IMPROVED B-SLICES DIRECT MODE CODING USING MOTION SIDE INFORMATION

Xiem Hoang Van\textsuperscript{a}, João Ascenso\textsuperscript{b}, Fernando Pereira\textsuperscript{a}

\textsuperscript{a}Instituto Superior Técnico – Instituto de Telecomunicações, Lisbon, Portugal

\textsuperscript{b}Instituto Superior de Engenharia de Lisboa – Instituto de Telecomunicações, Lisbon, Portugal

ABSTRACT

The so-called DIRECT coding mode plays an important role in the RD performance of predictive video coding such as the H.264/AVC and MPEG-4 standards because there is typically a large probability that the DIRECT mode is selected in B-slices by the rate-distortion optimization (RDO) process. Although the current H.264/AVC DIRECT coding procedure exploits the motion vectors obtained from the reference frames in a rather effective way, it may still be improved by considering better motion information such as motion data derived by the side information creation process typical of distributed video coding. Therefore, this paper proposes an improved DIRECT coding mode for B-slices by efficiently exploiting some motion side information available at both the encoder and decoder. Experimental results show that the proposed improved DIRECT coding mode provides up to 8\% bitrate saving or 0.46 dB PSNR improvement.

Index Terms— Direct mode coding, predictive video coding, decoder side information, B-slices coding

1. INTRODUCTION

Nowadays, the H.264/AVC standard\cite{1} dominates the video coding landscape with a wide range of applications and a huge number of customers. Despite this success, the ITU-T VCEG and ISO/IEC MPEG standardization bodies are working towards a new video compression standard, the High Efficiency Video Coding (HEVC) project. In fact, this joint collaborative team on video coding aims to develop a new generation of video compression technology that has substantially higher compression capabilities than the available H.264/AVC standard. The hybrid block-based motion compensated H.264/AVC coding architecture is still used but coding tools under development should more efficiently exploit the spatial and temporal correlations. For Inter picture prediction, a prediction frame is created based on powerful motion compensation techniques and only the residual between the original frame and the prediction is coded along with some auxiliary coded data, e.g. motion vectors and mode information, which have to be transmitted to the decoder.

A relevant H.264/AVC technique for improved performance exploiting the spatiotemporal correlation existing between adjacent macroblocks is the so-called SKIP and DIRECT modes for P and B slices. When used, these modes represent effectively the motion of a macroblock (MB) or $8 \times 8$ block as no motion data has to be transmitted at all, as required by the other inter-MB types\cite{1}. Specifically, for H.264/AVC B-slice coding, the DIRECT mode plays a major role due to its very small overhead (no motion vector data and reference picture index are transmitted) and also its high selection probability\cite{2}. For B-slices, there are two DIRECT mode coding types: the temporal DIRECT mode where the motion vector of the co-located position in a reference frame is used to create the prediction and the spatial DIRECT mode where the spatial correlation is exploited to predict the current MB\cite{3}. While the temporal DIRECT mode is efficient for sequences with high temporal correlation, the spatial DIRECT mode is typically chosen for sequences with scene changes and object occlusions. In the past, the DIRECT mode has been improved mainly by proposing new methods for the determination of the motion vector predictor from already coded data. In\cite{4}, an adaptive selection between temporal and spatial DIRECT modes is proposed along with new rules to select the reference picture in the image boundaries. In\cite{5}, a new direct mode is proposed with candidate motion vectors obtained from a spatiotemporal neighborhood and a cost function selecting the best motion vector predictor based on a boundary matching criterion.

Since late 2002, the so-called distributed video coding (DVC) paradigm has gained relevance, since it targets the exploitation at the decoder with motion estimation tools of most of the temporal correlation, a typical task of predictive video encoders\cite{6}. In DVC, the estimated side information is exploited to decode the frame under consideration to a certain quality level, by sending channel coding data which is able to correct the SI ‘estimation errors’. Naturally, there are many solutions already available in the literature to create the SI frame, following different approaches or fulfilling different requirements\cite{7, 8}.

In this context, the main paper objective is to improve the compression performance of H.264/AVC B-slice coding by proposing a novel B-slice DIRECT mode, inspired by the side information creation process typical of distributed video coding. In this case, a novel motion field is created at both the encoder and decoder for the frame being coded, which can be used with no rate penalty. This follows the same idea of the standard DIRECT mode where the motion vector is derived from spatially or temporally adjacent blocks. In this case, the motion field is directly estimated with a motion compensated frame interpolation framework (often used for SI creation in DVC codecs) which only makes use of temporally adjacent reference frames (no original frames are needed).

The rest of this paper is organized as follows. Section 2 makes a brief review of the state-of-the-art H.264/AVC standard, notably the standard B-slice DIRECT coding mode. After, Section 3 proposes the novel, motion SI inspired, B-slice DIRECT coding mode, after describing the adopted motion SI creation method and its integration in the H.264/AVC coding architecture. Next, Section 4 presents the test conditions and discusses the RD performance results, while Section 5, finally, summarizes the main conclusions and ideas for future work.

2. H.264/AVC STANDARD DIRECT MODE

As usual, H.264/AVC still specifies three main coding types for each slice, notably Intra (I), Predicted (P) and Bi-directionally predicted (B) slices\cite{1}. In addition, several partitions, reference frames and motion vectors may be selected for each macroblock,
Fig. 1: Proposed encoder architecture, highlighting the novel/changed modules.

3. IMPROVED B-SLICES DIRECT MODE CODING WITH MOTION SIDE INFORMATION

This section proposes an improvement to the H.264/AVC coding framework, notably regarding the DIRECT coding mode procedure for B-slices by exploiting a motion field available at both encoder and decoder that is derived by frame interpolation techniques, quite often used to create side information in DVC codecs. By improving the quality of the predicted motion vectors, the probabilities of using the very inexpensive DIRECT mode increase.

3.1. Architecture

The proposed encoder architecture is drawn in Fig. 1, highlighting the novel and improved coding modules. The main changes to the standard H.264/AVC coding architecture are: i) insertion in the encoder and decoder of a novel motion side information (SI) creation module with the purpose to create a new motion field for the frame being coded; ii) definition of a new B-slice DIRECT mode using the motion field provided by the motion SI creation module. If the motion SI creation process generates a more accurate motion field when compared to the H.264/AVC standard motion vector (MV) derivation process for the DIRECT mode, better RD performance can be achieved. The novel part in the encoding walkthrough for the proposed encoder architecture proceeds as follows:

1. **Motion SI field creation** – Two decoded (Intra or Inter) frames, one temporally in the past and another in the future, are given to the SI creation module to generate the motion field to be used in the next step.

2. **B-slice coding with new SI based DIRECT mode** – For each block, the proposed SI DIRECT mode uses the motion vector computed by the motion SI creation process (described in Section 3.2) as the MV predictor; the other coding modes are processed as usual [1].

3. **Coder control** – The typical RD optimization criterion selects the best coding mode from the set of possible modes; naturally, this set includes the MV prediction of the new SI DIRECT mode. In this case, the H.264/AVC spatial and temporal DIRECT coding modes cannot be selected since it is expected that the new SI DIRECT mode can provide better MV predictors.

The new SI DIRECT coding mode pays a rather small rate overhead (only the mode) since there is no bitrate cost for reference picture indexes or motion vectors as they are derived exactly by the same process at both the encoder and decoder.

3.2. Motion side information creation

This section describes the creation of the motion SI field which plays a major role in the proposed solution. The SI motion field accuracy directly affects the compression efficiency gains that may be expected with the proposed method.

The adopted side information creation technique works at the block level and extracts a piecewise smooth motion field capturing the linear motion between the backward (past) and the forward (future) reference frames. The SI creation solution consists in the following main steps inspired from [7]:

- **Forward motion estimation** - First, both reference frames are low pass filtered and used as references in a full search motion estimation algorithm using a modified matching criterion. This matching criterion favors motion vectors closer to the origin and regularizes the motion field.

- **Bi-directional motion estimation** - After, motion vectors for each SI block are selected and refined. A hierarchical approach is used with block sizes of 16x16 and 8x8. An adaptive search range is also dynamically computed for each block, taking into account the motion vectors of the neighboring blocks, notably the top, bottom, right and left blocks.

- **Weighted vector median filtering** - The weighted median vector filter improves the motion field spatial coherence by looking, for each SI block, for the candidate motion vectors at neighboring blocks which can better represent the motion trajectory. This filter is also adjustable by a set of weights, controlling the filter strength (i.e. how smooth becomes the motion field) and depending on the block distortion for each candidate motion vector.

The motion SI creation process makes available a motion field for the frame being coded and is performed in the same way at both the encoder and decoder since the SI creation techniques do not need the original frame as the traditional motion estimation process used in predictive video codecs.
3.3. Proposed Motion SI based DIRECT mode

This section proposes a new B-slices DIRECT mode based on the exploitation of the motion vectors obtained at the end of the motion SI creation process described in Section 3.2. This new mode will be referred as SI DIRECT mode and tries to overcome the main weakness of the temporal and spatial DIRECT modes. In these modes, motion is assumed to be spatially or temporal stationary, assumption that fails quite often in practice. For example, in the temporal DIRECT mode, the motion vector derived from the co-located position in the first frame of List-1 does not always express the motion of the current B-slice block, especially when motion is not continuous (or is irregular) and when the reference pictures involved are temporally far from each other. Moreover, when the co-located block is Intra, the poor zero motion vector prediction will reduce the percentage of blocks coded with the DIRECT mode which has a negative impact on the overall compression efficiency. Similar considerations can be made for the spatial DIRECT mode.

Basically, for each B-slice macroblock, two motion vectors $MV_{10}$ and $MV_{11}$ are computed (for the bi-predictive mode) depending on the temporal distance between each reference frame and the corresponding B-slice and the co-located motion vector $MV_{co}$ [2, 3]. As shown in Fig.2, the motion vector $MV_{10}$ points to a reference frame in List-0, the motion vector $MV_{11}$ points to a reference frame in List-1 and $MV_{co}$ corresponds to the motion vector of the co-located block in the first frame of the List-1 reference. When the co-located position is Intra coded, $MV_{co}$ is not available and thus a zero motion vector is used, i.e. the prediction for the current block corresponds to the average of co-located blocks in the first forward and backward pictures in the reference lists.

To overcome the weaknesses of the temporal and spatial DIRECT modes, a new SI DIRECT mode exploiting the motion vectors obtained from a typical DVC SI creation process, the frame interpolation framework described in Section 3.2, is proposed. The motion SI field corresponds to an accurate estimation of each B-slice block motion vector and does not require the availability of motion vectors in any neighboring block (both spatially or temporarily). Thus, if the SI motion vectors are accurate enough, the compression performance may be significantly increased. The novel SI DIRECT mode proceeds as follows for each block in the current B-slice:

- **Case 0**: If the co-located block in the first picture of List-1 refers to a picture in List-0 that is a long-term reference picture, the following applies:

  \[
  MV_{10} = MV_{co}, \quad MV_{11} = 0
  \]  

  This case is defined in the same way as the temporal DIRECT mode since long-term reference pictures typically have low temporal correlation with the B-slice being coded.

- **Case 1**: Otherwise, the motion vectors are obtained based on the motion SI data since this motion field corresponds to the best estimation for the block being coded:

  \[
  MV_{10} = MV_{co}, \quad MV_{11} = MV_{si}
  \]  

  where $MV_{co}$, $MV_{si}$ are the two motion vectors for the block under coding calculated for the two reference frames used in the motion SI creation proposed in Section 3.2. These motion vectors are symmetric, i.e. if $MV_{co} = (r, c)$, then $MV_{si} = (-r, -c)$ as shown in Fig.2.

4. PERFORMANCE EVALUATION

To evaluate the proposed video codec RD performance, the novel SI DIRECT mode has been implemented in the H.264/AVC reference software (JM 18.0).

4.1. Test conditions

The detailed test conditions are summarized in Table 1 and correspond to representative conditions, largely used in the literature. The compression efficiency gains are measured with Bjontegaard deltas for the bitrate (BD-Rate) and PSNR (BD-PSNR) [9]. The H.264/AVC Main profile is taken as the benchmark/anchor to compare with the proposed solutions.

4.2. Results and analysis

Tables 2 to 4 show the coding gains with the proposed SIDIRECT mode with respect to the temporal and spatial DIRECT coding modes. The B-slice rate savings ($\Delta BR$) and the increment on the number of DIRECT mode MBs in percentage ($\Delta NMB$) are computed as follows:

\[
\Delta BR = \frac{BR_A - BR_S}{BR_S}
\]

\[
\Delta NMB = \frac{NMB_S - NMB_A}{N} \times 100\%\]

where $BR_A$ is the B-slice rate when the spatial or temporal DIRECT coding modes are used (anchor), $BR_S$ is the B-slice rate for the new SI DIRECT mode. The number of macroblocks for which the RDO process chooses the DIRECT mode corresponds to $NMB_S$ for the new SI DIRECT mode and to $NMB_A$ for the anchor mode.

<table>
<thead>
<tr>
<th>Table 1: Test conditions details</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sequence name</strong></td>
</tr>
<tr>
<td><strong>Resolutions</strong></td>
</tr>
<tr>
<td><strong>Number of frames</strong></td>
</tr>
<tr>
<td><strong>GOP structure</strong></td>
</tr>
<tr>
<td><strong>QP</strong></td>
</tr>
<tr>
<td><strong>Encoder tools</strong></td>
</tr>
</tbody>
</table>
Table 2: Gains with the proposed SIDIRECT mode compared to the temporal DIRECT mode (IBI GOP structure).

<table>
<thead>
<tr>
<th>Sequences</th>
<th>BD Rate</th>
<th>BD PSNR</th>
<th>ΔBR</th>
<th>ΔNMB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Foreman</td>
<td>-4.38</td>
<td>0.27</td>
<td>37%</td>
<td>10%</td>
</tr>
<tr>
<td>Coastguard</td>
<td>-8.04</td>
<td>0.46</td>
<td>56%</td>
<td>15%</td>
</tr>
<tr>
<td>Soccer</td>
<td>-5.30</td>
<td>0.26</td>
<td>24%</td>
<td>17%</td>
</tr>
<tr>
<td>Hall monitor</td>
<td>-0.24</td>
<td>0.02</td>
<td>13%</td>
<td>1%</td>
</tr>
<tr>
<td>Average</td>
<td>-4.49</td>
<td>0.25</td>
<td>32.41%</td>
<td>10.75%</td>
</tr>
</tbody>
</table>

Table 3: Gains with the proposed SIDIRECT mode compared to the spatial DIRECT mode (IBP GOP structure).

<table>
<thead>
<tr>
<th>Sequences</th>
<th>BD Rate</th>
<th>BD PSNR</th>
<th>ΔBR</th>
<th>ΔNMB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Foreman</td>
<td>-1.10</td>
<td>0.05</td>
<td>3%</td>
<td>3%</td>
</tr>
<tr>
<td>Coastguard</td>
<td>-2.44</td>
<td>0.10</td>
<td>24%</td>
<td>5%</td>
</tr>
<tr>
<td>Soccer</td>
<td>-1.68</td>
<td>0.07</td>
<td>4%</td>
<td>4%</td>
</tr>
<tr>
<td>Hall monitor</td>
<td>-0.96</td>
<td>0.04</td>
<td>6%</td>
<td>1%</td>
</tr>
<tr>
<td>Average</td>
<td>-1.55</td>
<td>0.06</td>
<td>9.47%</td>
<td>3.36%</td>
</tr>
</tbody>
</table>

Table 4: Gains with the proposed SIDIRECT mode compared to the spatial DIRECT mode (IBP GOP structure).

<table>
<thead>
<tr>
<th>Sequences</th>
<th>BD Rate</th>
<th>BD PSNR</th>
<th>ΔBR</th>
<th>ΔNMB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Foreman</td>
<td>-7.42</td>
<td>0.32</td>
<td>27%</td>
<td>6%</td>
</tr>
<tr>
<td>Coastguard</td>
<td>-5.53</td>
<td>0.22</td>
<td>39%</td>
<td>4%</td>
</tr>
<tr>
<td>Soccer</td>
<td>-7.94</td>
<td>0.35</td>
<td>16%</td>
<td>7%</td>
</tr>
<tr>
<td>Hall monitor</td>
<td>-2.02</td>
<td>0.10</td>
<td>12%</td>
<td>1%</td>
</tr>
<tr>
<td>Average</td>
<td>-5.73</td>
<td>0.25</td>
<td>23.45%</td>
<td>4.67%</td>
</tr>
</tbody>
</table>

In summary, the proposed SI DIRECT coding mode efficiently exploits the motion vector field obtained by the side information creation process since it achieves significantly better RD performance, naturally more or less depending on the type of visual content.

5. FINAL REMARKS

This paper proposes to improve the performance of the H.264/AVC predictive video coding standard, notably the B-slice DIRECT mode coding by efficiently exploiting a motion field at both encoder and decoder, which is computed in a similar way as done at some distributed video decoders. The RD performance evaluation shows the superiority of the proposed coding solution with bitrate savings up to about 8% for both the IBI and IBP GOP structures, respectively. A topic for future work may be the simultaneous use of both the H.264/AVC standard and the proposed DIRECT modes for each B-slice, letting the RDO process find the best mode that maximizes H.264/AVC compression efficiency.

6. REFERENCES


