ERROR-PROPAGATION REDUCTION IN A BALANCED MULTIPLE DESCRIPTION VIDEO CODER

Nicola Franchi, Marco Fumagalli and Rosa Lancini

CEFRIEL – Politecnico di Milano, R. Fucini, 2 – 20133 Milano, Italy. E-mail: rosa@cefriel.it

ABSTRACT

Compressed video streams are very exposed to channel losses when they are transmitted through an unreliable network. The predictive encoding scheme is efficient but even a single packet loss can cause decoding error propagation through all the frames following the erasure. An error-resilience approach to tackle this problem is the multiple description coding (MDC) technique. Another important issue connected to this approach is the problem of the error-concealment (EC) used to reconstruct the information lost during the transmission.

In this paper we deal with sequence-based EC algorithm that performs the error-concealment by using the information coming from a certain number of frames in the future to minimize the effect of error propagation. Target of this work is applying such EC strategy to a balanced MDC system. Simulation results show the great improvement that a sequence-based EC algorithm can give compared to classical EC approaches.

1. INTRODUCTION

In recent years the volume of multimedia data transmitted over best-effort networks (such as the Internet) has continued to increase while packet loss and delay still remain an issue. Scientific literature and standards propose to encode the video using predictive coders. These kinds of coders are commonly used because they exploit the temporal source correlation to achieve better compression. The main drawback of a predictive coding scheme is that even a single packet loss can cause decoding error propagation through all the samples or frames following the erasure.

An approach to tackle the problem of error propagation is applying a multiple description coding (MDC) technique [1]. The MDC approach consists in generating two or more data streams containing descriptions of the source that, when combined, provide a good approximation of the source, but that can also be decoded independently. The source descriptions are sent to the receiver over different channels or, in the context of a best-effort network, they are sent over the same network but they are expected to experience independent losses. In balanced MDC the two descriptions are mutually refining while in unbalanced (UMDC) one description has a high quality and the other has low, but acceptable, quality. In this scheme the low-resolution description is primarily used as redundancy, to be used only in case of error in the high-resolution one.

Many approaches have been proposed to realize MD video coding [2], [3], [4], [5] which provide a strong robustness against packet loss especially at high loss rate.

Another important issue in robust video transmission is the problem of error-recovery. In fact it is necessary applying an error-concealment (EC) strategy in order to better reconstruct the information lost during the transmission.

The literature proposes two classical solutions: temporal and spatial EC. The former method consists on using the information coming from the previous received frame while the latter uses the information correctly received about the actual frame itself.

An alternative EC approach for video is proposed by Ortega *et al.* [4]. They present a sequence-based algorithm that performs the error-recovery by using the information coming from a certain number of frames in the future to minimize the effect of error propagation. This EC approach is effective in predictive encoders because it exploits the temporal correlation between the frames. To reach such a result, a side-information stream is needed. Intuitively we can think at this stream as an approximated version of the video source, e.g., a low-resolution stream. Thus we can associate each pixel in the video sequence to a correspondent approximated value of the side-information stream. Different systems are designed to obtain side-information in different ways. The work in [4] is built on an UMDC system and the side-information is carried by a low-resolution stream that takes a little percentage of the total bit-rate according to the packet loss rate of the channel. The results are interesting but the system suffers, among the others, the problem of the error-free transmission (extra-bandwidth) of the LR stream. The main drawback of this EC algorithms is to have an additional latency connected to the number of frames in the future used to perform the error recovery

0-7803-8554-3/04/$20.00 ©2004 IEEE.

and, e.g., in [4], the presence of an extra-bandwidth LR stream.

Target of this work is to apply a sequence-based error-concealment strategy to a Balanced MDC system. In fact the BMDC systems have the advantage that each description can be independently decoded. In this way, if a loss corrupts a description, then the other descriptions are error-free and can be used to create the side-information necessary to apply a sequence-based EC, without the addition of any other LR stream that would be necessary when using an UMDC approach like that reported in [4]. We think this technique can be effective in both the MDC schemes that use or not drift-compensation terms. We expect the EC strategy performs better when the number of descriptions is greater than two [5]. This last consideration is motivated by the chance, for these MDC schemes, to obtain a lower side-distortion and thus a higher quality side-information stream that increases the effectiveness of the EC algorithm.

In this work we refer to a system (called Independent Flow Multiple Description Video Coding, IF-MDVC [5]) based on a spatial polyphase down-sampler along columns to provide up to four descriptions per frame. We focus our attention on this system to show the effectiveness of sequence-based EC algorithm on BMDC system, but we believe that similar results can be achieved on other BMDC systems.

The rest of the paper is organized as follows. Section 2 presents in details the sequence-based EC strategy, while Section 3 is dedicated to the IF-MDVC system: review of the algorithm, creation of side-information stream from the correctly decoded descriptions, and application of the sequence-based EC algorithm. Section 4 shows the simulation results and then the conclusions.

2. SEQUENCE-BASED ERROR-CONCEALMENT ALGORITHM

Given a video source *v*, we consider coding it in predictional way. It means that each pixel *x* at the time *n* can be expressed as the sum of its reference value in the previous frame *n-1* and a prediction error *e* as in Equation (1)

$$x_i^v(n) = x^v_{i+MV_i^v(n)}(n-1) + e_i^v(n) \qquad (1)$$

where *i* refers to the pixel location and MV_i is the motion vector related to the pixel *i*, i.e., the location of the reference value in frame *n-1* of the pixel *i* is $i+MV_i(n)$.

Such a scheme provides excellent coding efficiency but it suffers by the error propagation drawback. In fact if at the frame *k* the prediction error $e_i^v(k)$ and the motion vector $MV_i^v(k)$ are lost, then, according to equation (1), $x_i^v(k)$ is unknown. In turn, the pixels in frame *k+1* that uses $x_i^v(k)$ as reference value cannot be correctly reconstructed, and so on in successive frames.

The goal of the sequence-based EC algorithm is to recover the pixel value $x_i^v(k)$ in the frame *k* where we had the loss, trying to minimize the effect of error propagation in a certain number *F* of frames in the future.

Before describing the algorithm, we state the sequence-based EC strategy needs at the decoder side (where the EC algorithm is performed) a side-information stream. Intuitively we can think of this stream as an approximated version of the video source, e.g., a low-resolution stream. Thus at the decoder side, we can associate each pixel in the video sequence with a correspondent approximated value coming from the side-information stream. In this paper we will try to obtain the side-information for a lost description from the other descriptions not corrupted by the losses.

Given the side-information $x_i^a(k)$, the sequence-based EC algorithm tries to restore the lost pixel value $x_i^v(k)$ with a value $\hat{x}_i^v(k)$ that minimizes the average distortion (in terms of MSE over *F* frames in the future) between the video sequence, that we obtain substituting the lost value with $\hat{x}_i^v(k)$ and bringing ahead the decoding process for *F* frames in the future, and the corresponding side-information stream. In other words the reconstructed pixel value in frame *k* minimizes the error propagation (considering *F* frames in the future) respect to the side-information. When the quality of the side-information increases, then the accuracy of the EC algorithm increases too. Moreover, there is the same effect when we enlarge *F*.

More formally, the sequence-based EC algorithm minimizes the mean square error, *E(F)*, as defined in equation (2) [4]. The first two terms of the sum in equation (2) represent the decoded video sequence starting from the reconstructed value and the corresponding side-information stream:

$$E(F)=\sum_{f=0}^{F-1}\left[\hat{x}^v(k)+\left(\sum_{s=1}^{f}e^v(k+s)\right)-x^a(k+f)\right]^2 \qquad (2),$$

where the contribution of the side-information and that of the prediction error in each frame are determined following the path indicated by the motion vectors. Equation (2) is then differentiated with respect to $\hat{x}^v(k)$ to determine the optimum reconstructed value of the lost pixel, which is shown in equation (3):

$$\hat{x}^v(k)=\frac{1}{F}\sum_{f=0}^{F-1}\left[x^a(k+f)-\left(\sum_{s=1}^{f}e^v(k+s)\right)\right] \qquad (3).$$

A different interpretation of equation (3) suggests the EC implementation scheme. In fact the reconstructed value can be seen as the average of F values that corresponds to the F side-information samples retro-propagated (up to the lost frame) through the right received error prediction values. Figure 1 shows the error-recovery functional scheme in the case of F=4.

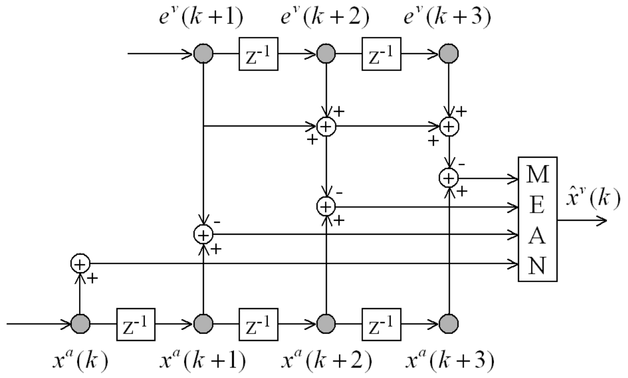

Figure 1. EC scheme for F=4 and loss on the flow *f* at frame *k*. e^v are the prediction errors and x^a the side-information samples.

3. SEQUENCE-BASED ALGORITHM APPLIED ON IF-MDVC SYSTEM

The IF-MDVC system proposes to split the video source in N separate flows [5]. For $N=4$ the input source is fed into the MD block that provides 4 independent flows for each frame by applying a polyphase down-sampler (*x2*) along rows and columns. Then each description is coded using conventional hybrid encoder architecture. The whole coding process can be seen as the application of a predictive coder to 4 distinct versions of the same video source.

At the decoder side, if the transmission is error-free, the four video flows are merged to restore the full resolution. If packet loss occurs, an EC algorithm is applied. As well in [5] it is shown how the spatial EC performs better than the temporal one in the case of $N=4$.

In the rest of this paper we consider the case of IF-MDVC, but we think similar results can be obtained by considering any other MDC system.

3.1. Side-Information Creation from the Correctly Decoded Descriptions

Let F be the number of frames considered in the future and $N=4$ the number of flows. For sake of simplicity, let us consider first the case of single loss (on the flow d at frame k) and only the usage of predictional coded frames inside the stream.

In this simple loss scenario, the only flow affected by error propagation is the flow d. The other three flows can be correctly decoded. What we propose is to obtain the side-information $x^a(k+j)$ for the flow d at the frame $k+j$ ($\forall\ 0 \le j < F$) by spatial interpolating the three correctly received descriptions at the frame $k+j$. We highlight that this interpolation process is a mere spatial operation and thus it proceeds frame by frame. In such way we can obtain an approximation of the pixel values for the description d (i.e., its side-information) for every frame from the loss on.

Let us consider now the more complicated case of multiple losses over the flows. The goal is to recover the single loss in the flow d at frame k. Two events are worth taking into account: first the loss of other data on the same flow d(in a frame $k+j$ with $0 \le j < F$) and second the loss of data within the other three flows that can prevent obtaining the side-information for flow d.

The problem is thus to estimate the maximum number F' of frames that is possible to consider in the sequence-based EC algorithm according to the experimented losses. We find that F' can be obtained by Equation (4). Let us briefly explain the meaning of the three terms:

$$F' = \min\{F, L, M\} \tag{4}$$

- F is the maximum number of frames that can be considered, assigned by the application. Therefore F' cannot exceed F.
- L is the number of frames in flow d between frame k (where we experimented the loss we are recovering) and the eventual successive loss (say frame h). L can be expressed as $L=h-k$. In fact, if we experiment a loss in frame h (we lose MVs and predictional error) we are not able to apply further the Equation (3) and we are obliged to stop it at frame $k + L$.
- M takes into account the losses in the other three flows different from d. In fact, if we experiment a loss in a flow m ($m \neq d$) at the frame $k+j$ then the description m cannot be decoded after that frame and thus flow m cannot be further used for the creation of the side-information from that frame on. M is the number of last frame where at least one description m ($m \neq d$) can be decoded.

Given that the proposed EC algorithm proceeds independently loss-by-loss, we describe its steps for a single loss (flow d, frame k):

- Determination of F' given the position of the experimented losses in the frames $k+j$ with $0 \le j < F$.
- Creation of the side-information for flow d for the frames $k+j$ with $0 \le j < F'$. According to the definition of F', we are certainly able to create a side-information for each frame in the considered interval.
- We restore the lost samples by using, pixel by pixel, the Equation (3), substituting F' to F.

4. SIMULATION RESULTS

In this section we present the simulation results when we apply a sequence-based EC algorithm on IF-MDVC system. We analyze the case of 4 descriptions.

We encode 'Foreman' and 'Akiko' sequences (QCIF, 7.5 fps) with IF-MDVC (N=4) with fixed quantizer at 20. We transmit the coded MD streams over an unreliable channel with variable packet loss rate form 0 to 20%. In different experiments we use different window lengths for the sequence-based EC algorithm (*F* from 1 to 10). We remind that *F*=1 means using, in the EC algorithm, only the frame where we experiment the loss and thus it is equivalent to the classical spatial EC.

Figure 2 presents the average PSNR for 'Foreman' sequence at different PLR and number *F* of frames used in EC algorithm. The top line (PLR=0%) refers to error-free situation where, of course, the window length of the EC algorithm does not affect the PSNR. On the contrary, it is evident the PSNR improves increasing *F* in error-prone situations (PLR=3-20%). The gain is more evident at high PLR where it is about 2.5 dB passing from *F*=1 (classical spatial EC) to *F*=10. It is interesting to note that the significant gain is done passing from *F*=1 to use few frames in EC algorithm (*F*=2 or 3) and then the performance tends to not increase very much using *F*>4. Intuitively, one reason it that, even if we set *F* to a high value, then the effective number of frames used in the EC algorithm is *F'*, as defined in (4), that strongly depends on the packet loss rate. In fact, e.g., at PLR=20%, using *F*=5 or 10 produces a very similar average values of *F'* due to the presence of high loss rate that leads to similar PSNR for both the cases. Similar consideration can be done for 'Akiko' sequence, whose performance is reported in Figure 3. The goodness of the approach is confirmed by Figure 4 that reports for 'Foreman' the PSNR tracks in the time at PLR=5% for several values of *F*.

5. REFERENCES

[1] V. A. Vaishampayan, "Design of multiple description scalar quantizer," IEEE Trans. Inf. Th., vol. 39, May 1993.

[2] Reibman, H. Jafarkhani, Y. Wang, M. Orchard, and R. Puri, "Multiple description coding for video using motion compensated temporal prediction," IEEE Trans. Circuits Syst. Video Technol., vol. 12, pp. 193–204, Mar. 2002.

[3] J. G. Apostolopoulos, "Error-resilient video compression via multiple state streams," in Proc. Int. Workshop Very Low Bit-rate Video Coding (VLBV'99), Kyoto, Japan, Oct. 1999.

[4] M. Fumagalli, A. Ortega, P. Sagetong, "Estimation of Erased Data in a H.263 Coded Stream by Using Unbalanced Multiple Description Coding," in Proc. ICME, Baltimore, US, July 2003.

[5] N. Franchi, M. Fumagalli, and R. Lancini, "A Real-Time N-Descriptions Video Coding Architecture," in proc. VLBV 2003, Madrid, Spain, September 2003.

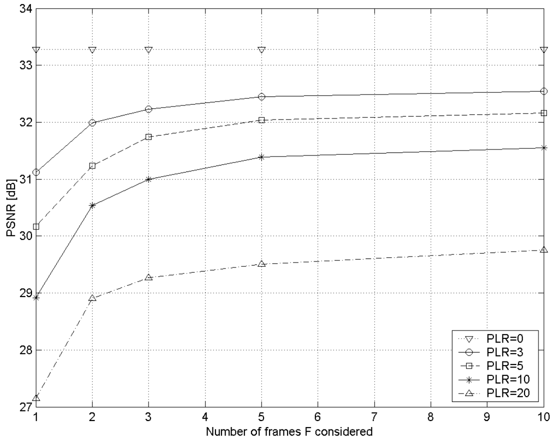

Figure 2. Average PSNR for 'Foreman' sequence at different PLR and number F of frames used in EC algorithm.

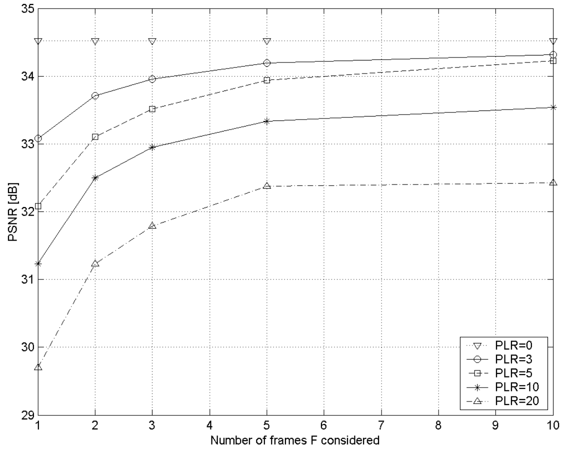

Figure 3. Average PSNR for 'Akiko' sequence at different PLR and number F of frames used in EC algorithm.

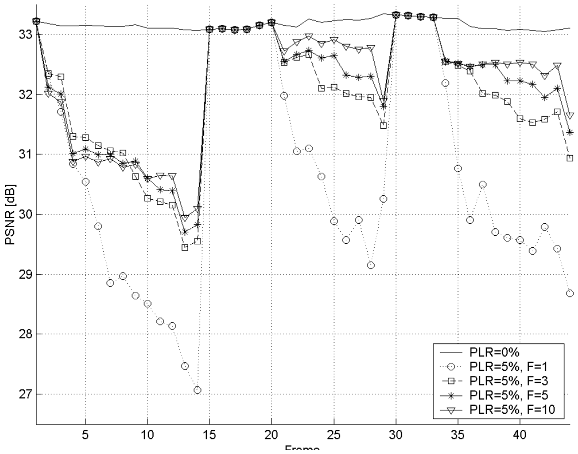

Figure 4. PSNR tracks in the time for 'Foreman' at PLR=5% for different values of F.