Content-Based Video Transmission Over Wireless Channels

Pankaj Batra and Shih-Fu Chang

Dept. of Electrical Engineering
and Center for Telecommunications Research
Columbia University, New York 10027
{pbatra, sfchang}@ctr.columbia.edu

Abstract

This paper describes new schemes for video transmission over wireless channels. Traditionally data was segmented using headers and data type (e.g. motion vectors, low and high frequency DCT coefficients). However, for wireless applications that have very low bandwidths available, a finer granularity of control may be essential. We propose content based approaches for further video segmentation. In particular, frame type, scene changes, and motion based procedures are used to provide finer level of control for data segmentation. We go on to argue that for transport over wireless channels, different video content requires different form of resources to be allocated to it. Joint source/channel coding techniques (particularly class/content/data dependent FEC/ARQ schemes) are used to do resource allocation after initial segmentation. FEC schemes are used to provide class/template dependent error robustness and ARQ techniques give delay control. An experimental simulation platform is used to test the objective (SNR) and subjective effectiveness of proposed algorithms. Based on our initial results from FEC control we argue that introduction of video content into video modeling allows us to define various classes of importance. We believe that by using such techniques we can get better visual quality under the very low bit rates available.

Keywords: wireless video, mobile-multimedia, content or object based video coding/segmentation.

1. Introduction

Lately, there has been a great demand for audio/visual services to be provided over wireless links. However, due to the severe bandwidth constraints, high error rates and time varying nature of these channels, the received video quality is often unacceptable. Transport of multimedia services over radio channels still remains an unsolved problem. Many new algorithms are needed to enable reliable, flexible, adaptive, and robust video communication in hostile wireless environments.

Traditional data filtering/segmentation methods have exploited the hierarchy in the coding structure. Both the H.26x and MPEGx suite of standards divide the coded data into many syntax layers. In MPEG2 (ISO/IEC 13818-2 Committee Draft), for example, the entire sequence is divided into group of pictures (GOPs). Each GOP has a fixed number of pictures/frames, and starts with an intra-coded I frame. I frames have no motion compensation performed on them. Also present are the P frames which are predictively coded from previous I and P frames. Between the anchor frames (I/P frames), are 2 bi-directionally predicted B frames. Each picture is further composed of slices, which in turn consist of macroblocks. 16×16 macroblocks consist of 4 8×8 blocks. Motion compensation is applied at a macroblock level, wherein a best matching block is found and a motion vector is sent in the bitstream.

Because of this hierarchy in coding, wireless errors affect video quality depending on where they hit. Errors occurring in headers (e.g. sequence startcodes, GOP headers, picture headers, slice headers) have the most detrimental effect. Next, errors in pictures depend on the kind of picture; the sensitivity being I>P>B. The perceived error sensitivity for picture data is typically motion vectors > low frequency DCT coefficients > high frequency DCT coefficients. Depending on the above, we can define a data segmentation template in decreasing order of priority (in terms of error robustness). The above is basically the philosophy behind the MPEG-2 data-partitioning method. Our approach differs primarily in that it uses video content [2,3] as a means of further data segmentation. We use an experimental simulation platform to test the effects of some video content features, particularly scene changes, frame type and motion. Incorporation of other artificial camera operations such as zooming, panning, camera motion, and object tracking are being investigated. A logical level figure (Figure 1) illustrating the idea is given below.

The remainder of this paper is organized as follows. Section 2 describes some previous approaches in this area. Section 3 explains the proposed approach being taken. In section 4 we present the simulation scheme and results. Section 5 concludes with suggestions for future work.

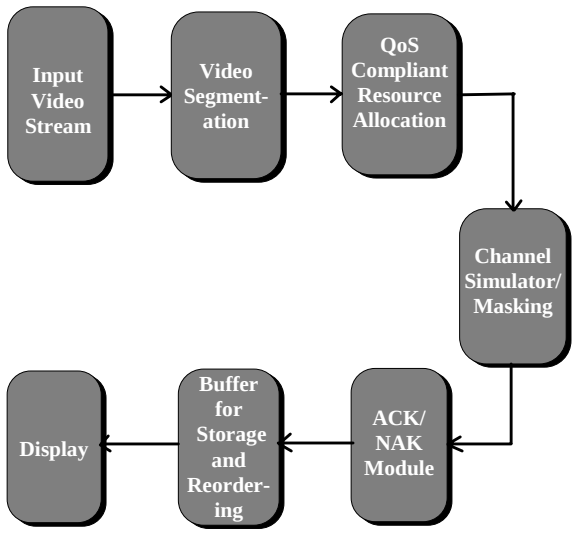


Figure 1. Logical diagram of Content-Based Wireless Video Transmission.

2 Previous Approaches

Earlier approaches on video over wireless have mainly concentrated on joint source-channel modeling [6], data-prioritization [1], optimal buffer allocation [5], and sub-stream level optimal power allocation [4] for wireless video transport. The layered transmission scheme presented in [6,4] addresses optimal power allocation to minimize end-to-end distortion over power constrained channels. In that scheme, a more important layer (e.g. a coarse resolution portion of multiresolution video [10]) is protected against channel error by allocating more power to it. A power control algorithm is presented in [4] that simultaneously minimizes interference and provides variable QoS contracts for different traffic types in a CDMA system. The algorithm uses substream power allocation and, in addition, adds or drops connections dynamically while ensuring that QoS specifications are satisfied. Some rate control schemes presented in [5] feed back channel information to the encoder. These are basically channel model based rate control schemes. Asynchronous video coding schemes presented in [10] use conditional replenishment based schemes that discard fine-resolution information (e.g. high frequency DCT coefficients). Their work will incorporate other known compression schemes such as vector quantization and motion compensation. Some conventional data partitioning schemes are given in [1] and effect of transmission errors is studied for H.263 codecs.

Current MPEG4 standardization efforts are trying to add the ability to quickly resynchronize or localize errors in a compressed video streams. Further schemes like duplicate information, two-way decode with reversible VLC, motion compensation using vector quantization (MCVQ) with Gray Code are being proposed [11]. System, syntax, and MUX level support will also be provided by work being done in MSDL Working Draft [12].

3. Proposed Approach

We use a human psychovisual basis for video segmentation. A typical video sequence can be divided into a set of independent scene shots. Each scene further has object motion, and global camera operations in it (Figure 2). We use these parameters to further segment video. This segmentation is applied in 'addition' to the conventional methods of segmentation described above.

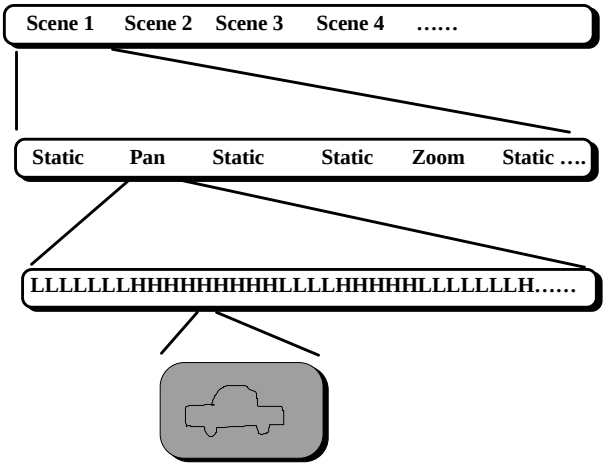


Figure 2. A typical video sequence.

We argue that different video content requires different form of resources to be allocated to it. The main QoS requirements that need to be satisfied for acceptable video quality are error robustness and low end-to-end delay. Delay and error robustness are the parameters which will be controlled in the following work. Delay as a parameter is controlled by ARQ based mechanisms, and variable FEC is used to provide segmented error resilience.

Resources are allocated at a substream level. The data segmentation and substream architecture is thus similar to techniques being proposed for low bit rate video applications. This allows us to transmit substreams with different QoS constraints such as PSNR, real time delay and subjective quality.

3.1 Scene change based segmentation

Automatic tools [3] are used for scene change detection and for motion based segmentation. For scene changes, the initial frames after the actual scene change forms a higher priority layer. Frames within scene shots form lower priority layer. This priority is in terms of error resilience. This kind of error protection is thus a form of source dependent channel coding. Another conventional alternative for this is substream power allocation, which is not being considered in the current framework. Both the above schemes (power allocation and unequal error protection), make the transmission more robust to channel errors.

3.2 Motion based segmentation

For motion, again the segmentation is done in two levels i.e. low and high motion. Motion is estimated at a frame level. For each scene, a global average motion vector size is calculated, which forms a threshold for that scene. For each frame, a frame level average motion vector is again calculated, and compared with the global average for that particular scene. This is then used to divide frames into high and low motion ones. Motion based segmentation can also be applied at a finer granularity e.g. at the macroblock level. In terms of error robustness, a high motion layer can tolerate higher errors as compared to a lower motion layer due to masking effect.

It must be also be noticed that for motion, delay is a more stringent constraint than error protection. Although a high motion scene can tolerate more errors due to masking, it requires lower delay requirements. A low motion frame, on the contrary can tolerate higher delay. Delay here is not simply the decoding delay, but the total end-to-end delay including that over the radio link. We control this by sending multiple copies of frames at a time and by restricting the actual number of times the receiver can ask for frames by sending NAKs (detailed description later).

3.3 FEC based resource allocation

Data loss is a major problem for transmitting video over wireless networks. Bit error rates of around 10^{-3} to 10^{-2} may arise for long periods of time especially in mobile environments. Data loss can be prevented if we used proper error correction codes; however using them increases the bandwidth requirements. This is not a feasible alternative in the already very low bandwidths available. Thus, it becomes all the more necessary to use error protection judiciously. Such unequal error protection schemes are also known to give graceful degradation over channels of varying quality [1].

The overall table showing class based segmentation for 'error-robustness' is shown in Figure 3. This has both the conventional (header, data type based) and content (frame type, motion, scene change) based schemes in it. Scene changes and motion (frame based) are considered independently. In either case, we can model the source as if switching state depending on the video content. Each state (row in table below) corresponds to a given set of allocated resources (in terms of the correction capability of the assigned code). Headers form the highest priority layer followed by picture type. Further it can be seen that the scene change algorithm is applied at an outer layer which has more priority than motion. Actual data (motion vectors and DCT coefficients) are at the innermost layer. It must be understood that decisions must be made not just about the granularity (i.e. the segmentation template), but also about it's hierarchical arrangement to satisfy particular needs (error robustness in the figure below).

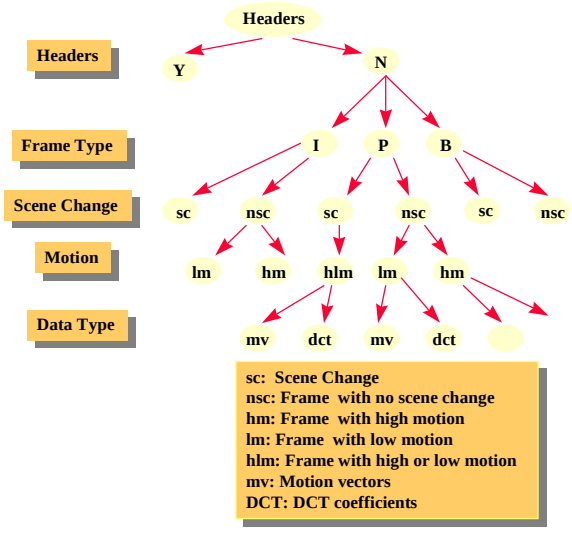


Figure 3. Template Assignment.

A wireless channel simulator [7] is used to generate bit error pattern files on a off-line basis. Error pattern files can be generated for different channel conditions (channel SNR), diversity combining techniques, error correction capabilities, and interleaving degrees. The coded video sequence is sequentially parsed, and depending on it's current content template, an appropriate error pattern file is used to mask it. To maintain randomness, 50 error pattern files (each of approximately 1.6 Mbits) are generated for each template (each row of Figure 3), and one is picked at random and used in masking.

3.4 ARQ for delay control

An ARQ based scheme is used to control latency requirements. A low motion frame can tolerate higher delay than a high motion scene. Consider a scenario in which there is a lot of motion in a basketball game, or if there is some fast object moving in a scene. If such a frame were to arrive late at the receiver, motion would become very jerky. This would give a very unpleasant effect to the eye, which tends to track motion. The high motion frames should therefore satisfy low delay constraints. On the other hand, if there isn't too much motion in a scene, and if the packet comes late, an earlier frame can be repeated in it's place without giving too bad an effect. The original packet can then be displayed on arrival (Figure 4).

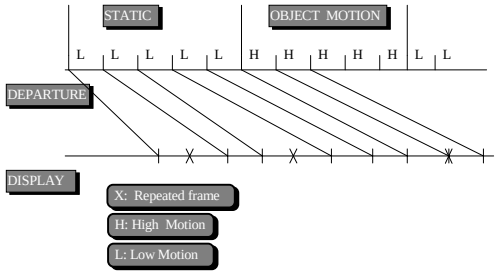


Figure 4. A Typical Scenario

Frames are segmented into low motion (Class C3 in Figure 5) and high motion (Classes C1 and C2 in Figure 5). The idea is to send multiple copies (2 here) of frames which cannot tolerate high delay; but, we restrict the number of times such packets can be sent if NAKed (no retransmissions). For a low motion frame (which can tolerate higher delay), a single copy is sent at a time, but more copies may be sent out again if the frame is NAKed. This introduces ARQ retransmission delay, but then these are the packets which can withstand it. In the high motion case, best of two arriving packets is picked by the receiver. Channel delay is modeled as a random variable (D) (currently deterministic) with some fixed mean value. The retransmission time-out of the timer at the sender is set at 'rtt+δ' where rtt is twice the mean value of the one way delay (mean(D)). We do not allow multiple outstanding packets (i.e. window size is 1). Thus, there are no reordering problems. Even with higher window sizes, we could transmit time-critical frames with higher time diversity and non-urgent frames with lower time-diversity. For 2-way interactive video, ARQ may not be suitable due to latency concerns. But, for 1-way video or VOD applications, the proposed scheme with ARQ can be used.

CLASS	CONTENT	MOTION
C1	New scene	hlm
C2	Old scene	hm
C3	Old scene	lm

Figure 5. Classes for delay

After a packet is passed to an upper layer, a decision has to be taken about display. This decision has to be based on whether the packet is late (this is the case if the arrival time + decoding time is later than it's target display time) or if it is early. The display axis is generally shifted with respect to the arrival axis by giving an initial startup delay. The following decision schemes can be used for late/early frames or packets (Figure 6).

When a packet arrives late:
For a packet (of Class C3) arriving late , an earlier frame can be repeated, and the original one can be redisplayed on arrival.

When a packet arrives early:
Here too, we use the same three frame types as above. These frames are stored in a buffer till their display time arrives. To avoid overflow, an earlier 'old scene, low motion' frame may be dropped (frame repeat is used for this frame).

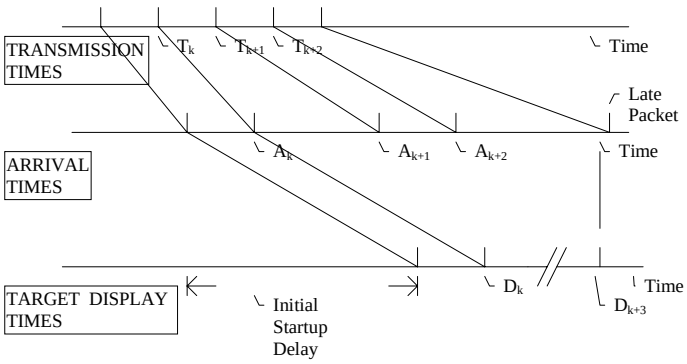


Figure 6. Transmission, arrival, and display

4. Simulation and Results

We use the wireless channel simulator described in [7]. The wireless channel is modeled using a Rayleigh fading multipath model. Personal Access Communications Services (PACS) system is used for the air-interface. The Rayleigh fading channel is simulated using the Jakes model [9]. The receiver is a coherent one, with perfect carrier recovery and optimal symbol timing and recovery. The final data with errors is compared with the original data to generate error masks.

BCH codes are used for error control. A BCH(n,k) code is capable of correcting a maximum of 't' symbols in an 'n' symbol code-word at the expense of '$2t$' symbols of overhead (here $k=n-2t$). We used n=40, and varied the error correction capability from 1 to 8 to form template dependent masks. An interleaving degree of 40 is used in the simulations. Interleaving is performed over 'n' symbol codewords, with each symbol of 64 bits. This introduces an additional buffer requirement of 40*64*40 bits (or a delay of 40*64*40/R), where R is the channel rate. For one-way communication, this can be remedied by providing an initial startup delay to the display process.

The algorithm is implemented using Columbia University's MPEG2 codec in C language. We used a movie sequence with a resolution of 608×224. The sequence is coded in 4:2:0 format with GOP=12 and M=3. An average channel SNR of 19 dB is used. Because the proposed algorithm is psychovisual in nature, subjective evaluations are made on the decoded stream. Another conventional alternative is the SNR of the decoded stream. The average SNR (over a frame) is compared for the conventional and content based schemes. The conventional schemes include error protection based on headers and data type (motion vectors and DCT coefficients). The content based schemes, in addition, used frame type, motion, and scene changes. This is also compared with the case when no data segmentation is done. To maintain fairness in comparisons, the effective overhead in each case was maintained the same. The total overhead due to channel coding is in all cases approximately 25%. Figure 7 compares the conventional error protection scheme with content based schemes. The unequal error protection schemes (both conventional, content based) are seen to perform much better than the equal error protection scheme (no segmentation). Further it is seen that using the content based approach gives an average gain of 0.62dB /frame. The subjective quality improvement is quite obvious, especially in the impaired image areas. Figure 8(b) shows a typical frame of the sequence after it is transmitted over the wireless channel. Figure 8(b) is the result if non-segmentation based schemes are used (the result is quite unacceptable). Figure 8(d) shows the result if content based schemes are used; it is seen to further improve quality over conventional data segmentation (Figure 8(c)). Work on the ARQ part of the algorithm is currently being conducted.

5. Conclusions and Future Work

Based on our results we argue that introduction of video content into video modeling allows us to define various classes of importance. We believe that by using such techniques we can get better visual quality under the very low bit rates available. Object based techniques for segmentation like those being proposed in the current MPEG4 standardization efforts are being investigated. Also redundancies due to camera motion, zooming/panning are being studied.

Acknowledgments

We would like to acknowledge Dr. M.T. Sun and Dr. Li-Fung Chang at Bellcore and Dr. A. Wong at Lucent Technologies for their help.

References

[1] H. Liu and Magda El. Zarki, "Data partitioning and unbalanced error protection for H.263 video transmission over wireless channels," Polytechnic University Symposium of Multimedia Communications and Video Coding, 1995.

[2] Paul Bocheck and Shih-Fu Chang, "A content based approach to VBR Video Source Modeling." IEEE 6th International Workshop on Network and Operating Systems Support for Digital Audio and Video, Apr. 1996.

[3] Jianhao Meng, Yujen Juan, and Shih-Fu Chang, "Scene change detection in a MPEG compressed video sequence," IS&T/SPIE Symposium Proceedings, Vol. 2419, Feb. 1995, San Jose, California.

[4] Louis C. Yun and D.G. Messerschmitt, "Variable Quality of Service in Systems by Statistical Power Control", Proc. IEEE ICC, 1995.

[5] A. Ortega and M. Khansari, "Rate Control for video coding over variable bit rate channels with applications to wireless transmission," ICIP Oct. 1995.

[6] M. Khansari and M. Vetterli, "Layer Transmission of signals over power-constrained wireless channels," ICIP 1995.

[7] T. Russell Hsing, Li-Fung Chang, A. Wong, Ming-Ting Sun, and T-C Chen. "A real-time software based end-to-end wireless visual communications simulation platform". Polytechnic University Symposium on Multimedia Communications and Video Coding, 1995.

[8] ISO/IEC 13818-2, MPEG-2, H.262, 1994.

[9] W.C. Jakes, Jr. Editor, "Microwave Mobile Communications", John Wiley and Sons, Jan. 1994.

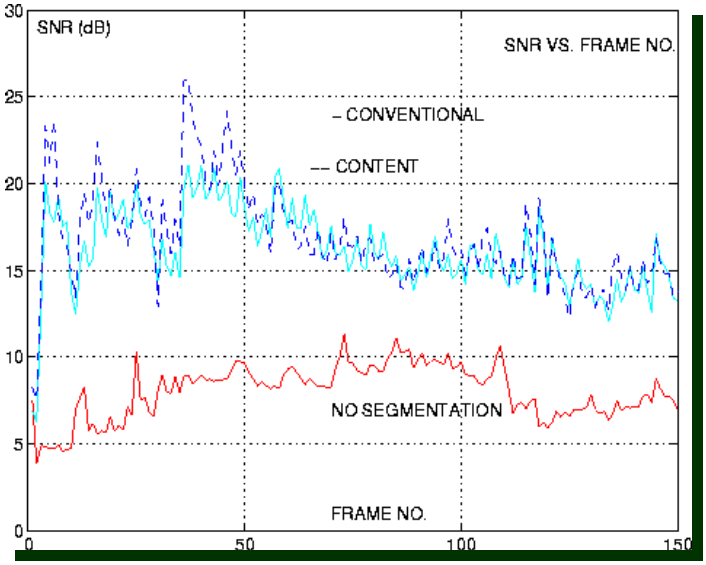
[10] D.G. Messerschmitt, J.M. Reason, A.Y. Lao, "Asynchronous video coding for wireless transport", Proc. IEEE Workshop on Mobile Computing, Dec. 1994.

[11] ISO/IEC JTC1/SC29/WG11 N1327, MPEG96, "Description of core experiments on error resilience aspects in MPEG-4 video", July 1996.

[12] ISO/IEC JTC1/SC29/WG11 N1331, MPEG96, "MSDL Specification Proposal", July 1996.

Figure 7.

Conventional Schemes
(Header, data type)
Content Based Schemes
(Header, data type, frame type, motion, scene change)



Figures 8 (a, b, c, d)
Channel SNR: 19dB, Resolution: 608x224,
Coded in 4:2:0 format, N=12, M=3,
B Frame, No Scene Change, High Motion

Original Sequence

No Data Segmentation

Conventional Data Segmentation
(Headers, Data Type)

Content-Based Data Segmentation
(Headers, Data Type, Frame type, Scene change, Motion)