Reference picture selection using checkerboard pattern for resilient video coding

This item was submitted to Loughborough University’s Institutional Repository by the/an author.

Citation: CARREIRA, J.F. ... et al, 2015. Reference picture selection using checkerboard pattern for resilient video coding. IN: Proceedings of the 2015 IEEE International Conference on Visual Communications and Image Processing (VCIP), 13th-16th December 2015, Singapore.

Additional Information:

- © 2015 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Metadata Record: https://dspace.lboro.ac.uk/2134/20919

Version: Accepted for publication

Publisher: © IEEE

Please cite the published version.
Reference picture selection using checkerboard pattern for resilient video coding

J. Carreira\textsuperscript{1,2}, P. Assuncao\textsuperscript{2}, S. Faria\textsuperscript{2}, E. Ekmekcioglu\textsuperscript{1}, A. Kondoz\textsuperscript{1}, H. Lim\textsuperscript{1}

\textsuperscript{1} Institute for Digital Technologies, Loughborough University in London, United Kingdom
\textsuperscript{2} Instituto de Telecomunica\c{c}\~{o}es; Instituto Polit\'ecnico de Leiria, Portugal

\texttt{joao.f.carreira@ieee.org}

\textbf{Abstract}—The improved compression efficiency achieved by the High Efficiency Video Coding (HEVC) standard has the counter-effect of decreasing error resilience in transmission over error-prone channels. To increase the error resilience of HEVC streams, this paper proposes a checkerboard reference picture selection method in order to reduce the prediction mismatch at the decoder in case of frame losses. The proposed approach not only allows to reduce the error propagation at the decoder, but also enhances the quality of reconstructed frames by selectively constraining the choice of reference pictures used for temporal prediction. The underlying approach is to increase the amount of accurate temporal information at the decoder when transmission errors occur, to improve the video quality by using an efficient combination of diverse motion fields. The proposed method compensates for the small loss of coding efficiency at frame loss rates as low as 3%. For a single frame-loss event the proposed method can achieve up to 2 dB of gain in the affected frames and an average quality gain of 0.84 dB for different error prone conditions.

\textbf{Index Terms}—HEVC, error resilience, error concealment, reference picture selection, motion compensation.

\section{I. INTRODUCTION}

Recent developments in video compression and networking technologies are contributing to develop new applications over the internet such as high resolution video streaming, multi-view video services, among others. Most of the upcoming applications and services will use the High Efficiency Video Coding (HEVC) \cite{1} standard, developed by the ITU-T Video Coding Experts Group (VCEG) and the ISO/IEC Moving Picture Experts Group (MPEG). HEVC addresses the demands of increased coding performance \cite{2}, by adopting a new block partition structure and improving the intra and inter coding methods. The new coding tools used in HEVC bring significant compression efficiency improvement, but some disadvantages at the same time, such as increased complexity and reduced error robustness. The specific characteristics of error robustness in HEVC were analysed in \cite{3}, and compared against the H.264/AVC. The results confirmed that HEVC has lower error robustness performance, despite its higher coding efficiency. Thus, HEVC bitstreams subject to transmission losses deliver lower objective and subjective quality \cite{4}.

Error robustness can be improved using resilient coding techniques to mitigate the artefacts caused by packet losses. These techniques can be grouped in four categories: localisation, data partitioning, redundant coding and concealment-driven techniques \cite{5}. In \cite{6} to reduce error propagation due to mismatch predictions, a transcoding operation is performed complemented with network feedback to select the correctly transmitted frames as reference. Although these techniques offer a more reliable control of the decoder distortion, they require feedback from the receiver side, which may not be suitable for real-time applications. Moreover, the error resilience techniques should be adapted to cope with different network technologies and protocols. In \cite{7} a rate distortion optimization for error resilience is presented, based on the distortion estimation using the Recursive Optimal Per-pixel Estimate (ROPE) method. In \cite{8} a optimization algorithm is used to select between short and long term references in order to increase the prediction distance and reduce error propagation. In the decoder side a concealment method can be used to take advantage of this prediction structure \cite{9}. Although these methods efficiently predict the decoder distortion in case of packet loss, they rely either on interpolated reference frames or long-term references, which might dramatically decrease the coding performance. Regarding HEVC, combining resilient coding to cope with network losses and efficient concealment of missing frames is still an open issue, without well established solutions.

Taking the decoder side into account, the problem of error resilient coding can also tackle the efficiency of possible error concealment methods used to minimise the impact of network losses. It is known that, in general concealment methods based on motion compensation are expected to perform better in comparison with spatial concealment. In \cite{10} motion vectors (MVs) of a previously received frame are projected onto the missing frame, in order to estimate the missing pixels. In \cite{11} the direct motion copy method \cite{12} is improved, by using a recursive difference between the MVs of successive frames, in order to refine the final MV accuracy. In order to increase the MV estimation accuracy bi-directional motion vector prediction may also be used as, shown in \cite{13}. The problem of error concealment was also investigated for HEVC by adapting previous methods to the new coding tools and data structures. For instance, in \cite{14} the block partitions of neighbouring frames are used to assist motion extrapolation in order to keep smooth object boundaries. In \cite{15} the same principle is used to prioritise the concealment through a boundary match algorithm. In order to select only the reliable motion information, the residue information can be exploited, as shown in \cite{16}, where the authors concluded that a higher residue implies less accurate MVs. In general, these methods focus on post-processing of past frames, which may lead to incorrect motion information recovery, as the motion is not constant neither purely translational.

In this paper a novel robust coding method to improve the overall error resilience of HEVC coded video is proposed. In order to facilitate concealment of lost frames and improve their reconstructed quality, a specific reference picture selection method based on checkerboard pattern is applied at the encoder side. Furthermore, a concealment method that takes advantage of the checkerboard pattern is also devised. It is shown that more accurate reconstruction of the motion field is achieved for missing frames. Moreover, due to the specific prediction structure used in the proposed method, it is possible to use bi-directional motion compensation to recover the missing pixels of lost P-frames.

This work was supported by the Funda\c{c}\~{a}o para a Ci\^{e}ncia e a Tecnologia, Portugal (PhD grant SFRH/BD/86368/2012) and Project UID/EIA/50008/2013.
The remainder of the paper is organised as follows. Section II describes the proposed method and Section III covers the experiments discussion of results. Finally, Section IV concludes this paper.

II. PROPOSED METHOD

The proposed method relies on a reference picture selection scheme to reduce error propagation when frame loss occurs in transmission networks. Moreover, an efficient concealment method is also developed to recover from frame loss events. The aim is to achieve increased resilience and higher quality in reconstructed frames. At the encoder side, a reference picture selection scheme based on a checkerboard pattern is devised to uniformly distribute the use of different reference pictures across the coding tree units (CTU) of the current frame. At the decoder side, a concealment method is used to take into consideration such reference coding structure whenever it is used. The proposed concealment method is essentially focused on efficient recovery of the lost motion field using information available in the neighbouring frames. It combines multiple MVs, in order to derive a more accurate motion field, which is then used to reconstruct the missing frame. Note that, in the proposed method standard compliance is maintained. The remainder of this section will briefly describe the coding structure used, and the strategy to recover the missing motion field.

A. Encoder side: checkerboard reference picture selection

To reduce the error propagation in case of frame loss, a distributed scheme for selecting reference pictures in the encoder is devised, aiming at decreasing the number of affected predictions in HEVC decoding. At the encoder side, such approach imposes different reference pictures for coding adjacent CTUs in order to avoid using the same reference picture in neighbouring CTUs. Therefore, the reference frames are alternated across each CTU within a frame, forming a checkerboard structure, as shown in Figure 2. Such scheme minimizes the probability of using the same reference picture for neighbouring CTUs. Although this method uses similar patterns as those of H.264/AVC FMO, it is applied to distribute reference frames instead of slice distribution. The underlying principle of this method is the following: let assume a coding scheme with two reference pictures, where each one is used to encode half of the pixels in the current frame. Then, if any single frame is lost, the subsequent frame is only partially affected by mismatch predictions because only half of its predictions come from the missing frame. Thus, the overall error propagation is reduced. The proposed method may also be extended to distribute the motion field over more than two reference frames in a controlled manner. This increases the probability of receiving reference frames without errors, which contributes to achieve higher error resilience in case of burst errors.

Finally, since the motion information available in the subsequent frame is referenced not only to the missing frame but also to past ones, many received MVs actually cross the missing frame providing increased accuracy when used for concealment. This fact is taken into account when recovering the missing motion field, as described in the next sub-section.

B. Decoder side: bi-directional motion vector extrapolation

At the decoder side, bi-directional MV extrapolation is used to recover the missing frames, referred to as BME. Figure 1 illustrates an example with a missing frame \( f_0 \) at instant \( t_0 \), which is recovered using the motion fields \( v_{-1} \) and \( v_1 \) from \( f_{-1} \) and \( f_1 \), respectively. In order to recover \( f_0 \), motion compensation from the previously decoded frames is performed. Therefore, the motion field of the missing frame, \( v_0 \), is firstly recovered, using the neighbouring information. In order to use the neighbouring MVs, the following procedure is first applied:

1) The MVs of the neighbour frame are extrapolated to compute the motion field \( v'_0 \) using:

\[
v'_0 = \frac{t_0 - t_C}{t_1 - t_R} \times v_{1}, \quad I = -1, 1,
\]

where \( t_C \) is the time instant of the frame used to reconstruct the missing one \( (t_{-1}, t_0, t_1) \), \( t_R \) is the time instant of the reference frame pointed by the original motion vectors \( v_{-1} \) and \( v_1 \).

2) Subsequently, the MV associated with position \( X_{f_t} \) in \( f_t \) may be used to recover the block in \( f_0 \), at the position given by:

\[
X_{f_0} = X_{f_t} + \left( \frac{t_0 - t_{f_t}}{t_1 - t_R} \times v_{f_1} \right).
\]

It should be noted that extrapolation is performed to the MVs from \( f_{-1} \) and \( f_1 \), which may result in more than one candidate for each block in the missing frame.

The final MVs are selected among the candidates from both neighbour frames. A low value of the residual energy \( E_r \) indicates a more accurate MV candidate \( v_i \). Therefore, the final MV, \( v_0^*(x, y) \), selected for motion compensated concealment from frame \( f_{-1} \), is obtained as follows:

\[
v_0^*(x, y) = \arg\min_{i=-1,1} E_r \{ v_i(x, y) \}.
\]

When motion extrapolation based on (1) and (2) does not produce MVs for all blocks, the co-located MVs from the previous frame are used. Finally, if all of these methods fail to find a MV for a certain region, then the median of the neighbours’ MVs is used. One should note that in case of B-slices the previous operations are performed.
As mentioned before, the encoder distributes the reference frames uniformly across all CTUs. Assuming that there might be regions in \( f_1 \) that are not affected by loss events, in the proposed method we use an extrapolated MV pointing to \( f_1 \) to apply bi-directional prediction in the reconstruction of the missing frame. Thus, the concealment takes into consideration information from future objects’ position.

### III. PERFORMANCE EVALUATION

The performance of the proposed method is evaluated in this section. The efficiency of the checkerboard reference picture selection (\( \text{CHKB} \)) is compared with the reference HEVC (\( \text{REF} \)), both in error free and lossy conditions. In this case both methods use the same concealment method which is only based on extrapolated MVs from the previous neighbouring frame (\( \text{MVE} \)). The performance of the checkerboard scheme is also evaluated with the concealment method described above to assess the benefits of joint resilient coding and adapted concealment. This is referred to as \( \text{CHKB (BMVE)} \).

Five well-known video sequences with 240 frames each and different resolutions were used in the experiments. Table I presents a summary of the sequences’ characteristics. These test sequences were selected to cover different types of motion and texture complexity. The experimental results were obtained using the HEVC reference software, version 16.2 using an IDR period of 32 frames. The sequences were encoded using different bitrates and two common HEVC test configurations were selected, i.e., Lowdelay and Random Access, in order to obtain results with and without B frames.

Since the checkerboard approach reduces the coding efficiency, Table II shows the Bjontegaard’s Delta PSNR (\( \Delta \text{PSNR} \)) using the checkerboard reference picture selection in error free conditions.

Fig. 3. Error propagation after frame #6 is lost for the Basketball Drill, Kendo and Race Horses sequences encoded at 2.0 Mbps.

Since a relevant performance indicator is the error propagation over temporally predicted frames, this was evaluated using a single frame loss in the Lowdelay coding configuration. Figure 3 shows the PSNR when Frame #6 is missing. One should note that both sequences were encoded at constant bitrate (2.0 Mbps), and the proposed approach does not affect intra-coded frames. The results indicate that the proposed checkerboard method is able to outperform the reference HEVC (\( \text{REF} \)) for different sequences, achieving up to 2 dB of gain in the reconstruction quality. This indicates that alternating the reference picture in adjacent CUs clearly leads to lower error propagation. When the concealment method described in the previous section is also used (i.e., \( \text{BME} \)), further quality gains are noticeable for Kendo and Race Horses sequence, revealing that using pixel information from future frames for reconstructing the missing ones actually contributes to increase the performance. Overall, the combination

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Resolution</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basketball Drill</td>
<td>832 × 480@50/50</td>
<td>High motion with several basketball players</td>
</tr>
<tr>
<td>Book Arrival</td>
<td>1024 × 768@30/30</td>
<td>Low translational motion with two moving persons</td>
</tr>
<tr>
<td>Kendo</td>
<td>1024 × 768@30/30</td>
<td>Moderate motion with two moving persons, and moving camera</td>
</tr>
<tr>
<td>Race Horses</td>
<td>832 × 480@30/30</td>
<td>Moderate motion with several horse riders</td>
</tr>
<tr>
<td>Tennis</td>
<td>1920 × 1080@32</td>
<td>High motion with one moving person in the scene</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Coding configuration</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Basketball Drill</td>
<td>Lowdelay</td>
<td>Random Access</td>
</tr>
<tr>
<td>Book Arrival</td>
<td>-0.88</td>
<td>-0.29</td>
</tr>
<tr>
<td>Kendo</td>
<td>-0.92</td>
<td>-0.48</td>
</tr>
<tr>
<td>Race Horses</td>
<td>-0.78</td>
<td>-0.32</td>
</tr>
<tr>
<td>Tennis</td>
<td>-0.61</td>
<td>-0.34</td>
</tr>
<tr>
<td>Average</td>
<td>-0.80</td>
<td>-0.37</td>
</tr>
</tbody>
</table>

TABLE I

TEST SEQUENCES USED IN THE EXPERIMENTS

TABLE II

BJONTEGAARD \( \Delta \text{PSNR} \) USING THE CHECKERBOARD REFERENCE PICTURE SELECTION IN ERROR FREE CONDITIONS.
of the checkerboard reference picture selection with an appropriate concealment is able to significantly increase the quality, achieving an average gain of 3 dB in the affected frames, in comparison with the reference HEVC bitstream.

Further tests were performed to evaluate the effectiveness of the proposed method under more realistic loss rates. For each test condition 30 trials were performed and the average results are analysed. Table III shows the ΔPSNR obtained using the proposed reference picture selection scheme for the both concealment methods tested, i.e., MVE and BME. The reference for the Bjontegaard’s Delta calculation is the HEVC with the MVE concealment method. Despite the reduction of coding performance in error-free transmission, the results in Table III show that the proposed alternate reference picture selection method, based on a checkerboard pattern, is able to outperform the reference HEVC frame loss rates as low as 3%. The increase in the quality clearly indicates that better error resilience is obtained for the proposed reference picture selection method in comparison with the reference HEVC. As shown before, combining the checkerboard-based reference picture selection scheme with a more efficient concealment can achieve up to 0.4 dB of quality gains.

In summary, the overall quality achieved with the combination of the checkerboard approach with an appropriate concealment method is able to improve the HEVC robustness under error-prone transmission. The small quality degradation due to loss of coding efficiency of the checkerboard approach is clearly compensated by the gains under lossy conditions. Overall, an average gain of 0.84 dB and (0.28 dB for Random Access) is achieved over different loss ratios and for different coding configurations.

### IV. Conclusion

In this paper a robust coding method is proposed to increase the error resilience and to enable efficient concealment techniques at the decoder. At the encoder side, a checkerboard reference picture selection method is devised. The results show that this approach significantly reduces the error propagation, by reducing the number of mismatch predictions. At the decoder side, a concealment method is developed to recover from full frame loss in the HEVC decoder. The concealment method combines multiple MV candidates in order to achieve a more efficient reconstruction. Since error propagation only affects certain regions of the leading frames, due to the robust coding approach used, the bi-directional prediction may be used to enhance the recovery quality. This results in higher reconstruction quality. The analysis of the results also indicates that further research should be carried out to find a more optimal selection of the reference frames, increasing the coding performance without compromising the overall error resilience.

### References


