JOINT ERROR CONCEALMENT AND ERROR RECOVERY FOR CONSECUTIVE FRAME LOSSES UNDER THE UNBALANCED MULTIPLE DESCRIPTION CODING ARCHITECTURE[a]

FENG HUANG, LIFENG SUN, BIN LI and YUZHUO ZHONG

Department of Computer Science and Technology, Tsinghua University, Beijing
hfbee@263.net,{sunlf, libin98, zyz-dcs}@mail.tsinghua.edu.cn

Real-time video transmission over error-prone wireless networks often experiences consecutive frame losses due to either temporary link outages or traffic congestion. Although error concealment (EC) techniques have been extensively studied, they usually cannot handle the problem. Thus, we envision using EC under the unbalanced multiple description coding (UMDC) architecture. UMDC has almost no coding delay and can produce two descriptions at any bit-rates adaptive to different path bandwidths. In this paper, we propose an iterative EC algorithm able to adaptively exploit both the high-resolution (HR) and low-resolution (LR) information via multi-hypothesis weights. It is applied to both lost HR frames and following undecoded ones. Considering error propagation, we design an interframe error recovery (ER) algorithm for the undecoded HR frames. It iteratively uses multi-frame recovery principle to frame-by-frame reduce error drift in the HR stream with respect to the intermediate information from EC. The joint design of EC and ER can be applied to most UMD approaches. Extensive experiments have been carried out under different conditions. The proposed EC technique exhibits high PSNR gains versus the usual ones under the UMDC architecture and the classical ones without the support of UMDC. The proposed ER technique is efficient in reducing error drift, especially in high motion scenes. In conclusion, joint EC and ER can provide satisfactory performance on both PSNR and visual quality.

Key words: Unbalanced Multiple Description Coding, Error Concealment, Error Recovery, Error Propagation
Communicated by: D. Taniar & K. Liu

1   Introduction

Real-time video transmission over mobile wireless networks is expected to experience packet loss due to either temporary link outages or traffic congestion. Such packet losses often occur in bursts which cause substantial degradation to the transmitted video quality. In order to save transmission overhead, one data packet usually carries a whole frame. Consequently, a burst of packet losses directly results in the loss of consecutive frames.

In order to deal with errors or losses in video transmission, lots of error resilience encoding techniques and decoder error concealment (EC) methods have been developed [17, 19, 20]. In general, decoder EC does not require any change to the encoder. It does not increase the bit rate, or the delay.

So it is preferable in conversational (low-delay) applications, such as video telephony and conferencing. Temporal EC is one of the most important EC techniques. It utilizes temporal neighbors, that is, the previous frame or the next frame, to conceal the errors in the current frame. Most of the temporal EC methods [1, 2, 12] assume that only a few macroblocks or slices in a video frame are lost. Some of them [4, 5, 14] can recover a whole missing frame. Although EC techniques have been extensively studied, they usually cannot handle the loss of consecutive video frames.

Forward Error Correction (FEC) codes can also be deployed to increase the robustness of the coded bit-stream [16]. While coding and interleaving over a large number of packets is one possible mechanism for dealing with bursty loss, it introduces delay due to the high encoding/decoding time. Another drawback of FEC codes is the 'cliff effect' that a rapid degradation of the performance when the number of losses exceeds the error-correction capacity of the code. An alternative approach to the problem is the use of a Multiple Description Coding (MDC) technique combined with multiple path transport (MPT). In the most common implementation [7, 15, 21], a MD codec generates two or more equal rate and equal importance descriptions. The source descriptions are individually packetized and sent to the receiver over different paths. As long as the two or more descriptions are not simultaneously affected by packet losses, an acceptable quality can be maintained. If all the descriptions are successfully received, the decoder will be able to reconstruct better video. So MDC can deal with bursty losses while achieving low coding delay for conversational applications. However, in such a Balanced MDC (BMDC) system, it is not flexible to control the amount of redundancy. An Unbalanced MDC (UMDC) approach [3, 6, 10, 13, 18] also exists that typically uses two descriptions: one description has a higher quality than the other. In this scheme the Low-Resolution (LR) description is primarily used as redundancy, and it is employed to conceal errors or losses in the High-Resolution (HR) description. The LR stream in the UMDC case has most in common with the redundancy added by the code in FEC approach. Unlike FEC, UMDC is able to produce an acceptable quality even if redundancy percentage is less than packet loss rate [10, 18]. Compared to BMDC, UMDC can produce two descriptions at different bit-rates adaptive to different path bandwidths.

Thus, it is promising to develop the EC algorithm based on UMDC for the problem of consecutive frame losses. In the literature, there are some related solutions. Nevertheless they do not further study how to handle this sort of losses. This paper presents an error concealment/recovery scheme aiming at this problem. We design an iterative EC algorithm able to adaptively exploit both the HR and LR information via multi-hypothesis weights. Based upon the concealed information coming from the EC algorithm, we focus on how to frame-by-frame reduce error drift and design an interframe error recovery (ER) algorithm for the undecoded HR frames following the lost ones. Historically, none of the EC algorithms continues process after concealing the lost frames. Error propagation results have shown that higher PSNR on a given concealed frame typically leads to propagation of a smaller error in the following frames. However, we can envision performing error recovery with respect to the redundant LR information under the UMDC architecture.

The remainder of the paper is organized as follows. After discussing related work in Section 2, we present the two-stage recovery procedure, the EC algorithm and the ER algorithm in Section 3. Section 4 includes details of experiments and exhaustive performance evaluations to demonstrate the superiority of the proposed methods to others. Finally, Section 5 offers conclusions.

2   Related Work

Techniques of estimating the lost motion vectors have been widely discussed to recover a whole missing frame. One straightforward method called temporal replacement (TR) is to repeat the last received frame with all zero motion vectors. TR works well at temporal stationary areas but fails at moving areas. MMA [4] is a pixel-based temporal EC method. It inversely tracks the motion of each pixel in a few past frames, and then averages the motion vectors in the trace to estimate the forward motion vector in the last received frame. Finally each pixel in the frame is projected onto an estimate of the missing frame. This method can smooth the boundaries of blocks but sometimes it behaves even worse that TR does. Block-based MVE [14] is another method that can combat the loss of a whole frame. The method first extrapolates motion vectors of macroblocks from the last received frame, and then estimates the overlapped areas between the damaged block and the motion extrapolation macroblocks. It selects the best extrapolated motion vector and conceals the damaged block using general motion compensation. This method overcomes the disadvantage of incorrect macroblock displacement, but the rough motion compensation of 8×8 pixel size usually causes block artifacts in large motion scenes. In order to overcome the above shortcomings, [5] proposes a pixel-based MVE method. In the case of recovering a loss of consecutive frames, whether these methods can be applied to all the lost frames should be further examined.

There are some related solutions [3, 6, 8, 10, 13, 18] of performing EC based on UMDC. Most of them provide localized EC techniques that only use the redundant information near the loss. The EC performances depend on the quality of the correctly received stream, but the redundant stream may be coded at a low bit-rate. The sequence-based EC algorithm [8] has revealed good performances for concealment. It uses multi-frame recovery principle to minimize the error propagation in the HR description with respect to the LR description. However, in order to conceal a loss in the current frame effectively, the algorithm has to await and examine a significant number of future frames. The more future frames examined, the greater the delay. Moreover, the algorithm assumes that in each of the successive frames there is only one pixel predicted from the previous one, and the prediction uses only integer pixel. This is not compatible with the mainstream video coding standards, i.e. H.264, which limits its deployment in practical systems.

3   Proposed Error Concealment/ Recovery Scheme

To make the proposed error concealment/recovery scheme suitable for most UMD approaches, we adopt a general UMDC approach. The HR and LR descriptions have the same frame-rate and intra refresh (IR) period. It is supposed that the HR and LR descriptions do not both encounter frame loss in one IR period if the IR period is set in proper length. Therefore, to analyze how lost consecutive frames of the HR stream can be concealed we can suppose the coded data of the corresponding LR ones are available. The proposed scheme is also compatible with the UMDC approaches [10, 13, 18] that use shorter IR period for LR in order to improve its resilience to bursts of frame losses. Before describing the core algorithms in Subsection 3.2 and 3.3, we give an example of illustrating the recovery procedure of the proposed scheme.

3.1 Two-stage Recovery Procedure

We denote the HR frame n of type x(I-,P-,B-) as $f^{HR}(n,x)$ and denote the LR frame n of type x(I-,P-,B-) as $f^{LR}(n,x)$. The UMD encoder sends $f^{HR}(1,I)$, $f^{HR}(2,P)$, $f^{HR}(3,P)$, $f^{HR}(4,P)$, $f^{HR}(5,P)$, $f^{HR}(6,P)$, $f^{HR}(7,I)$, $f^{HR}(8,P)$ and $f^{LR}(1,I)$, $f^{LR}(2,P)$, $f^{LR}(3,P)$, $f^{LR}(4,P)$, $f^{LR}(5,P)$, $f^{LR}(6,P)$, $f^{LR}(7,I)$, $f^{LR}(8,P)$. Each frame carries the frame number n in the packet header. Suppose that $f^{HR}(3,P)$ and $f^{HR}(4,P)$ are lost due to a bursting loss. Obviously, $f^{HR}(5,P)$, $f^{HR}(6,P)$ and $f^{HR}(7,SP)$ are undecoded due to predictive coding. The loss of $f^{HR}(3,P)$ and $f^{HR}(4,P)$ is detected using the inconsecutive frame numbers between $f^{HR}(2,P)$ and $f^{HR}(5,P)$. Then the UMD decoder involves the two-stage recovery procedure:

1. **Iterative error concealment for lost HR frames.** The decoded picture of $f^{HR}(2,P)$, together with the coded data (motion vectors and prediction errors) and decoded picture of $f^{LR}(3,P)$, is passed to the EC processor. Then the concealed picture $p^{EC}(3,P)$ is outputted for display. In a new round, the concealed picture $p^{EC}(3,P)$, together with the coded data and decoded picture of $f^{LR}(4,P)$, is sent back to the EC processor. Then the concealed picture $p^{EC}(4,P)$ is outputted for display. The same operations of iteratively yielding the concealed pictures $p^{EC}(5,P)$ and $p^{EC}(6,P)$ are performed. Instead of output for display, $p^{EC}(5,P)$ and $p^{EC}(6,P)$ are copied in the ER (error recovery) processor buffer for further use. They are intermediate results and serve for the next stage.

2. **Iterative error recovery for undecoded HR frames.** Although $f^{HR}(5,P)$ is predictive-encoded from the lost $f^{HR}(4,P)$ and undecoded, its coded data and the concealed picture $p^{EC}(5,P)$ are passed to the ER processor. Meanwhile, the last displayed picture $p^{EC}(4,P)$ is also sent to the ER processor. Then the recovered picture $p^{ER}(5,P)$ is outputted for display. In a new round, the coded data of $f^{HR}(6,P)$ and the concealed picture $p^{EC}(6,P)$ are passed to the ER processor. Meanwhile, the last recovered picture $p^{ER}(5,P)$ is also sent back to the ER processor. Then the recovered picture $p^{ER}(6,P)$ is outputted for display. The same operations of iteratively yielding the recovered pictures are performed until an I-frame, i.e. $f^{HR}(7,I)$, is available to restart the HR decoding process.

In summary, the UMD decoder outputs $f^{HR}(1,I)$, $f^{HR}(2,P)$, $p^{EC}(3,P)$, $p^{EC}(4,P)$, $p^{ER}(5,P)$, $p^{ER}(6,P)$, $f^{HR}(7,I)$ and $f^{HR}(8,P)$ for display.

3.2 Iterative Error Concealment with Multi-hypothesis Weight

First of all, it is assumed that frames $f^{HR}(n+k,P)(1\leq k\leq m)$ are lost and following frames $f^{HR}(n+k,P)(m+1\leq k\leq m+o)$ are undecoded. The corresponding LR ones are all correctly received and decoded. Obviously, concealing the lost HR frames by the corresponding LR pictures is a feasible method that presents an acceptable and stable performance. That is

$$x_i^{concealed}(n+k) = x_i^{LR}(n+k) \quad (1 \leq k \leq m) \tag{1}$$

where $x_i(n)$ denotes pixel i in frame n. However, the EC performance depends on the quality of the LR stream. It is poor when the LR description is coded at a low bit-rate to insure efficiency or adapt to the available path bandwidth. An alternative method is that, we can approximate the motion vectors $MV_i^{HR}(n+k)$ of the lost HR frames at the pixel level to the correspondent $MV_i^{LR}(n+k)$ from LR. Then for $f^{HR}(n+k,P)$, each pixel is reconstructed as follows:

$$x_i^{concealed}(n+k) = \begin{cases} x_{i+MV_i^{LR}(n+1)}^{HR}(n) & (k=1) \\ x_{i+MV_i^{LR}(n+k)}^{concealed}(n+k-1) & (2 \le k \le m) \end{cases} \qquad (2)$$

by referencing the last decoded HR picture or the last concealed picture. We find that in stationary or little motion scenes, the method works well. However, the rough LR motion vectors without residual information usually cause block artifacts, especially in large motion scenes.

Via the two methods, we can obtain two estimations of the lost HR frames. It has been found that weighted averaging of multiple candidate concealments can achieve more robust performance [1, 2, 5]. Accordingly, we propose a more effective EC algorithm by combining the two pixel-based methods. Then for $f^{HR}(n+k,P)$, each pixel is reconstructed as follows:

$$x_i^{concealed}(n+k) = \begin{cases} (1-\alpha)x_i^{LR}(n+1)+\alpha x_{i+MV_i^{LR}(n+1)}^{HR}(n) & (k=1) \\ (1-\alpha)x_i^{LR}(n+k)+\alpha x_{i+MV_i^{LR}(n+k)}^{concealed}(n+k-1) & (2 \le k \le m) \end{cases} \qquad (3)$$

where the pixel-based weight α is used to adjust the weights of the two methods. In the section, we focus on how to set α, namely developing a technique that can adaptively adjust weights. For the intra-coded pixels of the LR frames, α is set to 0 because $MV_i^{LR}(n+k)$ does not exist; for the succeeding pixels predicted from them, α is also set to 0 because $x_{i+MV_i^{LR}(n+k)}^{concealed}(n+k-1)$ is actually equal to $x_{i+MV_i^{LR}(n+k)}^{LR}(n+k-1)$. The following discussion only aims at the rest.

One principle is that more weights should be given to Equation (1) (that is, α should be less than 0.5) if the pixel possesses a larger prediction error $e_i^{HR}(n+k)$. $e_i^{HR}(n+k)$ is lost, but the absolute $e_i^{HR}(n+k)$ is not required for the purpose of adjusting weights. It is good enough to know whether $e_i^{HR}(n+k)$ is large or small, which can be estimated according to the correlation between $e_i^{HR}(n+k)$ and $|MV_i^{HR}(n+k)|$ (the absolute quantity of $MV_i^{HR}(n+k)$). Figure 1 and Figure 2 show that a large $|MV_i^{HR}(n+k)|$ usually implies a large $e_i^{HR}(n+k)$ for both large and small motion scenes. Since $MV_i^{HR}(n+k)$ is approximated to $MV_i^{LR}(n+k)$, a large $|MV_i^{LR}(n+k)|$ usually implies a large $e_i^{HR}(n+k)$. Therefore more weights are given to Equation (2) if a pixel possesses a lesser $|MV_i^{LR}(n+k)|$. We propose α can be calculated as follows:

$$\alpha_1 = 1 - \frac{|MV_i^{LR}(n+k)|}{\max(|MV_i^{LR}(n+k)|)} = 1 - \frac{|MV_i^{LR}(n+k)|}{\sqrt{2} \times search\_range} \quad (1 \le k \le m) \qquad (4)$$

Another principle comes from the optical flow concept [9]. In the hypothesis that motion in the sequence merely consists of pixels moving from one place to another, each pixel may move in all directions and tries to find a motion trajectory with the lowest energy consumption. If a pixel moves according to the motion trajectory in the LR description, the energy consumption of each movement is rough proportional to the absolute value of prediction error $|e_i^{LR}(n+k)|$. In other words, it does not move according to $-MV_i^{LR}(n+k)$ in the HR description if $|e_i^{LR}(n+k)|$ is large. Accordingly, more

weights are given to Equation (2) if the pixel possesses a lesser $|e_i^{LR}(n+k)|$. We propose α can also be calculated as follows:

$$\alpha_2 = \max(0, 1 - \frac{|e_i^{LR}(n+k)|}{T}) \quad (1 \le k \le m) \tag{5}$$

where T is a threshold. We find that the EC performance is not very sensitive to T and its value is set to 12 in practice.
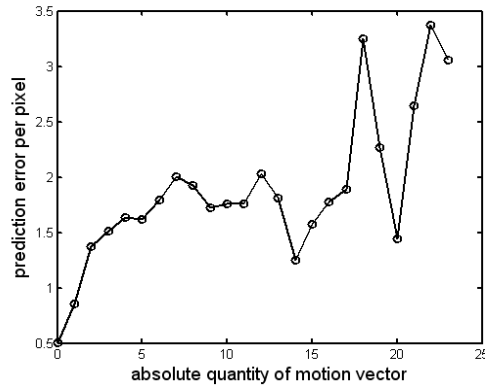

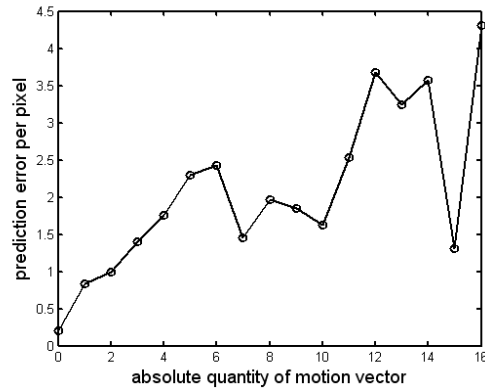Figure 1 Statistical correlation between motion vector and prediction error. QCIF Foreman sequence(large motion), bit-rate=91kbps(QP=28), 15fps(150frames), JM86

Figure 2 Statistical correlation between motion vector and prediction error. QCIF Mother-daughter sequence(small motion), bit-rate=31kbps(QP=28), 15fps(150frames), JM86

In the above statements, we have made some hypotheses and proposed α_1 or α_2 for Equation (3). However, the hypotheses come into existence in a certain probability. In the single hypothesis-based EC using either α_1 or α_2, the invalidation of the hypothesis reduces the accuracy of the weight. So, we can use multi-hypothesis to verify each other and improve the EC performance. Specifically, more weights are given to Equation (2) only when the pixel possesses not only a lesser $|MV_i^{LR}(n+k)|$ but

also a lesser $|e_i^{LR}(n+k)|$. For the rest cases more weights are given to Equation (1) whose performance is more reliable than the one of Equation (2). We propose that the weight α should be calculated instead of α_1 or α_2 as follows:

$$\alpha_3 = \alpha_1 \alpha_2 = (1 - \frac{|MV_i^{LR}(n+k)|}{\sqrt{2} \times search\_range}) \times \max(0, 1 - \frac{|e_i^{LR}(n+k)|}{T}) \quad (1 \le k \le m) \tag{6}$$

Let α_{opt} be the best α. Then the conditional probability $p_1 = P\{\alpha_1 < 0.5 | \alpha_{opt} < 0.5\}$ is the one that α_1 is less than 0.5 when lesser weights should be given to Equation (2). It is the accuracy of the weight α_1. The accuracy of α_2 is likewise measured by $p_2 = P\{\alpha_2 < 0.5 | \alpha_{opt} < 0.5\}$. We deduce the accuracy of $\alpha_1 \alpha_2$ as follows:

$$\begin{aligned}
&P\{\alpha_1 \alpha_2 < 0.5 | \alpha_{opt} < 0.5\} \\
&= 1 - P\{\alpha_1 \alpha_2 \ge 0.5 | \alpha_{opt} < 0.5\} \\
&= 1 - P\{\alpha_1 \alpha_2 \ge 0.5 \wedge \alpha_1 \ge 0.5 \wedge \alpha_2 \ge 0.5 | \alpha_{opt} < 0.5\} \\
&\ge 1 - P\{\alpha_1 \ge 0.5 \wedge \alpha_2 \ge 0.5 | \alpha_{opt} < 0.5\} \\
&= 1 - P\{\alpha_1 \ge 0.5 | \alpha_{opt} < 0.5\} \cdot P\{\alpha_2 \ge 0.5 | \alpha_{opt} < 0.5\} \\
&= 1 - (1 - P\{\alpha_1 < 0.5 | \alpha_{opt} < 0.5\}) \cdot (1 - P\{\alpha_2 < 0.5 | \alpha_{opt} < 0.5\}) \\
&= 1 - (1 - p_1)(1 - p_2) \\
&\ge 1 - (1 - p_{1,2}) = p_{1,2}
\end{aligned} \tag{7}$$

which is higher than p_1 and p_2. This accounts for the effectiveness of the multi-hypothesis weight $\alpha_1 \alpha_2$.

The proposed EC algorithm is also applied to the undecoded frames $f^{HR}(n+k,P)(m+1 \le k \le m+o)$, preparing the concealed pictures for ER (error recovery) described in the next subsection.

3.3 Interframe Error Recovery

As far as know, none of the EC algorithms continues process after concealing the lost frames. The following predictive frames are forced to be decoded by referencing the concealed picture. Error propagation results have shown that higher PSNR on a given concealed frame typically leads to propagation of a smaller error in the following frames. However, the redundant LR information can help to reduce error drift in the pseudo decoding process under the UMDC architecture. Here, we present an interframe ER algorithm considering pairs of interframe pixels where one depends on the previous one.

Before describing the ER algorithm, we remember it is assumed that frames $f^{HR}(n+k,P)(1 \le k \le m)$ are lost and following frames $f^{HR}(n+k,P)(m+1 \le k \le m+o)$ are undecoded. The ER algorithm has the advantage of working on a pixel-by-pixel basis. For the sake of simplicity we present its formalization referring to an undecoded HR pixel i with the pending value of $x_i^{HR}(n+k)$. For the intra-coded pixels of the undecoded HR frames and the succeeding pixels predicted from them, $x_i^{HR}(n+k)$ is available. The following discussion only aims at the rest. The motion vector $MV_i^{HR}(n+k)$ and the prediction error $e_i^{HR}(n+k)$ related to the pixel i are available, but its reference value $x_{i+MV_i^{HR}(n+k)}^{HR}(n+k-1)$ in the

previous frame is unavailable. Since the HR coded stream uses a motion compensated prediction loop, $x^{HR}_{i+MV^{HR}_i(n+k)}(n+k-1)$ can be expressed as:

$$x^{HR}_{i+MV^{HR}_i(n+k)}(n+k-1)=x^{HR}_i(n+k)-e^{HR}_i(n+k) \quad (m+1 \le k \le m+o) \tag{8}$$

At the decoder side the undecoded HR frames have an approximated version of the video source, i.e. concealed pictures or decoded LR pictures. Hence we can associate $x^{HR}_i(n+k)$ with a correspondent approximated value $x^{concealed}_i(n+k)$. In the same way the reference pixel $x^{HR}_{i+MV^{HR}_i(n+k)}(n+k-1)$ in the previous frame is associated with $x^{recovered}_{i+MV^{HR}_i(n+k)}(n+k-1)$ of the last recovered picture. As shown in Figure 3, the two pairs of dependent pixels $(x^{HR}_i(n+k), x^{HR}_{i+MV^{HR}_i(n+k)}(n+k-1))$ and $(x^{concealed}_i(n+k), x^{recovered}_{i+MV^{HR}_i(n+k)}(n+k-1))$ follow the same motion vector $MV^{HR}_i(n+k)$. The interframe ER algorithm tries to obtain an optimal approximation $x^{recovered}_i(n+k)$ of $x^{HR}_i(n+k)$ that minimizes the mean square error between the two pairs of pixels.
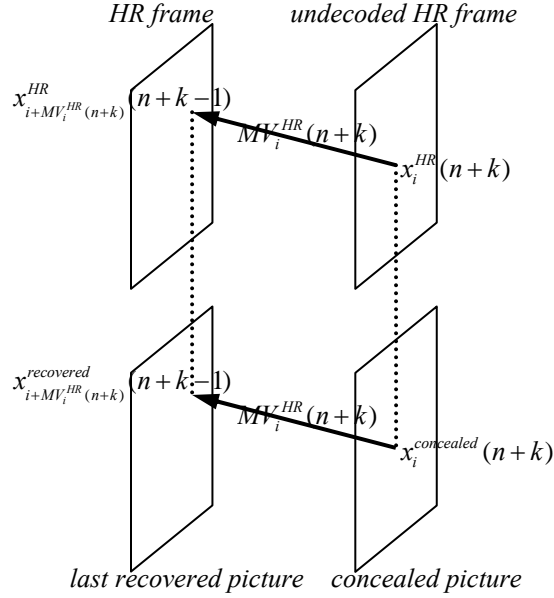

Figure 3 Two pairs of dependent pixels, as a computational unit of the interframe error recovery algorithm

More formally, the proposed ER algorithm minimizes the mean square error $D(x^{HR}_i(n+k))$ as defined in Equation (9), which turns to be a least squared estimation operation. It is assumed that we have in advance obtained the last recovered picture $p^{ER}(n+k-1,P)(m+1 \le k \le m+o)$ through the same iterative process. The beginning recovered picture is set as the concealed picture $p^{EC}(n+m,P)$ when $k=m+1$. $x^{recovered}_{i+MV^{HR}_i(n+k)}(n+k-1)$ is obtained via *k-m* ER processes and carries the concealment information of the past *k-m* frames. Naturally it contributes more to the estimation of $x^{HR}_i(n+k)$ than $x^{concealed}_i(n+k)$. As a result, we set a larger weight to the second square error.

$$D(x_i^{HR}(n+k)) \tag{9}$$

$$= \frac{1}{k-m+1}(x_i^{HR}(n+k)-x_i^{concealed}(n+k))^2 + \frac{k-m}{k-m+1}(x_{i+MV_i^{HR}(n+k)}^{HR}(n+k-1)-x_{i+MV_i^{HR}(n+k)}^{recovered}(n+k-1))^2$$

$$= \frac{1}{k-m+1}(x_i^{HR}(n+k)-x_i^{concealed}(n+k))^2 + \frac{k-m}{k-m+1}((x_i^{HR}(n+k)-e_i^{HR}(n+k))-x_{i+MV_i^{HR}(n+k)}^{recovered}(n+k-1))^2$$

Equation (9) is then differentiated with respect to $x_i^{HR}(n+k)$ to determine the optimum value $x_i^{recovered}(n+k)$. Equation (10) represents the result that we will use in the performance evaluation.

$$x_i^{recovered}(n+k) = \frac{x_i^{concealed}(n+k)+(k-m)(e_i^{HR}(n+k)+x_{i+MV_i^{HR}(n+k)}^{recovered}(n+k-1))}{k-m+1} \quad (m+1 \le k \le m+o) \tag{10}$$

To obtain $x_i^{recovered}(n+k)$, using only the previous result $x_{i+MV_i^{HR}(n+k)}^{recovered}(n+k-1)$ is equal to utilizing a series of past dependent pixels, but it is quite simple in computation.

4   Experimental Results

We applied the proposed EC and ER algorithms to our UMD codec based on the H.264 (JM86) codec. Then, we examine the effectiveness of EC on concealing the loss of consecutive frames and the effectiveness of joint EC and ER on reducing error drift. We test three different video sequences (QCIF, 15fps, 150frames), 'Foreman', 'Mother-daughter' and 'Coast-guard'. Our choice of test sequences represents a diverse set of source features, e.g. Foreman is a typical sequence, Mother-daughter has comparatively low motion, Coast-guard has a moving background and complex textures. The LR frames are encoded using the reconstructed data of the corresponding HR frames as input. The LR bit-rate is set 10% or 20% of the HR bit-rate without sacrificing too much coding efficiency. The search range for a motion vector is set to 16 in all the experiments.

4.1 Performance of Error Concealment Methods

We compare the proposed EC algorithm with the usual ones bases on UMDC(:). I) LR-EC: according to Equation (1); II) HR-EC: according to in Equation (2); III) half-weight EC: α is set to 0.5 in Equation (3). We also examine some temporal EC ones without the support of UMDC, such as TR, MMA and MVE. It is found out that MVE+TR (MVE for the first lost frame, TR for the rest) shows the best performance among their combinations. The upper bound of MVE is motion compensation (MC). In MC, it is assumed that the original motion vectors are correctly received while all the residual information is lost. So we present the representative SD(Single Description)-EC, namely MC+TR in the comparisons. The comparisons between SD's and UMD's are done at the same total bit-rate, i.e. the HR and LR bit-rates together equal the SD bit-rate.

The concealment performance of the five methods is first evaluated in Figure 4. The value of each point is obtained when the corresponding 3 HR frames $f^{HR}(n+k,P)(0 \le k \le 2)$ are lost while all the other frames are correctly received. The value of each point is calculated as the mean of PSNR over the 3 concealed HR frames. The HR bit-rate is 91kbps (QP=28), the LR bit-rate is 19kbps (QP=40) and the SD bit-rate is 111kbps (QP=27, 2 forced intra-coded macroblocks per picture). The proposed EC outperforms LR-EC, HR-EC, half-weight EC and SD-EC with *2.43dB*, *0.90dB*, *1.27dB* and *3.94dB* on average respectively. The concealment performance is further evaluated in Figure 5. The value of each point is obtained when the corresponding 6 HR frames $f^{HR}(n+k,P)(1 \le k \le 6)$ are lost while all the other

frames are correctly received. The proposed EC is also very effective and outperforms LR-EC, HR-EC and half-weight EC with *1.81dB*, *1.25dB* and *1.25dB* on average respectively. In Figure 6, our algorithm outperforms LR-EC, HR-EC, half-weight EC and SD-EC with *2.36dB*, *0.56dB*, *1.05dB* and *0.90dB* on average when the LR bit-rate is 11kbps (QP=45) and the SD bit-rate is 104kbps (QP=27). In the three groups of experiments, the concealment performance of HR-EC fluctuates more dramatically than ours although it is closest to ours. The performance of SD-EC fluctuates too dramatically although it sometimes exceeds ours. Finally in Table 1 we compare the average values for the three representative sequences when the LR bit-rate is around 20% of the HR bit-rate.
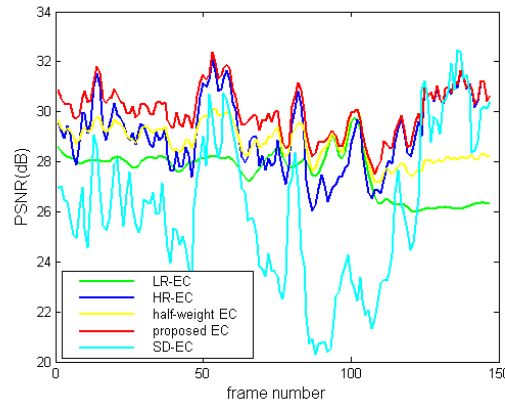
Figure 4 Concealment performance of five methods. QCIF Foreman sequence, HR bit-rate=91kbps, LR bit-rate=19kbps, SD bit-rate=111kbps, 3 consecutive frames per loss
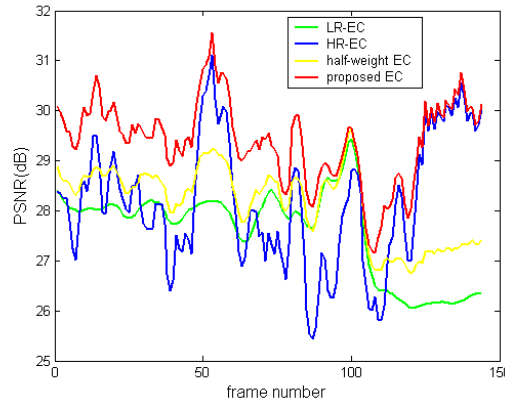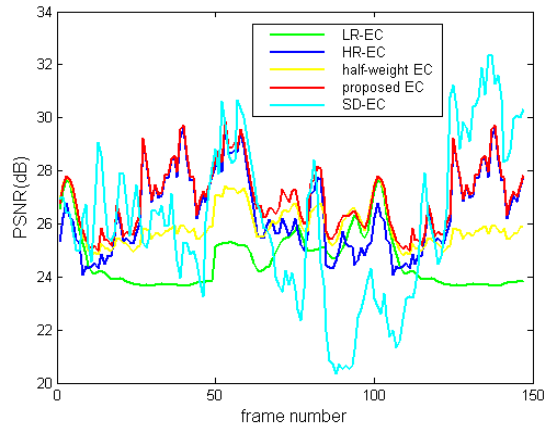

Figure 5 Concealment performance of four methods. QCIF Foreman sequence, HR bit-rate=91kbps, LR bit-rate=19kbps, 6 consecutive frames per loss

Table 1 PSNR comparison on the entire sequences (dB)

	LR-EC	HR-EC	half-weight	proposed	SD-EC
Foreman	27.59	29.12	28.75	30.02	26.08
Mother-daughter	29.69	33.37	31.41	33.62	33.91
Coast-guard	27.88	28.68	28.76	29.74	25.51

Figure 6 Concealment performance of five methods. QCIF Foreman sequence, HR bit-rate=91kbps, LR bit-rate=11kbps, SD bit-rate=104kbps, 3 consecutive frames per loss
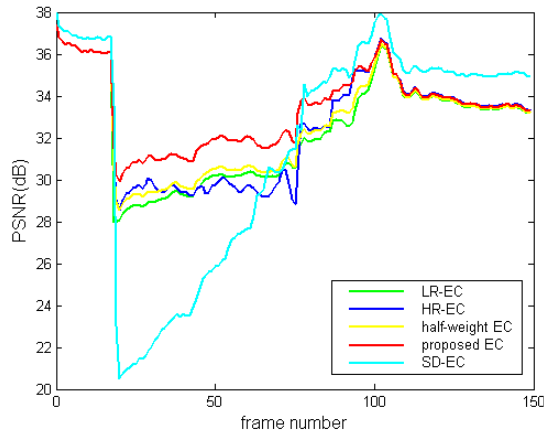
Figure 7 The recovery of PSNR after the lost 18th, 19th and 20th HR frames

Figure 7 illustrates the recovery of PSNR after a loss of consecutive frames occurs at the 18th, 19th and 20th HR frames of the Foreman sequence. Here, we do not continue performing ER after EC. We can observe that the error introduced by the concealment algorithms propagates over subsequent frames and the proposed EC is more efficient in stopping error propagation than the others. For subjective evaluation, one error-free luminance image and three images recovered by HR-EC, LR-EC and the proposed EC are shown in Figure 8. Image (b) recovered by HR-EC has obvious block artifacts in the face area, while image (c) recovered by LR-EC is blur. Comparatively, our algorithm provides a better presentation.
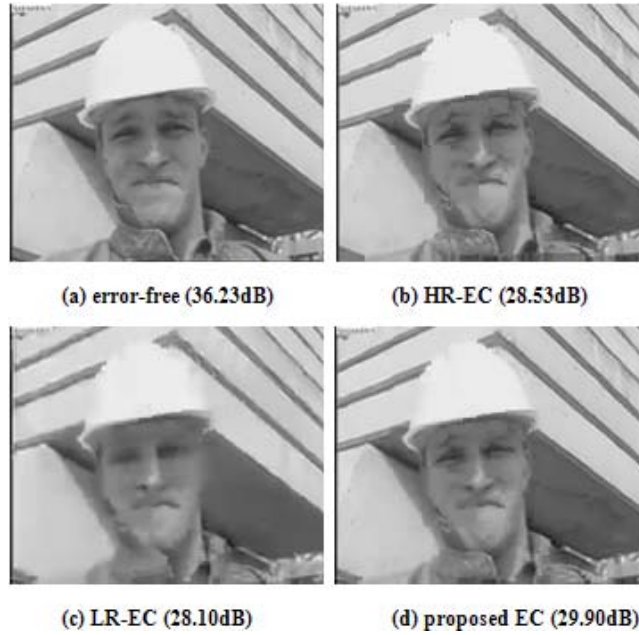
(a) error-free (36.23dB)   (b) HR-EC (28.53dB)

(c) LR-EC (28.10dB)   (d) proposed EC (29.90dB)

Figure 8 Subjective quality of the 20th frame when the 18th, 19th and 20th HR frames are lost

4.2 Performance of Joint Error Concealment and Error Recovery

We compare the proposed ER (P-ER) based on the proposed EC (P-EC) with the other combinations: I) LR-EC+P-ER: P-EC is displaced by LR-EC; II) P-EC+C(onventional)-ER: as in a conventional manner, the following HR frames are forced to be decoded after P-EC is applied to the lost ones. In fact, no ER is performed.
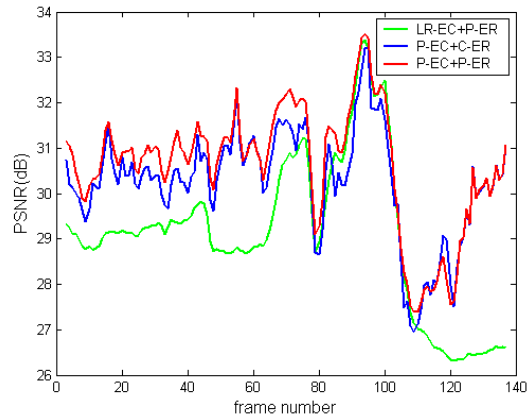

Figure 9 Recovery performance of three combinations. QCIF Foreman sequence, HR bit-rate=91kbps, LR bit-rate=19kbps, 3 consecutive frames per loss
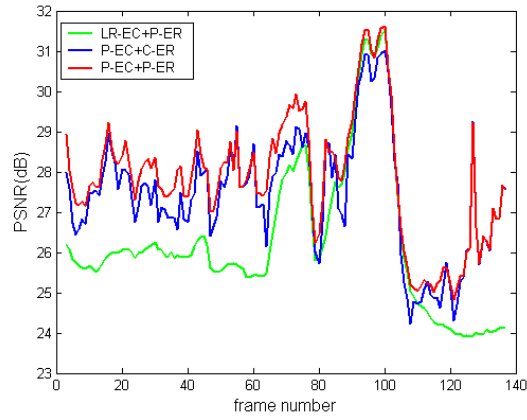
Figure 10 Recovery performance of three combinations. QCIF Foreman sequence, HR bit-rate=91kbps, LR bit-rate=11kbps, 3 consecutive frames per loss
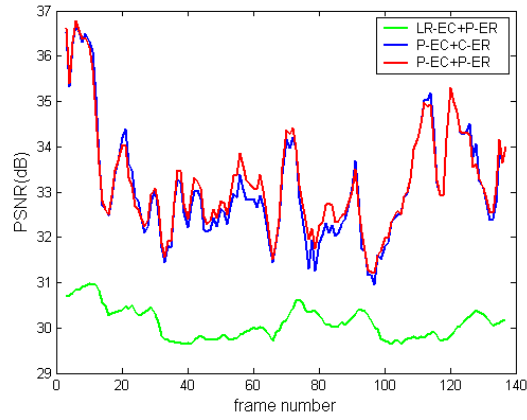

Figure 11 Recovery performance of three combinations. QCIF Mother-daughter sequence, HR bit-rate=31kbps, LR bit-rate=5kbps, 3 consecutive frames per loss
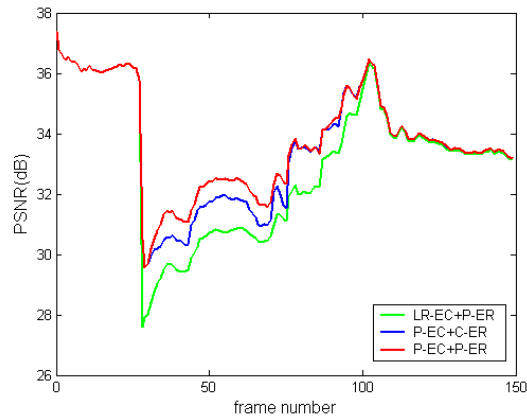

Figure 12 The recovery of PSNR after the lost 28th, 29th and 30th HR frames

The recovery performance of the three combinations is first evaluated in Figure 9. The value of each point is obtained when the corresponding 3 HR frames $f^{HR}(n-k,P)(0 \leq k \leq 2)$ are lost while all the other frames are correctly received. The value of each point is calculated as the mean of PSNR over the subsequent 10 undecoded HR frames. P-EC+P-ER outperforms LR-EC+P-ER and P-EC+C-ER with *1.52dB* and *0.36dB* on average respectively when LR bit-rate is 19kbps (QP=40). We can see that P-EC+P-ER is effective for those scenes that possess large motions, such as Foreman sequence. The same claim can be made for the case that LR bit-rate falls, as is Figure 10. P-EC+P-ER outperforms LR-EC+P-ER and P-EC+C-ER with *1.57dB* and *0.43dB* respectively when LR bit-rate is 11kbps (QP=45). On the other hand, for those scenes that have small motions, such as Mother-daughter sequence, P-ER just achieves *0.10dB* improvement to the conventional ER as shown in Figure 11.

Figure 12 illustrates the recovery of PSNR after a loss of consecutive frames occurs at the 28th, 29th and 30th HR frames of the Foreman sequence. We can observe that error decreases the fastest if the UMD codec uses P-EC+P-ER. For subjective evaluation, one error-free luminance image and three images recovered by P-EC+C-ER, LR-EC+P-ER and the proposed one are shown in Figure 13. The face area of image (b) recovered by P-EC+C-ER has some burrs, while image (c) recovered by LR-EC+P-ER is blur. Comparatively, our algorithm provides a better presentation.
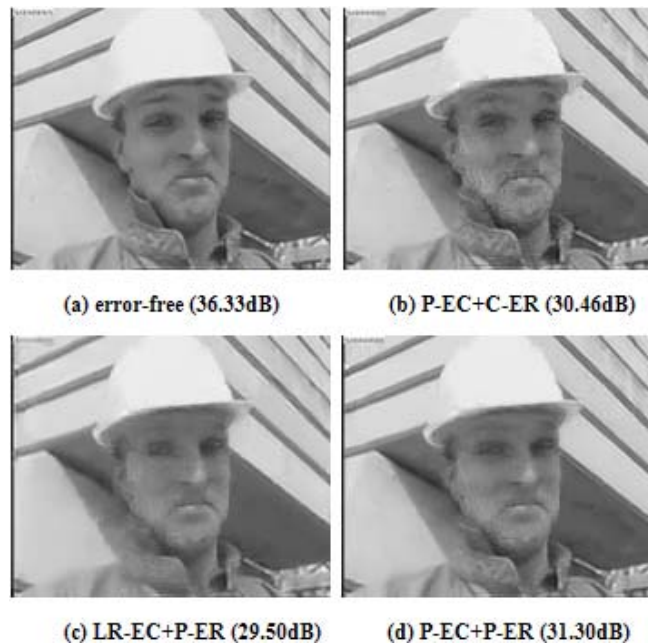

(a) error-free (36.33dB)   (b) P-EC+C-ER (30.46dB)

(c) LR-EC+P-ER (29.50dB)   (d) P-EC+P-ER (31.30dB)

Figure 13 Subjective quality of the 35th frame when the 28th, 29th and 30th HR frames are lost

5   Conclusions

The paper presents an error concealment/recovery scheme based on UMDC. The strength of the scheme is its ability to handle consecutive frame losses and perform error recovery after error

concealment. We design an iterative EC algorithm with multi-hypothesis weights and an interframe ER algorithm with respect to the intermediate information from EC. The joint design of EC and ER can be easily put into practice and applied to most UMD approaches. Moreover, the EC or ER algorithm can be displaced by other alternative to meet the given requirement in the joint, but flexible design. Extensive experiments have been carried out under different conditions. The proposed EC technique exhibits high PSNR gains versus the usual ones under the UMDC architecture and the classical ones without the support of UMDC. The proposed ER technique is efficient in reducing error drift, especially in high motion scenes. In conclusion, joint EC and ER can provide satisfactory performance on both PSNR and visual quality.

References

1. M. E. Al-Mualla, N. Canagarajah, and D. R. Bull. Temporal error concealment using motion field interpolation. IEEE Electronic Letters, 215-217, Feb. 1999.
2. M. E. Al-Mualla, N. Canagarajah, and D. R. Bull. Multiple reference temporal error concealment. in Proc. of IEEE ISCAS, 149-152, 2001.
3. J. G. Apostolopoulos, and S. J. Wee. Unbalanced Multiple Description Video Communication Using Path Diversity. in Proc. of IEEE ICIP, 2001.
4. S. Belfiore, M. Grangetto, E. Magli, and G. Olmo. An error concealment algorithm for streaming video. in Proc. of IEEE ICIP, 2003.
5. Y. Chen, K. Yu, J. Li, and S. Li. An Error Concealment Algorithm for Entire Frame Loss in Video Transmission. in Proc. of PCS, 2004.
6. S. Ekmekci, and T. Sikora. Unbalanced Quantized Multiple Description Video Transmission using Path Diversity. in Proc. of IS&T/SPIE's Electronic Imaging, 2003.
7. N. Franchi, M. Fumagalli, R. Lancini, and S. Tubaro. Multiple description video coding for scalable and robust transmission over IP. IEEE Trans. Circuits and Systems for Video Technology, 15(3):321-334, Mar. 2005.
8. M. Fumagalli, R. Lancini, and S. Tubaro. A Novel Error-Concealment Algorithm for an Unbalanced Multiple Description Coding Architecture. in Proc. of Packet Video, 2004.
9. B. Horn. Robot Vision. MIT Press, 1986.
10. F. Huang, L. Sun, and Y. Zhong. A Novel Unbalanced Multiple Description Coder for Robust Video Transmission over Ad hoc Wireless Networks. in Proc. of Multimedia Computing and Networking, 2006.
11. M. Karczewicz, and R. Kurceren. The SP- and SI-Frames Design for H.264/AVC. IEEE Trans. Circuits and Systems for Video Technology, 13:637-644, Jul. 2003.
12. S. H. Lee, D. H. Choi, and C. S. Hwang. Error concealment using affine transform for H.263 coded video transmissions. Electronics Letters, 37(4):218-220, Feb. 2001.
13. B. Li, F. Huang, L. Sun, and S. Yang. An Unbalanced Multiple Description Coding Scheme for Video Transmission over Wireless Ad hoc Networks. in Proc. of IEEE ICME, 2006.
14. Q. Peng, T. W. Yang, and C.Q. Zhu. Block-based temporal error concealment for video packet using motion vector extrapolation. in Proc. of IEEE Communications, Circuits and Systems and West Sino Expositions, 2002.
15. A. Reibman, H. Jafarkhani, Y. Wang, and M. Orchard. Multiple description video using rate-distortion splitting. in Proc. of IEEE ICIP, 978–981, 2001.
16. L. Rizzo. Effective Erasure Codes for Reliable Computer Communication Protocols. ACM Computer Communication Review, Apr. 1997.

17. P. Salama, N. Shroff, and E. J. Delp. Error concealment in MPEG video streams over ATM networks. IEEE Journal on Selected Ares in Communications, 18(6):1129-1144, 2000.
18. A. Vilei, G. Convertino, S. Oliva, and R. Cuppone. A Novel Unbalanced Multiple Description Scheme for Video Transmission over WLAN. in Proc. of ACM international workshop on WMASH, 2005.
19. Y. Wang, and Q. Zhu. Error control and concealment for video communication: A review. in Proc. of the IEEE, 86:974–997, May 1998.
20. Y. Wang, S. Wenger, J. Wen, and A. Katsaggelos. Error Resilient Video Coding Techniques. IEEE Signal Processing Magazine, 61-82, Jul. 2000.
21. Y. Wang, and S. Lin. Error resilient video coding using multiple description motion compensation. IEEE Trans. Circuits and Systems for Video Technology, 12:438–453, Jun. 2002.