RATE DISTORTION OPTIMIZATION FOR BIDIRECTIONAL SCALABLE MOTION MODEL

Hu Chen1, Meng-Ping Kao2, Zhao Liu3 and Truong Nguyen4

University of Electronic Science and Technology of China1,3, Qualcomm2, University of California, San Diego4

ABSTRACT

The fully scalable motion model (SMM) is proposed for scalable video codec by taking advantage of motion information scalability. In previous work, SMM has been improved to support hierarchical B frame and bidirectional or multidirectional motion estimation. It yields better results comparing to method using unidirectional motion estimation. However, the process of bidirectional or multidirectional rate distortion optimization motion estimation is very complex and is the most time-consuming task in the encoding process. We present several algorithms to reduce its complexity and show corresponding simulation to compare their performance.

Index Terms— Rate distortion optimization (RDO), motion estimation (ME), scalable video coding (SVC), scalable motion model

1. INTRODUCTION

Motion information is an important requirement for a fully scalable video codec, especially in low bit rate or small resolution decoding scenarios, for which the SMM [1] has been proposed. SMM embeds two dimensions, Variable Block Size (VBS) dimension and Accuracy dimension, for scalability and can integrate naturally with other scalabilities, such as spatial, temporal and quality, in a scalable video codec. Fig.1. summarizes this model. Reference [1] and [2] which implemented unidirectional and bidirectional ME, respectively, show its high coding efficiency property.

RDO ME plays a key role in a motion-compensation based video codec, such as H.264/AVC [3], H.264/SVC [4]. SMM used in a motion-compensation based video codec could significantly achieve higher coding efficiency by embedding VBS dimension and accuracy dimension in the RDO ME process. Though the two dimensions can improve the video quality remarkably, they also require substantial computational complexity to video encoder especially in bidirectional or multidirectional ME process. Therefore, the simplification of RDO ME in SMM becomes a very significant and challenging work. In [2], an assumption is used to simplify the RDO ME, and obtain good performance. However, this simplified cost function ignores the correlation between forward and backward RDO ME process.

This paper presents several algorithms to reduce the complexity in the RDO ME process. Section 2 introduces the proposed scalable motion model and its RDO estimation process. Section 3 discusses the proposed rate distortion optimized ME for bidirectional or multidirectional motion case and the corresponding algorithms are presented. Section 4 shows simulation results and comparison for the above proposed methods. Conclusions are given in Section 5.

2. BACKGROUND

Figure 1 summarizes the proposed SMM which has two dimensions (Accuracy Scalability Dimension and Variable Block Size (VBS) Scalability Dimension) that fine tune the motion quality [1]. Using this structure, quality scalability can be easily realized by choosing the decoding levels on each dimension. Lower decoding levels correspond to lower quality MVFs (Motion Vector Field), and vice versa.

Some notations have to be clarified before further description of our model. First, we assume that we have A accuracy layers, which are indexed by \( a, 0 \leq a < A \) with accuracy base layer \( a = 0, V \) VBS levels, which are indexed by \( v, 0 \leq v < V \) with the largest block size layer \( v=v_0 \). We also assume there are \( R \) resolution layers, which are indexed by \( r, 0 \leq r < R \) with the highest resolution image \( r=r_0 \) and \( a_r \) which are distortion multipliers for resolution \( r \). At last, we assume there are \( nref \) references and assign a multipliers \( \beta_i \) for reference \( i \).

In SMM, the RDO ME process is performed in the basis of sub blocks and the scanning order is shown in Fig.2 for an example of three VBS levels structure. As observed from Fig.2, the scan order is from top layer, \( v=0 \), to bottom layer \( v=V-1 \), with a raster scan in group of four sub blocks within the same VBS layer [1].

Thus for each sub block, our goal is to determine the best scalable motion vector difference with the following structure in the rate distortion sense:

\[
SMVD = (s, \text{ref} (0), \text{ref} (1), \ldots, \text{ref} (A-1))
\]

where \( -1 < s < A-1 \) denotes the starting accuracy level for current SMVD and \( \text{ref} (a), a=0, \ldots, A-1 \) are the refinement vectors for level \( a \). The cost function can be defined as follows:

\[
CF (SMVD) = RF (SMVD) + DF (SMVD)
\]

where \( SMVD \) is the scalable motion vector difference. \( RF (\cdot) \) is the rate cost that returns the actual coding bits of each component in \( SMVD \) and \( DF (\cdot) \) is the distortion function. For unidirectional motion case, the function becomes [1]

\[
CF (SMVD) = RF (SMVD) + DF (SMVD) = \lambda_s R (S) + \lambda_r R (\text{ref} (0)) + \sum_{a=0}^{A-1} \lambda_{SMVD(a)} R (\text{ref} (a)) + \sum_{r=1}^{R-1} \omega_r D_r (SMV (r, A-1))
\]
The RDO process is to find the best SMVD which minimizes the function in (3). It begins with determining the optimal pair \((\mathbf{s}, \text{ref}(0))\). Here \(\mathbf{s}\) and \(\text{ref}(0)\) should be jointly optimized to minimize (3). If the best \(\text{ref}(0)\) turns out to be 0, all refinements \{\text{ref}(a)\} are automatically set to 0 and that completes the RDO ME process. If not, since the optimal pair \((\mathbf{s}, \text{ref}(0))\) is already determined, the RDO ME process proceeds by tracing through \(a=1, \ldots, A-1\) to find out the rest of \(\text{ref}(a)\). \(\mathbf{s}, a=1, \ldots, A-1\) are sequentially determined according to the following rule: [1]

\[
\text{ref}(a) = \arg\min_{\mathbf{x}} \{CF(\mathbf{x}, \text{ref}(0), \ldots, \text{ref}(a-1), \mathbf{x}, 0, \ldots, 0)\}.
\]  

(4)

For multidirectional RDO ME, the cost function becomes

\[
CF_{\text{ref}}(\{\text{SMVD}\} | i = 0, \ldots, n_{\text{ref}} - 1) = R_F(\{\text{SMVD}\} | i = 0, \ldots, n_{\text{ref}} - 1) + DF(\{\text{SMVD}\} | i = 0, \ldots, n_{\text{ref}} - 1)
\]

\[
= \sum_{i=0}^{n_{\text{ref}}-1} R_F(\mathbf{s}, \text{ref}(0), \text{ref}(1), \ldots, \text{ref}(A-1))
\]

\[
+ \sum_{r=0}^{n_{\text{ref}}-1} \{\text{Orig}(r) - \sum_{j=r}^{A-1} \beta_j \text{MC}(\text{SMVD}(r, A-1))\}
\]

(5)

The RDO process is to find the best \{\text{SMVD}\} \(| i = 0, \ldots, n_{\text{ref}} - 1\) that minimizes the function in (5). This function is non-linear and we must find the best combination of \{\text{SMVD}\} \(| i = 0, \ldots, n_{\text{ref}} - 1\) from different references. Consequently, the complexity is huge.

In [2], we assume that if each \text{SMVD} for reference picture \(i\) is optimal, i.e. \(CF(\text{SMVD})\) is minimized among all possible \text{SMVD}, the combination of these \{\text{SMVD}\} \(| i = 0, \ldots, n_{\text{ref}} - 1\) is also optimal and therefore the cost function in (5) is simplified as,

\[
CF_{\text{ref}}(\{\text{SMVD}\} | i = 0, \ldots, n_{\text{ref}} - 1) = R_F(\{\text{SMVD}\} | i = 0, \ldots, n_{\text{ref}} - 1) + DF(\{\text{SMVD}\} | i = 0, \ldots, n_{\text{ref}} - 1)
\]

\[
= \sum_{i=0}^{n_{\text{ref}}-1} R_F(\mathbf{s}, \text{ref}(0), \text{ref}(1), \ldots, \text{ref}(A-1))
\]

\[
+ \sum_{r=0}^{n_{\text{ref}}-1} \{\text{Orig}(r) - \sum_{j=r}^{A-1} \beta_j \text{MC}(\text{SMVD}(r, A-1))\}
\]

\[
\approx \sum_{r=0}^{n_{\text{ref}}-1} \{R_F(\mathbf{s}, \text{ref}(0), \text{ref}(1), \ldots, \text{ref}(A-1)) - \beta_s \text{Orig}(r) + \sum_{j=r}^{A-1} \beta_j \text{MC}(\text{SMVD}(r, A-1))\}
\]

\[
= \sum_{i=0}^{n_{\text{ref}}-1} \{R_F(\{\text{SMVD}\}) - \beta_s \text{Orig}(r) + \sum_{j=r}^{A-1} \beta_j \text{MC}(\text{SMVD}(r, A-1))\}
\]

(6)

In summary, the RDO process for finding the best SMVDs includes \(n_{\text{ref}}\) independent sub processes and each sub process has the same search algorithm as in unidirectional ME process for each reference.

### 3. PROPOSED ESTIMATION RATE DISTORTION OPTIMIZATION

So far, the cost function and the RDO process to find the best SMVD have been illustrated and easily implemented in unidirectional and bidirectional ME. However, the simple assumption algorithm (SAA) for bidirectional motion estimation in (6) is a very coarse approximation for the original function in (5).

In this section, we investigate several ways to simplify the RDO process. To further illustrate the valid properties for improving coding efficiency, two subtopics will be discussed, covering the calculation and motion bit rate saved.

#### 3.1 Step by step search algorithm (SSA).

As mentioned earlier, in (6), the RDO process is partitioned into \(n_{\text{ref}}\) independent sub processes coarsely ignoring the correlations among these \(n_{\text{ref}}\) SMVDs. Therefore we improve the sub processes by adding correlation information into cost function. In other words, each sub search process includes the information from other SMVDs.

Some definitions and assumptions have to be clarified before further description of improved algorithm. First, we define two vectors \(x = (\mathbf{s}, \text{ref}(0), \ldots, \text{ref}(A-1))\) and \(\mathbf{0} = (-1, 0, \ldots, 0)\). We also define the motion compensation as \(MC(\text{SMVD}) = MC(\text{SMVD}(r, A-1) | \text{SMVD})\).

At last we give two assumptions \(R_F(\mathbf{0}) \equiv 0\) and \(MC(\mathbf{0}) \equiv \text{Orig}(r)\).

The improved algorithm is implemented step by step and begins with determining the first SMVD for the reference with index 0. The process is the similar to unidirectional estimation process. The cost function is as follows.

\[
\text{SMVD}_0 = \arg\min \{CF_{\text{ref}}(\mathbf{x}, \mathbf{0}, \ldots, \mathbf{0})\}
\]

\[
= \arg\min \left\{R_F(\mathbf{x}, \mathbf{0}, \ldots, \mathbf{0}) + DF(\mathbf{x}, \mathbf{0}, \ldots, \mathbf{0})\right\}
\]

\[
= \arg\min \left\{R_F(\mathbf{x}) + \sum_{r=0}^{n_{\text{ref}}-1} \beta_s \text{Orig}(r) + \sum_{j=r}^{A-1} \beta_j \text{MC}(\text{SMVD}(r, A-1))\right\}
\]

\[
\approx \arg\min \left\{R_F(\mathbf{x}) + \sum_{r=0}^{n_{\text{ref}}-1} \beta_s \text{Orig}(r) - \sum_{j=r}^{A-1} \beta_j \text{MC}(\text{SMVD}(r, A-1))\right\}
\]  

(7)
For SMVD, considering that the best SMVD and its corresponding motion compensated picture are available, the following process for SMVD should include the information \(RF_{ref}(SMVD_0)\) and \(MC(SMV_0(r,A-1))\) from SMVD_0. Consequently, the cost function is

\[
SMVD_0 = \arg\min_{x} \left( CF_{ref}(SMVD_0, x, 0, \cdots, 0) + DF_{ref}(SMVD_0, x, 0, \cdots, 0) \right)
\]

\[
= \arg\min_{x} \left( RF(SMV_0) + RF(x) + \sum_{i=0}^{n-1} \alpha_i (Orig(r) - \beta_i MC(SMV_0(r,A-1))) \right)
\]

\[
= \arg\min_{x} \left( RF(x) + \sum_{i=0}^{n-1} \alpha_i (Orig(r) - \beta_i MC(SMV_0(r,A-1))) \right)
\]

\[
= \arg\min_{x} \left( RF(x) + \sum_{i=0}^{n-1} \alpha_i (Orig(r) - \beta_i MC(SMV_0(r,A-1))) \right)
\]

\[
= \arg\min_{x} \left( RF(x) + \sum_{i=0}^{n-1} \alpha_i (Orig(r) - \beta_i MC(SMV_0(r,A-1))) \right)
\]

\[
= \arg\min_{x} \left( RF(x) + \sum_{i=0}^{n-1} \alpha_i (Orig(r) - \beta_i MC(SMV_0(r,A-1))) \right)
\]

(8)

When the search process for SMVD is performed, the previous n SMVDs are available and the cost function should include the motion information from these SMVDs.

\[
SMVD_n = \arg\min_{x} \left( CF_{ref}(SMVD_n, \cdots, SMVD_0, x, 0, \cdots, 0) \right)
\]

\[
= \arg\min_{x} \left( RF_{ref}(SMVD_n, \cdots, SMVD_0, x, 0, \cdots, 0) \right)
\]

\[
= \arg\min_{x} \left( RF(SMV_n) + RF(x) + \sum_{i=0}^{n-1} \alpha_i (Orig(r) - \beta_i MC(SMV_0(r,A-1))) \right)
\]

\[
= \arg\min_{x} \left( RF(x) + \sum_{i=0}^{n-1} \alpha_i (Orig(r) - \beta_i MC(SMV_0(r,A-1))) \right)
\]

\[
= \arg\min_{x} \left( RF(x) + \sum_{i=0}^{n-1} \alpha_i (Orig(r) - \beta_i MC(SMV_0(r,A-1))) \right)
\]

\[
= \arg\min_{x} \left( RF(x) + \sum_{i=0}^{n-1} \alpha_i (Orig(r) - \beta_i MC(SMV_0(r,A-1))) \right)
\]

\[
= \arg\min_{x} \left( RF(x) + \sum_{i=0}^{n-1} \alpha_i (Orig(r) - \beta_i MC(SMV_0(r,A-1))) \right)
\]

Consequently, the SMVD for the reference with index \(n\) \((n > 0)\) is based on the n SMVDs whose indices are less than \(n\). Compared to the previous method, this step by step algorithm (SSA) improves the accuracy of ME without computation increase.

### 3.2. The improvement of RDO process.

As shown in Fig.2, for a macro block (MB), there are 21 RDO ME processes (21 sub blocks) in unidirectional ME. For bidirectional or multidirectional ME, the simplified algorithm partitions the whole process into \(n_{ref}\) sub processes and there are total \(n_{ref} \times 21\) RDO ME processes. Therefore the calculation is very large for multidirectional ME.

Because the total number of RDO ME processes is constant, we can only speed up the ME process by decreasing the complexity of the calculation for a macro block. A simple way to achieve this goal is to reduce the search range. For example, if the original range is 16, we can reduce the range to half to reduce ME complexity. However, reducing the search range reduces motion vector accuracy and impacts the video quality. To resolve this problem, a possible way considering the motion trend is proposed and a corresponding example for bidirectional motion ME is shown in Fig.4.

As known from the structure of dyadic hierarchical coding structure (Fig.5), the forward reference and backward reference are symmetry temporally. For the block with biggest block size, it’s reasonable to assume that the magnitude and direction of motion are changed slightly. Therefore, if one directional motion has been obtained, the prediction of another direction can be generated by reversing the sign of the existing motion. The accuracy of this prediction is good except in the case of fast motion and therefore, we can reduce the search range of second reference. Consequently, this RDO ME complexity is reduced while preserving the reconstructed frame quality.

### 4. SIMULATION RESULTS

The evaluation of proposed SMM will be performed on the low complexity wavelet-based SVC framework [2], [5].

The main parameters are as follows: \(\alpha = [100 75 25]\), \(\beta = [0.5 0.5]\), three accuracy levels and three VBS levels.

The format of input testing sequences is CIF with 30fps and the SVC generates the scalable bitstream with maximum bit rate for various decoding scenarios. Our simulation will be compared by side with the proposed RDO ME methods in both CIF and QCIF decoding formats with different frame rates.

The comparison among the speeds of algorithms is listed in Table 2. Because the sourcecode of algorithms are not completely
The rate distortion curve for bus sequence in CIF size at 30 frame rate per second is shown in Fig 6(a). It is clear that the performance of SSA is better than SAA from low bit rate to high bit rate. The details of PSNR differences are listed in Table 3, along with the results from the Football and Mobile sequences.

As observed from Table 3, for bus and mobile sequence, although the improved RDO process reduces the searching range, the performance is still close to the original SSA. The SSA with 3/4 range performs similarly with SAA and SSA with half range performs little worse. However, for football sequence, the quality of reconstructed pictures decreases sharply with the range decreasing, especially when only using base motion layer in low bit rate. It’s reasonable because the football sequence has fast and unpredictable motion and the proposed reversed motion vectors are not as accuracy as estimating it directly.

Fig 6(b) and Fig 6(c) demonstrate another two decoding scenarios for (1) Reduced frame rate and (2) reduced spatial resolution. The results are similar to that in Fig 6(a). SSA performs better than SAA and the improved process maintains the performance while reducing the coding complexity in the case of general motion sequence. The details for Football and Mobile sequences can be found on [http://videoprocessing.ucsd.edu/~tiger](http://videoprocessing.ucsd.edu/~tiger).

### Table 2

<table>
<thead>
<tr>
<th>Time</th>
<th>Simplest</th>
<th>Step by step</th>
<th>½ range</th>
<th>¾ range</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>1</td>
<td>-2.3%</td>
<td>-37.5%</td>
</tr>
</tbody>
</table>

### Table 3

RD comparison for CIF at 30FPS (For each video, the first row shows the PSNR(in dB) of reconstructed pictures (RP) and corresponding MC using the original ME method and the subsequent rows show the difference in PSNR)

<table>
<thead>
<tr>
<th>Bit rate (kbps)</th>
<th>128</th>
<th>256</th>
<th>384</th>
<th>512</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bus</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSA</td>
<td>21.44</td>
<td>21.28</td>
<td>21.41</td>
<td>21.75</td>
</tr>
<tr>
<td>SAA</td>
<td>0.13</td>
<td>0.02</td>
<td>0.39</td>
<td>0.41</td>
</tr>
<tr>
<td>SSA ½ range</td>
<td>0.01</td>
<td>-0.01</td>
<td>0.00</td>
<td>-0.01</td>
</tr>
<tr>
<td>SSA ¾ range</td>
<td>-0.01</td>
<td>-0.01</td>
<td>-0.00</td>
<td>-0.00</td>
</tr>
<tr>
<td><strong>Football</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSA</td>
<td>23.62</td>
<td>23.57</td>
<td>23.53</td>
<td>27.29</td>
</tr>
<tr>
<td>SAA</td>
<td>0.13</td>
<td>0.32</td>
<td>0.59</td>
<td>0.34</td>
</tr>
<tr>
<td>SSA ½ range</td>
<td>-0.01</td>
<td>-0.01</td>
<td>-0.00</td>
<td>-0.01</td>
</tr>
<tr>
<td>SSA ¾ range</td>
<td>-0.01</td>
<td>-0.01</td>
<td>-0.01</td>
<td>-0.01</td>
</tr>
<tr>
<td><strong>Mobile</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSA</td>
<td>18.80</td>
<td>18.76</td>
<td>20.40</td>
<td>20.87</td>
</tr>
<tr>
<td>SAA</td>
<td>0.31</td>
<td>0.33</td>
<td>0.34</td>
<td>0.31</td>
</tr>
<tr>
<td>SSA ½ range</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>SSA ¾ range</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
</tbody>
</table>

5. CONCLUSIONS

In this paper, in order to improve the performance of bidirectional or multidirectional ME, the step by step algorithm is proposed. To reduce the cost of bidirectional or multidirectional motion calculation and maintain the performance in RDO process, an improved ME structure is presented. The combination of these techniques is implemented in the SMM and yields good performance. Simulations verify that these proposals can improve the coding efficiency and reduce the complexity of RDO process.

6. REFERENCES