SUB-PICTURE VIDEO CODING FOR UNEQUAL ERROR PROTECTION

Miska M. Hannuksela, Ye-Kui Wang**, and Moncef Gabbouj****

*Nokia Mobile Phones, Tampere, Finland
**Tampere International Center for Signal Processing (TICSP), Tampere, Finland
*** Tampere University of Technology, Finland

ABSTRACT

Unequal error protection is one of the key tools in video communication systems operating over error-prone networks. In order to allow unequal protection of a video bit-stream, codewords have to be categorized according to their importance to visual quality. The proposed sub-picture coding method allows partitioning images to regions of interest and helps to maintain a good image quality in the chosen regions. As an example, the sub-picture coding scheme is applied to multicast Internet streaming. It is shown that the overall subjective image quality and the objective foreground image quality are considerably better when compared to the selected conventional coding schemes.

1. INTRODUCTION

1.1. Slice-Based Coding

Modern video coding standards, such as ITU-T Recommendation H.263 and ISO/IEC MPEG-4 Part 2, allow division of coded pictures to slices. A slice consists of a number of consecutive macroblocks in raster scan order. Slices can be regarded as a way to split a bit-stream to transport packets that can be decoded independently. While spatial and syntactical prediction is disabled across slice boundaries, motion vectors may cross slice edges. This fact causes spatio-temporal error propagation, when motion vectors point to areas that are reconstructed incorrectly. In order to prevent this phenomenon, H.263 includes the optional independent segment decoding mode (H.263 Annex R). When this optional mode is in use, slice boundaries are treated as picture boundaries, and therefore no spatio-temporal error propagation over slice boundaries occurs. Due to restricted motion prediction, compression efficiency drops compared to normal slice-based operation. Furthermore, in order to avoid prediction confusions in slice boundaries, each slice must reside within one row of macroblocks, and its shape and position must be constant throughout the group of pictures.

1.2. Unequal Error Protection

Term unequal error protection (UEP) refers to techniques that protect part of the transmitted bit-stream in the transport system better than the rest. Examples of applicable UEP techniques include application-layer selective retransmission [1][2], transport-layer forward error control (e.g. RFC 2733 [3]), guaranteed network Quality of Service (e.g. QoS architecture of Universal Mobile Telecommunication System [4]), and Differentiated Services (DiffServ) [5][6].

In order to apply unequal error protection, video bit-streams have to be organized in portions of different importance in terms of visual quality. Techniques achieving this goal include data partitioning and scalable video coding.

Data partitioning refers to a technique where subjectively equally important codewords of all macroblocks in a slice are partitioned into a continuous block of data. Typically, macroblock headers and motion information form one partition and coded prediction error blocks form another partition.

Scalability refers to the capability of a compressed sequence to be decoded at different bit-rates. Scalability can be further categorized into temporal, SNR, and spatial scalability. Video coding schemes with temporal scalability have been reviewed in [7]. A spatially or SNR-scalable bit-stream [8] is organized into a base layer and one or more enhancement layers, and each additional enhancement layer improves perceptual quality. Fine granularity scalability (FGS) [8] is a specific type of SNR scalability. The FGS coding technique arranges the bit-stream into a base layer and an enhancement layer. Each coded picture in the enhancement layer can be truncated into any number of bits, and the number of enhancement layer bits is proportional to the image quality.

Data partitioning and scalable coding techniques generally treat an entire image equally in spatial domain. However, many images have distinct spatial regions of interest. These regions could have better error protection than other areas in order to obtain a better subjective quality compared to coding and transport schemes that

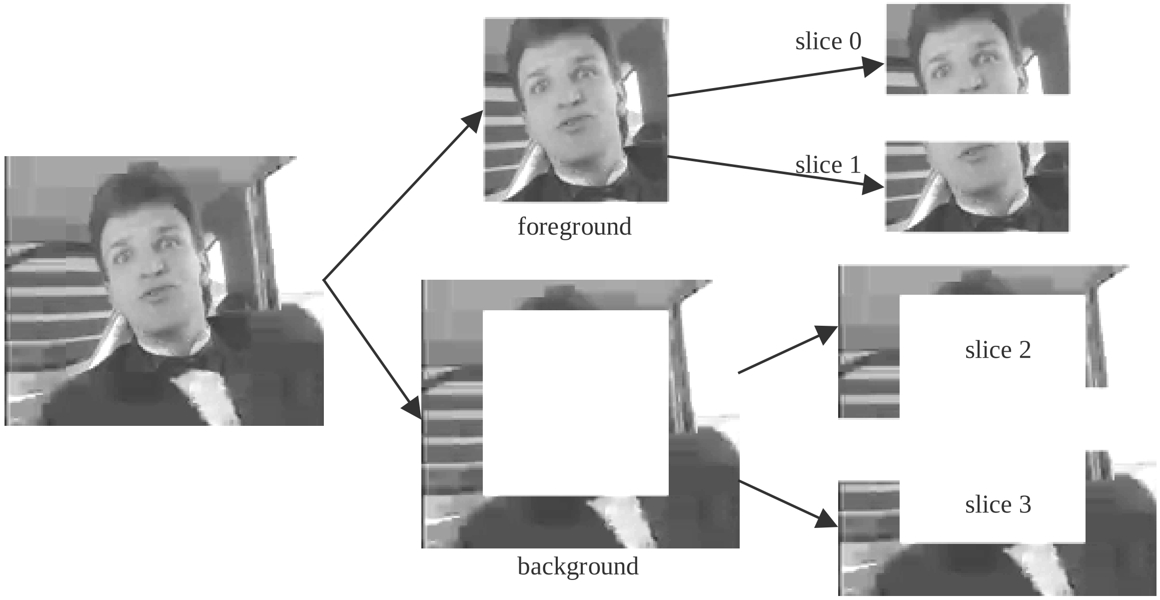


Figure 1. Example of picture, sub-picture, and slice structures.

treat all regions equally. Arbitrarily shaped objects [9], as defined in the MPEG-4 standard, can be used to extract the regions of interest. The object shape is represented by a binary alpha map and the object transparency can be represented by an alpha plane. Objects are then laid on top of each other to compose a picture to display.

1.3. Overview of the Paper

While MPEG-4 arbitrary shape coding has proven to be efficient in compression terms for many sequences, it often inherits relatively high processing requirements. This paper proposes a simple region-based coding method, called sub-picture coding, which is based on rectangular shapes. It can be considered as a simplification of arbitrary shape coding. While sub-picture coding does not provide means to define object boundaries accurately, it suits unequal error protection.

Section 2 of the paper presents the sub-picture coding method. The operation of the sub-picture coding method was simulated in multicast Internet streaming conditions, as explained in Section 3. The paper is concluded in Section 4.

2. SUB-PICTURE CODING

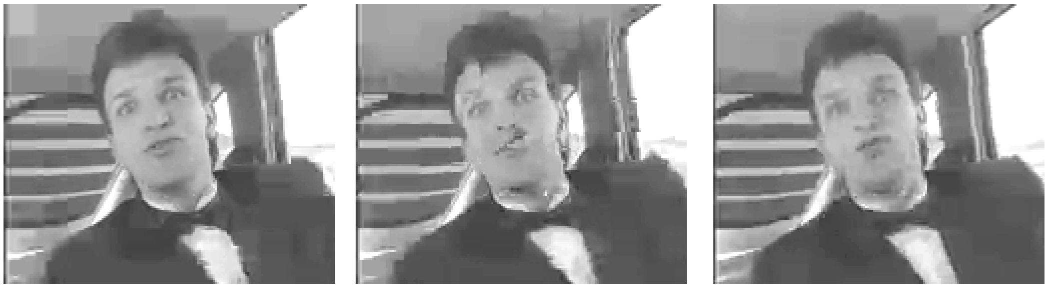
The proposal adds a sub-picture coding layer between picture and slice layers. Sub-pictures are rectangular except for the so-called background sub-picture, which consists of the picture area not falling to any of the rectangular sub-pictures. Rectangular sub-pictures are also referred to as foreground sub-pictures. Sub-pictures boundaries are aligned with macroblock boundaries, and sub-pictures do not overlap. A slice resides within one sub-picture only. A slice in a background sub-picture may not contain spatially adjacent macroblocks within a macroblock line, as it can be intervened by foreground sub-pictures. Figure 1 shows an example of a picture having one foreground sub-picture and one background sub-picture, both of which include two slices.

There are two coding modes associated with sub-pictures: normal prediction mode and independent sub-picture decoding mode. In the normal prediction mode, sub-picture boundaries are treated as slice boundaries. In the independent sub-picture decoding mode, boundaries of foreground sub-pictures are treated as image boundaries and sub-picture segmentation is static over a group of pictures similarly to H.263 Independent Segment Decoding. In other words, temporal and spatial prediction over sub-picture boundaries is prevented when coding rectangular sub-pictures. No such limitation exists when coding a background sub-picture, as the background sub-picture is considered to have a lower subjective importance and it is not protected against error propagation.

3. SIMULATIONS

3.1. Conditions

We selected multicast Internet streaming as a target application. As interactive error concealment cannot be used in large scale with IP multicast, forward error control methods have to be used. Thus, simulation conditions can



TML with sub-picture coding TML with ROI quantization TML with constant quantization

Figure 2. Example snapshots, 20 % packet loss rate.

be close to simulation conditions for IP video conferencing without interactive error control methods. We used the ITU-T Video Coding Experts Group (VCEG) common conditions for the low-delay Internet applications [10] as much as we considered appropriate. In addition, we applied transport coding level forward error correction according to RFC 2733 [3].

ITU-T VCEG has been working on the H.26L video codec project. A Joint Video Team (JVT) was recently established joining the experts from ITU-T VCEG and ISO/IEC Moving Picture Experts Group (MPEG). JVT develops a codec based on the draft H.26L design.

The sub-picture coding scheme was implemented in and tested with H.26L Test Model Long-Term (TML) version 8.6 [11][12]. A constant foreground sub-picture was selected, and a finer quantizer was used in the foreground region. The independent sub-picture decoding mode was in use. The scheme was compared to TML-8.6 with region-of-interest (ROI) quantization and to TML-8.6 with constant quantization. The quantization parameters for the ROI coding scheme were obtained in the same manner as for the sub-picture coding scheme but no sub-pictures were used.

The core of the simulator consisted of the following phases. First, the TML encoder was used to encode a sequence. The encoder generated an RTP packet stream. Then, a packet loss simulator erased some of the generated packets according to error patterns released in [13]. The resulting stream of packets was decapsulated and decoded using the TML decoder. Finally, the peak signal-to-noise ratio (PSNR) was calculated between each frame of the source sequence (at full frame rate) and the corresponding reconstructed frame.

Encapsulation into RTP packets was done as follows: In the sub-picture coding scheme, INTRA pictures were encapsulated into five packets. There were two packets for the foreground sub-picture: one packet contained odd macroblock rows and another packet contained even macroblock rows. This slice interleaving mechanism, introduced in [14], was used to obtain a better error concealment result. One parity FEC packet was generated for the two foreground packets according to RFC 2733. The background sub-picture was packetized into another two packets using slice interleaving method. Two consecutive INTER pictures consisted a group, and for each such group there were two foreground sub-picture packets, one parity FEC packet for the foreground packets, and two background sub-picture packets. A sub-picture packet contained data from two pictures: macroblocks from even rows of a certain frame and macroblocks from odd rows of the next frame or vice versa. When sub-picture coding was not in use, there were three packets for each INTRA and INTER frame: two packets for the entire picture (slice interleaving applied), and one parity FEC packet for the two packets.

As multicast Internet streaming was assumed, the coding and transport scheme was tailored for the worst expected case. We selected optimal coding parameters and encapsulation options by trial and error according to the PSNR performance in the 20 % packet loss rate. We tested several INTRA macroblock coding rates as well as several slice sizes to find the optimal ones. The packet stream obtaining the best PSNR performance was multicast virtually. The PSNR performance in 0, 3, 5, and 10 % packet loss rate was then simulated with the multicast packet stream. In order to obtain statistically significant results, each packet stream was virtually transported ten times for each packet loss rate. The first loss pattern position of a run starts from where the previous run ended.

3.2. Results

The experiments were done with the Carphone, Hall Monitor, and Coastguard sequences. We present the results for the Carphone sequence only due to lack of space. The results with the other sequences were consistent with the Carphone results.

Figure 2 shows an example snapshot of the Carphone sequence in 20 % packet loss rate. It can be seen that the sub-picture coding scheme maintains the best image quality. A clearly visible error hit the stream coded with the ROI quantization scheme. When constant quantization

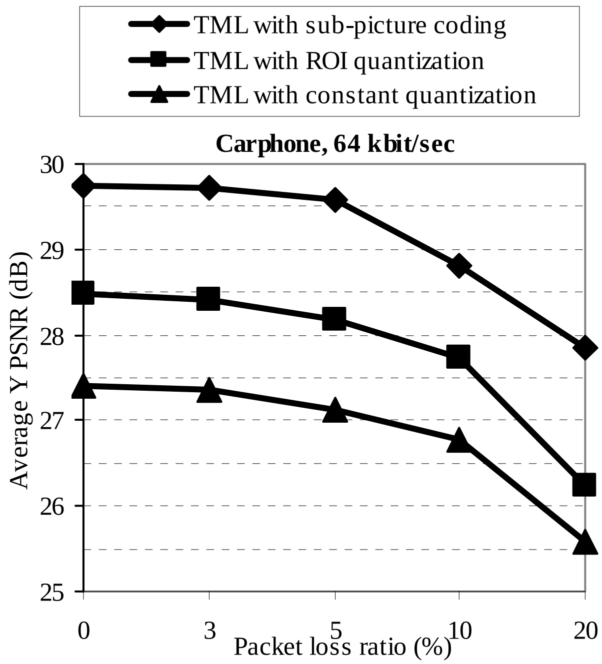


Figure 3. Average luminance PSNR of foreground.

was used, the foreground area was quantized more coarsely than in the other cases, which can be easily seen from Figure 2.

Figure 3 presents the average luminance PSNR of the foreground area of the Carphone sequence. These objective results confirm the subjective ones. Sub-picture coding improves the average luminance PSNR of the foreground area more than one decibel regardless of the packet loss rate. This improvement is gained at the expense of background image quality, which is degraded by coarser quantization and less error protection. In fact, the overall PSNR in the sub-picture coding case drops a little compared to the ROI quantization case and somewhat more compared to the constant quantization case. However, errors in the background are far less noticeable than errors in the foreground, and therefore the overall subjective quality is improved.

4. CONCLUSION

Many video communication systems can make use of unequal error protection to improve quality while maintaining or decreasing bit-rates. In order to apply unequal error protection, video bit-streams have to be organized in portions of different importance in terms of visual quality. Techniques achieving this goal include data partitioning scalable video coding, and region-based video coding. This paper introduced a simple region-based video coding method called sub-picture coding. It provides means for unequal error protection based on regions of interest. In sub-picture coding, a picture is partitioned to one or more rectangular foreground sub-pictures and to a hollow background sub-picture along macroblock boundaries, and sub-picture edges are treated as slice boundaries or as picture boundaries depending on the signaled coding mode. The sub-picture coding method was simulated in multicast Internet streaming conditions and it was found to be superior to conventional coding. Due to its simplicity and the obtained results, we think that sub-picture coding has some potential to become one of the key video coding tools for unequal error protection.

5. REFERENCES

[1] A. Miyazaki, H. Fukushima, K. Hata, T. Wiebke, R. Hakenberg, C. Burmeister, N. Takatori, S. Okumura, and T. Ohno, "RTP Retransmission Payload Format", IETF Internet Draft draft-ietf-avt-rtp-selret-02.txt, 18 July 2001.

[2] D. Leon, V. Varsa, "RTP retransmission framework", IETF Internet Draft draft-leon-rtp-retransmission-00.txt, July 2001.

[3] J. Rosenberg, H. Schulzrinne, "An RTP Payload Format for Generic Forward Error Correction" IETF RFC 2733, December 1999.

[4] 3GPP TS 23.107, "QoS Concept and Architecture (Release 4)", V4.2.0, October 2001, available via ftp://ftp.3gpp.org/Specs/2001-09/Rel-4/23_series/23107-420.zip.

[5] K. Nichols, S. Blake, F. Baker, and D. Black, "Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers", IETF RFC 2474, December 1998.

[6] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss, "An Architecture for Differentiated Services", IETF RFC 2475, December 1998.

[7] M. M. Hannuksela, "Simple Packet Loss Recovery Method for Video Streaming", International Packet Video Workshop PV2001, Kyongju, South Korea, 30 April – 1 May 2001.

[8] W. Li, "Overview of Fine Granularity Scalability in MPEG-4 Video Standard", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 11, No. 3, March 2001.

[9] N. Brady, "MPEG-4 Standardized Methods for the Compression of Arbitrarily Shaped Video Objects", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 9, No. 8, December 1999.

[10] S. Wenger, "Common conditions for wire-line, low delay IP/UDP/RTP packet loss resilient testing", ITU-T Video Coding Experts Group, document VCEG-N79r1, September 2001.

[11] T. Wiegand (editor), "H.26L Test Model Long-Term Number 9 (TML-9) draft0", ITU-T Video Coding Experts Group, document VCEG-N83d1, December 2001.

[12] H.26L Test Model Software, version 8.6, October 2001.

[13] S. Wenger, "Error patterns for Internet video experiments", ITU-T SG16 document Q15-I-16-R1, October 1999.

[14] S. Wenger and G. Côté, "Using RFC2429 and H.263+ at low to medium bit-rates for low-latency applications", International Packet Video Workshop PV99, April 1999.