MULTILEVEL CODING OF BROADCAST VIDEO OVER WIRELESS CHANNELS

Željko Čakareski, Nasir Ahmed, Aditya Dhar, and Behnaam Aazhang

Center for Multimedia Communication, Rice University, Houston, TX, 77005

ABSTRACT

This paper addresses the problem of broadcasting video over wireless channels. In such a scenario, different end users perceive the video signal through channels with different quality. We propose a source-channel coding system, that partitions the video data in such a way that maintains a baseline quality of service even for the user with the worst channel. The system consists of a layered source representation that is robust to error propagation effects, combined with a multilevel channel coding scheme that provides an unequal error protection to the different layers of the encoded video. Experimental results show that our system is robust and guarantees a baseline quality of service to all users over a wide range of channel conditions.

1. INTRODUCTION

This work addresses the problem of video communication over broadcast wireless channels. The problem under consideration is illustrated in Figure 1. This is a typical scenario found in many practical cases such as the multichannel and the local multipoint distributions services (MMDS and LMDS), which due to the considerable cost-reduction efforts may result in making wireless "last-mile" access technologies viable alternatives to always-on Internet broadband access technologies such as cable modems and digital subscriber lines [1].

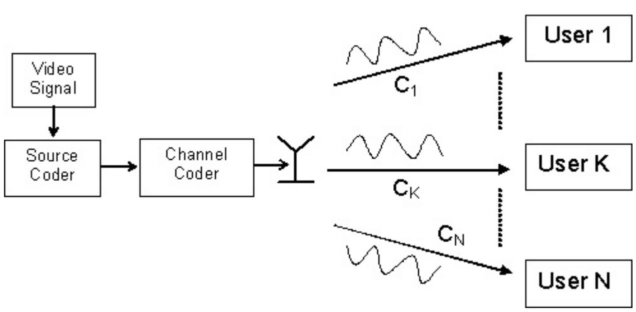

Fig. 1. A typical broadcast scenario.

In such a scenario, different users perceive the video signal through channels with different qualities. For example, a user nearby the local basestation might have an excellent reception of the transmitted signal, whereas the "far-away" user might experience a poor one. Hence, given the same broadcast service the users are provided with a different quality of service. Furthermore, there is no guarantee that each user will be able to maintain a baseline reproduction quality of the received video signal. To elaborate more on this, let R_0 be the transmission rate (in bps) at which information is being broadcasted and let C_i denote the throughput capacity (in bps) for a given SNR of the channel for user $i = 1, \ldots, N$, where N is the total number of users. Then for all users i such that $C_i < R_0$ the transmitted signal will be degraded by the channel due to the insufficient bandwidth. Thus, even though all of the transmitted information would be received, it would still be practically useless.

Due to its ubiquitous nature, the problem under consideration is clearly important and the above example motivates the need for a video broadcasting scheme that is more adaptive to the varying channel conditions at each user and provides flexibility to the users to decide how much of the transmitted information they could in fact decode and use subsequently, based on their channel conditions. In other words, each user should be given the opportunity to select an appropriate amount of the transmitted information, based on its channel state, that would still provide an acceptable degree of reproduction quality after decoding.

As a solution we propose a novel multilevel source-chan nel coding framework that is intuitive and as our results show, guarantees a baseline quality of service to every user. The framework consists of a source coding part, which provides a layered representation of the video signal, and a channel coding part based on the concept of multilevel codes, that provides unequal error protection (UEP) to the different layers of the video. Due to the fact that a broadcast environment is considered, the possibility of using other channel coding schemes, such as rate compatible punctured convolutional (RCPC) codes [2], for achieving UEP within our framework is excluded. Namely in a broadcast scenario, most of the coding complexity and decision making tend to be transferred to the receiver's end, since the transmission part is fixed and same for all users. Furthermore, no feedback is typically allowed in broadcasting, which is essential in RCPC applications for UEP.

Using the proposed framework, rather than broadcasting a single video bitstream to all the users with a rate R_0, our system broadcasts multiple, L, bitstreams each with a different rate R_j for $j = 1, \ldots, L$ such that:

$$R_0 = \sum_{j=1}^{L} R_j \tag{1}$$

Each user can then subscribe to one or more video bitstreams, based on its channel quality, such that:

$$\sum_{j \in S} R_j \leq R_0 \tag{2}$$

where S is the set of integer indices of bitstreams to which the user is subscribed. The broadcast system is designed such that all of the users could decode at least a single specially designed bitstream (henceforth denoted baselayer), that provides a baseline quality of service at each user. Naturally, the users with better channels will try to decode more than one bitstream (base + enhancement layers) and hence maximize the reproduction quality of the received video at their end.

To our knowledge, the most closely related contemporaneous works are [4, 5]. In [4] the authors present a scalable source representation, based on the spatio-temporal resolution pyramid, for robust video transmission over wireless channels using two priority classes. RCPC codes are used to provide UEP for the two video layers. In [5] the authors present a dual-rate H.261 video coder that adapts its coding rate based on the current conditions of the wireless channel. An ARQ (Automatic Repeat reQuest) error control technique is employed to retransmit erroneous data frames. Both of these solutions however, would fail to meet the desired requirements from a video broadcasting scheme discussed earlier, due to issues such as feedback and error propagation in predictive source encoding.

2. SYSTEM DESCRIPTION

Here we describe the proposed video broadcasting system and explain in details some of its constituent components. The block diagram of the end-to-end system is shown in Figure 2. In terms of the signal processing, the system can be seen as consisting of two sections: one done at the base station and the other at each receiver.

The video signal to be broadcasted is first encoded using a layered source encoding scheme. The partitioning into layers of the video signal is adapted to the characteristics of the broadcast channel, as explained later. The layered video is then passed on to a multilevel channel encoding block, which provides unequal error protection to the different layers. Finally, the channel encoded data is broadcasted by the basestation. At the other end, each user receives the transmitted signal with a certain amount of degradation, dependent on the characteristics of its channel (see Figure 1). The received signal is passed first through a multistage decoding block, which tries to correct for the channel effects. The layers of data output by the multistage decoder are then passed on to a source decoder which combines the different layers into a single data stream. This is the video signal observed by the user.

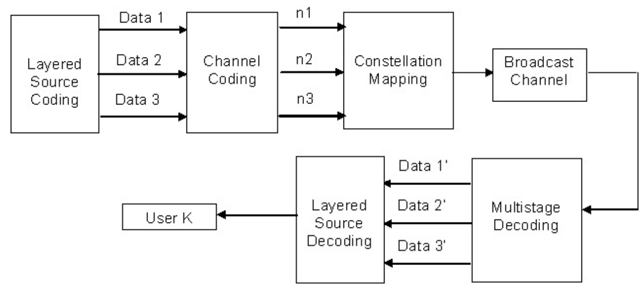

Fig. 2. Layered multilevel source-channel coding system.

2.1. Layered Source Coding

Our source coding framework consists of two constituent coding modules. The first one is a predictive motion compensation based video coder, compatible with the popular standard for low bit rate video coding MPEG-4 [7], used here for encoding the base layer. The second is based on a non-predictive scalable image coder, compatible with the popular SPIHT [8] coder of Said and Pearlman. The second module is used to code the enhancement layers. We decided to use two source coding modules in order to be able to exploit the individual advantages, as well as to avoid the individual drawbacks, of each of the coding paradigms they offer: predictive and non-predictive source coding. The concept of combining predictive and non-predictive coding of video has been studied first in [3] and is now part of the latest proposal for modifications of the MPEG-4 standard, to be adopted in April 2002. It has been shown that such combined coding systems perform better than the standard MPEG-4 predictive coders in terms of the end-to-end distortion when transmission over unreliable channels is considered, without losing much of the compression efficiency of the MPEG-4 coders. Due to the progressive nature of the SPIHT coder, the combined coding approach offers fine levels of scalability and thus has been termed fine granular scalability (FGS) [3].

Our source coding block is shown in Figure 3. For reasons of illustration only a single enhancement layer is shown. Given the original video, the baselayer is encoded first, at a rate equivalent to the estimate of the minimum bandwidth R_{min} available in the broadcast channel, as seen from the figure. Hence all the users will be able to decode at least the baselayer. Furthermore, the rest of the available bandwidth in the channel, $R_{max} - R_{min}$, is assigned to encode the only enhancement layer. Note that in practice, the partition of the variation in the bandwidth could be finer, i.e. more enhancement layers could be created.

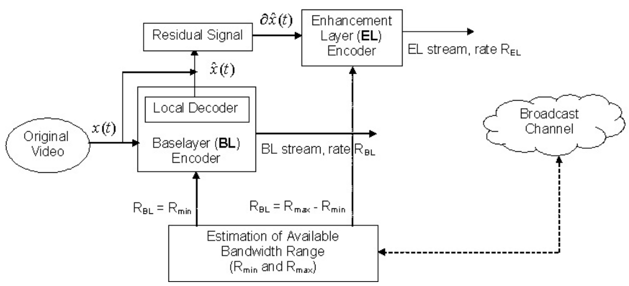

Fig. 3. Layered source encoder with a single enhancement layer.

2.2. Multilevel Channel Coding

Multilevel channel coding (MLC) using binary codes was first introduced by Imai et al. [9]. In an MLC scheme the input data to the channel is partitioned into multiple L levels which are channel coded separately. In our case, the data levels are in fact the video layers provided by the source encoder. Even or uneven channel code rates R_i could be applied to each of the layers. The coded data from each layer is then combined to map onto a constellation point which is transmitted over the channel. The method by which this is accomplished is denoted the labeling strategy.

In MLC, assuming equal coding rates R_i, UEP is achieved through the labeling strategy. The use of equal coding rates allows for simplicity at the decoder and more importantly in no error propagation in the multi-stage decoding module when block partitioning is used [6]. Our implementation uses equal coding rates and block partitioning.

To explain this labeling strategy, let assume first that only one bit is taken from each layer to select a constellation point (symbol) in the complex plane R^2. Hence the constellation size is 2^L. The bit taken from the first layer divides the constellation into two subsets. Based on its value $\{0, 1\}$, one of the subsets is selected. The bit from the second layer subdivides further the subset chosen by the first bit into two smaller subsubsets. Again, based on its value $\{0, 1\}$ one of the resulting subsubsets is selected. This dyadic partitioning proceeds further until the bit from last layer selects a symbol from a subset that consists of only two constellation points. This symbol would then be sent through the channel, and it has an L-bit label. Generalization of this scheme would be to use n_i bits from each video layer to select a constellation point. Note that then at the i^{th} partition, the n_i bits would further subdivide the subset selected by the $(i-1)^{th}$ partition, into 2^{n_i} smaller subsets. Let d_{min_i} denote the minimum Euclidean distance between subsets at the i^{th} partition level. Due to the construction, it is obvious that $d_{min_i} > d_{min_{i+1}}, \forall i$. This means that when a decision is made at the receiver as to what constellation point was transmitted, it is less likely to mistake two subsets at partition level i, than it is with subsets at partition level $i + 1$. Therefore, the data from video layer i would be more protected than the data from video layer $i + 1$ and hence UEP is achieved.

Each partition level in an MLC scheme can be thought of as an equivalent channel with information capacity C_i [6]. For the block-partitioning the following could be said regarding the information capacity at each level:

$$C_i \geq C_{i+1}, \qquad i = 1, \ldots, L-1. \qquad (3)$$

For a given user and SNR, the user can decode all L levels of the received signal if $C_i \geq R_i$ for $\forall i$. However, if one of the rates R_i exceeds the capacity C_i and $C_k \geq R_k, k < i$, then the user can decode error free only the first $i - 1$ levels, due to the equality of the coding rates and (3). This offers a certain degree of adaptivity at decoding. In other words using the SNR of the received signal, an approximation to the capacity can be made for the equivalent channels, and the receiver can decide how many of the levels to decode. This concept fits well within our framework since the data is simply broadcasted from the transmitter, and it is left to the receiver to decide how much of the noise corrupted data is indeed reliable.

3. EXPERIMENTAL RESULTS

We tested the end-to-end performance of our system through simulations. As a performance criterion we chose the luminance (Y) peak SNR averaged over all frames of the received video. The PSNR for a frame i is calculated as: $PSNR_i = 10log_{10}\frac{255^2}{MSE_i}$ where MSE_i is the mean squared error (MSE) between the original and the received frame i. Furthermore, the received video was subjected to visual tests for perceptual evaluation. A standard image sequence, the NEWS, was used in the simulations. The video consisted of 100 frames of 176x144 (QCIF) pixels spatial resolution and a temporal resolution of 10 frames/second.

In our *multilayer* system 3 video layers, a base and two enhancements (EN), were encoded each with data rate of 48 Kbps. A rate 1/2 convolutional code was used in the MLC scheme for all 3 layers. The performance of our system was compared with a standard *single layer* MPEG-4 system, encoded at data rate of 144 Kbps. Rate 1/2 convolutional code was also used here for channel protection. The PSNR performance of both systems is shown in Figure 4.

It can be seen that for good channel conditions (high SNR's), the single layer system provides a better performance, with a difference in PSNR of 3-4 dB. This is ex-

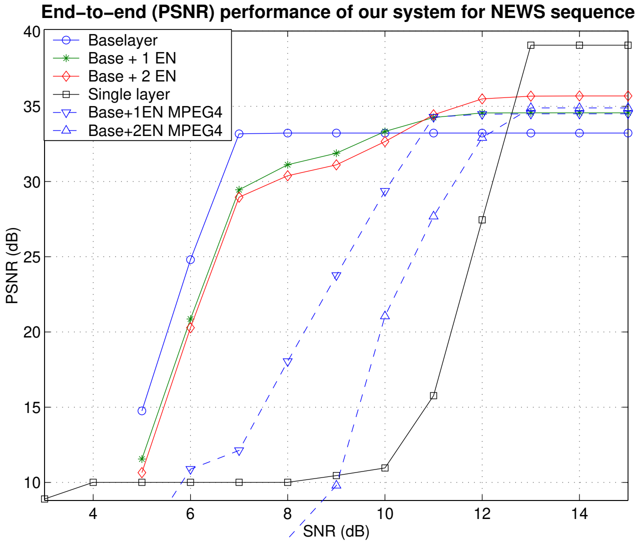

Fig. 4. End-to-end PSNR performance of our system (for $L = 3$).

pected since by layering the video, additional distortion is introduced into the video signal. However, as the channel starts deteriorationg (SNR $\leq$ 13 dB), the performance of the single layer case drops down abruptly and the received video simply becomes unusable. This is due to the error propagation effect present in MPEG-4 based systems, which was discussed earlier. On the other side, the multilayer system exhibits robust performance and provides a baseline reproduction quality even at low SNR's.

The baselayer displays a steady performance over a wide range of channel conditions due to the increased amount of protection against channel errors provided by the MLC. Due to the fact that for SNR values $\leq$ 10 dB the capacities of the equivalent channels introduced in Section 2.2 satisfy: $C_1 \geq R_1$, $C_2 < R_2$ and $C_3 < R_3$, decoding more than the baselayer provides a worse performance, compared to the case of decoding the baselayer only, as seen from the figure. On the other side, for SNR $\in$ (10,11] dB, it holds that $C_1 \geq R_1$, $C_2 \geq R_2$ and $C_3 < R_3$, so decoding base+1EN layer also make sense. Finally, for SNR $>$ 11 dB, decoding even the base+2EN layers is viable, since $C_3 \geq R_3$. Furthermore, due to the non-predictive encoding of the enhancement layers, the performance of the base+1EN and base+2EN layers in our system has significantly improved, for the same channel coding scheme (MLC), compared to their counterparts (Base+1(2)EN MPEG4, with dashed line in the figure) in a system where standard MPEG4 encoding is employed for all 3 layers. This is again due to the error propagation effect.

Finally, it should also be pointed out that for high SNR's the performance of our system does not improve for more than 1 dB for every additional enhancement layer used, as observed from the Figure 4. This is again due to the layering of the video signal, which introduces additional distortion. Nonetheless, the use of additional enhancement layers makes a significant difference in the perceptual quality of the received video as verified in our experiments.

4. CONCLUSIONS

In this work, we have considered the problem of video communication over a broadcast wireless channel. A new layered source/channel coding scheme has been proposed, which was shown to be robust and guarantee a baseline reproduction quality of the broadcasted video at every receiver. The user is provided with the flexibility to decide what would be the appropriate amount of information to decode from the received packets, based on an estimate of its channel conditions.

5. REFERENCES

[1] L. Wirbel, *LMDS, MMDS race for low-cost implementation*, EE Times Online edition, 30 November 1999.

[2] J. Hagenauer, *Rate-compatible punctured convolutional codes (RCPC codes) and their applications*, IEEE Trans. Commun., Vol. 36, No. 4, pp. 389-400, April 1988.

[3] H. Radha, Y. Chen, K. Parthasarathy, R. Cohen, *Scalable Internet video using MPEG-4*, Signal Processing: Image Commun., vol.15, no. 1-2, pp.95-126, Sep 1999.

[4] U. Horn, B. Girod, and B. Belzer, *Scalable video coding for multimedia applications and robust transmission over wireless channels*, 7-th International Workshop on Packet Video, Mar. 1996.

[5] M. Khansari, A. Jalali, E. Dubois, and P. Mermelstein, *Low Bit-Rate Video Trans. over Fading Channels for Wireless Microcellular Systems*, IEEE Trans. on Circ. and Syst. for Video Techn., vol. 6, no. 1, pp. 1-11, 1996.

[6] U. Wachsmann, R.F.H. Fischer, and J.B. Huber, *Multilevel Codes: Theoretical Concepts and Practical Design Rules*, IEEE Trans. on Info. Theory, July 1999.

[7] ISO/IEC JTC1/SC29/WG11, *MPEG4 video verification model*, 1997.

[8] A. Said and W. Pearlman, *A New, Fast, and Efficient Image Codec Based on Set Partitioning in Hierarchical Trees*, IEEE Trans. on Circ. and Syst. for Video Techn., 6(3):243-250, June 1996.

[9] H. Imai and S. Hirakawa, *A New Multilevel Coding Method using Error-Correcting Codes*, IEEE Trans. on Info. Theory, Vol. IT-23, No. 3, May 1977.