications, ETSI Final Draft prETS 300652.*
12. L. Hanzo and J. Stefanov, “The pan-european digital cellular mobile radio system-known as GSM.” In *Mobile Radio Communications*. R. Steele, Ed. : Pentech, London, pp. 677–773, 1992.

13. http://www.tiaonline.org/standards/sfg/imt2k/
14. E. G. Tiedemann Jr., Y.-C. Jou, and P. Odenwalder, "The evolution of IS-95 to a third generation system and to the IMT-2000 era." *Proc. ACTS Summit*, pp. 924–929, 1997.
15. E. G. Tiedemann Jr., "CDMA2000: The evolution of is-95 to third generation systems." *Proc. of the 3rd International Symposium on Multi-Dimensional Mobile Communications*, (Menlo Park, CA), 1998.
16. E. Dahlman, B. Gudmundson, M. Nilsson, and J. Skold, "UMTS/IMT 2000 based wideband CDMA." *IEEE Commun. Mag.*, vol. 39, pp. 70–80, Sept. 1998.
17. F. Adachi, M. Sawahashi, and H. Suda, "Wideband DS-CDMA for next-generation mobile communications systems." *IEEE Commun. Mag.*, vol. 39, pp. 56–69, Sept. 1998.
18. T. Ojanpera and R. Prasad, "An overview of air interface multiple access for IMT-2000/UMTS." *IEEE Commun. Mag.*, vol. 39, pp. 82–95, Sept. 1998.
19. L. Hanzo, "Bandwidth-efficient wireless multimedia communications." Proc. IEEE, vol. 86, pp. 1342–1382, 1998.
20. "Special issue on MPEG-4." *IEEE Trans. on Circ. and Syst. for Video Technol.*, Feb. 1997.
21. ITU-T, "Video coding for low bitrate communication." ITU-T Recommendation H.263; version 1, Nov. 1995, version 2, Jan. 1998.
22. T. H. Meng, A. C. Hung, E. K. Tsern, and B. M. Gordon, "Low-power signal processing system design for wireless applications." *IEEE Personal Commun.*, vol. 5, pp. 20–31, June 1998.
23. S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "Serial concatenated of interleaved codes: performance analysis, design, and iterative decoding." *IEEE Trans. Inform. Theory*, vol. 44, pp. 909–926, 1998.
24. M. Khansari, A. Jalali, E. Dubois, and P. Mermelstein, "Low bit-rate video transmission over fading channels for wireless microcellular systems." *IEEE Trans. on Circ. and Sys. for Video Technol.*, pp. 1–11, Feb. 1996.
25. E. N. Gilbert, "Capacity of burst-noise channel." Bell Syst. Tech. J., vol. 39, pp. 1253–1265, Sept. 1960.
26. M. Zorzi, R. R. Rao, and L. B. Milstein, "ARQ error control for fading mobile radio channels." IEEE Trans. on Veh. Tech., vol. 46, pp. 445–455, May 1997.
27. M. Zorzi and R. R. Rao, "ARQ error control for delay-constrained communications on short-range burst-error channels." Proc. IEEE VTC'97, 1997.
28. M. Zorzi, R. R. Rao, and L. Milstein, "On the accurancy of a first-order Markov model for data transmission on fading channels." *Proc. IEEE ICUPC'95*, 1995.
29. M. Zorzi, R. R. Rao, and L. Milstein, "A Markov model for block errors on fading channels." *Proc. PIMRC96*, 1996.
30. R. J. Clarke, *Digital Compression of Still Images and Video*, Academic Press, 1995.
31. J. Mitchell, W. Pennebaker, C. E. Fogg, and D. J. LeGall, *MPEG Video Compression Standard*. Chapman & Hall, New York 1997.
32. A. R. Reibman and B. G. Haskell, "Constraints on variable bit-rate video for ATM networks." *IEEE Trans. on CAS for Video Technol.*, vol. 2, pp. 361–372, Dec. 1992.
33. S.-W. Wu and A. Gersho, "Rate-constrained optimal block-adaptive coding for digital tape recording of HDTV." *IEEE Trans. on Circ. and Syst. for Video Technol.*, vol. 1, pp. 100–112, Mar. 1991.
34. A. Ortega, K. Ramchandran, and M. Vetterli, "Optimal trellis-based buffered compression and fast approximation." *IEEE Trans. on Image Proc.*, vol. 3, pp. 26–40, Jan. 1994.

35. W. Ding and B. Liu, “Rate control of MPEG video coding and recording by rate-quantization modeling.” *IEEE Transactions on Circ. and Syst. for Video Technol.*, vol. 6, pp. 12–20, Feb. 1996.
36. L.-J. Lin and A. Ortega, “Bit-rate control using piecewise approximated rate-distortion characteristics.” *IEEE Trans. on Circ. and Sys. for Video Technol.*, vol. 8, pp. 446–459, Aug. 1998.
37. A. Ortega and K. Ramchandran, “Rate-distortion techniques in image and video compression.” *IEEE Signal Processing Mag.*, Nov. 1998.
38. C.-Y. Hsu, A. Ortega, and A. Reibman, “Joint selection of source and channel rate for VBR video transmission under ATM policing constraints.” *IEEE J. on Sel. Areas in Commun.*, vol. 15, pp. 1016–1028, Aug. 1997.
39. J.-J. Chen and D. W. Lin, “Optimal bit allocation for coding of video signals over ATM networks.” *IEEE J. on Sel. Areas in Commun.*, vol. 15, pp. 1002–1015, Aug. 1997.
40. W. Ding, “Joint encoder and channel rate control of VBR video over ATM networks.” *IEEE Trans. on Circ. and Syst. for Video Technol.*, vol. 7, pp. 266–278, Apr. 1997.
41. C.-Y. Hsu, A. Ortega, and M. Khansari, “Rate control for robust video transmission over burst-error wireless channels,” *IEEE J. on Sel. Areas in Commun.*, 1999. To appear.
42. Z. Miao and A. Ortega, “Rate control algorithms for video storage on disk based video servers in *Proc. of Asilomar Confer. on Signals Syst. and Comp.*, Nov. 1998.
43. *ITU-T Recommendation H.261: Video codec for audiovisual services at p* $\times$ *64 Kbits*, Mar. 1993.
44. M. Liou, “Overview of the px64 kbit/s video coding standard.” *Comm. of the ACM*, vol. 34, pp. 59–63, Apr. 1991.
45. M. Kunt, A. Ikonomopoulos, and M. Kocher, “Second generation image coding techniques.” *Proc. of the IEEE*, vol. 73, pp. 549–579, Apr. 1985.
46. H. G. Musmann, M. Hötter, and J. Ostermann, “Object-oriented analysis-synthesis coding of moving images.” *Signal Processing: Image Commun.*, vol. 1, pp. 117–138, Oct. 1989.
47. J. Ostermann, “Object-oriented analysis-synthesis coding based on the source model of moving flexible 3D objects.” *IEEE Trans. on Image Proc.*, vol. 3, pp. 705–710, Sept. 1994.
48. N. Färber, E. Steinbach, and B. Girod, “Robust H.263 compatible video transmission over wireless channels.” In *Proc. PCS’96*, pp. 575–578, 1996.
49. *IEEE Personal Commun. Mag.*, vol. 5, no. 3. Special Issue on Energy Management in Personal Communications and Mobile Computing.
50. W. Pennebaker and J. Mitchell, *JPEG Still Image Data Compression Standard*. Van Nostrand Reinhold, 1994.
51. D. Lee, “New work item proposal: JPEG 2000 image coding system.” ISO/IEC JTC1/SC29/WG1 N390, 1996.
52. C. Chrysafis and A. Ortega, “Line Based Reduced Memory Wavelet Image Compression.” in *Proc. IEEE Data Compression Conf.*, pp. 308–407, IEEE Computer Society Press, Los Alamitos, California, 1998.
53. C. E. Shannon, “A mathematical theory of communication.” *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 1948.
54. T. Cover, “Broadcast channels.” *IEEE Trans. on Inform. Theory*, vol. IT-18, pp. 2–14, Jan. 1972.
55. K. Fazel and J. L. Huillier, “Application of unequal error protection codes on combined source-channel coding.” In *IEEE Int. Conf. Commun.*, pp. 320.5.1–6, Apr. 1990.

56. G. Sherwood and K. Zeger, "Progressive image coding for noisy channels." *IEEE Signal Processing Lett.*, vol. 4, pp. 189–191, Jul. 1997.
57. K. Ramchandran and M. Vetterli, Multiresolution Joint Source-Channel Coding for Wireless Channels. *Wireless Communications: A Signal Processing Perspective*, Editors: V. Poor and G. Wornell. Prentice-Hall, 1998.
58. D. Yu and M. W. Marcellin, "A fixed-rate quantizer using block-based entropy-constrained quantization and run-length coding." *IEEE Trans. on Image Proc.*, Aug. 1996. Submitted, Available in http://www-spacl.ece.arizona.edu/.
59. D. W. Redmill and N. G. Kingsbury, "The EREC: An error-resilient technique for coding variable-length blocks of data." *IEEE Trans. on Image Proc.*, vol. 5, pp. 565–574, Apr. 1996.
60. C. D. Creusere, "A new method of robust image compression based on the embedded zerotree wavelet algorithm." *IEEE Trans. on Image Proc.*, vol. 6, pp. 1436–1442, Oct. 1997.
61. Y. Yoo and A. Ortega, "Constrained bit allocation for error resilient JPEG coding." In *Proc. of Asilomar Confer. on Signals, Syst. and Comp.*, 1997.
62. R. Laroia and N. Farvardin, "A structured fixed rate vector quantizer derived from a variable-length scalar quantizers: Part I - Memoryless sources." *IEEE Trans. Inform. Theory*, vol. 39, pp. 851–867, May 1993.
63. J. Wen and J. Villasenor, "Reversible variable length codes for efficient and robust image and video coding." In *Proc 1998 IEEE Data Compression Confer.*, pp. 471–480, 1998.
64. S. McCanne, M. Vetterli, and V. Jacobson, "Low-complexity video coding for receiver-driven layered multicast." *IEEE J. on Sel. Areas in Commun.*, vol. 15, pp. 983–1001, Aug. 1997.
65. E. Steinbach, N. Färber, and B. Girod, "Standard compatible extension of H.263 for robust video transmission in mobile environments." *IEEE Trans. on Circ. and Sys. for Video Technol.*, vol. 7, pp. 872–881, Dec. 1997.
66. J. Hagenauer, "Source-controlled channel decoding." *IEEE Trans. on Commun.*, vol. 43, pp. 2449–2457, Sept. 1995.
67. A. A. El-Gamal and T. M. Cover, "Achievable rates for multiple descriptions," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 851–857, Nov. 1982.
68. J. C. Batllo and V. Vaishampayan, "Asymptotic performance of multiple description transform codes." *IEEE Trans. Inform. Theory*, vol. 43, no. 2, pp. 703–707, 1997.
69. V. A. Vaishampayan, "Design of multiple description scalar quantizers." *IEEE Trans. Inform. Theory*, vol. 39, no. 3, pp. 821–834, 1993.
70. A. Ingle and V. A. Vaishampayan, "DPCM system design for diversity systems with applications to packetized speech." *IEEE Trans. Speech and Audio Process.*, vol. 3, no. 1, pp. 48–58, 1995.
71. N. S. Jayant and S. W. Christensen, "Effects of packet losses in waveform coded speech and improvements due to an odd-even sample-interpolation procedure." *IEEE Trans. Commun.*, vol. COM-29, pp. 101–109, Feb. 1981.
72. S.-M. Yang and V. A. Vaishampayan, "Low-delay communication for rayleigh fading channels: An application of the multiple description quantizer." *IEEE Trans. Commun. Theory*, vol. 43, pp. 2771–2783, Nov. 1995.
73. V. H. M. A. Sasse, M. Handley, and A. Watson, "Reliable audio for use over the internet." In *Proc. INET*, 1995.
74. M. Podolsky, C. Romer, and S. McCanne, "Simulation of FEC-based error control for packet audio on the internet." *INFOCOM'98*, 1998.

75. M. T. Orchard, Y. Wang, V. Vaishampayan, and A. R. Reibman, "Redundancy rate-distortion analysis of multiple description coding using pairwise correlating transforms." *ICIP'97*, 1997.
76. V. K. Goyal and J. Kovacevic, "Optimal multiple description transform coding of large sc Gaussian vectors." *Proc. of IEEE Data Compression Confer.*, 1998.
77. S. D. Servetto, K. Ramchandran, V. Vaishampayan, and K. Nahrstedt, "Multiple description wavelet based image coding." *ICIP'98*, 1998.
78. W. Jiang and A. Ortega, "Multiple description coding via polyphase transform and selective quantization." In *Proc. of Visual Commun. and Image Process.*, 1999.
79. T.Kasami, S.Lin, V. Wei, and S. Yamamura, "Coding for the binary symmetric broadcast channel with two receivers." *IEEE Trans. on Inform. Theory*, vol. IT-31, pp. 616–625, Sept. 1985.
80. M.-C. Lin, C.-C. Lin, and S. Lin, "Computer search for binary cyclic UEP codes of odd length up to 65." *IEEE Trans. on Inform. Theory*, vol. 36, pp. 924–935, July 1990.
81. K. Ramchandran, A. Ortega, K. M. Uz, and M. Vetterli, "Multiresolution broadcast for digital HDTV using joint source-channel coding." *IEEE J. on Sel. Areas in Commun.*, vol. 11, pp. 6–23, Jan. 1993.
82. N. Phamdo, N. Farvardin, and T. Moriya, "A unified approach to tree-structured and multistage vector quantization for noisy channels." *IEEE Trans. on Inform. Theory*, vol. IT-39, pp. 835–850, May 1993.
83. N. Farvardin and V. Vaishampayan, "Optimal quantizer design for noisy channels: An approach to combined source-channel coding." *IEEE Trans. on Inform. Theory*, vol. IT-33, pp. 827–838, Nov. 1987.
84. I. Kozintsev and K. Ramchandran, "Robust image transmission over energy-constrained time-varying channels using multiresolution joint source-channel coding." *IEEE Trans. on Signal Process., Special Issue on Wavelets and Filter Banks*, vol. 46, pp. 1012–1026, Apr. 1998.
85. J. Hagenauer, "Rate-compatible punctured convolutional codes (rcpc codes) and their applications." *IEEE Trans. on Commun.*, vol. COM 36, pp. 389–400, Apr. 1988.
86. J. Hagenauer, N. Seshadri, and C.-E. W. Sundberg, "The performance of rate-compatible punctured convolutional codes for digital mobile radio." *IEEE Trans. on Commun.*, vol. 38, pp. 966–980, July 1990.
87. R. V. Cox, J. Hagenauer, N. Seshadri, and C. Sundberg, "Variable rate sub-band speech coding and matched convolutional channel coding for mobile radio channels." *IEEE Trans. on Signal Process.*, vol. 39, pp. 1717–1731, 1991.
88. M. Vetterli and J. Kovacevic, *Wavelets and Subband Coding*. Prentice-Hall, 1995.
89. D. Taubman and A. Zakhor, "Multirate 3-D subband coding of video." *IEEE Trans. on Image Process.*, vol. 3, Sept. 1994.
90. P. J. Burt and E. H. Adelson, "The laplacian pyramid as a compact image code." *IEEE Trans. on Commun.*, vol. 31, pp. 532–540, Apr. 1983.
91. K. M. Uz, M. Vetterli, and D. LeGall, "Interpolative multiresolution coding of advanced television with compatible sub-channels." *IEEE Trans. on CAS for Video Technol., Special Issue on Signal Process. for Advanced Television*, vol. 1, pp. 86–99, Mar. 1991.
92. J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients." *IEEE Trans. on Signal Process.*, vol. 41, pp. 3445–3462, Dec. 1993.
93. A. Said and W. A. Pearlman, A new fast and efficient image coder based on set partitioning in hierarchical trees. *IEEE Trans. Circ. and Syst. for Video Technol.*, pp. 243–250, June 1996.

94. A. Ortega and M. Khansari, Rate control for video coding over variable bit rate channels with applications to wireless transmission. *Proc. of the 2nd. Intl. Confer. on Image Process.,* 1995.
95. C.-Y. Hsu, A. Ortega, and M. Khansari, Rate control for robust video transmission over wireless channels. *Proc. of Visual Commun. and Image Process*, 1997.
96. C.-Y. Hsu, *Rate control for video transmission over variable rate channels*. Ph.D. thesis, University of Southern California, Aug. 1998.
97. R. H. M. Hafez and G. R. Rajugopal, Adaptive rate controlled robust video communications over packet wireless networks. *ACM/Baltzer Mobile Networks and Appl. J.*, vol. 3, no. 1, pp. 33–47, 1998.
98. H. Liu and M. El Zarki, Adaptive source rate control real-time wireless video transmission. *ACM/Baltzer Mobile Networks and Appl. J.*, vol. 3, no. 1, pp. 49–60, 1998.