ADAPTIVE PICTURE SLICING FOR DISTORTION-BASED CLASSIFICATION OF VIDEO PACKETS

E. Masala, D. Quaglia, J.C. De Martin*
Dipartimento di Automatica e Informatica/*IRITI-CNR
Politecnico di Torino, Italy
E-mail: [masala|davide.quaglia|demartin]@polito.it

Abstract -
We present a new algorithm to dynamically identify perceptually important regions in pictures of a video sequence. For each macroblock, the distortion that its loss would cause at the decoder is computed. High-distortion macroblocks are then grouped together into slices to be protected with forward error correction or to be sent as "premium" packets. We developed the approach for the case of video sequences encoded with the ISO MPEG-2 video coding standard. We then applied the algorithm to classify video packets within a 1-bit Differentiated Services architecture: slices were grouped either into *premium* packets, to be sent on a "virtual wire," or into *regular* packets, to be sent as best-effort traffic. In packet losses, the proposed distortion-based classification scheme outperforms source-transparent packet-marking techniques and provides substantially higher PSNR values than the regular best-effort case sending as little as 10% of the packets as premium traffic. Video samples are available at `http://multimedia.polito.it/mmsp2001/`.

INTRODUCTION

Reliable transmission of multimedia signals is required by many emerging applications, including Internet television, video/audio on-demand, web radio, videoconferencing and voice over IP. In many cases, multimedia data need to be transmitted over noisy channels with delay constraints that, in case of errors, do not allow data retransmission. In such cases, at the transmitter side some form of redundancy, usually in the form of error–correcting codes or multiple-descriptions, can be added to the bitstream to increase overall robustness, while at the receiver side, error concealment can be used to minimize the impact of residual errors. Since not all bits are perceptually equally important, unequal error protection (UEP) is often used (see, e.g., [1]); in such cases, bit classification is normally done at design time and then kept constant. A packet classification approach based on the characteristics of the current data block was recently proposed in [2]. In other cases, classification is performed transparently to content, e.g., randomly, to statistically achieve specified levels of quality of service (QoS) [3].

We propose to classify multimedia data packets *adaptively* depending on the desired level of perceptual quality of service and the distortion that the loss of each individual packet would introduce at the decoder. To do so, decoding, including error

0-7803-7025-2/01/$10.00 ©2001 IEEE.

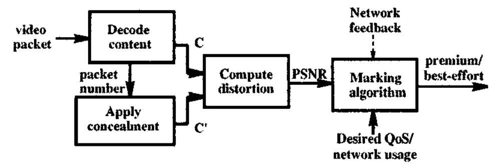


Figure 1: Block diagram of distortion-based packet marking.

concealment, is replicated at the encoder; a distortion measure is then computed and classification performed accordingly. The approach was recently applied to telephone-bandwidth speech in [4]; we now propose to extend it to classification and transmission of video sequences. In particular, we focus on ISO MPEG-2 video [5] with two main objectives: study of a distortion-based picture segmentation algorithm and analysis of transmission performance in the context of a network employing a 1-bit Premium Differentiated Services (DiffServ) model [6] [7].

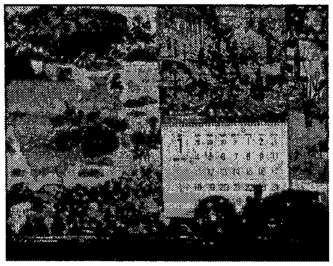
The paper is organized as follows. In the first Section, the distortion-based approach to multimedia classification for transmission over packet networks is explained; the approach is detailed for the specific case of the ISO MPEG-2 video coding standard. Results of tests comparing the proposed approach to current techniques are presented in the second Section. Finally, conclusions are presented in the third Section.

DISTORTION-BASED CLASSIFICATION OF MPEG-2 VIDEO

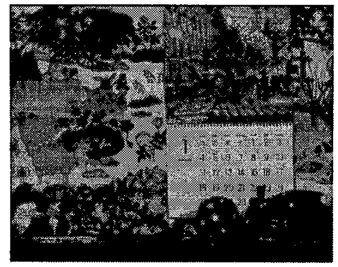
We propose to express the perceptual importance of a video data packet in terms of the distortion that would be introduced by its loss. Such distortion can be computed comparing the video sequence decoded using the correct data to the video sequence decoded using the replacement data generated by the concealment technique at the decoder. If the video data cannot be adequately concealed, the distortion will be high and the data consequently flagged as important. Based on the distortion value, the desired level of QoS and, if available, the current network status, a distortion-based classification algorithm will then mark the current packet, as shown in Figure 1.

We chose to implement this approach for the specific case of the MPEG-2 video format, Main Profile and Main Level (MP@ML). In the MPEG standard, an arbitrary number of consecutive macroblocks (MB) belonging to the same row is coded into a "slice." In the MPEG syntax the slice is the smallest unit which can be decoded independently. When MPEG video is transmitted over a packet network, packets should contain an integer number of slices since, in case of packet losses, a partial slice is generally useless [8].

Since slice subdivision can change for every row and at every picture, we chose to exploit such flexibility to develop an *adaptive slicing algorithm* aimed at identifying perceptually important picture regions. Before presenting the algorithm, the working scenario is described.

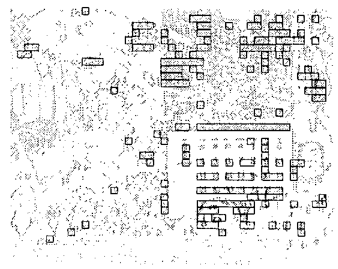


(a) n-th decoded picture.



(b) $(n + 1)$-th decoded image.

(c) Difference picture.



(d) Selected macroblocks.

Figure 2: Example of macroblock adaptative marking.

The ISO MPEG-2 decoder was extended to implement a simple temporal concealment at MB level: if a MB is lost, the corresponding picture area is filled with the pixels in the same position of the previously displayed picture (anchor area). The importance of each MB in a picture is, therefore, estimated by computing a distortion measure between a given area in the current picture and its anchor area in the previous picture. Figure 2.c shows the pixel-to-pixel difference between two successive displayed pictures (Figure 2.a and 2.b): dark pixels correspond to areas that have changed (e.g., the edges of a moving object).

As distortion measure, the mean squared error (MSE) between a MB area and its anchor was chosen. This metric is related to the PSNR, which is commonly used to estimate the quality of video after data loss (due to source coding or transmission).

A 1-bit classification scenario is assumed: perceptually important slices are marked as "premium," to be sent on a "virtual wire," or marked as "regular," to be sent on a best-effort (i.e., lossy) channel. Premium bandwidth, however, is a limited as well as an expensive resource. Therefore, the classification process should be constrained either in terms of minimum allowed QoS or in terms of maximum share of premium traffic. Absent clear guidelines matching absolute MSE values to well-known subjective quality levels, we preferred the latter approach, which also simplifies the analysis of network usage. A straightforward classification algorithm consists in creating a slice out of each MB in a picture, sorting them in decreasing order of MSE and marking the first k slices that lead to the desired premium share. This approach, however, maximizes the overhead due to slice headers, which can be quite significant, and does not permit differential encoding of adjacent MBs. We, therefore, opted for improved coding efficiency by grouping into the same slice adjacent MBs of similar perceptual importance. As explained below, the slicing process is iterative and stops when the desired premium share is reached.

Video data is classified into either premium or regular slices according to the following algorithm. Let b_0 be the desired premium share, i.e., the ratio between the number of bits sent as premium and the total number of bits. Let M be the set of all MBs of the picture. Let R and P be the set of regular and premium MBs, respectively. Let S be the set of slices in the picture, where each slice is identified by three numbers: first and last MBs, and a premium/regular flag. Given the set P, let $d(P)$ be the function returning the set S; slice subdivision is performed by grouping all adjacent MBs belonging to the same class of traffic. A premium slice is a slice of premium MBs. Given the set S, let $f(S)$ be the function that computes the resulting premium share for the picture. This function can be implemented by simulating the generation of the MPEG bitstream and counting the output bits. Finally, let $sum_{MSE}(P)$ be the function that computes the total MSE associated to premium MBs.

The adaptive distorsion-based picture-slicing problem can be then formulated as follows: for each pictures in the sequence, find the set P that maximizes $sum_{MSE}(P)$ under the constraint $f(d(P)) \leq b_0$. This is a well–known operational research problem that can be solved knowing $m()$, $d()$ and $f()$. The first two functions are easy to compute but, in general, we cannot estimate $f()$ a priori, since an increase in the premium MBs number does not necessarily lead to an increase in the premium share. For example, when we mark a MB in the middle of a regular slice, this MB becomes a premium slice and MBs which precede and follow it become two regular slices. Since the first and last MBs of a slice cannot be "skipped," the number of not-skipped, regular MBs may increase and, therefore, the overall premium share may, in fact, decrease. In our work, however, we used only I-pictures, in which skipped MBs are not allowed; in this case, the premium share grows almost linearly with the number of marked MBs. Because of this behavior, the problem is simplified and can be solved with the following algorithm:

```
mark sequence header as premium
for (each picture in the sequence)
    mark picture header as premium
    R = M           ; start with all MBs as regular
    P = ∅
    while (f(d(P ∪ {max_MSE(R)})) ≤ b0)
        P = P ∪ {max_MSE(R)}
        R = R − {max_MSE(R)}
    group slices into packets
```

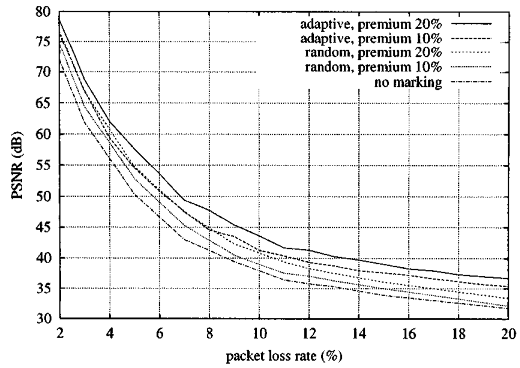


Figure 3: PSNR values for increasing packet loss rates.

where $max_{MSE}(R)$ returns the macroblock $\in R$ with the highest MSE. Figure 2.d shows an example of picture slicing performed according to the proposed distortion-based algorithm.

RESULTS

To test the proposed adative classification scheme we encoded the 720×576 (CCIR-601 resolution) 40-frame standard video sequence known as *Mobile*. The simulation was performed on the sequence concatened with itself 50 times, for a total of 2000 frames, to achieve statistical significance in packet loss conditions. Encoding was performed using the MPEG-2 Main Profile at a bitrate of about 5 Mb/s, using I-pictures only, to avoid to consider error propagation to future pictures. We used a modified version of the Test Model 5 Encoder by the MPEG Software Simulation Group [9]. For decoding, we used the MPEG-2 Reference Decoder version 1.2 by the same authors. As quality measure, we used the peak signal-to-noise ratio (PSNR) between a concealed picture and the corresponding one decoded from an error-free version of the bitstream; the PSNR value is averaged over all the pictures of the sequence.

We considered five cases. The first two represent the proposed classification algorithm with a premium share of 10% and 20%, respectively. In the third and fourth cases a slice included an entire row and each slice was transmitted in a different packet; packets were, then, randomly marked as premium using the same shares of case 1 and 2 (10% and 20%). The last test case did not use marking, i.e., all packets were subject to packet losses. Packet size was approximately 700 bytes, the same for all five cases. Packet-loss patterns for increasing packet loss rates were applied to the best-effort packets and then the resulting sequences were decoded. Figure 3 shows the performance results in terms of PSNR for the five cases under consideration and packet loss rates up to 20%. Distortion-based classification consistently outperforms random marking by up to approximately 3 dB. With respect to the regular best-effort case, even a modest premium share of 10% delivers a PSNR gain of approximately 4 dB, a gain that grows to about 6.6 dB for a premium share of 20%.

CONCLUSIONS

We have presented a distortion-based approach to classification of multimedia content. Data blocks are classified as either premium or regular depending on the distortion that their loss would introduce at the decoder. A technique for adaptive classification of MPEG-2 video sequences was implemented and tested. Transmission experiments in a 1-bit DiffServ packet scenario showed that distortion-based data classification consistently outperforms source-transparent techniques and provides substantially higher PSNR values than the unprotected case sending as little as 10% of the packets as premium traffic.

References

[1] J. Hagenauer and T. Stockhammer, "Channel Coding and Transmission Aspects for Wireless Multimedia," *Proceedings of the IEEE*, vol. 87, no. 10, pp. 1764–1777, October 1999.

[2] J. Shin, J.W. Kim, and C.C. Jay Kuo, "Relative Priority Based QoS Interaction Between Video Applications and Differentiated Service Networks," in *Proc. IEEE Int. Conf. on Image Processing*, 2000, vol. 3, pp. 536–539.

[3] W. Feng, D. Kandlur, D. Saha, and K. Shin, "Maintaining end-to-end throughput in a differentiated-services Internet," *IEEE/ACM Transactions on Networking*, vol. 7, no. 5, pp. 685–697, October 1999.

[4] J.C. De Martin, "Source-Driven Packet Marking For Speech Transmission Over Differentiated-Services Networks," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Salt Lake City, Utah, May 2001.

[5] ISO/IEC 13818-2 MPEG-2 Video Coding Standard, "Generic coding of moving pictures and associated audio information—Part 2: Video," *ISO*, 1995.

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

[7] K. Nichols, V. Jacobson, and L. Zhang, "A Two-bit Differentiated Services Architecture for the Internet," *RFC 2638*, July 1999.

[8] R. Hoffman, G. Fernando, V. Goyal, and M. Civanlar, "RTP Payload Format for MPEG1/MPEG2 Video," *RFC 2250*, January 1998.

[9] S. Eckart and C. Fogg, "ISO/IEC MPEG-2 software video codec," *Proc. SPIE*, vol. 2419, pp. 100–118, 1995.