A Multi-hypothesis Decoder for Multiple Description Video Coding

Mengyao Ma[†], Oscar C. Au[‡], Liwei Guo[‡], Xiaopeng Fan[‡], Ling Hou[‡] and S.-H. Gary Chan[†]
[†] Dept. of Computer Science and Engineering, [‡] Dept. of Electronic and Computer Engineering
Hong Kong University of Science and Technology
{myma, eeau, eeglw, eexp, eileenzb, gchan}@ust.hk

***Abstract*— *Multiple Description Coding* (MDC) can be used as an *Error Resilience* (ER) technique for video coding. In case of transmission errors, *Error Concealment* (EC) can be combined with MDC to reconstruct the lost frame, such that the propagated error to the following frames is reduced. In this paper we propose a novel algorithm based on a *Multi-hypothesis Decoder* (MHD), to improve the reconstructed video quality of MDC over packet loss networks. Both subjective and objective results show that MHD can help to achieve a better video quality than a traditional EC algorithm.**

I. Introduction

Error Resilience (ER) and *Error Concealment* (EC) techniques are very important for video transmission today, due to the use of predictive coding and *Variable Length Coding* (VLC) in video compression [1]. The conventional INTER mode approach is illustrated in Figure 1(a), where each P-frame is predicted from its immediate previous frame. Although the compression efficiency of this approach is high, it is vulnerable to errors in the transmission channel. If one frame is lost or corrupted (for example: P_4) during the transmission, the error in the reconstructed frame at the decoder will propagate to the remaining frames until the next I-frame (I_{11}) is received.

Several ER methods have been developed for video communication, such as *Forward Error Correction* (FEC) [2], *Layered Coding* [3], and *Multiple Description Coding* (MDC) [4]. Different from the traditional *Single Description Coding* (SDC), MDC divides the video stream into equally important *streams* (*descriptions*), which are sent to the destination through different channels. Error may occur in the channels. Suppose the failure probability of each channel is independently and identically distributed with probability p. If we use the conventional SDC method, the entire description will be lost with probability p; if we use M descriptions and send them on M different channels, the probability of losing the entire description is p^M, which is much less than p. One simple implementation of MDC is the odd/even temporal sub-sampling approach: an even (odd) frame is predicted from the previous even (odd) frame, as illustrated in Figure 1(b). Since the reference frames are farther in time, the prediction of such approach is not as good as the conventional codec and the compression efficiency is lower. On the other hand, since each stream is encoded and transmitted separately, the corruption of one stream will not affect the other. As a result, the decoder can simply display the correct video stream ($P_5P_7P_9\ldots$) at half of the original frame rate, or reconstruct the corrupted frame by some appropriate EC methods, e.g. *Temporal Interpolation* [5][6]. The objective of using temporal interpolation is that it can be well combined with temporal MDC methods. Recall that in Figure 1(c), when frame P_4 is corrupted during the transmission, its surrounding frames (P_3 and P_5) would be correct if stream 1 is error-free. So we can utilize P_3 and P_5 to interpolate P_4 with good quality.

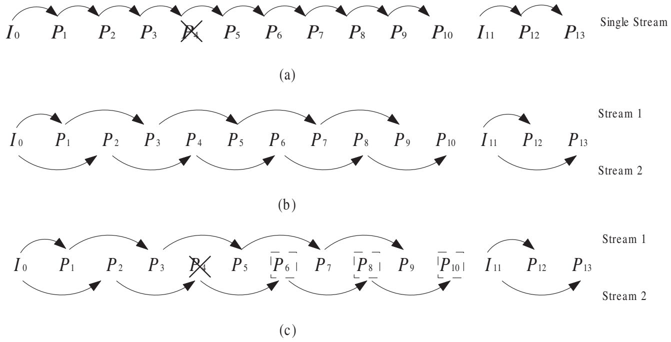
I_0 P_1 P_2 P_3 P_4 P_5 P_6 P_7 P_8 P_9 P_{10} I_{11} P_{12} P_{13}
Single Stream
(a)
Stream 1
Stream 2
(b)
Stream 1
Stream 2
(c)

Fig. 1. Illustration of different approach for video coding (the arrow means that the previous frame is used as the reference of the latter). (a) Conventional video coding; (b) Odd/even sub-sampling MDC; (c) Error occurs in (b).

In conventional EC algorithms, only the corrupted (lost) frames are error-concealed. The following frames are decoded as usual. Since error concealment may fail for the lost frame in some special cases, i.e. new objects appearing or old objects disappearing, a large initial error is generated and propagates to the following frames. In such circumstances, error-concealing the following frames may have a better quality than decoding them directly. In this paper we propose a novel algorithm based on a *Multi-hypothesis Decoder* (MHD), where error-concealed frame is used as an additional hypothesis to improve the reconstructed video quality of MDC. The rest of this paper is organized as follows. In Section 2, we describe the proposed approach (MHD). The comparison between an EC algorithm with MHD and without MHD is given in Section 3, by both subjective and objective results. Section 4 is conclusion.

II. Multi-hypothesis Decoder for MDC

When the odd/even sub-sampling is used in temporal MDC, an even (odd) frame is predicted from the previous even (odd)

978-1-4244-1684-4/08/$25.00 ©2008 IEEE

frame. These two streams (descriptions) are sent to the decoder through different channels. Consider the case of one frame loss during the transmission. By using some *Error Concealment* (EC) technique, this frame can be reconstructed at the decoder side with some error. Due to the use of motion compensation, this error will propagate to the following frames in the same stream (description). Define the frame at time n to be $\psi(n)$ and assume the loss occurs at time l_0. To improve the reconstructed video quality after the loss position, we propose an algorithm which is based on a *Multi-hypothesis Decoder* (MHD).

A. EC for the Lost Frame using Temporal Interpolation

Temporal interpolation was originally used to generate one or more frames between two received frames so as to improve the effective frame rate, and make the object motions in the video smoother. Usually both forward and backward motion estimations are performed to track motions of the objects between adjacent received frames [7]. This leads to high computational requirement. In [6], *Unidirectional Motion Compensated Temporal Interpolation* (UMCTI) is used, which performs only forward motion estimation and thus saves half of the computation time.

The objective of introducing temporal interpolation here is that it can be well combined with temporal MDC methods. Recall that in Figure 1(c), when frame P_4 is corrupted during the transmission, its surrounding frames (P_3 and P_5) would be correct if stream 1 is error-free, due to the independent failure probability of each channel. So we can utilize P_3 and P_5 to interpolate P_4 with good quality. In addition, the motion vector from P_5 to P_3 is conserved in stream 1 and thus helps us to skip the exhaustive motion estimation process. Since the main focus of this paper is to improve the reconstructed video quality after the loss position, we use the existing algorithm, i.e. UMCTI, to error-conceal the lost frame. One advantage of UMCTI is that the time for the interpolation is linear to the frame size, thus reducing the complexity of the multi-hypothesis decoder for the following frames. More details about the implementation of UMCTI can be found in [6].

B. Multi-hypothesis Decoding

In conventional EC algorithms, only the corrupted (lost) blocks are error-concealed. Although the following frames can be decoded as usual, error exists due to the use of temporal prediction. As shown in [8], spatial filtering in motion compensation can help to attenuate the propagated error energy. It can be an explicit loop filter, or implicitly brought by the bilinear interpolation for sub-pixel motion compensation [9]. Without generality, suppose $\psi(l_0)$ belongs to description 1 (D1). We propose to use two ways to reconstruct the following frames in D1: decoding directly as in the conventional codec, and interpolation using the same EC methods as that for $\psi(l_0)$. It seems at the first sight that the latter one is unnecessary, since the decoding process itself can decrease the propagated error. However, error concealment may fail for $\psi(l_0)$ in some special cases, i.e. new objects appearing or old objects disappearing, thus leading to a large initial error. In such circumstances, error-concealing the frames after $\psi(l_0)$ may have a better quality than decoding them directly.

Based on the previous discussion, we propose to reconstruct frame $\psi(l_0+2t)$ by a weighted sum of two hypotheses:

$$\hat{\psi}(l_0+2t) = h_1\psi^d(l_0+2t) + h_2\psi^c(l_0+2t), \qquad (1)$$

where $t \in [1, N]$ and $h_1+h_2 = 1$. $\psi^d(l_0+2t)$ and $\psi^c(l_0+2t)$ are the corresponding frames obtained by decoding and concealment, respectively. $2t$ is used here to specify the frames in the same description (D1) as $\psi(l_0)$. N is a constant specifying a *Time Interval* to apply the multi-hypothesis reconstruction. Note that if we set $h_1 = 1$ in (1) or use zero *Time Interval* ($N = 0$), MHD becomes a conventional decoder.

C. MHD with Adaptive Weights

For simplicity, the weights h_1 and h_2 in (1) can be constant for $t \in [1, N]$. They can also be adaptively determined based on the minimum mean square error (MMSE) criterion:

$$h_1 = \frac{\sigma_c^2}{\sigma_d^2+\sigma_c^2}, \qquad h_2 = \frac{\sigma_d^2}{\sigma_d^2+\sigma_c^2}, \qquad (2)$$

where $\sigma_d^2 = E\{(\psi^d(l_0+2t) - \tilde{\psi}(l_0+2t))^2\}$ and $\sigma_c^2 = E\{(\psi^c(l_0+2t) - \tilde{\psi}(l_0+2t))^2\}$; $\tilde{\psi}(l_0+2t)$ is the original reconstructed frame of $\psi(l_0+2t)$ at the encoder side. (2) is derived based on the assumption that $(\psi^d(l_0+2t)-\tilde{\psi}(l_0+2t))$ and $(\psi^c(l_0+2t)-\tilde{\psi}(l_0+2t))$ are uncorrelated random variables with zero mean.

Define error $\epsilon(t)$ to be the difference between $\psi^d(l_0+2t)$ and $\tilde{\psi}(l_0+2t)$. As stated previously, spatial filtering can attenuate the propagated error energy. This effect is analyzed in [8], where the decoder is regarded as a linear system and its impulse response is approximated as a gaussian filter. Based on the central limit theory, we also expect the impulse response of MHD to be gaussian. Using similar deriving process as [8], we can obtain

$$\sigma^2(t) \approx \frac{\sigma^2(0)}{1+\gamma t}, \qquad (3)$$

where $\sigma^2(t)$ is the variance of $\epsilon(t)$. γ is a parameter describing the efficiency of the loop filter to attenuate the error; typically $\gamma \in (0, 1)$. Based on (3), we can obtain:

$$\sigma_d^2 = \frac{\sigma^2(0)}{1+\gamma t}. \qquad (4)$$

Since the same error concealment method is used to interpolate the lost frames, the error variance of $\psi^c(l_0+2t)$ approximates to that of $\psi^c(l_0)$. In other words,

$$\sigma_c^2 \approx \sigma^2(0). \qquad (5)$$

By using (2), (4) and (5), the values of h_1 and h_2 can be obtained:

$$h_1 = \frac{1+\gamma t}{2+\gamma t}, \qquad h_2 = \frac{1}{2+\gamma t}. \qquad (6)$$

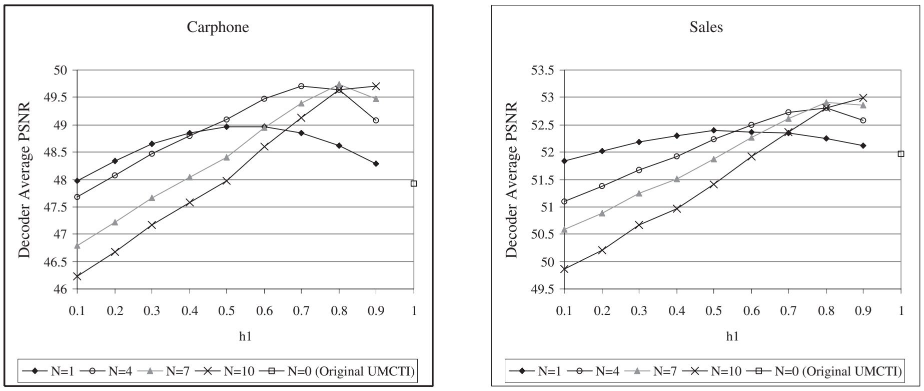

Fig. 2. The average PSNR at the decoder side for CMHD with different weight h_1. The packet loss rate is $P = 3\%$.

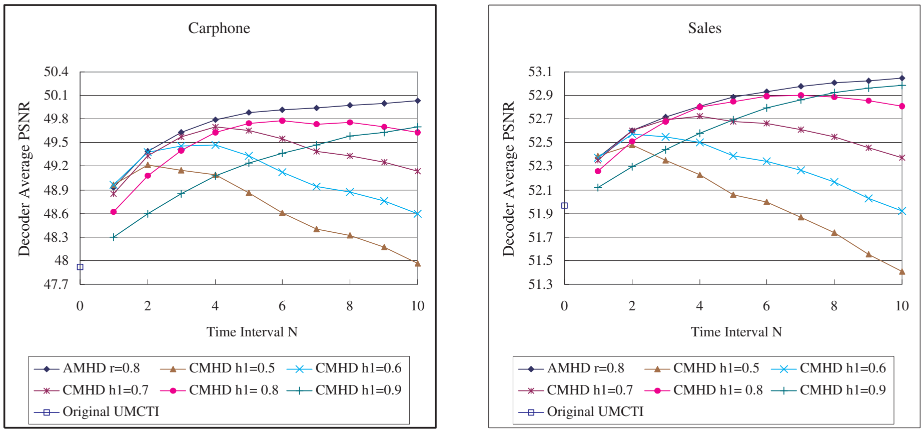

Fig. 3. Comparison between CMHD and AMHD with $\gamma = 0.8$. The packet loss rate is $P = 3\%$.

III. Simulation Results

In the simulation, we compare the performance of MHD to the original UMCTI algorithm, by both subjective and objective results [6]. MHD with constant weights (CMHD) and MHD with adaptively determined weights (AMHD) are both simulated. The value of parameter γ in AMHD is trained to be 0.8. For UMCTI, only the lost frames are error-concealed and the following frames are decoded as usual. We use the JVT reference software version 8.2 (baseline profile) for the simulations [10]. The first 300 frames of video sequences *Carphone* and *Sales* (QCIF) are encoded at 15fps, and only the first frame is I frame. Fixed QP is used: for *Carphone*, 28 is used for I frame and 30 for P frame; for *Sales*, 27 is used for I frame and 29 for P frame. To generate two descriptions, ref_idx_l0 is specified for each P frame to simulate the odd/even sub-sampling MDC. For the I frame, we just send it twice to the two streams, since the main focus of the simulation is to compare the error resilience properties, instead of the compression efficiency of MDC. To further improve the coding of I frame, method in [11] can be employed.

We first test the effect of weighting parameter h_1 on the performance of MHD. Suppose the two video streams are transmitted though two packet loss channels, and the failure probability of each channel is independent and identically distributed with probability P. One packet contains the information of one frame, and the loss of one packet will lead to the loss of one entire frame. Four different *Time Intervals* (N) are used. For each combination of h_1 and N, we transmit the video sequence 100 times. The average PSNR is obtained at the decoder side and plotted in Figure 2. For the comparison, the PSNR obtained by the original UMCTI algorithm is also plotted. As shown in the figure, an optimal h_1 can be obtained for a specific N, which has the maximum PSNR in the corresponding curve; the larger N is, the bigger the optimal h_1 is. For $N = 1$ and $h_1 = 0.5$ in *Carphone*, about 1dB gain can be obtained compared to the original UMCTI. When N is larger, more gains can be achieved with an optimal h_1. Note that in this paper, we use the encoder reconstructed frame (the error-free one) as the reference in the calculation of PSNR. Similar behaviors can be observed if the original frame (the uncompressed one) is used as the reference.

In Figure 3, the comparison between AMHD and CMHD is given, for different *Time Interval*. From the figure we can see that the PSNR of AMHD is higher than CMHD for

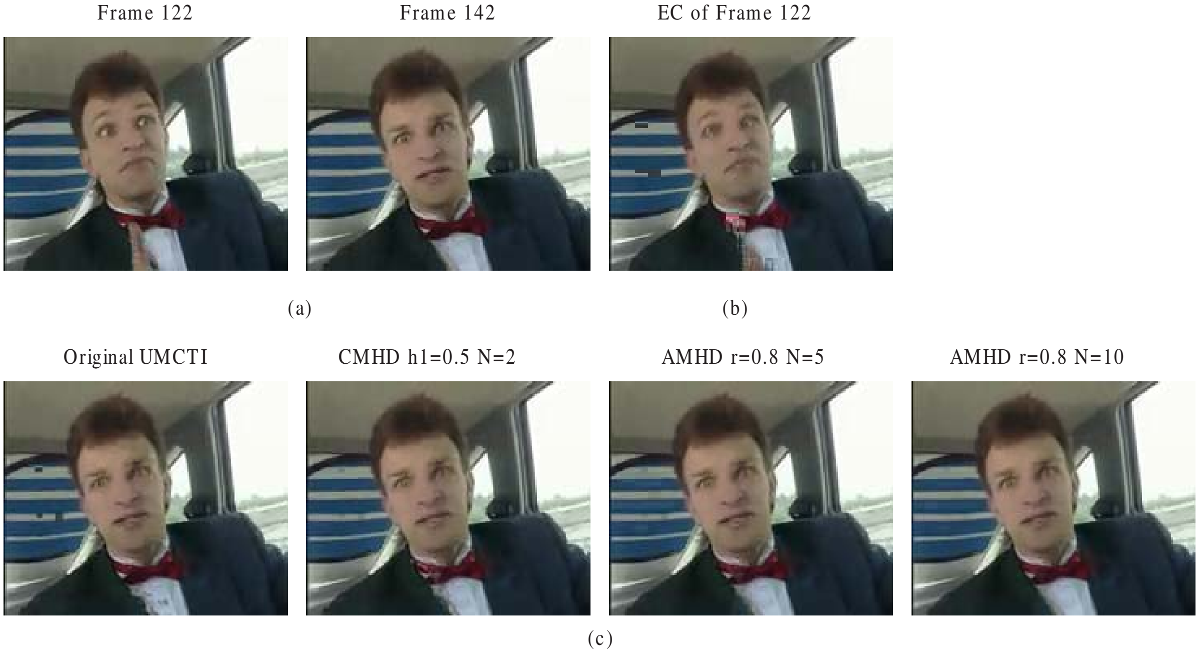

Fig. 4. The visual results of applying UMCTI and MHD on *Carphone*, for one frame loss (frame 122). (a) The original encoded frames without loss; (b) The error-concealed frame 122 using UMCTI; (c) The reconstructed frame 142 using different methods.

almost all the compared N. The larger N is, the higher the PSNR of AMHD is. Although for a small N, CMHD with an appropriate h_1 can obtain a higher PSNR than AMHD, its performance (PSNR) decreases when N is larger. In these situations, i.e. when $N > 2$, AMHD is preferred to get a better performance.

Figure 4 illustrates the visual quality after applying UMCTI and MHD on *Carphone*, for one frame loss (frame 122). In the first row, the first two frames are the original reconstructed frames at the encoder side, and the third one is the error-concealed frame 122 using UMCTI. Since the finger enters the scene with a large motion, the interpolation works badly around this region. Then the following frames are reconstructed by different methods, and the 10^{th} one in the same description is shown in Figure 4(c). From the figure we can see that the original EC method gives the worst visual quality, since the frames after loss are just decoded as usual without using the additional hypothesis. CMHD with $N = 2$ can improve the quality a little, but the boundary between the shirt and the coat is still ambiguous. Much improvement can be achieved by AMHD. As in Figure 3, a longer *Time Interval* N helps to make the result better.

IV. Conclusion

In this paper we propose a novel algorithm based on a *Multi-hypothesis Decoder* (MHD), to improve the reconstructed video quality of MDC over packet loss networks. Both subjective and objective results show that MHD can help to achieve a better video quality than a traditional EC algorithm. In the current work, the weight of MHD is fixed for a whole frame. To further improve the reconstructed video quality, block or pixel level adaptation can be used to adjust the weight. We take this as a future work.

Acknowledgement

This work has been supported by the Innovation and Technology Commission of the Hong Kong Special Administrative Region, China (project no. GHP/033/05).

References

[1] Y. Wang and Q. F. Zhu, "Error control and concealment for video communication: a review," in *Proc. IEEE*, May 1998, pp. 974 – 997.
[2] Y. Mei, W. Lynch, and L. N. Tho, "Joint forward error correction and error concealment for compressed video," in *Proc. IEEE ITCC*, Apr. 2002, pp. 410 – 415.
[3] C.-M. Fu, W.-L. Hwang, and C.-L. Huang, "Efficient post-compression error-resilient 3D-scalable video transmission for packet erasure channels," in *Proc. IEEE ICASSP*, Mar. 2005, pp. 305 – 308.
[4] Y. Wang, A. Reibman, and S. Lin, "Multiple description coding for video delivery," in *Proc. IEEE*, Jan. 2005, pp. 57 – 70.
[5] J. Apostolopoulos, "Reliable video communication over lossy packet networks using multiple state encoding and path diversity," in *Proc. SPIE VCIP*, Jan. 2001, pp. 392 – 409.
[6] C.-W. Tang and O. Au, "Unidirectional motion compensated temporal interpolation," in *Proc. IEEE ISCAS*, June 1997, pp. 1444 – 1447.
[7] C.-K. Wong and O. Au, "Fast motion compensated temporal interpolation for video," in *Proc. SPIE VCIP*, May 1995, pp. 1108 – 1118.
[8] N. Farber, K. Stuhlmuller, and B. Girod, "Analysis of error propagation in hybrid video coding with application to error resilience," in *Proc. IEEE ICIP*, Oct. 1999, pp. 550 – 554.
[9] B. Girod and N. Farber, "Wireless video," in *Compressed Video Over Networks*, M.-T. Sun and A. R. Reibman, Eds. Marcel Dekker, 2000.
[10] Jvt reference software, version 8.2. [Online]. Available: http://iphome.hhi.de/suehring/tml/download/
[11] Y. Wang, M. Orchard, and A. Reibman, "Multiple description image coding for noisy channels by pairing transform coefficients," in *IEEE MMSP*, June 1997, pp. 419 – 424.