A SEQUENCE-BASED ERROR RECOVERY ALGORITHM FOR A MULTIPLE DESCRIPTION P2P VIDEO TRANSMISSION SYSTEM

Marco Fumagalli (*), Rosa Lancini (*), Stefano Tubaro (†)

(*)CEFRIEL - Politecnico di Milano, via R. Fucini 2 - 20133 Milano, Italy {fumagall,rosa}@cefriel.it
(†)Dip. di Elet.e Inf., Politecnico di Milano, P.zza L. Da Vinci, 32 - 20133 Milano, Italy. tubaro@elet.polimi.it

ABSTRACT

Peer-to-peer (P2P) networks are distributed systems that aggregate a large amount of independent nodes known as peers. This model is expected to resolve observed limitations in current centralized video streaming solutions and to improve in a significant way their performance.
Multiple description coding (MDC) is a video encoding technique that presents interesting features when applied to P2P video streaming system, e.g., it eases the management of variable bandwidth/throughput by transmitting a suitable number of descriptions, it provides high resilience to packet losses improving decoded video quality as the peer connecting probability increases.
The literature shows that MDC could be an effective approach to realize effective P2P networks for video streaming.
In fact a key issue in video streaming is the problem of error-recovery at the decoder side, i.e., the error-concealment (EC) strategy. The main contribution of this paper is the design of a sequence-based EC 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. Target of this work is to apply such EC strategy to a balanced MDC system. Simulation results show the great improvement that a sequence-based EC algorithm gives to the system performance compared to common EC approaches.

1. INTRODUCTION

In recent years the volume of video/audio data transmitted over best-effort networks (such as the Internet) is increased in an extraordinary way. Peer-to-peer (P2P) traffic will take accordingly a non negligible amount of the global Internet exchange in the near future.
P2P networks [8] are distributed systems that aggregate a large amount of heterogeneous and independent nodes known as ‘peers’. Such a system encompasses interesting characteristics like self configuration, self adaptation and self organization. These intrinsic characteristics make the P2P model a potential candidate to solve some of the problems of today multimedia streaming over the Internet; in fact P2P networks overcome the setback of bottleneck around centralized server due to its distributed design and architecture. Moreover, it facilitates to manage dynamically the available resources in the networks since they scale with the number of peers in the systems.
Although P2P technology gives novel opportunities to define an efficient multimedia streaming application, at the same time it brings a set of technical challenges and issues due to its dynamic and heterogeneous nature. Even though the problem has been already studied in the literature [9, 10], works on P2P media streaming systems is still in the early stages.
One of the key issues in designing P2P systems is the choice of video coding technique and related challenges in order to optimize the resource usage and to improve substantially the overall video quality. Packet loss and error propagation, which occur frequently over current IP networks, can dramatically reduce the video quality at the receiver end. Hence, error resilience and handling packet loss are critical issues in the streaming applications. Several coding solutions have been developed to tackle these issues, to enhance the overall quality and to protect multimedia traffic against severe network congestion conditions. These systems mainly use two techniques for video encoding: multiple description coding (MDC) and layered coding (LC). LC provides a scalable representation that enhances rate control but it is sensitive to transmission losses. On the other hand, MDC provides increased resilience to packet losses by creating multiple streams that can be decoded independently.
Both MDC and LC techniques provide interesting features when applied to P2P video streaming system. They facilitate the management of variable bandwidth/throughput by transmitting a suitable number of descriptions/layers. Despite the “standard P2P video systems” where the decoding process starts when all the sub streams are fully received, the systems based on MDC and LC techniques can allow a fast start of the decoding process and the decoded video quality improves as the peer connecting probability increases. Moreover both MDC and LC techniques give a scalable representation of the source providing videos with different quality to users equipped with different bandwidth connections.
The difference between MDC and LC lies in the dependency among descriptions/layers. In fact in LC paradigm the enhancement layers can be decoded only starting from a decoded base layer. In contrast to LC, in MDC each description can be decoded individually to get the base qual-

ity. This feature gives MDC a slight advantage over LC technique.

In this paper we deal with MDC techniques [1] and in particular with the problem of error-concealment (EC). In fact it is necessary applying an EC strategy in each P2P client in order to reconstruct the information eventually lost during the transmission or when the peer connecting probability is too high and part of some sub stream arrives late.

The literature proposes two classical EC solutions, the temporal and the spatial one. Moreover some MDC approach presents a specific EC algorithm. One of the motivations of this work is that, even though MDC is particularly useful in error-prone environments, the schemes proposed in literature present only a "local" EC strategy (in terms of time) that does not deeply exploit the redundant information of the sub streams [2].

An attractive EC approach for video is proposed by Ortega et al. [3]. 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 of the frames. To reach such result, a side-information stream is needed. Intuitively, we can think at this stream as a more approximated version of the video source, e.g., a low-resolution stream. Therefore, we can associate each pixel in the video sequence to a correspondent more approximated value of the side-information stream. The work in [3] is built on an unbalanced MDC (UMDC) system and the side-information is carried by a low-resolution (LR) stream that takes a percentage of the total bit-rate according to the channel conditions. The results are interesting but the system suffers the problem of the extra-bandwidth transmission of the LR stream.

A similar approach is given in [4]. It proposes an MPEG-compatible spatial and temporal EC algorithm that uses multi-frame recovery principle to minimize the error propagation. In this case the side-information is obtained exploiting the intrinsic spatial/temporal correlation of nearby pixels in the video source.

Aim of this work is to apply a sequence-based error concealment strategy to a balanced MDC (BMCD) system. In fact the advantage of the BMDC descriptions of being independently decodable is a very welcome feature for a P2P video streaming system. If a loss corrupts a description, then the other descriptions are error-free and can be used to create, without any other additional LR stream, the side-information necessary to apply a sequence-based EC. We think this approach can be particularly effective in those MDC schemes that do not use drift compensation terms (e.g., [2,5]) and where the number of descriptions is greater than two [5]. This consideration is motivated by the chance, for these schemes, to obtain a lower side-distortion and thus a higher quality side-information that increases the effectiveness of the EC algorithm.

In this work we refer to IF-MDVC system (Independent Flow Multiple Description Video Coding) [5] based on a spatial polyphase down-sampler along rows and 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 similar results can be achieved on other BMDC systems.

The rest of the paper is organized as follows. Section 2 presents 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. Conclusions are given in Section 5.

2. SEQUENCE-BASED ERROR-CONCEALMENT ALGORITHM

Given a video source, we consider prediction-based coding. Such a scheme provides excellent coding efficiency but it suffers from the error propagation drawback. The goal of the sequence-based EC algorithm is to recover the lost pixel values due to packet losses trying to minimize the effect of error propagation in a certain number F of frames in the future.

Before describing the algorithm, we remind that the sequence-based EC strategy needs at the decoder side (where the EC algorithm is performed) a side-information stream that can be intuitively thought of a coarser approximation of the video source. In this paper we will obtain this side-information for a lost description from the other descriptions not corrupted by the losses.

Given the side-information $x_i^a(k)$ (pixel i at time k; a is for auxiliary, i.e. side-information stream), 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 mean square error 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 at time 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 as well. Moreover, there is the same effect when we enlarge F.

More formally, the sequence-based EC algorithm minimizes the mean square error between the decoded video sequence starting from the reconstructed value and the corresponding side-information stream. The two contributions in each frame are determined following the path indicated by the MVs. The optimum reconstructed value of the lost pixel is shown in Equation (1):

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

The form of Equation (1) 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 predictional error values. Figure 1 shows the error-recovery functional scheme in the case of F=4. More detailed explanation can be found in [6].

Till now we have considered for simplicity the case of integer precision ME/MC coding. Here we analyze the impact of using a half-pixel precision in the proposed sequence-based EC algorithm. This choice is motivated by the fact that the ME/MC in the standard coders is typically carried out with this accuracy.

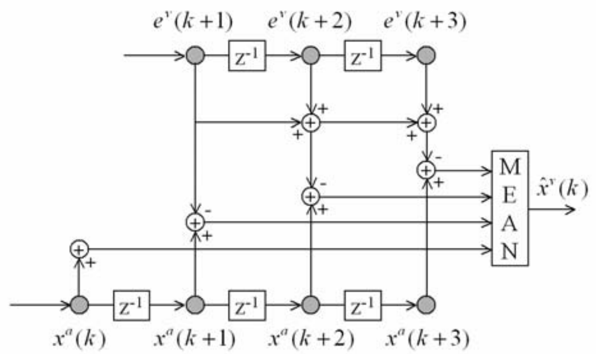

Figure 1. EC scheme for F=4 and loss on the flow f at frame k. e^v are the prediction errors and x^a the side-information samples.

The main issue in this extension is the difficulty to handle the MC-sequences within the video in the case of not-integer precision MVs. In fact in the case of integer MVs, each pixel is predicted from one pixel in the previous frame and thus the prediction relationship between temporal consecutive pixels (i.e., the MC-sequences) are simply individuated. On the contrary, if the MVs have half-pixel precision, the prediction of each pixel comes from the average of more than one pixel in the previous frame. Thus the MC-sequences become complicated structures whose handling turns to be unpractical. However it is easy to show that the key-problem that has to be solved in order to extend the sequence-based EC approach to half-pixel precision MVs is the opportunity to invert the MC operation, i.e., to estimate the pixels of the previous frame given the MVs and the half-pixel prediction frame.

We present a simplified example to explain this idea. Let us consider two frames as in Figure 2: $X(t)$ is a N pixels block in frame t that has a half-pixel precision MV. Thus its prediction $\hat{X}(t)$ comes from the $N+1$ pixels block $X(t-1)$ in frame t-1 by averaging pair-by-pair the near pixels (e.g., $\hat{x}_1(t)=[x_2(t-1)-x_3(t-1)]/2$). The problem we need to solve is to estimate the values of the $N+1$ pixels of $X(t-1)$ given the value of the MV, $X(t)$ and $E(t)$.

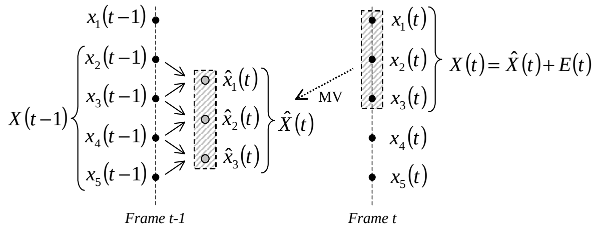

Figure 2. Two consecutive frames where the MV has half-pixel precision.

We formalize the problem as in Equation (2), where T is a $(N+1)xN$ matrix. The goal is to estimate the elements of T. The MC operation imposes the constraint of Equation (3), where S is a $Nx(N+1)$ matrix. Combining Equation (2) and (3) we find Equation (4), which is the first constraint for the variable T: it insures the perfect reconstruction of the prediction $\hat{X}(t)$.

$$X(t-1)=T\cdot\left(X(t)-E(t)\right)=T\cdot\hat{X}(t) \qquad (2)$$

$$\hat{X}(t)=S\cdot X(t-1) \quad \text{with}$$
$$S=\begin{bmatrix} 0.5 & 0.5 & 0 & \dots & 0 \\ 0 & 0.5 & 0.5 & \dots & 0 \\ \dots & & & & \dots \\ 0 & \dots & 0 & 0.5 & 0.5 \end{bmatrix} \qquad (3)$$

$$S\cdot T=I \qquad (4)$$

It is evident that the Equation (4) does not uniquely determinate matrix T; in fact it is impossible to uniquely invert the average operation. Therefore we need to add an additional constraint on T. The idea, depicted in Figure 3, is to impose $X(t-1)$ to be as similar as possible to $\hat{X}(t)$, as the ME/MC operation requires. We apply on $\hat{X}(t)$ an averaging operation similar to that used in MC operation in order to evaluate $\hat{X}(t)$ in the same spatial position of $X(t-1)$. To reach such objective we define $\overline{X}(t-1)$ by pruning the first and the last sample of the block $X(t-1)$; in the same way we define $\overline{T}$ by pruning the first and the last rows of T and $\overline{S}$ by pruning the last row and the last column of S (Equation (5)). In order to impose similarity between $X(t-1)$ and $\hat{X}(t)$ the Equation (6) states the second constraint for T.

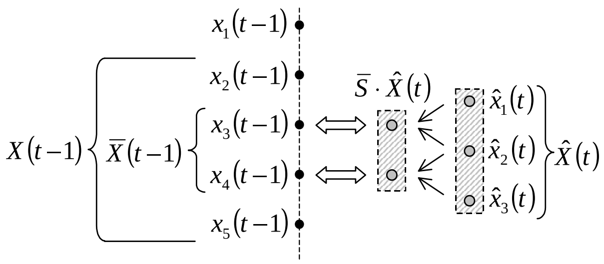

Figure 3. An additional constraint increases the similarity between the half-pixel precision prediction and the previous frame.

$$\overline{X}(t-1) = \overline{T}\cdot\hat{X}(t) \qquad (5)$$

$$\begin{aligned} T &= \arg\min\left(\overline{S}\cdot\hat{X}(t) - \overline{X}(t-1)\right) = \\ &\arg\min\left(\left(\overline{S}-\overline{T}\right)\cdot\hat{X}(t)\right) \\ &\Rightarrow T = \arg\min\left(\overline{S}-\overline{T}\right) \end{aligned} \qquad (6)$$

The matrix T is obtained, e.g., by applying the linear programming method [7] to Equation (4) and (6). It can be shown that the resulting matrix T is a non-polarized accurate operator that inverts the MC operation and it can be advantageously applied in the sequence-based EC algorithm.

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, in a error-free transmission case, the four video flows are merged to restore the full resolution images. If packet loss occurs, an EC algorithm is applied. In the rest of this paper we consider the case of IF-MDVC, but 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 the sake of simplicity, let us consider first the case of single loss (on the flow d at frame k) and only the usage of prediction-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 to obtain 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 (7)

$$F' = \min\{F, L, M\} \qquad (7).$$

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 prediction error) we are not able to apply further the Equation (1) 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). The algorithm steps are:

- 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 $0 \le j < F'$.
- Recovery of the lost samples by using, pixel by pixel, the Equation (1), substituting F' to F.

4. SIMULATION RESULTS

In this section we present the simulation results when the proposed sequence-based EC algorithm is applied on a IF-MDVC system.

We code 'Foreman' sequence (QCIF, 7.5 fps) with IF-MDVC (N=4 descriptions) with fixed quantizer at 20 (97.3 kbps) and a GOP length of 15. We transmit the coded multiple description streams over an erasure channel with variable packet loss rate (PLR) form 0% to 20%. In different experiments we use different values of F for the sequence-based EC algorithm (from 1 to 10). We recall 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.

To show the effectiveness of the algorithm, Figure 4 presents the average PSNR for 'Foreman' sequence at different PLR and number F of frames used in EC algorithm. 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 3 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 be stable using *F*>5. 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 (7), 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 a high loss rate that leads to similar PSNR for both the cases.

The goodness of the approach is confirmed by Figure 5 that reports the PSNR tracks in time at a PLR=5% for different values of *F*. Simulations on other video sequences confirm these results.

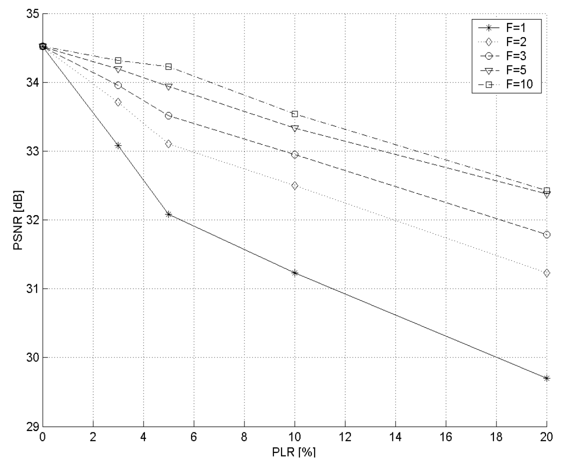

Figure 4. 'Foreman' (QCIF, 7.5 fps) - IF-MDVC (*N*=4). Average PSNR at different PLR and number *F* of frames.

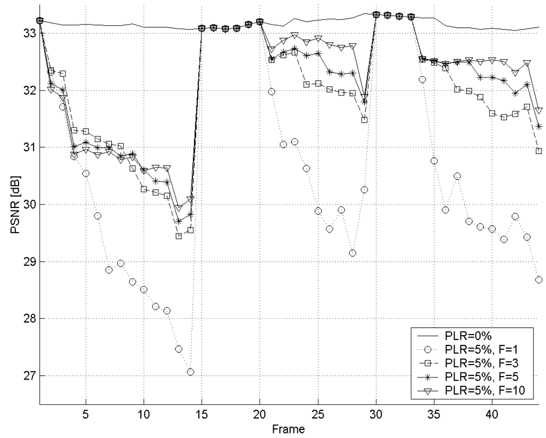

Figure 5. PSNR tracks in the time (50 frames) at PLR=5% for different values of *F*. Reference PSNR in error-free environment.

5. CONCLUSIONS

In the filed of peer-to-peer video streaming system design, this paper presented a novel sequence-based error-recovery algorithm applied to a balanced MDC system. The proposed method 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. The additional delay introduced by the proposed approach is reasonable compared to the buffering time required by a "traditional" P2P system model.

Simulation results proved the great improvement that the proposed sequence-based error-recovery algorithm gives to the system performance compared to common EC approaches.

REFERENCES

[1] V. K. Goyal, "Multiple description coding: Compression meets the network," in IEEE Signal Processing Mag., vol. 18, pp. 74–93, Sept. 2001.

[2] 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.

[3] 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.

[4] Y. C. Lee, Y. Altunbasak, and R. M. Mersereau, "Multi-frame error concealment for MPEG-coded video delivery over error-prone networks," IEEE Trans. Image Proc., vol. 11, Nov. 2002.

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

[6] N. Franchi, M. Fumagalli, and R. Lancini, "Error-Propagation Reduction in a Balanced Multiple Description Video Coder," in Proceedings of IEEE International Conference on Image Processing, Singapore, 2004 October 24-27.

[7] Bertsimas, Dimitris and Tsitsiklis, "Introduction to Linear Optimization," Athena Scientific, 1997 (ISBN 1-886529-19-1).

[8] D-E. Meddour, M. Mushtaq, T. Ahmed, "Open Issues in P2P Multimedia Streaming," in Proc. of MULTICOMM 2006, June 2006, Istanbul, Turkey.

[9] V. N. Padmanabhan, H. J. Wang, P. A. Chou, "Resilient Peer-to-Peer Streaming", IEEE ICNP 2003, Atlanta, GA, USA, November 2003.

[10] M. Zhou and J. Liu, "A Hybrid Overlay Network for Video-on-Demand, IEEE International Conference on Communications (ICC'05), Seoul, Korea, May 2005.