# ROBUST VIDEO CODING — AN OVERVIEW

*Sheila S. Hemami*

School of Electrical & Computer Engineering, Cornell University, Ithaca, NY 14853

## ABSTRACT

Robust video coding is creating coded video data which provides received and decoded video of application-appropriate quality when transmitted over an imperfect channel. Video contains an enormous amount of redundancy, both from signal processing and psychological perspectives, which effective compression attempts to remove. However, ideal source coding leaves a coded stream which is extremely vulnerable to errors, and therefore compression and transmission provisions that avoid catastrophic failure caused by lost, delayed, or errant data are imperative. Robustness in video coding is obtained through source coding, channel coding, and error concealment following decoding. This paper first discusses issues specific to video in providing robustness and then poses the problem with two extremes of providing robustness and an example. Common strategies employed in robust video coding are then briefly described, and the specific application of these strategies in session papers are presented. The paper concludes with a discussion of future directions in robust video coding.

## 1. INTRODUCTION

This overview paper considers *robust video coding* in the context of *transmission*: robustness to bit and burst errors, to packet losses, and/or to variability in channel conditions. Robust video coding is creating coded video data which provides received and decoded video of application-appropriate quality when transmitted over an imperfect channel. Typically the channel characteristics are known to some extent during the design and/or execution of the encoding algorithm. These characteristics can be as general as simple knowledge of average channel bit rates and the fact that the channel is imperfect, or as specific as bit error rates and packet loss rates for multiple available channels. They may be assumed constant or time varying, and they may provide feedback data about their states. The challenge of robust video coding is to provide high quality video in a bandwidth- and computationally efficient manner, in spite of channel imperfections. (In this paper, robust video coding and robust video transmission will be used interchangeably.)

Video contains an enormous amount of redundancy, both from signal processing and psychological perspectives, which effective compression attempts to remove. An ideal source coder removes all redundancy, effectively equalizing the importance of bits in the coded stream. A single bit error in this resulting stream can then cause catastrophic decoder failure, and therefore such a stream requires perfect transmission. However, transmission systems may not be able to provide the necessary quality-of-service guarantee for such compressed video. Therefore compression and transmission provisions that avoid catastrophic failure caused by lost, delayed, or errant data are imperative.

Robustness in video coding is obtained through source coding, channel coding, and error concealment following decoding. Source coding can be modified from the ideal by increasing the redundancy in the video representation and by making the encoded bitstream itself more robust to errors. Channel coding adds controlled redundancy in exchange for source coding rate. Error concealment is a post-processing step which mitigates the effect of errors which were not compensated by the channel coding. These three approaches need not be implemented independently, but rather provide the best performance when strategically combined. For example, redundancy can be included in the source coding to optimize the performance of a particular concealment algorithm; source and channel coding can be jointly selected to satisfy an overall rate constraint while minimizing a distortion metric or achieving a throughput measure.

The latest video coding standards (MPEG-2, MPEG-4, H.263+, H.264, etc.) have included various features to improve robustness. Use of resynchronization markers, reversible variable-length codes, and other strategies limit error propagation and facilitate use of the standards in less-than-perfect transmission environments. However, they still require channel coding in most practical applications. Work beyond the standards has developed robust video coding techniques which are appropriate for a variety of transmission environments and applications.

Clearly, a complete review with citations of approaches to robust video coding is not possible in this short paper. References given are intended to be examples only, and do not constitute a complete survey of the literature. The reader is directed to two excellent reviews considering standards-compatible robustness enhancements [1] and general robust video coding [2], both of which cover the field to the end of the 1990's, and to a special journal issue on error-resilient image and video transmission [3] This paper first discusses issues specific to video in providing robustness and then poses the problem with two extremes of providing robustness and an example. Common strategies employed in robust video coding are then briefly described, and the specific application of these strategies in session papers are presented. The paper concludes with a discussion of future directions in robust video coding.

## 2. CONSIDERATIONS FOR IMPROVING ROBUSTNESS

Providing robust transmission of video is not best achieved via a straightforward application of robust image transmission approaches [4]; video has greater structure than images, it is inherently time sensitive, and it is perceived differently than images. Appropriate consideration of these features of video suggests obvious and efficient techniques for improving robustness.

**Video data structure.** The extreme temporal redundancy of video is the greatest difference between still images and video; the majority of frames have extremely high correlation with those immediately before and after them. Efficient coding reduces this redundancy through temporal coding.

Consider a motion-compensated video coder which employs both spatial and temporal coding. (Note that the standards as well the newer $2D + t$ motion compensated wavelet coders fit into this category.) Data comprising a compressed video sequence possess extreme inequality of importance both at the frame level and within a frame.

As an example, consider an MPEG-like coder. I, P, and B frames have decreasing importance within a group-of-pictures (GOP); the B frames require both the P and I frames to be decoded; the P frame requires the I-frame, and without the I-frame, the entire GOP cannot be decoded. Within a P or B frame, loss of motion vectors results in a completely non-decodable frame, while loss of difference information will result in a possibly distorted but still decodable frame.

**Delay.** Video data is time and hence delay sensitive; if information is decoded after its display time, it is useless. This implies limitations on the strength and extent of channel coding techniques, and also allows the decoder to abort frame decoding and concealment when necessary.

**Human perception**. The human visual system responds differently to errors in video data than in still image data. It is well known that the average MSE of individual frames comprising what is perceived as a sequence of acceptable quality can be substantially below that of an image considered of acceptable quality. Additionally, the HVS can tolerate one or several frames with extreme errors without impacting the observer's overall perception of the quality of the sequence [5].

While structure and delay constraints are commonly used in robust video coding techniques, exploiting perception is less common due to the challenges posed by accurately measuring perceived quality and appropriately adapting to video content.

## 3. TWO EXTREMES FOR ROBUST VIDEO CODING, AND AN EXAMPLE

To appropriately understand the video transmission problem, consider two extremes of video transmission over unreliable channels which allow lost or errant data to be recovered from received data.

**The first extreme** is an information theoretic result given by Shannon's well known separation theorem: a stochastic process can be optimally transmitted over a channel if the source coding and channel coding are performed independently and optimally. Zero redundancy is placed in the source coding, and maximum redundancy is placed in the channel coding. Recovery from transmission errors is possible, providing that restrictions placed by the channel coding on the errors are not exceeded. Note that knowledge of the channel is required to select an appropriate channel code.

**As an example**, digital satellite television broadcasters employ this extreme to deliver MPEG-2 coded broadcast and cable television to their subscribers. While each vendor's specific code selections vary, the overall approach is similar. MPEG-2 source encoding is applied to raw digital video. The coded video, appropriately packetized, is channel coded using both Reed-Solomon encoding and convolutional coding. The resulting data streams are transmitted to users over a satellite link. The channel coding is sufficient to deliver perfectly decoded video when the receiving antenna has line-of-sight to the satellite. The channel codes are selected to provide "robust video" except in the case of extreme precipitation; some satellite television providers increase their transmit power to areas of heavy rain. Clearly this approach is sufficient for this system, as approximately 24 million American households paid sometimes well over $30/month for these services in 2004.

**A second extreme** exists in which knowledge of the channel is not required to provide robust video transmission. The uncoded video is simply transmitted and the redundancy present in the image is used to compensate for lost data. In this case, raw data can be corrupted, but uncoded video has sufficient redundancy to allow successful concealment of the errors using the received data at the decoder. The reconstructed video will not be pixel-for-pixel equivalent to the original, but it will be visually equivalent, which is as well as the first extreme performed anyway because in the first extreme, the data was first source-coded via lossy compression to achieve visual but not exact equivalence.

The first extreme is far more efficient with respect to the total channel bandwidth required, so the second is only of hypothetical interest. But the second extreme suggests the existence of a continuum between the two. Separate source and channel coding is surely the simplest, and is also clearly functional. However, it is not necessarily the most efficient. The next section discusses common strategies to provide robust video coding.

## 4. COMMON STRATEGIES EMPLOYED IN ROBUST VIDEO CODING

This section describes common strategies for providing robust video coding in both source and channel coding as well as joint source-channel techniques and error concealment. Typically, the techniques described below are not used independently. Rather, they are applied in conjunction with each other, as will be seen in the other papers in the session.

### 4.1 Forward Error Correction

To combat the effects of both bit and burst errors, channel coding providing forward error correction (FEC) is commonly applied to video coded for transmission. The FEC can be applied to all coded video data equally, yielding equal error protection (EEP), or unequally based on the importance of the information in generating a decoded stream, yielding unequal error protection (UEP). Achieving UEP includes not only specific channel code selection but also power management in at the physical layer in wireless transmission applications.

Knowing the channel characteristics *a priori* is necessary to select an appropriate FEC code. Interleaving can be and is often used to lessen the effect of burst errors, though delay constraints must be considered. Selection of appropriate FECs depends on both decoder complexity as well as the delay constraints.

### 4.2 Layered Coding & Scalability

*Layered coding* is a source coding strategy and refers to an encoding strategy where information of differing importances is transmitted with differing priorities [6]. Generally, the highest-priority data is encoded in a self-contained form and transmitted over the most reliable channel. Many robust video applications assume or require that the most reliable channel is indeed perfect. Note that layered coding can be assumed without an explicit mechanism for providing the desired reliabilities. In practice, appropriate selection of FEC channel codes suffices.

Layers can be formed by ranking coded data in order of its importance in decoding. For example, in hybrid video coding, within a motion-compensated frame, motion vector data is generally considered to be more important than texture data (i.e., difference frames). Within an MPEG group-of-pictures, I frames are more important than P-frames. These examples, however, simply rank data within a stream. In many applications, it is important that the lowest layer be independently decodable. This leads to the concept of *scalability.*

A *scalable* video compression algorithm provides an encoded stream which is decodable at multiple rates. The *base layer* provides the lowest rate, and one or more *enhancement layers* augment the data to improve the video quality. Providing for such scalability incurs an overhead in source rate relative to simply coding the full-resolution video. However, the robustness provided is argued to outweigh the overhead cost. Scalable coding provides natural layers which are different than those given previously in that the base layer is decodable and subsequent layers improve the general quality of the decoded video. Base layer data itself, however, can also be layered.

As examples in the standards, MPEG-2 and H.264 provide for *spatial scalability* (providing video of differing frame sizes), *temporal scalability* (providing video of differing frame rates), and *quality* or *PSNR scalability* (providing video of differing quality).

### 4.3 Robust Entropy Coding

Most source coders employ a final entropy coding stage to further reduce the data rate. Transmission errors can cause catastrophic decoder errors when data has been encoded with a variable length code (VLC). Even a single bit error left uncorrected by the channel code can render the remainder of the bitstream useless. One way to ensure that random bit or burst errors will not catastrophically affect decoding of the VLC through loss of synchronization is to use fixed-length rather than variable-length codes, but this is often at the expense of compression efficiency. Perhaps the simplest technique to deal with errors in VLC streams is to employ *resynchronization flags* which are assigned to a source symbol that serves as a positional marker and whose reception insures the correct placement of subsequently decoded data. Such flags are called restart markers or synchronizing codewords. Use at a shorter interval improves robustness but decreases compression efficiency.

More sophisticated techniques to provide robustness for VLC-coded data include both packetization strategies and specially designed VLCs. A packetization strategy to provide robustness is the *error-resilient entropy code* (EREC), which is applicable to block coding strategies in which the input signal is split into blocks which are coded as variable-length blocks of data; EREC produces negligible overhead [7]. *Reversible variable-length codes* are uniquely decodable both forward and backward and are useful both for error location and for maximizing the amount of decoded data [8]; they also incur negligible overhead and are included in the MPEG-4 standard. *Resynchronizing variable-length codes* allow rapid resynchronization following bit or burst errors and are formed by designing a resynchronizing Huffman code and then including a restart marker at the expense of slight non-optimality of the resulting codes [9]. Error correcting arithmetic codes have also been proposed [10]. These techniques provide negligible redundancy.

### 4.4 Joint Source-Channel Coding

Optimizing or adjusting the source coding for channel characteristics is known as joint source-channel coding. This can involve solely the source coding if no channel coding is used, or a joint distribution of resources (fractions of the channel bit-rate) between the source and channel coding. The latter is most commonly used and requires accurate models of the effects of channel errors on the decoded video quality; such models can be developed theoretically or empirically.

For example, in [11], bits are distributed among source and channel coding of subbands in a 3-D subband decomposition such that the expected distortion is minimized; UEP is also applied to the subbands. More recently, Wyner–Ziv coding (i.e., lossy compression with receiver side information) has been proposed as an approach to joint source-channel coding [12].

### 4.5 Multiple Description Coding

Multiple description (MD) coding is a relatively new approach to practical video coding. In its most common instantiation, MD coding is a dual to layered coding: multiple representations of the same data (descriptions) are formed such that any single description provides a decodable stream, and reception of more descriptions increases the quality of the decoded stream. (Note, however, that MD coding does not require that the single descriptions are themselves of equal quality.) The extreme redundancy in video suggests many techniques for forming single descriptions. A challenge in providing MD coding of video is in synchronization of any prediction loops in the individual descriptions. A thorough review of this issue and others in MD video coding is given in [13].

### 4.6 Error Concealment

In spite of source and channel coding strategies to provide robust video transmission, errant data will inevitably appear at the decoder input. Error concealment (obviously along with error detection) attempts to mitigate the effects by reconstructing lost or clearly incorrect data. Both signal processing techniques such as spatial or temporal interpolation and decoding strategies to select the most likely data from a finite set are employed. The success of both of these techniques can be facilitated by the strategic addition of redundancy in the source coding either through layered coding or through joint source-channel coding.

## 5. APPROACHES IN THE SPECIAL SESSION

The papers in the special provide several examples of use of multiple strategies to create a robust video coder or to analyze robust coding.

Chiang et al. propose adding redundancy to the source coding by using oversampled filter banks, and then exploiting this redundancy in error correction. A joint source-channel approach is employed to encode the filter bank outputs, which requires knowledge of the channel BER. Layered coding is a part of their solution, as headers and motion vectors are assumed to be perfectly received via use of strong channel codes. Additional robustness to bit errors is provided by avoiding entropy codes.

Heng et al. also add redundancy to the source coding by providing an adaptive multiple description technique for lossy packet networks in the context of an H.264-like coder. Mode selection on a macroblock basis is performed in a rate-distortion optimized manner in which the encoder estimates the distortion induced by both source coding and transmission errors. The optimization requires an expected burst length for packet losses, and strategic packetization is employed to minimize the effects of packet losses.

Petrisor et al. explore a specific approach to adding redundancy at the source coding level, by investigating techniques for oversampling spatio-temporal wavelet decompositions.

Scaglione and van der Schaar assume a layered coding strategy and develop a framework to minimize the total distortion for all users subject to a capacity constraint. The analysis assumes known fading parameters and that strong channel coding eliminates all bit errors within a packet.

Lastly, Flierl and Vandergheynst investigate not robust coding but coding of video signals in the presence of correlated side information, as might be obtained from cameras providing two views of the same scene.

These examples demonstrate how multiple strategies can be simultaneously employed to provide robust video coding in specific transmission environments.

## 6. FUTURE DIRECTIONS

Robustness *for transmission* has been discussed in this paper and session, and is by far the most commonly addressed form of robustness in the video coding literature today. However, video coding to provide robustness to other variations should also be explored. For example, robustness in the context of *source content* can address efficient and appropriate coding of unexpected, highly non-stationary data in the video stream. Robustness in the context of the *user* should consider end-user-specific characteristics such as viewing conditions, decoder configurations, and desired quality. Many realistic video transmission systems must deal with these issues, and a methodical and intelligent approach may push further developments in general video coding.

While robust video coding has been an active research topic for many years and remains one, it is worthwhile to consider that commercial "robust video coding" is performed regularly today in many forms. Digital satellite television broadcasters provide MPEG-2 to their subscribers without error *most of the time,* and the without-error signal is of excellent quality. However, subscribers to satellite television services are familiar with decoded video quality degradations during severe weather. Users of streaming video over high-traffic internet connections see the frame-dropping effects of delays and congestion. Both of these applications have been designed with channel degradations in mind, but when the transmission channel delays and/or errors exceed the design points, difficulties in decoding the received video signals still ensue. However, the designers and users of these systems accept the occasional decoder error or system failure. These applications provide examples of application-appropriate quality, in which quality is defined in terms of probability of outage, rather than average PSNR. Such a quality definition is in sharp contrast to the quality measures employed in most literature, typically measuring average PSNR or PSNR vs. time. A reconsideration of the community's perspective on appropriate quality measures may be beneficial.

To conclude, robust video coding has provided a tantalizing research problem for over a decade, and in spite of its age as a problem, innovative techniques continue to be developed. With the continuing consumer demand for complete multimedia connectivity, the need for efficient, robust video coding will continue to drive new research.

## REFERENCES


[1] J. D. Villasenor, Y.-Q. Zhang, J. Wen, "Robust video coding algorithms and systems," *Proc. IEEE*, Vol. 87, No. 10, pp. 1724-33, October 1999.

[2] Yao Wang, Qin-Fan Zhu, "Error control and concealment for video communication: a review," *Proc. IEEE*, Vol. 86, No. 5, pp. 974-97, May 1998.

[3] "Error resilient image and video transmission," special issue of *IEEE Journal on Selected Areas in Communications*, ed. C. W. Chen, P. Cosman, N. Kingsbury, J. Liang, J. W. Modestino, Vol 18, No. 6, June 2000.

[4] S. S. Hemami, "Robust image communication over wireless channels," *IEEE Communication Magazine*, Nov. 2001.

[5] M. Masry and S. S. Hemami, "CVQE: A metric for continuous video quality evaluation at low rates," in *SPIE Human Vision and Electronic Imaging*, Santa Clara, CA, Jan. 2003.

[6] M. Ghanbari, "Two layer coding of video signals for VBR networks," *IEEE Journal on Selected Areas in Communications,* Vol. 7, No. 5, pp. 771-81, June 1989.

[7] D. W. Redmill, N. G. Kingsbury, "The EREC: An error-resilient technique for coding variable-length blocks of data," *IEEE Trans. Image Processing*, Vol. 5, No. 4, April 1996, pp. 565-74.

[8] J. Wen, J. D. Villasenor, "Reversible variable-length codes for robust image and video transmission," *Proc. 31st Asilomar Conf. on Signals, Systems, and Computers*, Vol. 2, Nov. 1997, pp. 973-9.

[9] S. S. Hemami, "Robust image transmission using resynchronizing variable-length codes," *IEEE Journal on Selected Areas in Communications, Special Issue on Error Resilient Image and Video Transmission,* Vol. 18, No. 6, pp. 927-39, August 2000.

[10] M. Grangetto, E. Magli, G. Olmo, "Robust video transmission over error-prone channels via error correcting arithmetic codes," IEEE Communication Letters, Vol. 7, No. 12, pp. 596-8, December 2003.

[11] G. Cheung, A. Zakhor, "Bit allocation for joint source/channel coding of scalable video," *IEEE Trans. on Image Processing*, Vol. 9, No. 3, pp. 340-56, March 2000.

[12] B. Girod, A. M. Aaron, S. Rane, D. Rebollo-Monedero, "Distributed Video Coding," *Proc. IEEE*, Vol. 93, No. 1, pp. 71-83, January 2005.

[13] Y. Wang, A. R. Reibman, S. Lin, "Multiple Description Coding for Video Delivery," *Proc. IEEE*, Vol. 93, No. 1, pp. 57-70, January 2005.