Joint Source-Channel Content-based Multistream Video Coding Scheme

J. Chen and K. J. R. Liu
Department of Electrical Engineering, Univ. of Maryland at College Park

Abstract

In this paper, we propose a joint source-channel content-based multistream video coding scheme to combat the transmission errors. By coupling with our ground breaking innovation frequency-hopping OFDM at the physical layer, our proposed multistream design provides better error protection than the conventional design. Note that the frequency-hopping OFDM provides a significantly superior platform to all other existing technologies including the cdma2000 and WCDMA for the wireless services. Simulations demonstrate that our design outperforms those single stream layered design by $3 - 7dB$ under the harsh network conditions.

I. Introduction

Currently, the focus of IP transport continues to be the data-oriented via the packet switches while it used to be speech-oriented via the circuit switch infrastructures. With the advances in digital compression technology and the steady deployment of broadband networks such as the fiber optics, cable, xDSL and third-generation wireless CDMA systems, multimedia services such as the packets (data/voice/video over IP) through the broadband networks have been emerging as a new technologies for the new millennium. However, before we can realize the full potential of those multimedia services, we have to address the challenges of how to deliver the multimedia applications over networks cost effectively, ubiquitously, and with sufficient quality.

Due to the large variety of existing network technologies, it is most likely that the hybrid networks are used to support video services as shown in Fig. 1. Many error control and

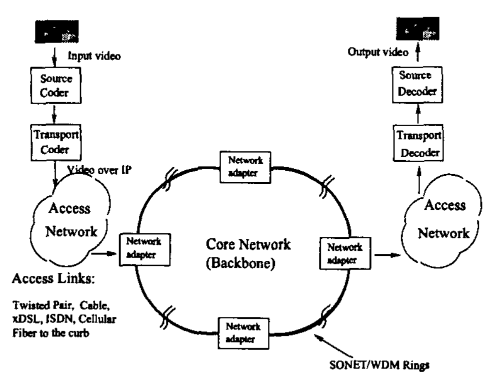


Figure 1: Hybrid networks needed to support the streaming video.

concealment techniques have been proposed for video communication such as *Error concealment by preprocessing* [1], *Error concealment by postprocessing* [2], and *Error concealment by interaction* [3]. However different networks have different characteristics. To optimize the performance of those multimedia systems with the given Quality of Service (QoS) requirements, we should *JOINTLY* consider the video compression and delivery schemes based on the network alternatives, capacities, and characteristics. For the multimedia services, the real-time/interactivity requirements introduce another level of difficulties in multimedia transmission because the intolerance of extra delay sometimes excludes the deployment of some well-known error-recovery techniques such as Automatic Repeat-Request (ARQ). In addition, issues such as the audio-visual synchronization and multipoint communications further complicate the problem of the error recovery. However, as the video segmentation techniques become more mature [4], it enables us to view, access, and manipulate video objects rather than the frame of pixels, which enables the great error robustness at a large range of bit rates. In this paper, we will address the challenge of delivering video over IP or *packet video* in a bandwidth efficient and error resilient manner. And, our contributions in this paper is:

- A joint source-channel content-based multistream video coding scheme built on top of our ground breaking innovation frequency-hopping OFDM to combat transmission errors under the harsh network conditions.

The remainder of the paper is organized as follows. In Section 2, we describe our joint source-channel content-based multistream video coding scheme to combat transmission errors. The simulation results are then revealed in Section 3. Finally, we conclude our work in Section 4.

II. Design Details

Our frequency-hopping OFDM is designed specifically for the delivery of advanced Internet services. This delivery is enabled by the vertical integration of layers 1 (physical layer) to 3 (network layer), while layers 3 and above are purely IP-based infrastructures to allow internetworking with the peer networks and applications using the standard IP protocols. In OFDM [5], the spectrum is divided into a number of equally spaced subcarriers or tones, and each carries a portion of a user's information. The OFDM tones enjoy the property of being orthogonal, in that the individual tones do not interfere with each other. In contrast, the direct sequence CDMA codes used in the 2G/3G systems do not have this nice property, and are not orthogonal under the multipath. Our system couples the inherent advantages of OFDM with Frequency-Hopping Spread-Spectrum, as well as a jointly source-channel design of multistream video coding that is highly efficient for the data communications and its usage of the spectrum. Our design advantages are summarized as following:

- Packet-over-the-air architecture that is ideal for the IP-based communications,

- Frequency re-use of one and no intra-cell interference. In addition, the fast hopping minimizes inter-cell interference,
- Frequency diversity and multistream video coding reduce the impact of fading and improve the quality of service,
- Low signaling overhead using compressed PPP (Point-to-Point Protocol) [6] header leads to high efficiency and system throughout.

The following paragraphs describe the design in more detail.

A. Physical Layer Design

At the physical layer, a fast frequency hopped version of Orthogonal Frequency Division Multiplex (OFDM) is employed. Fast hopping turns our design into a spread spectrum technology that maintains CDMA's average interference between cells and frequency re-use of one. At the same time, it has no in-cell interference (same as the TDMA system), making the overall system superior since it combines the most significant advantages of both its predecessors. The frequency-hopping OFDM system is designed for the paired frequency division duplex (FDD) operation and supports the spectrum allocations in the multiples of 1.25 or 5 MHz bandwidth. The system is frequency-independent and supports the practical mobile applications in the range of frequencies between 220 MHz and 3.5 GHz. In our simulation, we target in the 700-900 MHz band, which is the license-free spectrum. In each cell, the same spectrum is divided into a

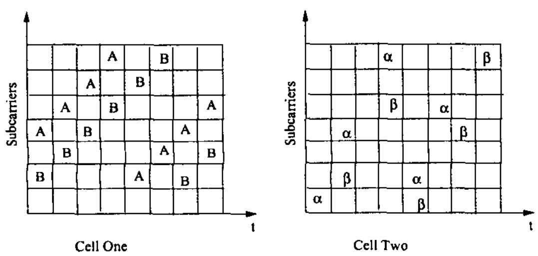


Figure 2: The Physical layer design.

set of subcarriers, and time is divided into the symbols as shown in Fig. 2. A number of those subcarriers are assigned to each user when there is data to send or receive. The subcarriers that comprise a user's channel hop over the entire 5 MHz or 1.25 MHz band as time goes by: the tone pattern in time/frequency space is referred to as the hopping sequence. Tone hopping provides frequency diversity, which helps reduce the effects of fading. The salient feature of this physical layer is the absence of in-cell interference, due to the orthogonality of user tones being preserved even in the presence of multipath. Inter-cell interference is caused by the tones being re-used from cell to cell. However, this interference is averaged across cells, since user tones employ fast hopping. For example, the frequency-time squares labeled 'A' in the cell one is rarely transmitted in the same frequency-time squares labeled 'α' in the cell two. These nice properties of averaged out-of-cell interference and zero in-cell interference lead to improved the physical layer spectral efficiency (three times that of the wideband CDMA), and more efficient support of bursty traffic for packet-over-the-air architecture.

B. The Layer 3 and above Design

By taking advantage of the content-based video coding, here we propose a joint source-channel multistream video coding technique to combat the transmission errors. In principle, the video objects are encoded into the different IP video streams (multistream) as shown in Fig. 3. Each video

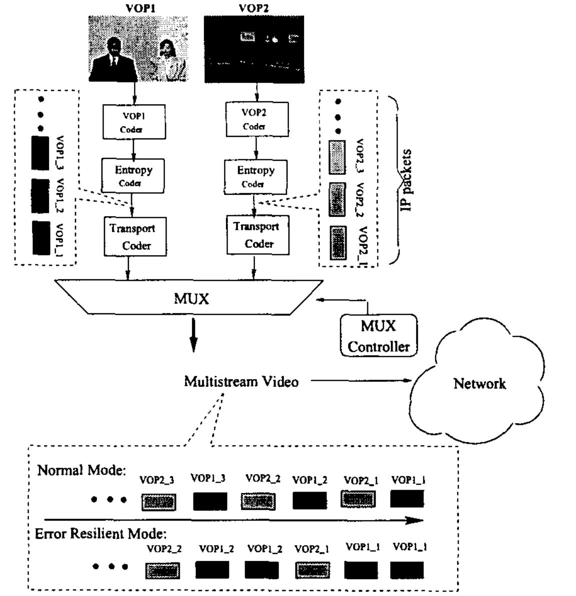


Figure 3: Illustration of multistream video over IP (Here the drawing is not proportional to the real packet size).

stream is encoded differently based on its perceptual importance. The perceptual importance of a video sequence can be determined based on the outputs of segmentation [4]. For instance, in "Akiyo" test sequence as shown in Fig. 6, we simply refer the *primary video stream* to the foreground video object in the sequence. The transport coder with the *"transport prioritization"* used in Fig. 3 refers to an ensemble of devices performing channel coding, packetization and/or modulation. In other words, the term transport prioritization here refers to the various mechanisms to provide different quality for different video streams in transport, including using the unequal error protection [7], and assigning different priorities to different video streams. Therefore, the router will make the best effort to deliver those high priority packets associated with the primary video stream under the harsh network conditions. The *MUX controller* in Fig. 3 controls whether those video packets are transmitted in the *normal mode* or in the *error resilient mode* based on the network conditions. Here the feedback information about the network condition can be obtained by using the delay and loss-rate statistics at the decoder [8], [9]. Under the *normal mode*, those IP packets are multiplexed and sent out alternatively. When the network condition deteriorates then *the MUX controller* switches to *error resilient mode*, and the

IP packets of primary video stream are **repeated once** and then sent out as shown in Fig. 3. The rest of secondary video streams are coarsely quantized, less error protected coded and transmitted at the low priority. Sometimes, we can even stop transmitting those secondary video information to give the resource to the primary video stream.

This error resilient scheme works based on the fact that there are several parallel paths between the source and destination such as the wireless multihop network or the multipath fading in mobile wireless channels. The probability that all paths simultaneously experience losses is small. Even when only one single physical path exists between the source and destination, the path can be divided into several virtual channels by using the time interleaving scheme as shown in Fig. 3. Under the adverse channel conditions, the decoder therefore can still reconstruct the original video sequence depending on the stream which is received correctly or least distorted. Unlike the conventional multiple-description coding [10], we apply the content-based video coding instead of the layered coding (In the layered video coding approach, we treat all video objects within a frame equally important and allocate the bandwidth evenly to transmit those coded information). In our content-based video coding, the video information is partitioned into more than one video object and encoded into multiple different video streams with different transport priorities. Under the harsh network condition, we allocate more resource to error protect those important video information. In principle, we trade the quality of the least significant video objects for the quality of the most significant ones. As a result, we can reconstruct a close approximation of the original video (at least for the primary video object) and make the output video at the decoder least objectionable to human eyes.

Head Compression

Our video over IP packet is shown in Fig. 4, which includes both the payload and PPP header. In contrast to

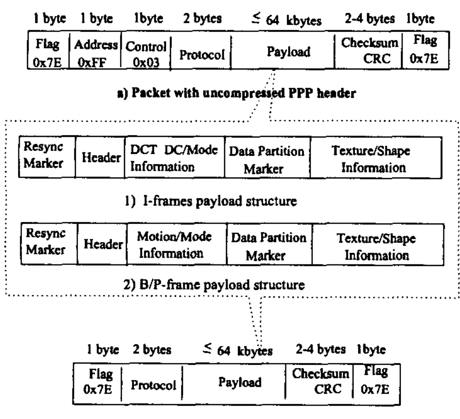


Figure 4: Compressed vs. uncompressed IP packet.

the uncompressed IP packet as in Fig. 4 (a), we propose the compressed IP packet as in Fig. 4 (b) to save the bandwidth for video transmission. The difference between these two formats is that the address (0xFF) and control (0x03) are eliminated from the PPP header in the compressed IP packet. Based on the recent studies of data traffics over Internet [11], nearly half of data streams have the packet size of 40 to 44 bytes. With the average IP packet size of 40 bytes, we can save up to 5% of bandwidth using the compressed PPP header. The reason why we can use the compressed IP packet is as following: the PPP supports the multiprotocol encapsulation and we can classify the incoming data stream into two groups: the *control signals* and *real datum*. Those two types of traffics can be distinguished based on the protocol numbers embedded in the PPP header. For the control signals, which are transmitted prior to the real data traffic, are not allowed to strip those address and control fields off to save bandwidth because the information is critical for establishing the connections and not retrievable once they are discarded. Once the connections established, the real datum such as video bit-stream can flow through the networks. For those IP packet, the control and address fields are no longer important. Therefore, we can periodically strip those fields off to save the bandwidth.

III. Simulation Results

To test our joint source-channel content-based multistream video coding technique built on top of our frequency-hopping OFDM to combat the transmission errors, we ran several experiments using the MPEG-4 test sequences at both the CIF and QCIF resolutions (Each sequence contains 300 video frames). In our simulation setup as shown in Fig. 5, the

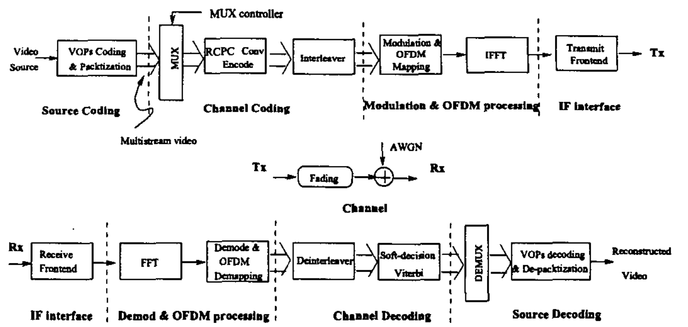
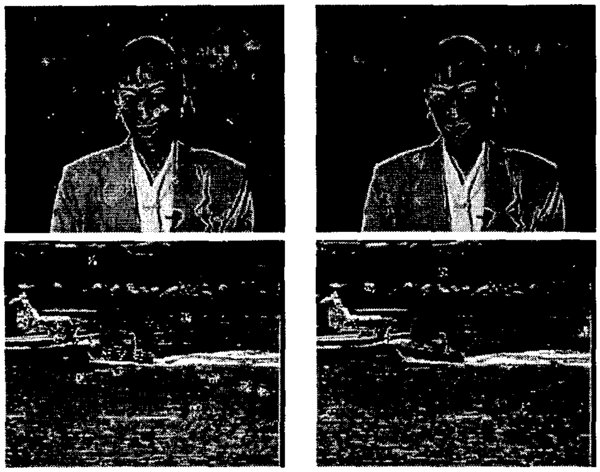


Figure 5: The simulation setup to test the performance our multistream video coding scheme for the wireless multimedia communication.

quantization parameters and channel coding rates are chosen so that the combined output rate of video streams is constant (around 128 kb/s). The rate-compatible punctured convolutional (RCPC) [12] code is employed in our simulations to provide an efficient means of implementing a variable-rate error control for different video streams so that only a single encoder/decoder pair is needed. In addition to that, our fast-hopping OFDM is used for the modulation and it achieves the robustness against the frequency selective fading or narrowband interference. At the decoder side, the soft-decision Viterbi decoding scheme is utilized to correct the erroneous subcarriers. Our simulations are performed under both the added white Gaussian noise (AWGN) and

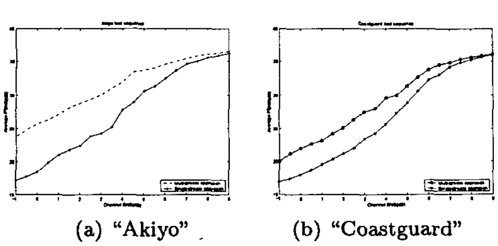


(a) single stream (b) multistream

Figure 6: Transmission errors affect the video quality under different coding schemes.

fading channel conditions (Here we adopt the Rayleigh fading channel to model the bursty channel condition caused by multipath fading [13]).

The transmission errors affect the video quality by using different coding schemes as shown in Fig. 6. For simple cases such as "Akiyo" and "Mother and Daughter" test sequences in which the foreground scenes change while the background scenes remain the same, we allocate all resource under the harsh network condition to transmit the primary video object because it dominates the whole video scene. Our multistream works extremely well for those simple test cases. For the complicate cases such as "Coastguard" test sequence in which both the foreground and background scenes change, we allocate more resource to transmit the primary video object than the secondary ones as network condition deteriorates. Compared to the conventional layered coding mechanism, our multistream works better in terms of error resilience as shown in Fig. 7. From the simulation results,



(a) "Akiyo" (b) "Coastguard"

Figure 7: Comparison of proposed multistream design to the single stream design in terms of Signal to Noise ratio.

we observe that the multistream approach outperforms the single stream approach by $3-5$ dB under noisy channel conditions. Under the multipath fading condition (we use the carrier frequency of 900 MHz and the mobile velocity of $10-60$ miles/hour with upto 6 propagation paths), the simulation demonstrates that our multistream works better than the single stream design by up to $5-7$ dB with the data throughput rate at $128kb/s$. As a result, the decoder can reconstruct a better original video sequence depending on the stream which is received correctly or least distorted by taking the advantage of multipath between source and destination.

IV. Conclusion

In this paper, we propose a joint source-channel content-based multistream video coding scheme to combat the transmission errors for the delay-sensitive multimedia services. Simulations demonstrate that our multistream design outperforms those single stream layered design in terms of the error resilience under the harsh network conditions because, in our design, the IP packets of primary video stream are repeated once and then sent out under harsh network condition. And, the dual description of primary video stream independently travels through the different fading paths. Since the chance that all paths simultaneously experience information losses is small, the sum of the signal level at the receiver antenna in our multistream design is, therefore, stronger than that in the single stream design thus better reconstructed video quality. In addition, our frequency-hopping OFDM eliminates the intra-cell interface and averages the inter-cell interface. Overall, our design provides a robust airlink and the packet-over-the-air design is ideal for the IP-based communications.

References

[1] Y-Q Zhang, Y. J. Liu, and R. L. Pickholtz, "Layered image transmission over cellular radio channels", *IEEE Trans. Veh. Technology*, vol. 43, no. 8, pp. 786–796, Aug. 1994.

[2] Q-F Zhu, Y. Wang, and L. Shaw, "Coding and cell loss recovery for DCT-based packet video", *IEEE Trans. Circuits and Systems for Video Technology*, vol. 3, no. 6, pp. 248–258, June 1993.

[3] M. Ghanbari, "Postprocessing of late cells for packet video", *IEEE Trans. Circuits and Systems for Video Technology*, vol. 6, no. 12, pp. 669–678, Dec. 1996.

[4] P. Salembier and F. Marques, "Region-based representations of image and video: Segmentation tools for multimedia services", *IEEE Trans. Circuits and Systems for Video Technology*, vol. 9, no. 8, pp. 1147–1169, Dec. 1999.

[5] R. V. Nee and R. Prasad, *OFDM for wireless multimedia communications*, Artech House Publishers, 1994.

[6] J. Carlson, *PPP Design and Debugging*, Addison Wesley, 1998.

[7] ISO/IEC 13818-2, *Information technology - generic coding of moving pictures and associated audio information: video*, ISO/IEC, Geneve Switzerland, 1996, MPEG2.

[8] S. Cen, C. Pu, R. Staehli, C. Cowan, and J. Walpole, "A distributed real-time MPEG video audio player", in *Proc. 5th Int. Workshop Network and Operating System Support for Digital Audio and Video*, Durham, NH, Apr. 1995, pp. 151–162.

[9] Z. Chen, S-M Tan, R. H. Campbell, and Y Li, "Real time video and audio in the world wide web", in *Proc. 4th World Wide Web Conf.*, Oct. 1995, pp. 326–329.

[10] Y. Wang, M. T. Orchard, and A. R. Reibman, "Multiple description image coding for noisy channels by pairing transform coefficients", in *Proc. IEEE 1st Workshop Multimedia Signal Processing*, Princeton, NJ, Mar. 1997, pp. 419–424.

[11] *Packet Size Distributions*, National Lab for Application Network Research, 1998.

[12] J. Hagenauer, "Rate-compatible punctured convolutional codes (RCPC codes) and their applications", *IEEE Trans. Communications*, vol. 36, no. 4, pp. 389–400, Apr. 1988.

[13] W. C. Jakes, *Microwave mobile communications*, IEEE Press, 1994.