MAP based blocking artifact reduction for DCT domain distributed video coding

Liu Hongbin(刘鸿彬)**, Zhao Debin*, Ma Siwei**, Gao Wen**
(∗School of Computer Science and Technology, Harbin Institute of Technology, Harbin150001, P.R.China)
(**School of Electronics Engineering and Computer Science, Peking University, Beijing100871, P.R. China)

Abstract
This paper proposes a maximum a posteriori (MAP) based blocking artifact reduction algorithm for discrete cosine transform (DCT) domain distributed video coding, in which the SI and the initial reconstructed Wyner-Ziv (WZ) frame are utilized to further estimate the original WZ frame. Though the MAP estimate improves quality of the artifact region, it also leads to over-smoothness and decreases quality of the non-artifact region. To overcome this problem, a criterion is presented to discriminate the artifact and the non-artifact region in the initial reconstructed WZ frame, and only the artifact region is updated with the MAP estimate. Simulation results show that the proposed algorithm provides obvious improvement in terms of both objective and subjective evaluations.

Key words: distributed video coding (DVC), blocking artifact reduction, maximum a posteriori (MAP)

0 Introduction

Current video compression standards usually perform computationally intensive motion field estimation including motion estimation and mode decision for inter-pictures coding at encoder to efficiently exploit the temporal correlation, e.g. MPEG-x and H.26x. As a result, the encoder is much more complex than the decoder. This is reasonable for broadcasting or for the systems where videos are compressed once and decoded many times. However, in scenarios where the encoder is not so powerful, such as the sensor network and the mobile communication, the dual complexity allocation may be required. Distributed video coding (DVC) provides an opportunity to afford the applications with simple encoder and powerful decoder.

The foundation of DVC is the theoretic results of Slepian and Wolf[1] on lossless coding of correlated sources with SI available at the decoder, which was extended to lossy coding by Wyner and Ziv [2]. In DVC[3-10], frames are usually classified into key frames and Wyner-Ziv frames. A traditional intra codec is employed to encode key frames, based on which the SI of the WZ frames can be generated. The SI can be considered as the corrupted WZ frame, and the errors in the SI are corrected by the parity bits sent from the encoder. By generating the SI at the decoder, the temporal correlation is exploited by the decoder and the computational complexity is also shifted to the decoder. DVC framework includes two categories: pixel domain DVC (PDVC)[3] and DCT transform domain DVC (TDVC)[4]. TDVC is more complex than PDVC, but brings significant performance improvement, thus it becomes the prevalent architecture and is chosen as the platform in this paper. In TDVC, reconstruction quality of the WZ frame is largely influenced by the SI quality, especially when the quantization is coarse. When medium to large motion occurs, the SI quality is usually poor and blocking artifact will appear on the WZ frames. In this case, post-processing aiming at reducing the blocking artifact is required.

Various methods have been proposed to reduce blocking artifact in the block DCT based image and video compression, such as: the filtering approaches[11,12], the iteration approaches based on projections onto convex sets (POCS)[13,14] and the MAP probability approaches[15-17] and so on. Unfortunately, few papers contribute to the specific blocking artifact reduction problem in TDVC. As far as we know, [18] is the only work reporting the deblocking technique for TDVC. In[18], deblocking filter of H.264/AVC is extended to TDVC. Because there is only one coding mode in TDVC and the quantization matrixes (QM) in TDVC are different from that of the H.264/AVC, definition of the filter strength cannot be used directly in TDVC. Ref.[18] proposes a new definition for the filter strength to tackle this problem. However, Ref.[18] brings only little improvement in the group of picture (GOP) = 2 case, which is the frequently used GOP setting in the TDVC. Therefore, new blocking artifact reduction algorithm in TDVC is still urgently required.

1 Supported by the National Natural Science Foundation of China (No.60672088 , No.60736043) , and the National Basic Research Development Program of China (2009CB320905)
2 To whom correspondence should be addressed. E-mail:liuhongbinhit@qq.com
Received on May 11, 2010
In this paper, a new MAP based post-processing technique is proposed to reduce the blocking artifact of the WZ frames in TDVC. We choose to follow the MAP probability approach, because both the SI and the reconstructed WZ frame can be included in it very conveniently. In the proposed algorithm, the WZ frame model is based on Huber-Markov random field (HMRF), which is extensively used in the image processing [15-17]. On the other hand, Laplace distribution is used to model the residual image (between WZ frame and SI) as usual. As the MAP estimate usually leads to over-smootheness and thus decreases the quality of the non-artifact region, a criterion is also presented to discriminate the artifact and the non-artifact region in the initially reconstructed WZ frame, and only the artifact region is updated with the MAP estimate.

1 Framework of the traditional TDVC codec

In this section, the framework of the TDVC codec proposed in [4] is reviewed. The block diagram of the codec is shown in Fig.1, and the work flow of both the encoder and the decoder are described as follows.

1.1 Encoder of the TDVC codec

At the TDVC encoder, frames are encoded into either key frames or WZ frames. The key frames are encoded using the H.264/AVC intra encoder. For each WZ frame, a non-overlapped 4×4 block based DCT is performed firstly. Then the 16 DCT positions are classified into several groups according to the pre-selected quantization parameter. Coefficients in the same group of all the blocks are raster scanned and organized into one band, and then pipelined to one uniform quantizer. Different bands are quantized utilizing independent uniform quantizers. The employed QMs are shown in Fig.2. The \( M_i \) block defines the reserved quantization level of each coefficient band.

1.2 Decoder of the TDVC codec

At the decoder, the key frames are decoded by the H.264/AVC intra decoder. For the WZ frame, motion compensated interpolation/extrapolation is used to generate its SI. Then as the encoder does, 4×4 DCT transform, band organization and band quantization are performed on the SI. After this, turbo decoding will be performed on DCT bands in zigzag order, and the bitplane of each band is decoded from the MSB to the LSB. During the decoding of one bitplane, multi-stage decoding may be required. In each stage, the parity bits transmitted from the encoder and the corresponding SI band will be forwarded to the channel decoder to start the iterative decoding. If the decoding error ratio converges to a value which is below a threshold, for example 0.001, the decoding process is stopped; otherwise, the decoder will request more parity bits from the encoder through a request channel, and a new decoding stage will start. After decoding all bitplanes of one band, they will be used to reconstruct the band with the help of the corresponding SI band. If some bands are not encoded, the corresponding SI bands will be used as the reconstruction. Finally, when all bands are reconstructed, they will be inv-organized and inverse DCT (IDCT) transformed to reconstruct the WZ frame.

2 MAP estimate of the WZ frame

At the TDVC decoder, besides the reconstructed WZ frame, the SI is also known. In this section, the MAP estimate is extended to include the SI. The employed WZ frame model and residual model are also described. Besides, the iterative solution of the MAP estimate is also presented.

2.1 MAP estimate

To describe the MAP estimate of the WZ frame, some notations are defined. Define \( X \) as the \( N \times M \) WZ frame, \( Y \) as its SI, \( X' \) as the initially reconstructed WZ frame and \( \Pr(\cdot) \) as the probability density function. Image pixels take integer values (LSB). To achieve high compression efficiency, parity bits of each bitplane are successively transmitted to the decoder until a satisfied decoding error ratio is achieved or all parity bits are sent.
located in the range \( R = [0 \, 255] \). The MAP estimate of \( X \) can be formulized as

\[
\hat{X} = \arg \max_{X \in R^{x \times y}} \Pr(X \mid X', Y)
\]

\[
= \arg \max_{X \in R^{x \times y}} \log \Pr(X \mid X', Y)
\]

The log likelihood function can be rewritten as:

\[
\log \Pr(X \mid X', Y) = \log \Pr(X) + \log \Pr(Y \mid X) + \log \Pr(X' \mid X, Y) - \log \Pr(X', Y)
\]

The last term of the right side can be dropped because it is constant. Notice that \( X' \) is determined by \((X, Y) \) in TDVC according to:

\[
X' = \text{IDCT} \left( F \left( Q \left( \text{DCT}(X) \right), \text{DCT}(Y) \right) \right)
\]

Here, \( \text{DCT} \) and \( \text{IDCT} \) refer to the DCT and IDCT transform respectively. \( F \) is the reconstruction function and \( Q \) is the quantization function. Consequently,

\[
\Pr(X' \mid X, Y) = \begin{cases} 1 & \text{if } X' = \text{IDCT} \left( F \left( Q \left( \text{DCT}(X) \right), \text{DCT}(Y) \right) \right) \\ 0 & \text{otherwise} \end{cases}
\]

and

\[
\log \Pr(X' \mid X, Y) = \begin{cases} 0 & \text{if } X' = \text{IDCT} \left( F \left( Q \left( \text{DCT}(X) \right), \text{DCT}(Y) \right) \right) \\ -\infty & \text{otherwise} \end{cases}
\]

Therefore, Eq. (1) can be rewritten into:

\[
\hat{X} = \arg \max_{X \in S} \log \Pr(X) + \log \Pr(Y \mid X)
\]

Here, \( X \) is confined to:

\[
S : \left \{ X \mid \text{IDCT} \left( F \left( Q \left( \text{DCT}(X) \right), \text{DCT}(Y) \right) \right) = X', X \in R^{x \times y} \right \}
\]

In this way, the SI is included into the MAP estimate.

**2.2 WZ frame model**

The WZ frame model is based on HMRF, which is verified to depict the image distribution very well\(^{[15-17]} \). To facilitate the description, \( X \) is organized into an \((N \times M) \times 1\) column vector in raster scan order. HMRF is given in Eq.(7) without much discussion, and interesting readers can refer to\(^{[15]} \).

\[
\Pr(x) = \frac{1}{Z} \exp \left( -\frac{1}{\lambda} \sum_{t=0}^{N-1} \rho_t (d_t \cdot x) \right)
\]

Here, \( Z \) is a normalizing constant and \( \lambda \) is a regularization parameter. \( c \) is a local group of pixels called clique (a clique consists of a pair of pixels, a pixel and one of its neighbor pixels), and \( C \) is the set of all such cliques, which depends on the neighborhood structure of HMRF. The Huber function \( \rho_t(\bullet) \) is given by

\[
\rho_t(x) = \begin{cases} x^2 & \text{if } |x| \leq T \\ T^2 + 2T(|x| - T) & \text{if } |x| > T \end{cases}
\]

The \( \rho_t \) function permits some inconsistency of Eq. (7) when the difference between a pixel and its neighbor is larger than \( T \). \( T \) is a threshold which controls the amount of allowed inconsistency with Eq.(7).

In Eq.(7), \( d_t \) is a \((N \times M) \times 1\) column vector which is defined to extract the difference of the two pixels in one clique, and thus each clique corresponds to a different \( d_t \). \( d_t' \) is the transpose of \( d_t \). Using the \( d_t \) defined above, Eq.(7) can be rewritten into

\[
\Pr(x) = \frac{1}{Z} \exp \left( -\frac{1}{\lambda} \sum_{t=0}^{N-1} \sum_{m,n} \rho_t (x[m] - x[n]) \right)
\]

Here, \( \Psi_m \) is the index set of the \( m \)th pixel’s neighbors. The neighbors include the available (not out of the frame range) upleft, up, upright, left, right, downleft, down and downright pixel of the \( m \)th pixel.

**2.3 Residual image model**

Denote \( y \) as the column vector form of the SI \( Y \), then \( y \) can be seen as the original WZ frame \( x \) corrupted by an additive noise \( n \).

\[
y = x + n
\]

In DVC, \( n \) is usually assumed to follow the Laplace distribution\(^{[3-4]} \). That is to say:

\[
\Pr(y \mid x) = \exp(-\alpha |x - y|) \cdot \left(\frac{\alpha}{2}\right)^{N \times M}
\]

Here, \( \alpha = \sqrt{2} / \delta \) and \( \delta^2 \) is the variance of the noise \( n \).

**2.4 Iterative solution**

Substitute Eqs(9) and (11) into Eq.(6) and drop the constants, and multiply Eq.(6) by \(-\lambda \), we can get:

\[
\hat{X} = \arg \min_{X \in S} \sum_{n=0}^{N \times M-1} \sum_{m,n} \rho_T(x[m] - x[n]) + \alpha \cdot \lambda \cdot |x - y| \]

Classic algorithms\(^{[11]} \) for convex functional minimization are based on iterative techniques, and the update monotonically decreases the functional value at each iteration. Let \( x(k) \) denote the estimate of \( x \) at the \( k \)th iteration, and the gradient projection method is used to find the next estimate:

\[
x^{(k+1)} = x^{(k)} + \alpha^{(k)} \cdot p^{(k)}
\]

Here the \( p^{(k)} \) is the steepest direction and \( \alpha^{(k)} \) is the step size:

\[
p^{(k)} = \sum_{n=0}^{N \times M-1} \sum_{m,n} \rho_T^* (d_t^* \cdot x^{(k)}) d_t^* d_t^* + \alpha \cdot \lambda \cdot \text{sgn}(x - y)
\]

\[
\alpha^{(k)} = \frac{\sum_{n=0}^{N \times M-1} \sum_{m,n} \rho_T^2 (d_t^* \cdot x^{(k)}) d_t^* d_t^* p^{(k)}}{\sum_{n=0}^{N \times M-1} \sum_{m,n} \rho_T^2 (d_t^* \cdot x^{(k)}) d_t^* d_t^*}
\]

Here, \( p^*(\bullet) \) and \( p^*(\bullet) \) are the first and second order derivative of \( \rho_t(\bullet) \), and \( p^{(k)} \) is the transpose of \( p^{(k)} \). \( \text{sgn}(x - y) \)
returns a sign vector including only -1, 1 and 0, which denotes the sign of each element in $x$. $d_x^{(n)}$ is defined to extract the difference between the $n$th pixel and the $m$th pixel, and $d_x^{(m)}$ is its transpose. The step size may be large and leads to the functional value increase sometimes, therefore, it is necessary to verify that the functional value does reduce at each iteration. If it does not, we reduce the step size by half until it does reduce the cost function.

After obtaining $x^{(k+1)}$ and thus $X^{(k+1)}$, the $X^{(k+1)}$ will be DCT transformed, and each DCT coefficient $x$ in it will be projected to the valid range defined by the decoded quantization bin $[l, u]$ according to:

$$x = \begin{cases} 
l & x < l \\
l \leq x < u & \\
u & x \geq u 
\end{cases} \quad (15)$$

Here, $l$ and $u$ are the lower and upper bound of $x$. The projection guarantees that the obtained solution at each iteration is in the set $S$ defined in subsection A.

The iteration procedure stops when $\|d_x^{(k)}p^{(k)}\| < 10$ or the iterations is larger than a given number.

## 3 Proposed WZ frame blocking artifact reduction system

The block diagram of the proposed WZ frame blocking artifact reduction system for GOP=2 is shown in Fig.3. New modules highlighted in the dotted window are added in the decoder of the traditional TDVC codec. The encoder is the same with the traditional TDVC encoder, thus the proposed algorithm does not increase the encoder complexity.

At the decoder, the MAP estimate described in Section 3 is employed to generate a new reconstruction of the WZ frame. However, as the new reconstruction may be over-smooth and degrades the quality of the non-artifact region, we should only update the artifact region with it. To achieve this aim, MAP estimator module, artifact detector module and a fuser, which can be considered to be the non-artifact block and the artifact block respectively. $N$ and $M$ denotes the size of the WZ frame. $C$ is a threshold which is empirically set to 6 in this paper.

### 3.1 Artifact detector

In TDVC, the blocking artifact is usually caused by the coarse quantization and the poor SI quality. To find out the artifact regions, we should find out the regions with poor SI.

The SI quality can be measured by the difference between the SI and the initially reconstructed WZ frame to some extent. This is because the larger difference between them usually means that the more errors in SI are corrected, thus corresponds to a poorer SI, and vice versa. In this paper, the mean absolute difference (MAD) between the SI and the initially reconstructed WZ frame is calculated and the artifact map is generated according to:

$$M(n) = \begin{cases} 
0 & \text{if } MAD_n < C \\
1 & \text{otherwise} 
\end{cases} \quad (16)$$

Here, $n$ is the $4 \times 4$ block index and $MAD_n$ is the MAD of the $n$th block. $M(n) = 0$ and 1 means the $n$th block is considered to be the non-artifact block and the artifact block respectively. $N$ and $M$ denotes the size of the WZ frame. $C$ is a threshold which is empirically set to 6 in this paper.

### 3.2 Fuser

It should be noted that the artifact region is identified based on the $4 \times 4$ block, which is different from the filtering approach[12] where only the boundaries of the blocks are filtered. This is because the poor SI in TDVC is usually caused by an error motion vector (MV) rather than an imprecise MV, which can lead to the poor reconstruction of the whole block. After generating the artifact map, the initially reconstructed WZ frame which is denoted as $X'$ and the MAP estimated frame which is denoted as $\hat{X}$ are merged. For the $n$th block, if $M(n)$ equals to 0, the corresponding block in $X'$ is used as the final reconstruction; otherwise, the corresponding block in $\hat{X}$ is used as the final reconstruction.

## 4 Experimental results

To evaluate the performance of the proposed algorithm, simulation results and performance comparisons are provided in this section. TDVC without MAP post-processing (TDVC-NMAP)[3], TDVC in which the MAP estimated reconstruction is used as the final reconstruction (TDVC-MAP) and TDVC in which the initially reconstructed WZ frame is merged with the MAP estimated reconstruction (TDVC-MMAP) are compared.

Fig. 3 Block diagram of the proposed WZ frame blocking artifact reduction system.
interpolation (MCI)\(^{(4)}\) is employed to generate the SI for the WZ frames. In the experiment, \(T\) in Eq.(8), \(\lambda\) in Eq.(7) and \(\delta\) in Eq.(11) are set to 3, 2 and 2 respectively. Though \(\delta\) varies with different residual image and \(T\) and \(\lambda\) also change with different situations, it is shown that the empirical values work well for a wide range of conditions though they may be not optimal. Besides, Eq.(13) converges to a good result with no more than 3 iterations in the implementation. The subjective quality is measured by the well known structural similarity (SSIM)\(^{(19)}\), and the objective quality is measured by peak signal-to-noise ratio (PSNR). Both the subjective quality and the objective quality are compared only on the luma component as usual. The bit rate is not provided since the proposed algorithm does not change it in the GOP=2 case. All QMs (see Fig.2) of the WZ frames are tested, and quantization parameter (QP) of the key frames is chosen to achieve similar average quality between them and the initially reconstructed WZ frames (i.e., WZ frames which are not post-processed).

Average PSNR and average SSIM of the WZ frames are compared for TDVC-NMAP, TDVC-MAP and TDVC-MMAP in Table 1 and Table 2 respectively, in which the best performance for each QM and the average performance of each method are highlighted in bold. Several phenomena can be observed: first, when compared with TDVC-NMAP, TDVC-MMAP can always achieve better performance on all tested sequences (both subjectively and objectively). Second, TDVC-MAP achieves best performance on Football but worst performance on News and Hall Monitor sequence. This is because Football contains large motion and the blocking artifact on it is severe, thus a thorough deblocking is preferred. However, when motion is not fast and blocking artifact is weak, such as on News and Hall Monitor, deblocking should be weak to maintain the details of the WZ frames. In this case, a thorough deblocking will lead to performance loss. Finally, TDVC-MMAP can achieve larger improvement when motion is larger and thus blocking artifact is more severe.

It also can be observed from Table 1 and Table 2 that: TDVC-MMAP can improve the PSNR by up to 0.81dB and can improve the SSIM by up to 0.015. On the other hand, TDVC-MAP can improve the PSNR by up to 0.85dB and can improve the SSIM by up to 0.0162. However, TDVC-MAP also degrades the PSNR by up to 3.47dB and decreases the SSIM by up to 0.0385. To give an intuitive impression, 32nd frame of Football and 54th frame of News (for QM=\(M^{2}\)) are compared in Fig.4 and Fig.5. It can be seen that both TDVC-MAP and TDVC-MMAP can reduce the blocking artifact effectively (see Fig.4). Meanwhile, TDVC-MAP provides a stronger blocking artifact reduction capability (see Fig.4) but loses more details (see Fig.5) than TDVC-MMAP.

To sum up, though TDVC-MAP works slightly better than TDVC-MMAP on sequences with jerky motions, it also leads to performance loss on sequence with medium to slow motion. On the other hand, TDVC-MMAP brings obvious performance improvement both objectively and subjectively on sequence with large motion, and it always achieves better performance than TDVC-NMAP. Therefore, TDVC-MMAP is more robust than TDVC-MAP, and is employed by the proposed system.

5 Conclusion

In this paper, a MAP based blocking artifact reduction algorithm is proposed for the WZ frames in TDVC. To improve the subjective quality and the objective quality on sequences with both low and high motion, artifact detection criterion is also presented to discriminate the artifact and the not-artifact region, and only the artifact region is updated with MAP estimated reconstruction. It is shown that the proposed TDVC-MMAP algorithm provides obvious improvement in terms of both subjective and objective evaluations when compared with TDVC-NMAP, and it is more robust than TDVC-MAP which leads to performance loss on sequence with slow motion.

References


Liu Hongbin  born in 1984. He received his B.S. degree in School of Computer Science and Technology of Harbin Institute of Technology in 2001. He is currently pursuing the Ph.D. degree at the Department of Computer Science, Harbin Institute of Technology. His research interests include distributed video coding, image and video coding/processing.

### Table 1  PSNR (dB) comparison of: TDVC-NMAP, TDVC-MAP and TDVC-MMAP.

<table>
<thead>
<tr>
<th>QM</th>
<th>Football</th>
<th>Soccer</th>
<th>News</th>
<th>Hall Monitor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TDVC-NMAP</td>
<td>TDVC-MAP</td>
<td>TDVC-MMAP</td>
<td>TDVC-NMAP</td>
</tr>
<tr>
<td>M1</td>
<td>33.80</td>
<td>34.43</td>
<td><strong>34.46</strong></td>
<td>35.52</td>
</tr>
<tr>
<td></td>
<td>31.15</td>
<td><strong>32.00</strong></td>
<td>31.96</td>
<td>33.09</td>
</tr>
<tr>
<td>M3</td>
<td>30.41</td>
<td><strong>31.15</strong></td>
<td>31.05</td>
<td>32.23</td>
</tr>
<tr>
<td></td>
<td>30.29</td>
<td><strong>31.03</strong></td>
<td>30.92</td>
<td>32.10</td>
</tr>
<tr>
<td>M5</td>
<td>28.49</td>
<td><strong>28.16</strong></td>
<td>29.04</td>
<td>30.61</td>
</tr>
<tr>
<td></td>
<td>28.05</td>
<td><strong>28.68</strong></td>
<td>28.56</td>
<td>30.14</td>
</tr>
<tr>
<td>M7</td>
<td>26.49</td>
<td><strong>29.94</strong></td>
<td>26.79</td>
<td>28.89</td>
</tr>
<tr>
<td>Average</td>
<td><strong>29.81</strong></td>
<td><strong>30.48</strong></td>
<td><strong>30.40</strong></td>
<td><strong>31.80</strong></td>
</tr>
</tbody>
</table>

### Table 2  SSIM comparison of: TDVC-NMAP, TDVC-MAP and TDVC-MMAP.

<table>
<thead>
<tr>
<th>QM</th>
<th>Football</th>
<th>Soccer</th>
<th>News</th>
<th>Hall Monitor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TDVC-NMAP</td>
<td>TDVC-MAP</td>
<td>TDVC-MMAP</td>
<td>TDVC-NMAP</td>
</tr>
<tr>
<td>M1</td>
<td>0.8664</td>
<td>0.8713</td>
<td><strong>0.8748</strong></td>
<td>0.8591</td>
</tr>
<tr>
<td></td>
<td>0.7763</td>
<td>0.7883</td>
<td><strong>0.7913</strong></td>
<td>0.7751</td>
</tr>
<tr>
<td>M3</td>
<td>0.7429</td>
<td><strong>0.7583</strong></td>
<td>0.7552</td>
<td>0.7406</td>
</tr>
<tr>
<td></td>
<td>0.7417</td>
<td><strong>0.7572</strong></td>
<td>0.7541</td>
<td>0.7400</td>
</tr>
<tr>
<td>M5</td>
<td>0.6696</td>
<td><strong>0.6858</strong></td>
<td>0.6832</td>
<td>0.6811</td>
</tr>
<tr>
<td></td>
<td>0.6632</td>
<td><strong>0.6787</strong></td>
<td>0.6769</td>
<td>0.6655</td>
</tr>
<tr>
<td>M7</td>
<td>0.5881</td>
<td><strong>0.6011</strong></td>
<td>0.5976</td>
<td>0.6195</td>
</tr>
<tr>
<td>Average</td>
<td><strong>0.7212</strong></td>
<td><strong>0.7344</strong></td>
<td><strong>0.7333</strong></td>
<td><strong>0.7258</strong></td>
</tr>
</tbody>
</table>
Fig. 4  32nd frame of Football for QM= $\hat{M}_2$. Left up: original, right up: TDVC-NMAP (PSNR: 30.19, SSIM: 0.7892), left down: TDVC-MAP (PSNR: 31.69, SSIM: 0.8302) and right down: TDVC-MMAP (PSNR: 31.41, SSIM: 0.8148).

Fig. 5  54th frame of News for QM= $\hat{M}_2$. Left up: original, right up: TDVC-NMAP (PSNR: 36.97, SSIM: 0.9364), left down: TDVC-MAP (PSNR: 34.04, SSIM: 0.9024) and right down: TDVC-MMAP (PSNR: 37.24, SSIM: 0.9372).