Error Protection with extended Dual Frame Motion Compensation

Da Liu, Li Wang, Chunyan Li and Yuncai Hao
Beijing Institute of Control Engineering
Beijing, China
lovda@outlook.com

Fang Yin
Beihang University
Beijing, China

Abstract — In this paper, an error resilient coding structure is proposed. Firstly an extended dual frame motion compensation (DFMC) coding structure is proposed, and then an end-to-end distortion model with error resilient filter is used for mode decision in Macroblock (MB) level rate distortion cost decision, finally the number of header bits packet in HQF is determined in frame level rate cost decision. Experimental results show that the proposed method can achieve better performance than previous schemes.

Keywords—Error resilient; dual frame; motion compensation; rate distortion model.

I. INTRODUCTION

With the growing demand of wireless communications, the requirement of high data rate transmission become more and more important. Although wireless channel are able to transmit high bit-rate video data, robust video transmission over wireless networks is still a challenging task due to the channel quality fluctuation. This requires novel video coding technique to improve video robustness.

Dual frame motion compensation (DFMC) is a kind of video encoding which has its own advantages in error protection video coding. In DFMC, two reference buffers are utilized for motion estimation and compensation, as shown in Fig. 1, the first reference buffer contains the most recently decoded frame, which is called short-term reference frame (STR), and the second one contains a reference frame from the past, which is named as long-term reference frame (LTR). In jump updating DFMC, LTR is periodically updated. It remains static for \( N \) frames, and jumps forward to be the frame at a distance 2 back from the frame to be encoded. An example is shown in Fig. 1, for the frames from time instant \( i-N+1 \) to \( i \), the LTR is \( i-N-1 \). When the encoder moves on to encoding frame \( i+1 \), the STR will slide forward by one to frame \( i \), and the LTR will jump forward by \( N \) namely from frame \( i-N-1 \) to frame \( i-1 \).

In jump updating DFMC, LTR is usually allocated more bits to have high quality and the frame is called high quality frame (HQF), such as the \( i \)-th frame \( f_{i+1}^{H} \) and the \( i-N \)-th frame \( f_{i-N+1}^{H} \) in Fig. 1; the other has relatively lower quality and is called low quality frame (LQF), such as the \( i+1 \)-th frame \( f_{i+2} \) in Fig. 1, which have similar quality and will be utilized as STR for the next frame.

A number of DFMC related error resilience and error concealment have been done. In [5], the LTR mode is determined by the recursive optimal per-pixel estimate (ROPE) algorithm. In [8], the drift errors are reduced by using feedback in mode decisions in DFMC, the error propagation is reduced with the proposed method. In [9], a binary decision tree designed by the classification and regression trees (CART) algorithm was utilized to choose among various error concealment choices in DFMC. It has better performance than just concealing using the short-term median MV block.

Some other error resilient work for HEVC has also been done. In the paper [11], the behavior of HEVC video coding standard over diverse networking scenarios is studied. It is shown that HEVC is less error resilient compared to H.264/AVC, especially in the scenes with high amount of motion, in exchange of a significantly reduced bandwidth. The paper [12] addresses the problem caused by motion vector coding dependencies on the error resilience performance of the emergent High Efficiency Video Coding (HEVC) standard. A method based on the prediction dependency of motion vectors (MV) is proposed to select the most relevant ones for redundant coding with reduced overhead. The paper [13] proposed a method to make intelligent use of temporal motion vector (MV) candidates during the motion estimation process, in order to decrease the temporal dependency, and improve the error resiliency without penalising the rate-distortion performance. This paper [14] described the HEVC error resilience in end-to-end multimedia systems, video transport and delivery such as broadcast, television over the Internet Protocol, Internet streaming, video conversation.
In this paper, an error resilient DFMC is proposed for video transmission over error-prone channels. Firstly, an extended DFMC is presented. Then MB level rate distortion cost decision with error resilient filter is utilized to MB level mode decision. Finally, frame level rate distortion cost decision is performed in HQF to determine header transmission times.

The rest of the paper is organized as follows. In Section II, the proposed methods are given. In Section III, experimental results are provided. Finally, Section IV concludes this paper.

II. ERROR RESILIENT DFMC

A. Extended dual frame motion compensation

In the dual frame motion compensation, HQF is utilized as long-term reference frame for frames following the HQF. In the work, an extended dual frame motion is proposed. The HQF is utilized as long-term reference frames for frames before and following it, as shown in Fig.2. In the extended DFMC, each HQF and following LQFs comprise a GOP. Since the GOP length and bits allocation is not the emphasis of the paper, the bit allocation in HQF is 4 times the average bit allocation in following LQFs, and the GOP length is set as 8.

![Figure 2. Extended DFMC.](image)

B. End-to-end distortion model

The end-to-end distortion model is a further deduction from the model proposed in [15]. The difference is that, two data partitions A and B are assumed in the paper. A contains the header information such as MB types, quantization parameters, and motion vectors, which are more important than the remaining texture data. B contains texture coefficients.

Some notations used in the derivation is defined as follows. Let \( f_{i}^{n} \) be the original value of pixel \( i \) in frame \( n \), and let \( \hat{f}_{i}^{n} \) and \( \tilde{f}_{i}^{n} \) be the reconstructed values in the encoder and decoder, respectively. Let \( r_{i}^{n} \) be the reconstructed residue in the encoder, i.e., \( \hat{f}_{i}^{n} = \tilde{f}_{i}^{n} + r_{i}^{n} \), when it references pixel \( j \) in frame \( ref \). When the current pixel is lost in the decoder, it copies from pixel \( k \) in frame \( n-1 \).

1. Suppose the transmission error rates of partitions A and B are separately \( p \) and \( q \), and the header information transmit \( m \) times. Then we can represent \( \hat{f}_{i}^{n} \) as

\[
\hat{f}_{i}^{n} = \begin{cases} 
\tilde{f}_{i}^{ref} & \text{w.p. } (1-p)^2 \\
\hat{f}_{i}^{ref} & \text{w.p. } (1-p)p \\
\hat{f}_{i}^{n-1} & \text{w.p. } p/m 
\end{cases}
\] (1)

In obtaining the \( ref \) and \( n-1 \), since the prediction value \( ref \) and \( n-1 \) may be polluted by transmission error, loopfilter is adopted to reduce the error propagated from reference frame. For each block in LQF, the variance in two reference block is calculated and separately named as \( B_1 \) and \( B_2 \). If the difference of \( B_1 \) and \( B_2 \) is larger than a threshold \( T \). Then the block with larger variance has propagated error. In the case, error resilient filter is done in the block to reduce propagated error, the filter is defined as

\[
g(i,j) = \frac{\sum_{k,l} f(k,l) \omega(i,j,k,l)}{\sum_{k,l} \omega(i,j,k,l)}
\] (2)

where \( \omega(i,j,k,l) \) is determined by distance factor and pixel value factor

\[
\omega(i,j,k,l) = F_d \times F_p
\] (3)

and

\[
F_d = \{(i-k)^2 + (i-l)^2\}
\] (4)

\[
F_p = \exp(|| f(i,j) - f(k,l) ||^2).
\] (5)

\( f(i,j) \) in the block which has the larger variance is replaced with the pixel value in the same location of the block which has the small variance. In the work, \( T \) is 1.5 times the min value of \( V1 \) and \( V2 \).

And then, the expectation of end-to-end distortion in the decoder is

\[
d(n,i) = E((f_{i}^{n} - \hat{f}_{i}^{n})) = (1-p)^2 E((f_{i}^{n} - (\tilde{f}_{i}^{n} + r_{i}^{n} + f_{i}^{ref})^2)) + (1-p)pE((f_{i}^{n} - \hat{f}_{i}^{ref})^2) + (p/m)E((f_{i}^{n} - \hat{f}_{i}^{n-1})^2)
\] (6)

\[
= (1-p)^2 E((f_{i}^{n} - \hat{f}_{i}^{n})^2) + (1-p)^2 E((f_{i}^{n} - \hat{f}_{i}^{ref})^2) + (p/m)E((f_{i}^{n} - \hat{f}_{i}^{n-1})^2)
\] (7)

\[
d_s = d_s + d_{ep},
\]

where \( d_s \) represents the source distortion and is calculated as
In a block can be defined as the sum of distortions (\( d(n,i) \)) in all contained pixels. Then in the MB level mode decision, the best mode is calculated by minimum cost \( J \), where

\[
J = D + \lambda R.
\]  

In \( J \), \( R \) is the encoded bit rate in the block, \( D \) is the overall distortion which include source distortion.

### C. HQF Header transmission times determination

The overall distortion \( D \) in HQF can be calculated as

\[
D_{\text{all}} = D_s + D_{\text{influence}}
\]  

where \( D_s \) is the overall distortion in HQF, \( D_s \) is the source distortion and is calculated as the sum of source distortions \( d_s \) in all contained pixels of the frame. \( D_{\text{influence}} \) is the overall influenced distortion in the HQF. \( D_{\text{influence}} \) will be propagated into the LQFs before and following the HQF which utilize the HQF as reference frame. Since the bi-prediction is adopted in every LQF, then error propagation in every LQF influenced by the HQF is \( 1/2 \times D_{\text{influence}} \). Before and after the HQF, 16 LQFs utilize the HQF as reference frame, then the overall influenced distortion is \( 1/2 \times D_{\text{influence}} \times (8+8) = 8 \times D_{\text{influence}} \).

In the proposed error resilient JU-DFMC, lots of reference blocks in LQF are come from HQF, so HQF is very important than LQFs. If one error occurs at HQF, the quality in following LQFs and HQFs will all be influenced. In the case, header information in HQF becomes a very important factor. If header information is not lost, even errors occur in HQF, the quality will not be influenced too much.

In the study, header transmission time in LQFs is 1. In HQFs, header transmission time is determined by frame level rate distortion cost decision. In the frame level rate distortion cost, the distortion is not only include the end-to-end distortion in HQF itself, but also the end-to-end distortion in following LQFs which is influenced by HQF,

\[
D_{\text{all}} = D_{\text{HQF}} + 8 \times D_{\text{influence}}
\]

where \( D_{\text{HQF}} \) is the end-to-end distortion in HQF itself, \( D_{\text{influence}} \) is the overall influenced distortion in LQFs before and following HQF.

After coding the HQF, a frame level rate distortion cost decision is performed to decide the header transmission times in the current HQF,

\[
J = D_{\text{HQF}} + 8 \times D_{\text{influence}} + \lambda \times (R_{\text{HQF}} + (m-1) \times R_{\text{header}})
\]

where \( D_{\text{HQF}} \) is the end-to-end distortion in the HQF, in calculating \( D_{\text{influence}} \), the LQFs after the HQF is not encoded and then the influenced distortion in LQFs after the HQF is predicted from that in LQFs before the HQF which utilized the HQF as reference frame, \( R_{\text{HQF}} \) is the bit rate in the current HQF, \( R_{\text{header}} \) is the bit rate of the header information in the current HQF, \( m \) is the header transmission time. With the change of \( m \), \( D_{\text{influence}} \) changes as well. When \( J \) is the smallest, the header transmission time \( m \) is the best. In the first HQF, \( m \) is initialized as 3.

### III. EXPERIMENTAL RESULTS

The proposed method is integrated into the H.265 reference software HM16.0. The test sequences are all encoded at a frame rate of 30 fps. Each sequence has 200 frames. In motion estimation, the search range is \( \pm 16 \). The entropy coder is context-adaptive binary arithmetic coding (CABAC). Each row of MB comprises a slice and is transmitted in a separate packet. To evaluate the average PSNR, the method [15] is employed.

The performance of the proposed scheme is evaluated against two other approaches: reference HEVC in which TMVP is enabled and the TIDR method as proposed in [13]. In the anchor coding scheme HEVC HM16.0, 8 B frames are in a GOP, and two reference frames for each B frame. The QPs in frames 1-8 of a GOP are slightly adjusted as +3,+2,+3,+1,+3,+2,+3,+0 based on the fixed QP. The experimental results of the error protection method in decoder are repeated 100 times using the bit error sequences which are transmitted from the encoder via the error-prone channels [16].

In Table 1, the experimental results of the proposed error resilient method (Proposed Method), HEVC, TIDR method proposed in [13] are listed and compared. The bit-stream of each scheme has the same bit rates of 493.32, 651.90, 1220.51, 1321.21 and 863.37 kbps in sequences Kimono, ParkScene, Cactus, BasketballDrive and BQTerrace respectively. Compared to HEVC and TIDR [13], the...
proposed error resilient method can separately achieve a maximum gain of 2.32dB and 1.98dB on ParkScene sequence at a packet loss rate of 20%, and can separately achieve an average gain of 0.53dB and 0.96dB on all sequences under the different packet loss rates. This shows that the proposed error resilient method can bring better performance than previous schemes.

The performance gain comes from three aspects. Firstly, the error propagation in LQFs of a GOP is terminated by the extended DFMC prediction structure. Secondly, the filter in end-to-end distortion model improves the prediction performance in decoder after video transmission over error channel. Thirdly, HQF Header is transmitted multiple times, then error propagation in HQFs is reduced.

IV. CONCLUSION

In this paper, an error resilient coding structure has been proposed. Firstly an extended DFMC was proposed. Then an end-to-end distortion model with filter was utilized for MB level mode decision. Finally header transmission times in HQF were determined by frame level rate distortion cost decision. Experimental results show that the proposed method can achieve better performance than previous schemes. The proposed scheme can be widely used in network video communication. In the future, adaptive LQF header transmission times will be further exploited.

V. ACKNOWLEDGEMENT

This work was supported by the National Natural Science Foundation of China under grants 61302177 and 61174004.

REFERENCES


<table>
<thead>
<tr>
<th>Sequence</th>
<th>Schemes</th>
<th>PSNR of different packet loss rates (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0%</td>
</tr>
<tr>
<td>Kimono (1080p) 493.32kbps</td>
<td>HEVC</td>
<td>34.43</td>
</tr>
<tr>
<td></td>
<td>TIDR [13]</td>
<td>34.21</td>
</tr>
<tr>
<td></td>
<td>Proposed Method</td>
<td>34.27</td>
</tr>
<tr>
<td></td>
<td></td>
<td>31.50</td>
</tr>
<tr>
<td></td>
<td>HEVC</td>
<td>31.90</td>
</tr>
<tr>
<td></td>
<td>Proposed Method</td>
<td>31.51</td>
</tr>
<tr>
<td>ParkScene (1080p) 651.90kbps</td>
<td>HEVC</td>
<td>32.17</td>
</tr>
<tr>
<td></td>
<td>Proposed Method</td>
<td>31.99</td>
</tr>
<tr>
<td></td>
<td>HEVC</td>
<td>31.01</td>
</tr>
<tr>
<td></td>
<td>TIDR [13]</td>
<td>30.97</td>
</tr>
<tr>
<td></td>
<td>Proposed Method</td>
<td>31.01</td>
</tr>
</tbody>
</table>

TABLE I

PERFORMANCE COMPARISON OF THE PROPOSED METHOD WITH OTHER SCHEMES.