Efficient Multi-Hypothesis Error Concealment Technique for H.264

Kwanwoong Song, Taeyoung Chung, and Chang-Su Kim
School of Electrical Engineering, Korea University
Email: {kwsong71, lovelool17, changsukim}@korea.ac.kr

Young-O Park, Yongdeok Kim, Younghun Joo, and Yunje Oh
Telecommunication R&D Center, Samsung Electronics
Email: {youngo.park, gokyd, mrzoo, yunje}@samsung.com

Abstract—An efficient multi-hypothesis error concealment algorithm for H.264 video is proposed in this work. The proposed algorithm temporally conceals a lost block by combining several hypothesis blocks in the previous frame. We investigate the error recovery performance according to the number of hypotheses and the weighting coefficients. Simulation results demonstrate that the proposed algorithm provides better performance than the conventional error concealment algorithm, although the additional complexity requirement is negligible.

I. INTRODUCTION

Recently, the bandwidth expansion of wired or wireless networks has brought widespread consumption of multimedia data. Especially, mobile video applications have grown rapidly with the development of wireless communication technologies. However, packets are vulnerable in wireless networks, and their losses can degrade video quality severely or even break down the decoder [1].

Several approaches have been proposed to protect the quality of reconstructed video signals against packet losses, which can be classified into three categories according to the roles of the encoder and the decoder [1]: First, in forward error correction, the encoder plays the primary role of error protection by inserting redundant information into compressed bitstreams. Second, error concealment (EC) methods attempt to hide the effect of lost blocks at the decoder side by using the information in intact neighboring blocks. Third, interactive methods can provide better reconstruction results, if there is a feedback channel from the decoder to the encoder. Among the three categories, the error concealment methods are widely used in various applications, since they do not require additional information or feedback channels.

In this work, we propose a temporal error concealment algorithm for H.264 video. Based on the multi-hypothesis theory, the proposed algorithm enables the decoder to reconstruct lost blocks reliably. Specifically, the proposed algorithm combines more than two candidate (or hypothesis) blocks to reconstruct a lost block. However, in contrast to the conventional multi-hypothesis error concealment schemes [2]–[4], the proposed algorithm requires negligible additional complexity to find the multiple candidates. Simulation results demonstrate that the proposed algorithm provides an excellent concealment performance.

The paper is organized as follows. Section II briefly reviews conventional error concealment techniques. Section III describes the proposed algorithm, and Section IV evaluates its performance. Finally, Section V gives concluding remarks.

II. CONVENTIONAL EC TECHNIQUES

EC techniques exploit the high correlation between a lost block and its neighboring blocks. Spatial EC techniques interpolate the pixel values of a lost block using the pixel values in spatially neighboring blocks. On the other hand, temporal EC techniques replace a lost block with a block in adjacent frames. Although spatial techniques are simpler to implement, their reconstruction performance is often inferior to temporal techniques. This is because the temporal correlation is generally much higher than the spatial correlation in real video sequences.

In temporal EC techniques, a lost block is replaced by the block in the previous or the following frames, whose location is specified by the motion vector. If the motion vector is also lost, it is interpolated from the motion vectors of the spatially adjacent blocks [5]. Alternatively, the decoder can directly estimate the motion vector of the lost block using the neighboring pixel values, as done in the decoder motion vector estimation (DMVE) algorithm [6], [7] and the block matching algorithm (BMA) [8]. In DMVE, the motion vector of the neighboring pixels is estimated and then used for the lost block. In BMA, the motion vector is selected such that it minimizes the side matching distortion between the lost block and the neighboring pixels.

Multi-hypothesis EC is a more advanced approach, which combines more than two candidate blocks to reconstruct a lost block. In [2], Park et al. proposed an algorithm, which finds a candidate block in each of the $N$ previous frames and combines the $N$ candidate blocks to form the concealed block. In [3], Kung et al. proposed an adaptive EC algorithm, which combines multiple candidate blocks to minimize error propagation. Also, Lee et al. proposed an algorithm to fuse the spatial EC results and the temporal EC results based on the Markov process modeling. All these algorithms yield better EC performance than the single-hypothesis approaches. However, their main disadvantage is that they require about $N$ times higher computational complexity to select $N$ hypotheses.
III. PROPOSED ALGORITHM

In multi-hypothesis EC, the decoder combines several candidate blocks to form the reconstructed signal for a lost block. The underlying assumption of multi-hypothesis EC is that the individual reconstruction errors of the candidate blocks cancel one another, and thus the energy of the final reconstruction error is stochastically lower than that of the reconstruction error of each hypothesis. However, as mentioned previously, conventional multi-hypothesis EC algorithms require much higher computational complexity than the single-hypothesis approaches.

In this work, we propose a multi-hypothesis EC algorithm, which selects multiple candidate blocks efficiently with only slight increase of computational complexity. Fig. 1 illustrates the proposed algorithm. To conceal a lost block in the kth frame \( F_k \), the proposed algorithm scans the search window in the previous frame \( F_{k-1} \) to find several candidate blocks. As in the DMVE method \([6],[7]\), we first estimate the motion vector for the lost block, which minimizes the sum of absolute differences (SAD), given by

\[
\text{SAD}(x, y) = \sum_{(x, y) \in N} |F_k(x, y) - F_{k-1}(x+v_1, y+v_1)|, \tag{1}
\]

where \( F_k(x, y) \) denotes the \((x, y)\)th pixel in \( F_k \), and \( N \) denote the set of neighboring pixel coordinates. In this work, we define \( N \) to include the left two lines, the upper two lines, and the lower two lines, as shown in Fig. 2. We assume that a slice consists of rows of macroblocks. Thus, when a macroblock is lost, its right block is also lost. Therefore, the right pixels are not included in the computation of the SAD.

The conventional DMVE selects the motion vector, yielding the smallest SAD, and replaces the lost block with the block in the previous frame specified by the motion vector. However, note that the computation of the SAD is based on the neighboring pixels, not on the lost block itself. Therefore, the estimated motion vector is not reliable and may not effectively represent the motion of the lost block. In other words, the concealed block may not be faithful to the error-free block.

To improve the reliability of the concealment result, we find the \( N \) motion vectors, which provide the smallest SAD, the second smallest SAD, and so forth. Let \((v_1^i, v_2^i)\) denote the motion vector, which provides the \(i\)th smallest SAD. Then, the proposed algorithm reconstructs the pixel value \( F_k(x, y) \) in the lost block by

\[
\hat{F}_k(x, y) = \sum_{i=1}^{N} w_i F_{k-1}(x + v_1^i, y + v_2^i), \tag{2}
\]

where \( w_i \)'s are weighting coefficients, satisfying the normalizing constraint \( \sum_{i=1}^{N} w_i = 1 \). As will be shown in the later section, double hypotheses are enough in general. In such a case, (2) can be rewritten as

\[
\hat{F}_k(x, y) = w_1 F_{k-1}(x + v_1^1, y + v_2^1) + w_2 F_{k-1}(x + v_1^2, y + v_2^2), \tag{3}
\]

where \( w_1 + w_2 \) = 1. Notice that, since SAD\((v_1^1, v_2^1)\) is higher than SAD\((v_1^2, v_2^2)\), \( F_{k-1}(x + v_1^2, y + v_2^2) \) is a less accurate estimate of \( F_k(x, y) \) than \( F_{k-1}(x + v_1^1, y + v_2^1) \) in general. However, on average, we can reconstruct \( F_k(x, y) \) more accurately by combining those two estimates as in (2), than by using either of the individual estimates.

The weights \( w_1 \) and \( w_2 \) in (3) can be fixed or adapted dynamically. For example, they can be fixed to \( w_1 = w_2 = \frac{1}{2} \). However, to achieve better EC performance, \( w_1 \) can be fixed to a higher number than \( w_2 \), since it is associated with the more accurate estimate \( F_{k-1}(x + v_1^1, y + v_2^1) \). Furthermore, assuming that the reliability of a motion vector \((v_1, v_2)\) is inversely proportional to its corresponding SAD, \( \text{SAD}(v_1, v_2) \), the weighting coefficients can be adaptively determined by

\[
w_1 = \frac{\text{SAD}(v_1^1, v_2^1)}{\text{SAD}(v_1^1, v_2^1) + \text{SAD}(v_1^2, v_2^2)},
\]

\[
w_2 = \frac{\text{SAD}(v_1^2, v_2^2)}{\text{SAD}(v_1^1, v_2^1) + \text{SAD}(v_1^2, v_2^2)}. \tag{4}
\]

In this way, as \( \text{SAD}(v_1^2, v_2^2) \) becomes larger, \( w_1 \) gets higher while \( w_2 \) gets lower. The performance of the proposed algorithm according to the weighting coefficients will be investigated further in the next section.

It is worthwhile to point out that the additional complexity due to the multiple hypotheses is minimal in this work. In the conventional DMVE, for each candidate motion vector in the search window, the SAD is computed and then the
best motion vector with the smallest SAD up to that point is recorded. Instead of recording only the best motion vector, the proposed algorithm should maintain the list of the \( N \) best motion vectors. However, the computational complexity of the list maintenance is negligible as compared with the SAD computations.

IV. EXPERIMENTAL RESULTS

A. Test Conditions

We evaluate the performance of the proposed algorithm on various test sequences, including the “Foreman” and “City” sequences, which are compressed by the H.264 standard. Table I summarizes the encoding parameters.

In H.264, a slice is the minimum independent decoding unit and consists of a number of macroblocks. Due to the variable length coding, when an error occurs in a slice, all the remaining information till the end of the slice is useless. Thus, we adopt the slice loss model, where each slice is lost or not randomly according to the slice loss rate. To simulate severe error conditions, the slice loss rate is set to 10\%, which corresponds to the loss of about two slices per frame. The locations of erroneous slices affect the quality of the reconstructed video significantly. Therefore, we use 10 different error patterns for each test and present the average PSNR performances.

B. Test Results

Fig. 3 shows the reconstructed “Foreman” 93th frames. Fig. 3 (a) is the error-free reconstruction, and Fig. 3 (b) shows the locations of erroneous slices. Fig. 3 (c) is obtained by the conventional DMVE method, while Fig. 3 (d) is reconstructed by the proposed algorithm. In this test, double hypotheses are used and the weighting coefficients are fixed to \( w_1 = w_2 = \frac{1}{2} \).

We see that the proposed algorithm provides more faithful image quality than the conventional DMVE method. Table II compares the average PSNR performances over all the frames and all the error patterns. The proposed algorithm provides 0.23 dB and 0.64 dB better performance on the “Foreman” sequence and the “City” sequence, respectively.

Fig. 4 plots the PSNR of each frame in the “City” sequence. Every thirtieth frame is encoded in the intra mode. It is assumed that those intra frames are not corrupted by packet losses. After each intra frame, the inter frames are corrupted by slice losses and their error propagation. However, it is observed that the proposed algorithm provides up to 1 dB better performance than the conventional DMVE method.

Next, Fig. 5 investigates the performance of the proposed algorithm according to the number of hypotheses \( N \). In this test, the weighting coefficients are fixed to \( \frac{1}{N} \). On both the sequences, the proposed algorithm provides the best PSNR performance when we use double hypotheses (\( N = 2 \)). It does not improve the quality of reconstructed videos, but increases the computational complexity, to use more than three hypotheses. This indicates that the reliability of the third and higher-order hypotheses are not high enough to provide a positive contribution.

Fig. 6 shows the performance of the proposed algorithm according to the weighting coefficients, when \( N = 2 \). The weight \( w_1 \) in (3) is fixed to

\[
   w_1 = \frac{\alpha}{\alpha + 1},
\]

where \( \alpha \) is called the weighting factor. We see that the performance of the proposed algorithm is optimized when \( \alpha \) is about 2, or equivalently when \( w_1 = \frac{2}{3} \) and \( w_2 = \frac{1}{3} \). Since the first hypothesis is more reliable, its weighting coefficient is twice larger than that of the second hypothesis. Notice that the case \( \alpha = 2 \) provides about 0.2 dB better performance than the equal weight case \( \alpha = 1 \).

Finally, we test the performance of the adaptive weighting scheme, which determines the weighting coefficients dynamically via (4). Table III compares the performance of the

<table>
<thead>
<tr>
<th>Sequence</th>
<th>DMVE method</th>
<th>Proposed algorithm</th>
<th>Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Foreman</td>
<td>28.19</td>
<td>28.42</td>
<td>0.23</td>
</tr>
<tr>
<td>City</td>
<td>29.45</td>
<td>30.09</td>
<td>0.64</td>
</tr>
</tbody>
</table>

TABLE III

Authorized licensed use limited to: Korea University. Downloaded on July 21, 2009 at 01:04 from IEEE Xplore. Restrictions apply.
fixed weighting scheme with that of the adaptive weighting scheme. The adaptive weighting scheme does not improve the performance. Instead, we see that the proposed algorithm provides the best performance when the weighting factor is fixed to 2.

These simulation results indicate that the proposed algorithm provides excellent error recovery performance, although the additional complexity is negligible as compared with the conventional DMVE method.

V. CONCLUSIONS

We proposed an efficient multi-hypothesis error concealment algorithm for H.264 video. To conceal a lost block, the proposed algorithm combines several hypothesis blocks in the previous frame. We investigated the error recovery performance according to the number of hypotheses and the weighting coefficients. Simulation results demonstrated that the proposed algorithm provides better performance than the conventional error concealment algorithm, but the additional complexity requirement is negligible. Further research issues include the development of more effective adaptive weighing scheme.

VI. ACKNOWLEDGEMENTS

This research was supported by the MIC, Korea, under the ITRC support program supervised by the IITA (IITA-2006-C1090-0603-0017).

REFERENCES