IMPROVED BIT ALLOCATION IN AN ERROR-RESILIENT SCHEME BASED ON DISTRIBUTED SOURCE CODING

M. Fumagalli*, M. Tagliasacchi†, S. Tubaro†

(*) CEFRIEL-Politecnico di Milano,
Via R. Fucini 2 - 20133 Milano, Italy
fumagall@cefriel.it

(†) Dip. di Elet.e Inf., Politecnico di Milano,
P.zza L. Da Vinci, 32 - 20133 Milano, Italy.
{tagliasacchi,tubaro}@elet.polimi.it

ABSTRACT

In this work we propose an error-resilient scheme that allows enhancing the robustness of a video stream. Based on distributed source coding (DSC) principles, an auxiliary stream is sent in parallel to the main stream as a redundant representation of the sequence that is used to correct errors at the decoder, thus reducing the impact of drift. In order to perform an optimal bit allocation in the auxiliary stream, the encoder needs to compute a reliable estimate of the expected video distortion observed at the decoder side due to channel loss. This paper proposes an algorithm to calculate the expected distortion of decoded DCT-coefficients (dubbed EDDD) and its application to the bit allocation problem in a DSC based auxiliary stream.

1. INTRODUCTION

Sending a video sequence over a network that does not provide any QoS guarantee (e.g., IP network) is nowadays a very common application. If errors occur, some information does not reach the decoder and error-concealment (EC) techniques cannot completely avoid error propagation due to inter-frame dependencies introduced by predictive encoding.

Recently, error-resiliency tools based on the principles of distributed source coding (DSC) have appeared in the literature [1][2][3]. The work in [1] employs a Wyner-Ziv bitstream in an error protection framework to provide improved error-resilience. The Wyner-Ziv bitstream contains parity bits corresponding to a coarsely quantized video description, which are exploited when the decoded video is corrupted by channel errors, ensuring that the video quality degrades slowly as the probability of transmission errors increases.

In [3] a redundant representation is encoded according to DSC principles and it is sent through an auxiliary channel to correct errors at the decoder. An important point in the design of this system is to optimally allocate bits assigned to the auxiliary representation. To this end, the encoder needs to characterize in statistical terms the induced channel noise between the original sequence and the reconstructed sequence at the decoder. Intuitively, the higher is the estimated noise level, the larger is the number of bits that need to be spent in the auxiliary channel in order to correct errors. Since the induced channel noise is neither spatially nor temporally stationary, the encoder needs to estimate the expected distortion for each DCT-coefficient.

In the literature, the ROPE algorithm (recursive optimal per-pixel estimate) [4] is an efficient technique that allows the encoder to calculate the sample-by-sample expected distortion of the decoded video due to channel loss. In the case of packet-switched networks, an estimate of the network packet loss rate and the EC technique used at the decoder side are the only required parameters. In [3], a simple approximation of the DCT-coefficients distortion is given by an extension of ROPE to the transform domain. Simulation results [5] reveal that this approach leads to a coarse approximation of the estimated distortion and to suboptimal bit allocation in the DSC-based auxiliary stream.

In this work we propose a DSC-based error-resilient system based on the work in [3] where the novelty is the introduction of a more accurate algorithm to estimate the expected decoded distortion of each DCT-coefficient (based on the work in [5]). The novelty is that this improved estimate is then used to carefully allocate the bits in the auxiliary stream.

This paper is organized as follows: in Section 2 the DSC error-resilient system proposed in [3] is briefly reviewed. Section 3 deals with the estimate of the expected distortion of a decoded video: first, a brief overview of the original ROPE algorithm is presented, followed by the proposed video distortion estimation algorithm (EDDD) in the DCT domain. In Section 4 we present the proposed DSC error-resilient system. Finally Section 5 shows some simulation results and final remarks are given in Section 6.

2. DSC BASED AUXILIARY CHANNEL

Ramchandran et alt. [3] propose a method to reduce the drift in predictive codecs by sending extra information over an auxiliary channel using DSC principles. The basic scheme is depicted in Figure 1 where X is the original DCT-coefficient to be encoded while $\tilde{X}$ is the noisy reconstruction obtained at the decoder after EC. The auxiliary channel is used to correct $\tilde{X}$ into $\hat{X}$ which is a cleaner version of X, thus reducing the effect of drift. In order to achieve this, the encoder quantizes X and sends the index of the coset the quantized version of X belongs to. In Figure 2 it is shown an example where the source codebook (i.e. the set of all quantization bins) is partitioned into four cosets. The auxiliary channel communicates to the decoder that X belongs to the coset *1*. At the decoder, $\hat{X}$ is obtained as the codeword closest to $\tilde{X}$ that belongs to coset *1*.

In order to optimally allocate the bit budget devoted to the auxiliary channel, the encoder needs to decide the number of cosets to be used for each DCT-coefficient, which is related to the expected distortion $E\left[(X-\tilde{X})^2\right]$. In [3] it is assumed that the correlation between X and $\tilde{X}$ can be modeled as $X = \tilde{X} + Z$, where the components of the correlation noise vector Z are independent and Gaussian. A modified version of ROPE, dubbed DCT-ROPE, is used to estimate the induced channel noise $E\left[(X-\tilde{X})^2\right]$ directly in the DCT domain.

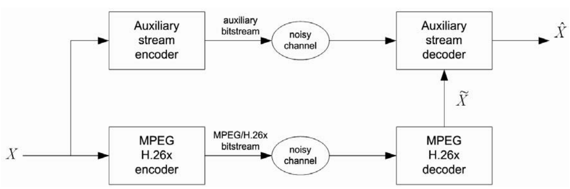

Figure 1. DSC error-resilient scheme proposed in [3]

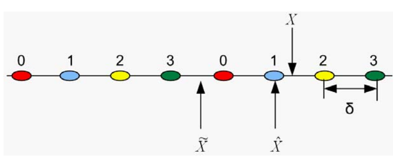

Figure 2. Partitioning of the quantization lattice into cosets.

3. ESTIMATE OF THE EXPECTED DISTORTION OF THE DECODED VIDEO

3.1. ROPE Algorithm Overview

The recursive optimal per-pixel estimate (ROPE) algorithm allows the encoder to calculate the sample-by-sample expected distortion of the decoded video due to channel loss. The only required parameters are an estimate of the network packet loss rate (PLR) and the decoder EC technique. The encoder does not know the specific loss pattern; hence it has to characterize the actual reconstruction of a sample value operated by the decoder as a random variable. A detailed explanation of ROPE algorithm can be found in [4]. The original algorithm [4] is based on the assumption that the video sequence is encoded with integer-pixel MVs. Since most of the common encoders implement a sub-pixel precision motion estimation, the works in [5][6][7], propose extensions of the algorithm in order to obtain an accurate solution for this case. Moreover the work in [8] further extends the algorithm in the case of H.264/AVC coding technique.

In [3] an extension of ROPE that works directly in the DCT domain is proposed. The basic idea is to use the recursive equation of the original ROPE algorithm on the DCT-coefficients instead of working on pixel values. This straightforward solution (that we call in this work DCT-ROPE) has to cope with the difficulty of managing the motion compensation phase in the transform domain. In fact, in order to map DCT-coefficients from the current to the reference frame, MVs are quantized with a step size equal to the block side length in such a way that each MB of the current frame is matched with the nearest MB of the reference frame. This coarse approximation of MVs leads to a loss of accuracy in the distortion estimation [5].

3.2. EDDD Algorithm

This section presents the expected distortion of decoded DCT-coefficients (EDDD) algorithm (more details are given in [5]). The proposed algorithm allows estimating the distortion of the reconstructed video sequence in the DCT domain, yet retaining the accuracy of the original ROPE algorithm. The basic idea is to run ROPE in the pixel domain and then to estimate the distortion in the DCT domain, block-by-block, using the corresponding spatial information. This is rather different from the approach presented in [3], where the estimate was obtained directly in the DCT domain.

Following the same approach as ROPE, the EDDD algorithm considers the decoded value of each pixel j as a random variable x_j whose statistics are expressed by the estimated first and second moments of the pixel value after the reconstruction at the decoder side. The two expected quantities are calculated by recursive equations. In a similar way, for each block, EDDD represents each i-th DCT-coefficient as a random variable y_i. In order to characterize the expected distortion, both the first ($E[y_i]$) and the second ($E[y_i^2]$) moments need to be estimated.

Given w_j^i (the j-th elements of the i-th DCT basis function) the random variable y_i can be written as in Eq. (1).

$$y_i = \sum_J w_j^i \cdot x_j \tag{1}$$

Using Eq. (1) EDDD algorithm obtains the expression of $E[y_i]$ and $E[y_i^2]$ as reported in Eq. (2) and Eq. (3).

$$E[y_i] = \sum_J w_j^i \cdot E[x_j] \tag{2}$$

$$E[y_i^2] = \sum_Z \sum_J w_z^i \cdot w_j^i \cdot \left(E[x_z x_j] - E[x_z] \cdot E[x_j]\right) + E[y_i]^2 \tag{3}$$

Eq. (2) is of easy calculation while Eq. (3) presents an unknown term that represents the expected value of the product of pixels pairs within the considered block. In [7] we showed that a recursive equation for $E[x_z x_j]$ (where z and j are adjacent pixels) is needed when ROPE is extended to work with half-pixel precision MVs. In the EDDD case we need to compute $E[x_z x_j]$ also for all non-adjacent pixels z and j. A simple, yet efficient approximation consists of approximating $E[x_z x_j]$ with the product of the expected values of the products of the adjacent pixels that connect z with j. This latter approximation is the only cause of inaccuracy of EDDD algorithm.

4. PROPOSED SYSTEM

The system proposed in this paper is based on the work in [3] and implements the scheme depicted in Figure 1. The goal is to replace the correlation estimation algorithm DCT-ROPE with EDDD in order to improve the bit allocation in the auxiliary channel. The encoding phase of the auxiliary channel proceeds as follows (for each block): (1) initialize the ROPE algorithm by setting the PLR and the EC technique used; (2) estimate the expected distortion in the DCT domain with DCT-ROPE / EDDD; (3) computes the number of cosets as:

$$l_i = \max\left(0, \left\lceil \log_2 \alpha \frac{\sigma_i}{\delta} \right\rceil\right) \tag{4}$$

where $\sigma_i^2 = E[(y_i - \tilde{y}_i)]$ is the expected distortion for the DCT-coefficient i, δ is the quantization step size and α is a tuning factor that determines the probability of decoding errors; (4) determine the coset index c_i of the quantized version of y_i; (5) encode/send l_i and the coset index c_i until the bit budget is used.

Decoding proceeds as follows: (1) decode the current block applying EC; (2) compute the DCT transform to obtain $\tilde{y}_i$ for each DCT-coefficient; (3) correct $\tilde{y}_i$ as $\hat{\tilde{y}}_i$ based on the received coset index c_i. It can be noticed that the decoding process does not depend on the algorithm used to estimate the distortion in the DCT domain

5. EXPERIMENTAL RESULTS

This section describes two tests. First, we compare the EDDD estimation algorithm with DCT-ROPE used in [3]. In order to assess the accuracy and consistency of both approaches, they are compared with the actual measured distortion of the reconstructed sequence averaged over several network simulations (NS) with different error patterns. Figure 3 illustrates the PSNR tracks as a function of the frame index of the decoded '*Foreman*' sequence (QCIF, 30 fps, 256 kbps) affected by a random packet loss rate of 10%. The network simulations track is the average PSNR over 60 different realizations, while pixel-domain ROPE (PEL-ROPE) and EDDD have the same track due to the fact that EDDD is obtained from PEL-ROPE via a unitary transform. The inaccuracy of DCT-ROPE approach is evident. Similar results have been obtained over all video sequences.

Figure 4 shows the accuracy of EDDD and DCT-ROPE algorithms in the DCT domain. Each 8x8 DCT block is divided into four not-overlapping frequency bands: DC coefficient, AC-3 (the three coefficients adjacent to DC), AC-12 (the twelve coefficients around AC-3), and AC-48 (the remaining coefficients). For these bands we calculate the expected distortion (MSE). The proposed EDDD approach presents a significant accuracy improvement with respect to the DCT-ROPE approach. Moreover the estimated distortion is rather close to the actual average distortion measured in the network simulations. Tests on other sequences at different PLRs lead to the same results.

The second set of experiments deals with the improvement of the DSC based error-resilient scheme using EDDD in the auxiliary channel bit allocation instead of DCT-ROPE. In this work we do not consider how to encode the auxiliary channel data (number of cosets and coset indexes) as this is topic of current research activities. For the auxiliary channel we impose the maximum number of uncoded bits for each DCT block: two different tests are provided, 150 or 24 bits/block (in this way about 5 or 25 DCT-coefficients can be corrected using the corresponding coset index). Each frame is divided into non-overlapping slices and each slice fits into a packet (9 packet/frame for a QCIF video sequence). The simulation results refer to '*Foreman*' and '*Stefan*' sequences (QCIF, 30 fps, 500 kbps) affected by a random PLR of 15% (for both H.263 and side-channel stream). Figure 5 reports the PSNR tracks for the two systems DCT-ROPE and EDDD considering a maximum of 150 bits/block. Figure 6 refers to a maximum of 24 bits/block. In both the cases the improvement of the proposed approach is evident.

Finally Figure 7 presents the frames 65 and 60 for the two sequences and the two approaches. Most of the details that the codec proposed in [3] are not able to reconstruct are well recovered by the proposed system.

6. CONCLUSIONS

This paper presents a distributed source coding (DSC) error-resilience video coder based on the scheme proposed by Ramchandran et alt. [3]. The knowledge of the decoded video distortion that the receiver is expected to suffer turns to be crucial for the side-channel rate allocation problem.

This work proposes an extension of ROPE algorithm (called EDDD) that is able to estimate the expected distortion in the DCT domain. The accuracy of the EDDD is validated by a significant improvement of accuracy in auxiliary channel bit allocation in a DSC-based error-resilient scheme.

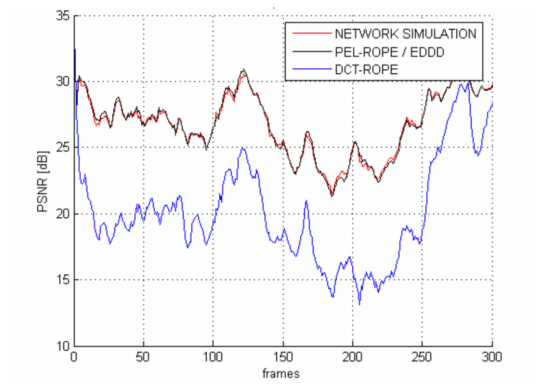

Figure 3. Foreman: PSNR tracks for EDDD, DCT-ROPE and NS.

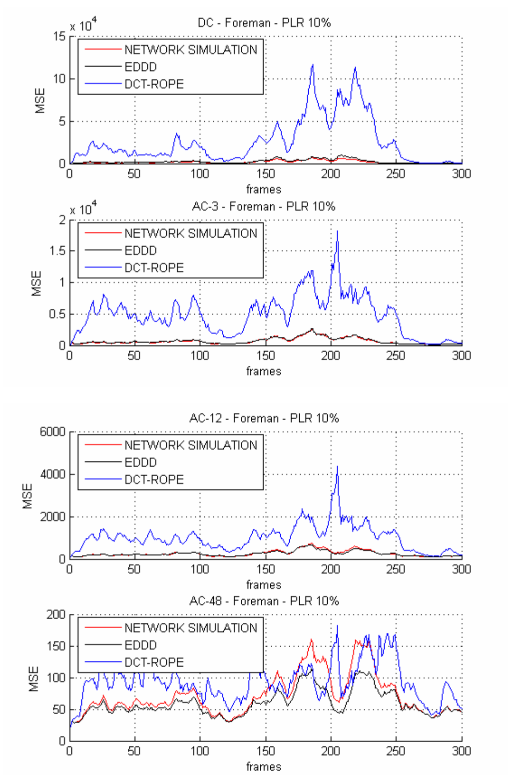

Figure 4. Foreman: MSE in the four not-overlapped DCT bands.

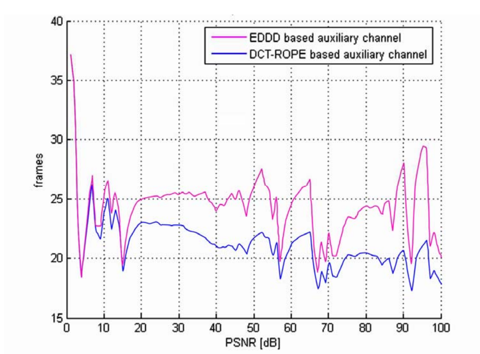

Figure 5. 'Stefan' QCIF sequence, 30 fps, 500 kbit/s, PLR=15%, 150 bits/block: DSC codec based on DCT-ROPE and EDDD.

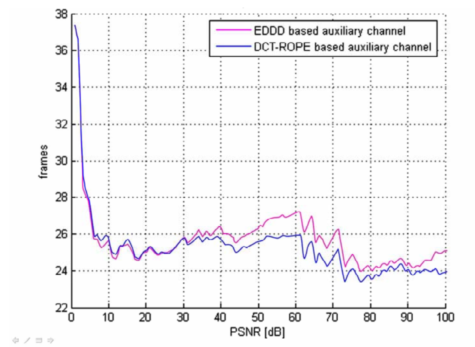

Figure 6. 'Foreman' QCIF sequence, 30 fps, 500 kbit/s, PLR=15%, 24 bits/block: DSC codec based on DCT-ROPE and EDDD.

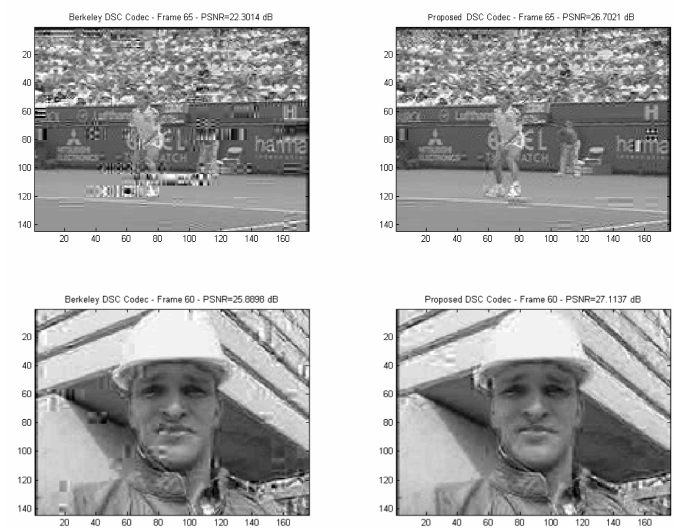

Figure 7. Frames 65 and 60 that refer respectively to Figure 5 and 6.

REFERENCES

[1] S. Rane, A. Aaron, B. Girod, "Systematic lossy forward error protection for error-resilient digital video broadcasting - A Wyner-Ziv coding approach," Proc. ICIP 2004, Singapore, Oct. 2004

[2] A. Sehgal, N, Ahuja, "Robust Predictive Coding and the Wyner-Ziv Problem", DCC, Snowbird, Utah, Oct. 2003

[3] A. Majumdar, J. Wang, and K. Ramchandran, "Drift Reduction in Predictive Video Transmission using a Distributed Source Coded Side-Channel," ACM Multimedia, October 2004.

[4] R. Zhang, S. L. Regunathan, K. Rose, "Video coding with optimal inter/intra mode switching for packet loss resilience," in IEEE J. S. A. Comm., vol. 18, no. 6, June 2000.

[5] M. Fumagalli, M. Tagliasacchi, S. Tubaro, "Expected Distortion of Video Decoded DCT-Coefficients," International Workshop VLBV, Sardinia (Italy), Sept. 2005

[6] A. Leontaris, and P. Cosman, "Video compression for lossy packet networks with mode switching and a dual-frame buffer," Image Proc. , IEEE Trans. on, vol. 13, no. 7, July 2004.

[7] V. Bocca, M. Fumagalli, R. Lancini, S. Tubaro, "Accurate Estimate of the Decoded Video Quality: Extension of ROPE Algorithm to Half-Pixel Precision," PCS, San Francisco, Dec. 2004.

[8] H. Yang and K. Rose, "Advances in Recursive Per-Pixel Estimation of End-to-End Distortion for Application in H.264," Proc. ICIP 2005, Genova (Italy), Sept. 2005.