Efficient Residual Prediction with Error Concealment in Extended Spatial Scalability

Muhammad Shoaib, Anni Cai
Multimedia Communication Center
Beijing University of Posts and Telecommunications
Beijing, China
moshoaib@gmail.com, annicai@bupt.edu.cn

Abstract—In this paper, inter-layer residual prediction’s trade-off between concealing lost macroblock (MB) and removing visual artifact in extended spatial scalability (ESS) is investigated. In order to improve visual artifacts, various schemes are proposed in the literature, including higher distortion to the residual prediction used in ESS enhancement layer for non matching residuals. Whereas, residual prediction is also used to conceal lost MBs in inter-layer error concealment methods. We propose an efficient use of residual prediction to prevent those artifacts as well as to conceal lost MBs, exploiting features in homogenous characteristics in video objects for error resilient coding. Simulation results show that the proposed scheme achieves up to 0.29 dB PSNR gain, with overall average of 0.1 dB PSNR gain at the decoder for various tested sequences compared to JVT-W123 under testing conditions specified in JVT-V302.

Keywords- Extended spatial scalability; error resilient coding; scalable video coding; inter layer residual prediction

I. INTRODUCTION

Scalable video coding has been designed to provide a mechanism for reusing an encoded lower resolution version of an image sequence for the coding of a corresponding higher resolution sequence, from which appropriate sub bit-streams can be extracted to cope different preferences of rate and resolution [1]. In its basic version, scalable video coding (SVC) is designed for dyadic spatial scalability. However, in ESS which introduces new cropping and non-dyadic scaling scalability mode, picture width and height of successive spatial layers can have any possible configurations [2].

SVC allows prediction tools inherited from H.264/AVC along with new inter-layer predictions using corresponding base layer. In joint scalable video model (JSVM), it is possible that an enhancement layer MB is coded relative to the base layer, even though several base layer MBs may be needed to form the prediction. Information used from base layer MBs as part of this inter-layer coding process may include motion vectors, residual predictions and, if available, reconstructed pixels. The mechanism of using the base layer residual to predict the enhancement layer residual is called inter-layer residual prediction. Residual prediction is among the new coding tools in Annex G of JSVM [3], and is an important feature that provides efficient inter-layer prediction with reasonably low coding complexity.

Recently, number of efforts has been made to improve the visual quality at ESS enhancement layer with residual up-sampling from base layer. Woo proposed smoothed reference prediction as an extension of intra-base prediction [4]. Ye et al. in [5] proposed a fix in [4] to the residual up-sampling process so that the bilinear interpolation can be performed across a base layer block edge if the edge falls within an enhancement layer transform block. However, disabling smoothing filter introduces some visual artifacts for certain ESS scaling ratios. Wang et al. in [6] proposed a fix in [5] to improve the subjective quality caused by residual prediction with non matching residuals.

When a video stream passes through noise channel, packet loss often happens. To conceal lost MBs, various error concealment techniques have been adopted in JSVM, involving motion compensation in inter-prediction, and residual and motion vector up-sampling in inter-layer prediction. In inter-layer error concealment algorithm, up-sampling residual from the base layer is used to conceal lost MBs in enhancement layer. Hence, disabling up-sampled residuals, which may be adopted in some circumstances for visual artifact reduction, will have a severe impact to error concealment in ESS. To our knowledge, the efficient use of residual prediction to prevent artifacts has never been investigated with its impacts to conceal lost MBs in error concealment.

In this paper, we will extend the previous method, which was originally proposed by our lab to improve the efficiency of inter-layer prediction on MB mode in ESS [7], to error prone transmission conditions, and investigate the efficient use of up-sampling residuals to compromise between visual artifact reduction and inter-layer error concealment in ESS.

The rest of this paper is organized as follows. First, overview of the residual prediction in ESS is presented in Section II. The proposed implementation is presented in detail in Section III. Simulation results and related discussions are presented in Section IV. Finally, Section V concludes the paper.

II. OVERVIEW OF RESIDUAL PREDICTION IN ESS

ESS in JSVM is used to handle all the cases where the edge alignments of base layer MBs and enhancement layer MBs are not maintained. This can happen when scaling ratio is not 1 or...
2, or the picture of lower spatial layer may represent a cropped area of its enhancement layer pictures with higher resolution.

In JSVM, residual prediction is a mechanism that involves using base layer prediction residual to predict enhancement layer prediction residual. This mechanism is mainly designed for inter-coded MB so that a feature called single loop decoding can be enabled. The attraction of single loop decoding is that it permits an enhancement layer video stream to be decoded with only one motion compensation loop at the enhancement layer. No motion compensation needs to be performed at the base layer, since the base layer does not need to be fully decoded and reconstructed. This assists in reducing decoder complexity.

According to the specification in JSVM, intra-coded MBs from the base layer are fully decoded and reconstructed so that they may be up-sampled and directly used to predict the current enhancement layer being decoded. However, inter-coded MBs from the base layer are not fully decoded and reconstructed. Instead, only the prediction residual of each base layer inter-coded MB is decoded and may be used to predict enhancement layer residual, but motion compensation is not performed on the base layer inter-coded MB itself. With this limitation, JVT-W123 adopted in JSVM proposed a fix to remove the artifacts caused by residual prediction with non matching residuals [8]. The MBs having non matching residuals are identified so that a different rate-distortion (RD) measure may be applied to prevent those artifacts.

In ESS, the number of up-sampled base layer 4x4 blocks that may cover a single enhancement layer 4x4 block varies from 1, 2 and 4 according to both block location and the spatial up-sampling ratio. Fig. 1 depicts two consecutive spatial layers where BL1, BL2, BL3 and BL4 represent four base layer blocks having corresponding motion vectors MV1, MV2, MV3 and MV4 respectively, whereas EL0 with blue lines represents sub-sampled enhancement layer block having motion vector MV0. In JSVM, a motion vector applies to 4x4 block size at minimum, so that every pixel in enhancement layer EL0 must share the same motion vector. The applicable motion vector of EL0 is derived from just one up-sampled base layer 4x4 block, in general. However, in ESS when multiple base layer blocks cover a given enhancement layer block, different approach is used as defined in Annex G of JSVM to determine which one of the base layer blocks is used for motion vector derivation. In this example, it is likely that EL0 derives its motion vector from BL1, as the projection of the central point of EL0 locates in BL1 which also covers the largest area among up-sampled base layer blocks.

In ESS, it is possible that the candidate base layer blocks may have quite different motion vectors. Assume that the motion vectors MV1, MV3 and MV4 are equal, while MV2 is a quite different motion vector as shown in Fig. 1. If the motion vector of EL0 is derived from BL1, it follows that the motion vector of EL0 would also be quite different to that of BL2. As shown in Fig. 1, the base layer residuals used to form the residual prediction for EL0 do not come from just one of the four covering base layer blocks, but rather from all four of them. Thus for the enhancement layer pixels covered by base layer BL2 (the gray area shown in the Fig. 1), the base layer residual is formed using one motion vector, but applies in the enhancement layer using a very different motion vector.

Instead, only the prediction residual of each base layer inter-coded MB is decoded and may be used to predict enhancement layer residual, but motion compensation is not performed on the base layer inter-coded MB itself. With this limitation, JVT-W123 adopted in JSVM proposed a fix to remove the artifacts caused by residual prediction with non matching residuals [8]. The MBs having non matching residuals are identified so that a different rate-distortion (RD) measure may be applied to prevent those artifacts.

In general, residual prediction works when an enhancement layer block and its base layer share the same or similar motion vectors. As a result, in Fig. 1 relatively larger prediction error can be expected for the pixels in the grey area when residual prediction is applied to EL0. In addition, because the grey area is relatively a small fraction of EL0, the prediction error is often poorly captured after transform coding. The consequence is that the reconstructed picture has a large and noticeable distortion in the grey area. Such an unmitigated distortion within a coding block causes the visual artifact.

For error concealment in ESS based resilient coding, up-sampling residuals is used to conceal loss MBs. Exploiting the spatial homogeneity and the temporal stationary characteristics in video objects with different resolutions but the same contents [9]-[10], efficient use of residual prediction can be achieved without compromising on coding efficiency.

III. THE PROPOSED IMPLEMENTATION

The approximate motion/texture characteristics of H.264/AVC video frames can generally be analyzed via mode-motion map [11]. In this paper, we fellow the technique in [7], using the homogenous and temporal stationary characteristics in video objects, to differentiate between homogenous region and non homogenous region.

In [9], it is observed that mode distribution between the base layer and its enhancement layers have certain correlations. In spatial scalability, for each MB at the base layer, the corresponding up-sampled MBs at enhancement layers tend to have the similar mode partition. In case of temporal scalability, the mode partition of MBs in the current frame is most similar to the mode partition of the MBs in its reference frames. Motivated by these observations, we use the approach in [7] to find MB-Mode-Complexity (MMC) parameter derived from the mode context of local MBs in the base layer to estimate the motion/texture characteristics of the collocated MBs in the enhancement layers.
Fig. 2 depicts two consecutive spatial layers: base layer and enhancement layer (also showing scaled up-sampled base layer MB “C”), where $W_{new}, H_{new}$ and $W_{enh}, H_{enh}$ represent the widths and heights of the base layer picture and the enhancement layer picture respectively. The base layer picture is a sub-sampled version of the region with dimensions $W_{new}$ and $H_{new}$, partially or totally inside the enhancement layer picture, positioned at coordinates $(x_{up}, y_{up})$. $W_{new}/W_{new}$ and $H_{new}/H_{new}$ are actually up-sampling factors between the base layer picture and the extracted region in the enhancement layer picture [12].

Let $MB^e$ be the current enhancement layer MB and $uMB^b$ (block C) be the up-sampled base layer MB in which the projection of the center point of $MB^e$ locates as shown in Fig. 2. In ESS, an enhancement layer MB may be covered by 1, 2, or 4 up-sampled base layer MBs [6]. Let A, B, C and D are the up-sampled blocks of base layer blocks a, b, c and d respectively, whereas up-sampled base layer blocks A, B and D are the neighboring blocks of current up-sampled base layer block $uMB^b$ “C”. Let $uMB^b_i$ (for example A in Fig. 2) be the neighboring up-sampled MBs of $uMB^b$ which overlap with current $MB^e$. In [12], high layer MBs are classified into four classes: Corner, Hori, Vert and Center. When $MB^e$ belongs to Corner, $uMB^b$ has no neighboring MBs and the sub-partition $n$ of $uMB^b$ is assigned to 0. When $MB^e$ belongs to Hori/Vert, $uMB^b$ has one neighbor $uMB^b_{e}$ ($n=1$). When $MB^e$ belongs to Center, $uMB^b$ has three neighbors $uMB^b_{e}$ ($n=3$). The MB-Mode-Complexity (MMC) of $MB^e$ derived from up-sampled base layer $uMB^b$ is defined as [7]:

$$MMC = \begin{cases} A_e W_0 & \text{for } MB^e \in \text{Corner} \\ A_e W_0 + A_n W_n \quad (1+n) & \text{for } MB^e \in \text{Hori/Vert} \\ A_e W_0 + \sum A_n W_n \quad (1+n) & \text{for } MB^e \in \text{Center} \end{cases}$$

where $W$ is MB mode factor of up-sampled base layer macroblock $uMB^b$ defined in Table I. $A$ is the area factor which is equal to the ratio of the overlapped area of the up-sampled base layer $uMB^b$ of $MB^e$ to the area of $MB^e$ itself.

### A. Derivation of MB Mode in Non Homogenous Region

Generally the larger the mode factor $W$, the more complex the MB is. Since diverse MB partitions can be referred as non homogenous region, so to determine whether or not $MB^e$ belongs to the non homogenous region, a threshold $T_{non-Hom}$ is set on MMC. The criterion is given as:

$$MB^e \in \text{Non Homogenous} \quad \text{if} \quad MMC > T_{non-Hom}$$

$$MB^e \in \text{Homogenous} \quad \text{if} \quad MMC \leq T_{non-Hom}$$

### Table I. Mode Factors Assigned to Each Mode

<table>
<thead>
<tr>
<th>MB Mode</th>
<th>SKIP</th>
<th>16x16</th>
<th>16x8</th>
<th>8x16</th>
<th>8x8</th>
<th>8x8ref</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>W</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>BLK Mode</td>
<td>8x8</td>
<td>8x4</td>
<td>4x8</td>
<td>4x4</td>
<td>SKIP</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$W = \frac{W_{enh}}{W_{new}}$ and $H = \frac{H_{enh}}{H_{new}}$ are the neighboring up-sampled MBs of $MB^e$ which overlap with current $MB^e$. In ESS, an enhancement layer MB may be covered by 1, 2, or 4 up-sampled base layer MBs [6]. Let A, B, C and D are the neighboring blocks of current up-sampled base layer block $uMB^b$ “C”. Let $uMB^b_i$ (for example A in Fig. 2) be the neighboring up-sampled MBs of $uMB^b$ which overlap with current $MB^e$. In [12], high layer MBs are classified into four classes: Corner, Hori, Vert and Center.

For homogenous region, the derivation of MB partition is early terminated and extensive search for each sub-partition can be avoided using proposed scheme. For each $MB^e$ belongs to non homogenous region, its constituent 4x4 blocks are checked one by one. Assume a given enhancement layer 4x4 block is covered by a number $N$ of base layer 4x4 blocks, $N=1,2$ or 4. If $N$ is greater than one, difference between the motion vectors of $N$ covering base layer 4x4 blocks is checked. So, an average motion vector $MV_{ave}$ is calculated as:

$$MV_{ave} = \frac{1}{N} \sum_{i=1}^{N} MV[i]$$

where $T_{non-Hom} = 0.36$. For homogenous region, the derivation of MB partition is early terminated and extensive search for each sub-partition can be avoided using proposed scheme. For each $MB^e$ belongs to non homogenous region, its constituent 4x4 blocks are checked one by one. Assume a given enhancement layer 4x4 block is covered by a number $N$ of base layer 4x4 blocks, $N=1,2$ or 4. If $N$ is greater than one, difference between the motion vectors of $N$ covering base layer 4x4 blocks is checked. So, an average motion vector $MV_{ave}$ is calculated as:

$$MV_{ave} = \frac{1}{N} \sum_{i=1}^{N} MV[i]$$

where $MV[i], i=1,...,N$ are the motion vectors of the base layer blocks that cover the enhancement layer 4x4 block. Now, the motion vectors of each of these base layer blocks $MV[i]$ are compared against the average motion vector $MV_{ave}$ to determine the similarity of motion vectors. In [6], JVT-W123 proposed a method in which weighted distortion measures are applied for base layer 4x4 blocks having absolute difference of motion vectors (MVs) greater than threshold set in JSVM tool. Whereas, residual prediction can be enabled/disabled on MB basis, even a single base layer 4x4 block having dissimilar MV can make residual prediction less favorable in an MB with existing method. However, a more relaxed criterion may be applied. In general, for adjacent blocks, their MVs are similar as long as they belong to one homogeneous video object [10]. In this case, it is helpful to apply usual RD measure and make residual prediction favorable on base layer 4x4 blocks, even though their MV absolute difference is larger than the MV threshold set in JVT-W123. Therefore, to evaluate the...
The similarity of MVs of neighboring 4x4 blocks by considering the motion/texture characteristics in video objects, absolute difference of MVs proposed in [6] is modified as:

\[
MMC \ast \left(\|\text{Diff}_{(x,y)}\|^2\right) \leq T_e
\]

\[
\text{Diff}_{(x,y)} = MV_{1(x,y)} - MV_{2(x,y)}
\]

where \(\|\|\|^2\) represents square of \(l_2\) norm of MV difference \(\text{Diff}_{(x,y)}\). \(MV_{1(x,y)}\) and \(MV_{2(x,y)}\) are the motion vectors of two 4x4 blocks respectively, \((x, y)\) are the horizontal and vertical indices of \(MV_i\) and \(MV_j\). \(T_e\) is the threshold set in JSVM tool. To remove visual artifacts, we only need to check base layer 4x4 blocks having mode factor \(W = 11\) from Table I, so MMC factor (product of \(W\) and \(A\)) used in (4) is much larger than \(T_{\text{ann-hom}}\) = 0.36, resulting that product of MMC and \(\text{Diff}_{(x,y)}\) is almost comparable with existing method. Since \(T_e\) is a fixed value, the similarly between \(MV_i\) and \(MV_j\) is adapted with MMC, which reflects the motion characteristic of enhancement layer MB. For any base layer 4x4 block having \(T_e\) greater than the threshold set in JSVM, we set enhancement layer MB contains problematic areas. When a MB contains potentially problematic areas, a weighted distortion measure is used in the rate-distortion evaluation of the mode decision process, so that the unbalanced distortion associated with residual prediction mode is weighted more than usual in calculating the overall distortion of the MB. As a result, residual prediction is less likely to be selected for the MB.

B. Error Concealment in ESS

In order to conceal lost \(MB_e\) at ESS enhancement layer using inter-layer error concealment method, the following operations need to be performed at the decoder. As shown in Fig. 3 residuals and MVs of lost \(MB_e\) are derived from up-sampled residuals and MVs of base layer \(uMB_b\) using inter-layer prediction, whereas motion compensation of \(MB_e\) is performed from the reference frame in ESS enhancement layer using inter-prediction.

The BLSkip error concealment method drives MVs of the blocks in the lost enhancement layer picture as if the blocks are coded in base layer skip mode. Consider a bit stream consisting of I and P pictures in which each P picture is motion compensated from its previous picture. Let the \(n^{th}\) P picture of enhancement layer is denoted by \(P^e_n\). Consider the current enhancement layer picture \(P^e_n\) containing \(K\) blocks \(P^e_n = (B^e_1, B^e_2, ..., B^e_N)\). Let \(R^e_{b_k}\) and \(D^e_{k}\) be the reconstructed and residual blocks of \(B^e_{k}\) respectively. Then, the temporal prediction in the enhancement layer can be represented by [13].

\[
R^e_{b_k} = S(MC_k(P^e_{n-1}) + D^e_{k})
\]

where \(S(.)\) is a deblocking filter operation and \(MC_k(.)\) indicates the motion compensation operations using the reference picture \(P^e_{n-1}\) and decoded MV \(V^e_k\). When the current enhancement layer picture \(P^e\) is lost in transmission, the original MV cannot be decoded. Thus, the motion compensation for the lost block should be performed by using the estimated MV \(V^e_k\) instead of the original MV \(V^e_k\). In BLSkip method, the estimated MV \(V^e_k\) is calculated as:

\[
V^e_k = \eta \cdot V^b_k
\]

where \(V^b_k\) is the MV of the base layer block \(B^b_k\) corresponding to \(B^e_k\) and \(\eta\) is the up-sampling ratio of ESS enhancement layer and corresponding base layer. In the error concealment process, the error-free MC operation \(MC_k(P^e_{n-1})\) is replaced by \(MC_k(P^e_{n-1})\) which is performed using \(V^e_k\). In (5), since \(D^e_{k}\) is also not available when the current picture is lost, the BLSkip method estimates \(D^e_{k}\) exploiting the inter-layer correlation. Let \(D^e_{k}\) be the residual block of \(B^e_{k}\) and then \(D^e_{k}\) can be approximated as follows:

\[
D^e_{k} = D_b^e - U(D^e_{k})
\]

where \(U(.)\) represents an optional up-sampling filter operation and \(D_b^e\) is the estimated residual. The concealed block \(R^e_{b_k}\) is reconstructed as:

\[
R^e_{b_k} = S(MC_k(P^e_{n-1}) + U(D^e_{k}))
\]
Finally, for any of these base layer blocks $B_k^b$ having $T_\infty$ less than the threshold set in JSVM, up-sampling residuals $U(D_k^b)$ from base layer blocks are activated, along with estimated motion vector $\hat{V}_k^e$, prediction residual $\hat{D}_k^e$ and reconstructed block $\hat{R}_k^e$ to conceal lost MB in ESS enhancement layer. Thus, it can be expected that our method will effectively trade-off residual prediction between removing artifacts and concealing lost MB in non homogenous region.

IV. SIMULATION RESULTS AND RELATED DISCUSSIONS

To demonstrate the performance of the proposed method 200 frames of various sequences (YUV, 4:2:0, progressive, intra period-1, group of pictures 4) with testing conditions as specified in [14] are used. After generating bit stream, packet-loss simulator is used to simulate packet loss during transmission. Error pattern is used in simulator as proposed in [15]. Other simulation conditions are as follows:

- JSVM version 9.8 [16].
- Spatial scaling ratios: 2:3 (224x192 to 336x288) for foreman, bus and football sequences @15fps; (352x288 to 640x480) for down-sampled version of harbour and crew sequences @30fps; and 3:4 (528x432 to 704x576) for soccer and city sequences @30fps as proposed in [14] are used.
- QP setting: quantization parameter of $QP_0 = (28, 32, 36, 40)$ and $QP_1 = (30, 34, 38, 42)$ are used for base layer and enhancement layer respectively.
- BLSkip method is performed for error concealment at the decoder [17].
- Threshold for MV difference is selected as $T_\infty = 30$.
- Two layers sliding window base layer and enhancement layer are used, as error concealment at the decoder supports maximum two layers only.
- Bit-stream is decoded after passing through simulator with 3% packet loss rate for base layer and 20% packet loss rate for enhancement layer.

The comparison of JVT-W123, referred as “Anchor”, is made with “Proposed” algorithm in error prone environment. The testing parameters in our experiments include Bjontegaard delta PSNR (BD-PSNR) and Bjontegaard delta bit-rate (BD-BR) [18]. Reliabilities of BD-PSNR and BD-BR are measured as assigned in reliability metric [19]. BD-PSNR and BD-BR are used to represent the average PSNR and bit-rate differences between the RD curves derived from anchor and the proposed algorithm respectively. It should be noted that positive value in PSNR gain and negative value in bit-rate saving show improvement; whereas reliability value ranges from 0 to 1, and value approaching 1 shows better reliability.

In Table II, average gain and reliability with corresponding PSNR-Y and bit-rate saving for various sequences are listed. From Tables II, we can observe that our proposed scheme achieves up to 0.29 dB gain with overall average 0.1 dB gain for all sequences which shows efficient use of residual prediction.

<table>
<thead>
<tr>
<th>Sequences</th>
<th>BD-PSNR Gain</th>
<th>BD-BR Gain</th>
<th>BD-PSNR Reliability</th>
<th>BD-BR Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Foreman</td>
<td>0.0173</td>
<td>-0.9052</td>
<td>0.9968</td>
<td>0.9668</td>
</tr>
<tr>
<td>Bus</td>
<td>0.2923</td>
<td>-29.3871</td>
<td>0.9966</td>
<td>0.7129</td>
</tr>
<tr>
<td>Football</td>
<td>0.2118</td>
<td>-5.4382</td>
<td>0.9942</td>
<td>0.9364</td>
</tr>
<tr>
<td>Harbour</td>
<td>-0.0059</td>
<td>0.0275</td>
<td>0.9989</td>
<td>0.8825</td>
</tr>
<tr>
<td>Crew</td>
<td>0.1537</td>
<td>-14.0780</td>
<td>0.9995</td>
<td>0.8054</td>
</tr>
<tr>
<td>Soccer</td>
<td>0.0152</td>
<td>-0.6525</td>
<td>0.9942</td>
<td>0.8262</td>
</tr>
<tr>
<td>City</td>
<td>-0.0100</td>
<td>1.0021</td>
<td>0.9961</td>
<td>0.9543</td>
</tr>
</tbody>
</table>

Since visual artifact and effect of error concealment are the subjective quality measures, the comparison of quality degradation for football sequence with (top) anchor and (bottom) proposed scheme is shown in Fig. 4. It is clearly observed that error concealment has severe adverse impact as compare to artifact in the visual quality degradation, which
shows the efficient use of residual prediction’s trade-off. Fig. 5 shows RD curves for football sequence (top) having 336x288 resolutions, harbour sequence (middle) having 640x480 resolutions and soccer sequence (bottom) having 704x576 resolutions. As far as complexity reduction is concerned, it shows remarkable achievement. In homogenous region, the derivation of MB partition is early terminated, whereas for non homogenous region, residuals are performed for only those constituent 4x4 blocks having threshold $T_e$ less than that set in JSVM tool. Otherwise, residual prediction is less likely to be selected, and weighted distortion is applied for the MB.

V. CONCLUSIONS

In this paper, by exploiting spatial homogeneity and temporal stationary characteristics in video objects, efficient use of residual prediction to conceal lost macroblock and to improve visual artifacts in a packet loss environment is achieved. Comparing with JVT-W123, the proposed algorithm achieves not only average PSNR gains of 0.1 dB but also reduces coding complexity as MB mode derivation in homogenous region is early terminated and extensive search for dissimilar motions are only performed in non homogenous region. Future work is required for more efficient use of inter-layer residual prediction with temporal stationary characteristics in non homogenous region.

REFERENCES


