

# EXPECTED DISTORTION OF DCT-COEFFICIENTS IN VIDEO STREAMING OVER UNRELIABLE CHANNEL

M. Fumagalli<sup>\*</sup>, M. Tagliasacchi<sup>†</sup>, S. Tubaro<sup>†</sup>

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

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

## ABSTRACT

The Recursive Optimal per-Pixel Estimate (ROPE) algorithm allows the encoder to estimate the pixel-by-pixel expected distortion of the decoded video sequence due to channel loss. The algorithm requires in input an estimate of the packet loss rate and the knowledge of the error concealment technique used at the decoder with no need to perform any comparison between original and decoded frames.

Although the ROPE algorithm computes the expected distortion in the pixel domain, in some applications it is important to have access to the expected distortion in the DCT domain, e.g., for an accurate allocation of the redundancy bits in error-resiliency schemes.

This paper presents the extension of the ROPE algorithm in the transform DCT domain that allows estimating the expected distortion of the decoded video sequence for each DCT coefficient.

## 1. INTRODUCTION

Nowadays, sending a video sequence over a network that does not provide any QoS guarantee (e.g., IP network) is a very common application. If errors occur, some information does not reach the decoder and error-concealment techniques cannot completely avoid error propagation due to inter-frame dependency introduced by predictive encoding. In the design of error-resilient approaches, it is interesting for the sender to estimate the decoded video distortion that the receiver is expected to suffer.

The Recursive Optimal per-Pixel Estimate (ROPE) algorithm [1] allows the encoder to calculate the pixel-by-pixel expected distortion of the decoded video due to channel loss. In the case of packet-switched networks, the only required parameters are an estimate of the network Packet Loss Rate (PLR) and the error concealment technique used at the decoder side. The encoder does not know the loss pattern, hence it has to characterize the actual reconstruction of a pixel value operated by the decoder as a random variable. The algorithm in [1] is based on the assumption that the considered video sequence is encoded with integer-pixel motion vectors.

Since most of the common encoders implement a sub-pixel precision motion estimation, the work in [5] proposes an extension of the ROPE algorithm in order to obtain an accurate solution for this case.

The knowledge of the expected distortion in the reconstructed video sequence is a valuable information in several applications e.g., when comparing different encoding techniques in various scenarios (in [3] the competitors are MDC and optimized one-layer encoding), tuning of the encoder parameters (in [1], [2], and [4] an estimate of the decoded video distortion, due to packet losses, is used for mode decision and to improve the error resilience of the stream), etc. In all these approaches the expected distortion is computed in the pixel domain and applied at macro-block level.

In contrast with the above-mentioned approaches, in several emerging applications, it can be interesting to have access to an estimate of the expected distortion directly in the DCT domain. This information can be used, e.g., for an accurate allocation of the redundancy in error-resiliency schemes. More recently, error resiliency tools based on the principles of distributed source coding have appeared in the literature [6][7][8]. In [8] an estimate of the expected distortion for each DCT coefficient is used to drive the rate allocation of the side channel, i.e. a redundant representation that is used to correct errors at the decoder, thus stopping drift. The fundamental idea behind the work in [8] is that the encoder needs to characterize in statistical terms the induced channel noise between the original sequence and the reconstructed sequence at the decoder (the so-called side-information in distributed source coding jargon). Intuitively, the higher is the estimated noise level, the larger is the number of bits that need to be spent for the side channel in order to correct errors. In [8], a simple approximation of the DCT coefficients distortion is given by an extension of ROPE algorithm to the transform domain. Simulation results reveal that this approach leads to a coarse approximation of the estimated distortion. In this work we propose a more accurate algorithm to estimate the expected decoded distortion of each DCT coefficient.

This paper is organized as follows: Section 2 presents a brief overview of ROPE algorithm and its extension to half-pixel precision motion estimation. In Section 3 the DCT extension of ROPE presented in [8] is briefly summarized. Section 4 describes the proposed video distortion estimation algorithm (EDDD) in the DCT domain. Section 5 shows some simulation results and the conclusions are given in Section 6.

## 2. ROPE ALGORITHM OVERVIEW

In this section the original ROPE algorithm [1] is briefly summarized together with its extension to half-pixel precision motion-estimation as proposed in [5].

### Integer-pixel ROPE

The expected distortion at the decoder for pixel  $i$  in frame  $n$  can be expressed as in Eq. (1)

$$d_n^i = E\left\{\left(f_n^i - \tilde{f}_n^i\right)^2\right\} = \left(f_n^i\right)^2 - 2f_n^i E\left\{\tilde{f}_n^i\right\} + E\left\{\left(\tilde{f}_n^i\right)^2\right\} \quad (1)$$

where  $f_n^i$  is the value of pixel in original video, and  $\tilde{f}_n^i$  is the decoder reconstruction;  $p$  is the probability of a packet to get lost. We assume this quantity is known at the encoder side. According to Eq. (1),  $\tilde{f}_n^i$  first and second moments are required to compute the distortion. We assume a simple motion-compensated temporal error-concealment scheme, i.e. the lost MB is replaced with the one in the previous frame pointed by the motion vector (MV) of the above MB; if not available, a simple zero-motion replacement is used.

For an Inter-coded MB (similar equations can be written for the Intra case [1]), the ROPE algorithm computes recursively the first and second moments, frame after frame, for each pixel:

$$E\left\{\tilde{f}_n^i\right\} = (1-p)\left(\hat{e}_n^i + E\left\{\tilde{f}_{n-1}^j\right\}\right) + p(1-p)E\left\{\tilde{f}_{n-1}^k\right\} + p^2E\left\{\tilde{f}_{n-1}^i\right\} \quad (2)$$

$$E\left\{\left(\tilde{f}_n^i\right)^2\right\} = (1-p)\left(\left(\hat{e}_n^i\right)^2 + 2\hat{e}_n^iE\left\{\tilde{f}_{n-1}^j\right\} + E\left\{\left(\tilde{f}_{n-1}^j\right)^2\right\}\right) + p(1-p)E\left\{\left(\tilde{f}_{n-1}^k\right)^2\right\} + p^2E\left\{\left(\tilde{f}_{n-1}^i\right)^2\right\} \quad (3)$$

$\hat{f}_n^i$  is the correct reconstruction of the considered pixel and  $\hat{e}_n^i$  is the reconstruction of the prediction error ( $i$  index refers to actual pixel location,  $j$  to the location pointed by the actual MV and  $k$  by the reconstructed MV).

If the video is encoded at half-pixel precision, the algorithm proposed in [1] rounds the half-pixel precision MVs to integer values. This solution leads to a sub-optimal distortion estimate that is often unacceptable.

### Half-pixel ROPE

In [5] the ROPE algorithm is extended to deal with half-pixel precision MVs. The expression of a half-pixel  $j$ ,

interpolated horizontally or vertically from two adjacent integer-pixels  $j_1$  e  $j_2$ , is reported in Eq. (3)

$$f_{n-1}^j = \left\lfloor \frac{r + f_{n-1}^{j_1} + f_{n-1}^{j_2}}{2} \right\rfloor \quad (3)$$

where  $\lfloor \cdot \rfloor$  represents the integer part and  $r$  is a polarization terms (to obtain a zero-mean quantization error). Eq. (4) and (5) present the expressions of the first and second moments of Eq. (3), neglecting the integer rounding:

$$E\left\{\tilde{f}_{n-1}^j\right\} = \frac{r + E\left\{\tilde{f}_{n-1}^{j_1}\right\} + E\left\{\tilde{f}_{n-1}^{j_2}\right\}}{2} \quad (4)$$

$$E\left\{\left(\tilde{f}_{n-1}^j\right)^2\right\} = \frac{1}{4}\left(r^2 + E\left\{\left(\tilde{f}_{n-1}^{j_1}\right)^2\right\} + E\left\{\left(\tilde{f}_{n-1}^{j_2}\right)^2\right\}\right) + 2r\left(E\left\{\tilde{f}_{n-1}^{j_1}\right\} + E\left\{\tilde{f}_{n-1}^{j_2}\right\}\right) + 2E\left\{\tilde{f}_{n-1}^{j_1}\tilde{f}_{n-1}^{j_2}\right\}. \quad (5)$$

In Eq. (4) and (5) every term is known but the last one in Eq. (5), which represents the expectation of the product of two adjacent integer pixels<sup>1</sup>. If the two pixels  $j_1$  and  $j_2$  of frame  $n$  belong to the same MB, in every possible packet loss pattern they have the same MV. The work in [5] proposes to calculate the expected value of the product of two integer adjacent pixels as a linear combination of the various expected values of products of the adjacent pixels from whom they are predicted in the different loss scenarios (analogue equations for the Intra case can be found in [5]):

$$E\left\{\tilde{f}_n^{i_1}\tilde{f}_n^{i_2}\right\} = (1-p) \cdot \left(\hat{e}_n^{i_1}\hat{e}_n^{i_2} + \hat{e}_n^{i_1}E\left\{\tilde{f}_{n-1}^{j_2}\right\} + \hat{e}_n^{i_2}E\left\{\tilde{f}_{n-1}^{j_1}\right\} + E\left\{\tilde{f}_{n-1}^{j_1}\tilde{f}_{n-1}^{j_2}\right\}\right) + p(1-p)E\left\{\tilde{f}_{n-1}^{k_1}\tilde{f}_{n-1}^{k_2}\right\} + p^2E\left\{\tilde{f}_{n-1}^{i_1}\tilde{f}_{n-1}^{i_2}\right\} \quad (6)$$

If the MV associated to the pair of integer pixels  $i_1$  and  $i_2$  is a half-pixel precision MV, the reference pair of pixel is formed by half-pixels (called  $j_1$  and  $j_2$ ). A possible configuration is shown in Figure 1:



Figure 1 - Example of two horizontally adjacent pixels ( $j_1$  and  $j_2$ ) that are used as references in MC operation. Their values come from interpolation of the consecutive pixels  $a$ ,  $b$ , and  $c$ .

In order to calculate the expected value of the product of half-pixels  $j_1$  and  $j_2$ , given the known expected products between integer ( $a, b$ ) and ( $b, c$ ), the product can be developed substituting  $j_1$  and  $j_2$  with their expression as in Eq. (3). In this way a new unknown term is needed: the expected value of the product between not adjacent

<sup>1</sup> The work in [2] proposes to approximate the expectation of the product of two adjacent integer pixels with its superior limit indicated by Cauchy-Schwartz inequality.

integer pixels  $a$  and  $c$ . The problem is tackled using the following well-known relation

$$E\{\tilde{f}_{n-1}^a \tilde{f}_{n-1}^c\} = \rho_{a,c} \sigma_a \sigma_c + \mu_a \mu_c \quad (7)$$

where  $\mu$  e  $\sigma$  represent the expected value and the standard deviation of reconstructed pixel (known by Eq. (2) and (3)) and  $\rho_{a,c}$  is the correlation coefficient between  $\tilde{f}_{n-1}^a$  e  $\tilde{f}_{n-1}^c$ ;  $\rho_{a,c}$  is then approximated (assuming an autoregressive linear model between adjacent pixels) by:

$$\rho_{a,c} = \rho_{a,b} \cdot \rho_{b,c} \quad (8)$$

In case pixels  $i_1$  and  $i_2$  belong to different MBs (it happens less than 10% of the cases), they are supposed to be uncorrelated and the expected values of products of the adjacent pixels are approximated by the product of the expected values. The results presented in [8] show the accuracy of this solution on the whole sequence.

### 3. DCT-ROPE

In [8] 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 [1] on the DCT coefficients instead of working on pixel values. This straightforward solution (that we call here DCT-ROPE) has to cope with the difficulty of managing the motion compensation phase in the transform domain.

In order to map DCT coefficients from the current to the reference frame, motion vectors are quantized with a step size equal to the block side length in such a way that each macroblock of the current frame is matched with the nearest macroblock of the reference frame. This coarse approximation of motion vectors leads to a loss of accuracy in the distortion estimation as shown by the simulation results reported in Section 5.

### 4. EXPECTED DISTORTION OF DECODED DCT-COEFFICIENTS (EDDD)

This section illustrates the proposed algorithm that 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 of 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 [8], where the estimate was obtained directly in the DCT domain.

As we saw in the previous section, the ROPE 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 similar way, for each image block, we represent each of the  $i$ -th DCT coefficient as a statistical variable  $y_i$ . In order to characterize the expected distortion we need to estimate both the first ( $E[y_i]$ ) and the second ( $E[y_i^2]$ ) moments. In the following, we refer to the proposed approach as Expected Distortion of Decoded DCT-coefficients (EDDD).

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. (9).

$$y_i = \sum_j w_j^i \cdot x_j \quad (9)$$

Using Eq. (9) we obtain the expression of  $E[y_i]$  and  $E[y_i^2]$  as reported in Eq. (10) and Eq. (11).

$$E[y_i] = \sum_j w_j^i \cdot E[x_j] \quad (10)$$

$$E[y_i^2] = \sum_z \sum_j w_z^i \cdot w_j^i \cdot (E[x_z x_j] - E[x_z] \cdot E[x_j]) + E[y_i]^2 \quad (11)$$

Eq. (10) is of easy calculation while Eq. (11) presents an unknown term that represents the expected value of the product of pixels pairs within the considered block. Although this quantity is not needed in the full pixel precision version of ROPE [1], the work in [5] introduces an extension of ROPE to work with half-pixel precision, as briefly summarized in Section 2. In particular, the additional recursive Eq. (6) gives the expected value of the product of adjacent pixels in both horizontal and vertical directions. The values of  $E[x_z x_j]$  for not adjacent pixels  $z$  and  $j$  are obtained by the combination of the expected values of the product of the adjacent pixels that connect  $z$  with  $j$ . This latter approximation is the only cause of inaccuracy of the proposed EDDD algorithm.

Besides its accuracy, the proposed approach has the significant feature of energy preservation over each block, i.e., the distortion calculated by ROPE algorithm over each block in pixel domain is exactly the same distortion calculated by EDDD algorithm in the DCT domain. In fact it is easy to show that this feature is not affected by the aforementioned simplification and thus it is always true.

### 5. EXPERIMENTAL RESULTS

This section compares the proposed EDDD approach with the ROPE algorithm applied directly in the DCT domain (DCT-ROPE). In order to assess the absolute accuracy and consistency of both approaches, they are compared with the actual distortion of the reconstructed sequence averaged over several network simulations with different error patterns.

Figure 2 illustrates the PSNR tracks at frame level of the decoded ‘Foreman’ sequence (QCIF, 30 fps, 256 kbps, Intra MB refresh 10%) subjected to 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 energy conservation feature of EDDD. The inaccuracy of DCT-ROPE approach is evident. Similar results can be obtained over other video sequences.

Figure 3 shows the accuracy of EDDD and DCT-ROPE algorithms in the DCT domain. Each 8x8 DCT block is divided into four not-overlapped 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). As we can see, 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 video sequences at different PLRs lead to the same results.

To conclude, we showed that the proposed EDDD approach in DCT domain presents the same expected distortion of ROPE at frame level and that it estimates with considerable accuracy the expected distortion in the DCT domain.

## 6. CONCLUSIONS

This paper presents an extension of ROPE algorithm that is able to estimate the expected distortion in the DCT domain. The basic idea of the proposed EDDD approach is to run the original ROPE algorithm in the pixel domain and then to estimate the DCT domain, block-by-block, using the corresponding spatial information. The accuracy of the EDDD is validated by several simulation results.

## REFERENCES

[1] R. Zhang, S. L. Regunathan, and 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.

[2] 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.

[3] A. Reibman, “Optimizing multiple description video coders in a packet loss environment,” in Proc. of 12th PV Workshop, April 2002, Pittsburgh, PA, US.

[4] R. Zhang, S. L. Regunathan, and K. Rose, “Switched error concealment and robust coding decisions in scalable video coding,” in Proc. ICIP 2003 Sept. 14-17, Barcelona (Spain).

[5] V. Bocca, M. Fumagalli, R. Lancini, S. Tubaro, “Accurate Estimate of the Decoded Video Quality: Extension of ROPE Algorithm to Half-Pixel Precision,” in Proceedings of PCS 2004, December 15-17, 2004, San Francisco (CA).

[6] S. Rane, A. Aaron and B. Girod, “Systematic lossy forward error protection for error-resilient digital video broadcasting - A Wyner-Ziv coding approach,” Proc. IEEE International Conference on Image Processing, ICIP-2004, Singapore, Oct. 2004

[7] A. Sehgal, N. Ahuja, “Robust Predictive Coding and the Wyner-Ziv Problem”, Data Compression Conference, Snowbird, Utah, Oct. 2003

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

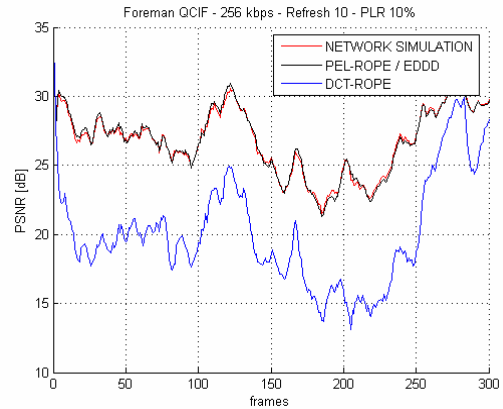


Figure 2. PSNR tracks for EDDD, DCT-ROPE and NS.

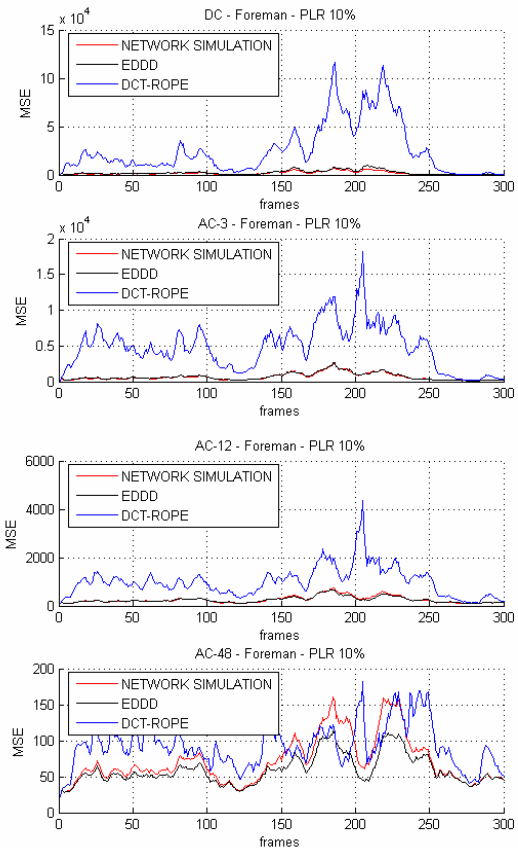


Figure 3. MSE in the four not-overlapped DCT bands.