Concealment of Whole-Picture Loss in Hierarchical B-Picture Scalable Video Coding

Xiangyang Ji, Debin Zhao, and Wen Gao, *Senior Member, IEEE*

***Abstract*—H.264/AVC scalable video coding (H.264/AVC SVC), as the scalable extension of H.264/AVC, offers the flexible adaptivity in terms of spatial, temporal and SNR scalabilities for the generated bitstream. However, such compressed video still suffers from the bad playback quality when packet loss occurs over unreliable networks. In this paper, we present an error concealment algorithm to tackle the whole-picture loss problem in H.264/AVC SVC when hierarchical B-picture coding is used to support temporal scalability. In the proposed algorithm, by taking advantage of the temporal relationship among the adjacent video pictures, the motion information of the lost picture is derived simply and efficiently based on the principle of temporal direct mode. Utilizing the derived motion information, the lost picture is concealed by performing motion compensation on the correctly received temporally previous and future video pictures. The experimental results demonstrate that as a post-processing tool, the proposed error concealment algorithm is able to significantly improve both the objective and subjective qualities of the decoded video pictures in the presence of packet losses when compared to the error concealment algorithm used in H.264/AVC SVC reference software. The proposed method can also be applied to H.264/AVC with hierarchical B-picture coding for error concealment.**

***Index Terms*—Error concealment, H.264/AVC, hierarchical B-picture, motion compensation, scalable video coding.**

I. Introduction

H.264/AVC scalable video coding (H.264/AVC SVC) [1] is the scalable extension of H.264/AVC [2], which can yield the bitstream with spatial, temporal (frame rate) and SNR (quality) scalabilities. It is a desired solution to adapt video transmission on the time-varying networks, such as a wireless channel and the Internet. However, in video communication, transmission errors still result in a severe distortion for the reconstructed video.

Error concealment algorithm, usually as a post-processing tool at decoder, can conceal the erroneous region due to transmission errors according to the correctly received information. Temporal error concealment is one of the most important approaches to combat transmission errors. The simplest temporal error concealment is temporal replacement (TR) [3], in which each damaged macroblock is directly replaced by the co-located one in the temporally previous picture with zero motion. It usually only performs well for concealment of the erroneous region without motion. TR can be improved by a boundary matching algorithm (BMA) [4], in which a suitable motion vector for a damage macroblock can be selected from a set of candidate motion vectors based on side match distortion measure. In some cases, BMA algorithm is still unsatisfactory since the boundary information is not enough to accurately recover the lost motion vectors. Thus, the improved BMA algorithms are further studied in [5], [6]. In addition, the error concealment algorithms based on spatial and spatio-temporal interpolations have also been investigated in [7]–[11].

In the aforementioned BMA based error concealment algorithms, to conceal the damaged macroblock, its neighboring macroblocks need to be correctly received. Thus, it is desirable to split the current macroblock and its neighboring macroblocks in a picture into different slices since one slice data typically fits one packet and bit errors usually make a whole packet useless, especially in today's wireless network. This can be fulfilled by a flexible slice organization format with the error resilience tool like flexible macroblock ordering (FMO) in H.264/AVC. However, in H.264/AVC, FMO is only supported for Baseline and Extended Profiles not for Main Profile. Meanwhile, this kind of approach may degrade coding efficiency due to the close correlation broken between neighboring macroblocks and the extra overhead information for small packets [12]. On the other hand, traffic congestion may still lead to burst consecutive packet losses and thus, it is possible that all or some of the macroblocks surrounding a damaged macroblock are also lost. This situation becomes even worse when the packet loss rate increases. In addition, in low bitrate packet-based video communication, the size for bitstream packetization of video pictures may be even smaller than the minimum transfer unit of the network used for video communication [13]. For example, when transmitting a low bitrate bitstream over a universal mobile telecommunications system (UMTS) link, a picture typically will be coded in one slice and fit in one packet. Therefore, a transmission error will lead to the whole picture loss. As a result, the aforementioned BMA based error concealment approaches will no longer perform well in these situations.

An underlying solution to tackle this problem is to use the error concealment for the whole-picture loss, which can work well by only taking advantage of high temporal correlation

Manuscript received September 05, 2007; revised August 06, 2008. Current version published January 08, 2009. This work was supported by National Science Foundation of China (60736043). The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Deepa Kundur.

X. Ji was with the Institute of Computing Technology, Chinese Academy of Sciences and the Graduate School of Chinese Academy of Science, 100080 Beijing, China. He is now with Broadband Networks & Digital Media Laboratory of Automation Department, Tsinghua University, 100084 Beijing, China (e-mail: xyji@mail.tsinghua.edu.cn, xyji@jdl.ac.cn).

D. Zhao is with the Department of Computer Science, Harbin Institute of Technology, Harbin 150001, China (e-mail: dbzhao@jdl.ac.cn).

W. Gao is with the Institute of Digital Media, Peking University, Beijing 100871, China (e-mail: wgao@pku.edu.cn).

Digital Object Identifier 10.1109/TMM.2008.2008874

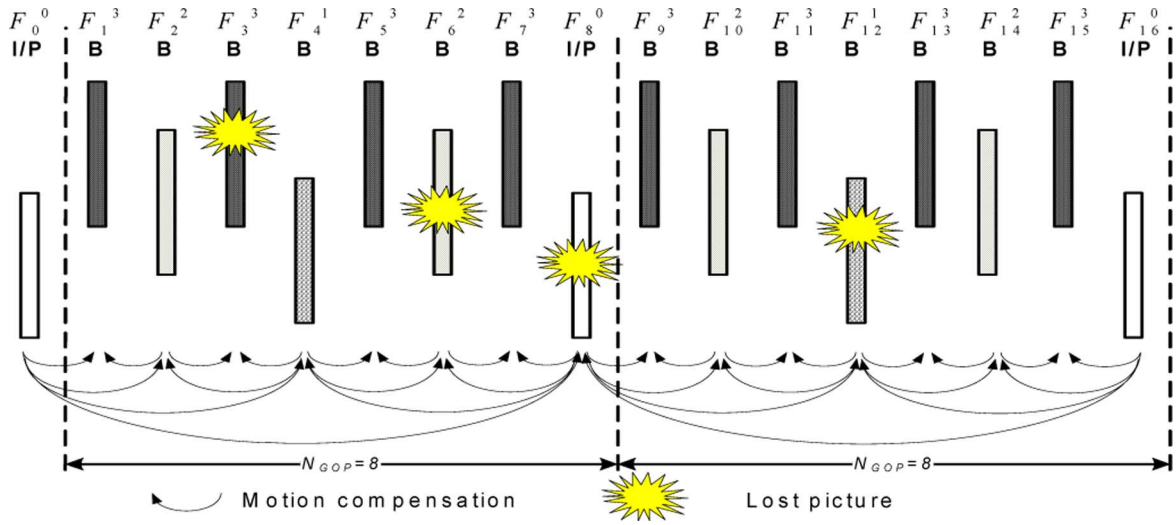


Fig. 1. Dyadic hierarchical B-picture coding with temporal decomposition number $M = 4$.

among the adjacent pictures. A motion vector extrapolation (MVE) has been presented in [14] to combat the whole-picture loss. It can extrapolate the motion vectors of the damaged macroblock from the last received picture and estimate the overlapped areas between the damaged macroblock and the motion extrapolation one. Another error concealment algorithm has also been proposed to combat the whole-picture loss for wireless low bitrate video coding based on the multi-frame optical flow estimation [13]. These error concealment algorithms mainly deal with P-picture with forward motion field estimation and are not specified for B-picture since B-picture in typical H.26x and MPEG-x hybrid video coding standards is not referenced by the subsequent pictures and thus, does not result in error propagation when it is lost. So they are not applicable to the error concealment for hierarchical B-picture coding in H.264/AVC SVC.

A temporal error concealment approach based on temporal direct mode [15], [16] has been adopted to combat the B-picture loss for hierarchical B-picture coding in H.264/AVC SVC [17], [18]. It derives the motion vector for each block in the lost B-picture according to the motion vector of the co-located block in the temporally subsequent reference picture and the temporal distance relationship between the lost picture and the reference pictures involved. This method has low computational complexity due to no motion estimation. However, its error concealment efficiency is usually unsatisfactory. In this paper, we present a new error concealment of whole-picture loss in hierarchical B-picture scalable video coding. Firstly, inspired by the work on the enhanced temporal direct mode coding for the hierarchical B-picture coding in H.264/AVC [19], in which the motion vectors for each block are allowed to be derived from the temporally not only subsequent reference picture, as in H.264/AVC but also previous reference picture, in the proposed algorithm, the motion vectors of each block in the lost picture can be recovered from the motion vector of the co-located block in the temporally subsequent or previous reference picture. Furthermore, the appropriate motion vector scaling techniques are also introduced when the picture pointed by the motion vector of the co-located block in the reference picture is not available in the reference buffer. In addition, an improved motion vector derivation is also proposed to yield better motion vectors for the lost B-picture not at the maximum temporal level, in which the motion vectors of the damaged block can be derived from the motion vectors of the co-located blocks in the temporally neighboring left and/or right B-pictures at next higher temporal level. Meanwhile, this method could also be used to conceal lost key picture as a P-picture.

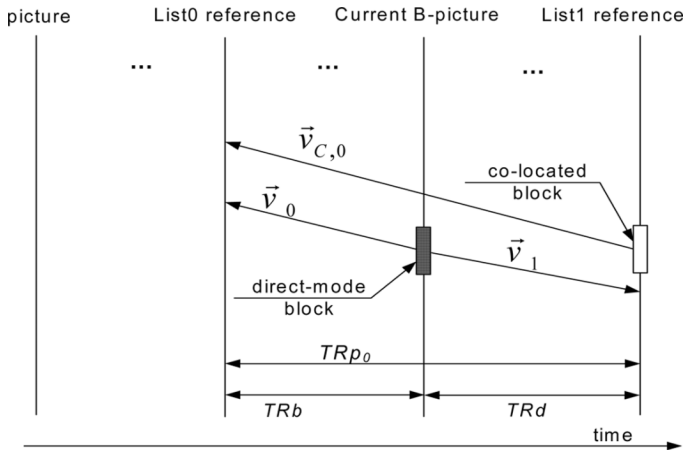


Fig. 2. List_0 and List_1 motion vectors derivation for the TDM block from the List_0 motion vector of the co-located block in the List_1 reference picture.

The rest of this paper is organized as follows. Section II describes the proposed error concealment algorithm for hierarchical B-picture coding in detail. Section III gives the experimental results in terms of both objective and subjective qualities. Finally, Section IV concludes this paper.

II. Proposed Error Concealment Strategy

A. Hierarchical B-Picture Coding

In dyadic hierarchical B-picture coding, the temporal decomposition number M can be calculated by

$$M = \log_2^{N_{\mathrm{GOP}}} + 1. \tag{1}$$

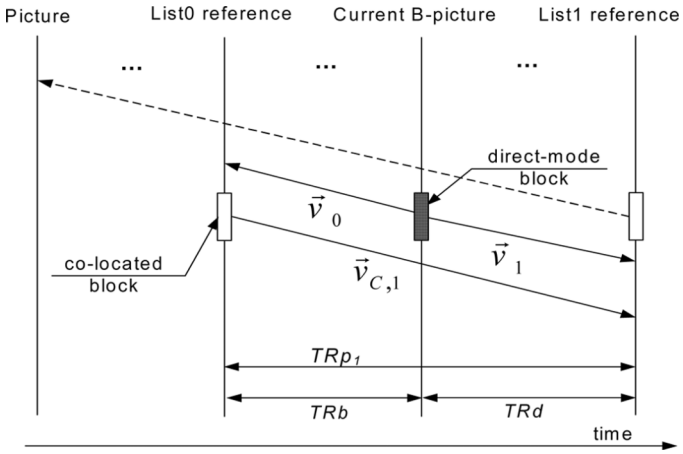


Fig. 3. List_0 and List_1 motion vectors derivation for the TDM block from the List_1 motion vector of the co-located block in the List_0 reference picture.

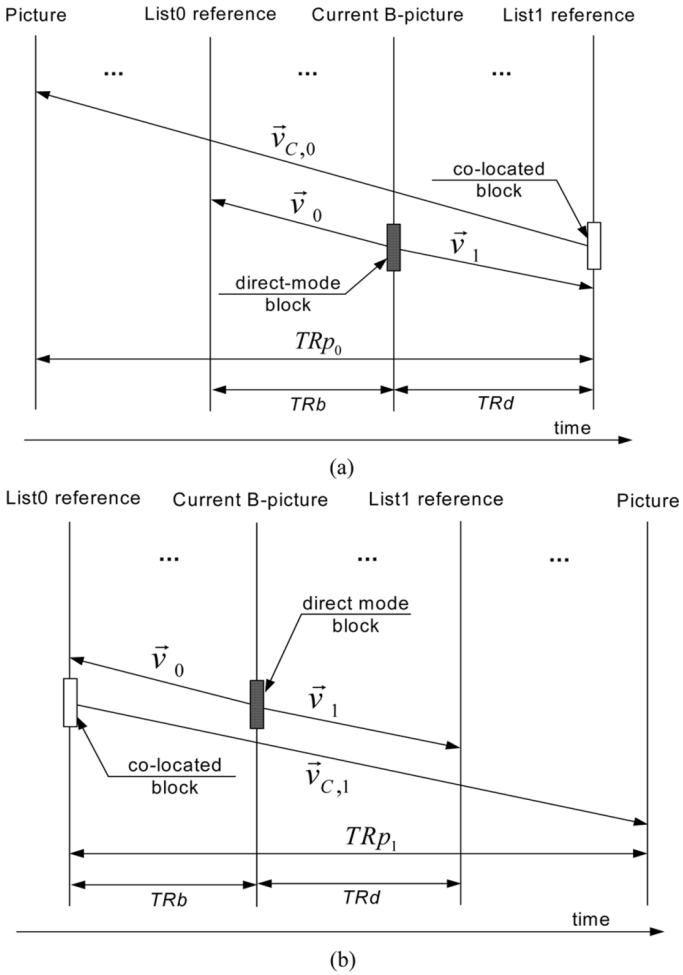


Fig. 4. List_0 and List_1 motion vectors $\vec{v}_0$ and $\vec{v}_1$ derivation according to the List_0 and List_1 motion vectors of the co-located blocks (a) in the List_1 reference picture and (b) in the List_0 reference picture with a new scaling technique when they point to the pictures which the current TDM block can not access.

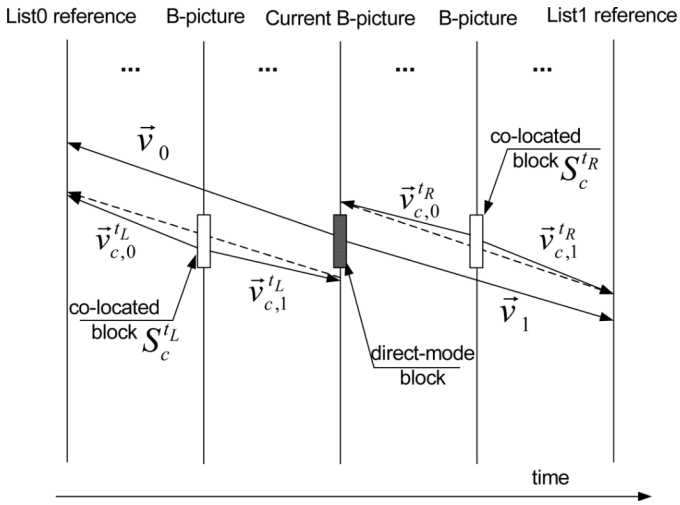


Fig. 5. Motion vectors $\bar{v}_0$. and $\bar{v}_1$ for the TDM block are derived from both of the co-located blocks $S_c^{t_L}$ and $S_c^{t_R}$ in the left and right neighboring B-pictures at next high temporal level.

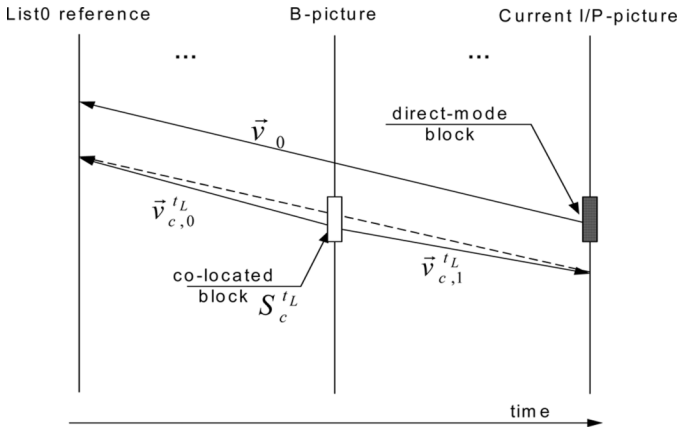


Fig. 6. Forward motion vectors $\vec{v}_0$ for the TDM block in the key picture is derived from the motion vector of the co-located block $S_c^{t_L}$ in the left neighboring B-picture at next high temporal level.

Here, N_{GOP} is the number of pictures in one group of pictures (GOP). Fig. 1 illustrates the dyadic hierarchical B-picture coding structure with a GOP size of 8 when only neighboring pictures at coarser temporal level are used for motion compensation prediction. In Fig. 1, I, P and B are used to denote the pictures coded in I-picture, P-picture and B-picture, respectively. The picture at the lowest temporal level 0 will be coded as a key picture, which is usually coded by P-picture. In general, multiple reference prediction can also be enabled for the hierarchical B-picture coding in H.264/AVC SVC. Furthermore, it is also allowed to be employed with non-dyadic temporal decomposition for the flexible applications [20], [21]. For simplicity, the error concealment algorithm presented in this paper is only restricted to the dyadic hierarchical B-picture coding. However, it can be extended to the non-dyadic decomposition.

For a bi-prediction block S in the B-picture F_t^m at time instant t with temporal level m, at encoder, its motion-compensated prediction signal can be obtained by

$$P(\vec{x}) = w_0 \tilde{F}_{t_0}(\vec{x} + \vec{v}_0') + w_1 \tilde{F}_{t_1}(\vec{x} + \vec{v}_1'). \qquad (2)$$

Here, $\vec{x}$ is the spatial coordinate of the pixel in the block S. w_0 and w_1 are the weighting values. In H.264/AVC SVC, due to the use of generalized B-picture, the bi-prediction block in

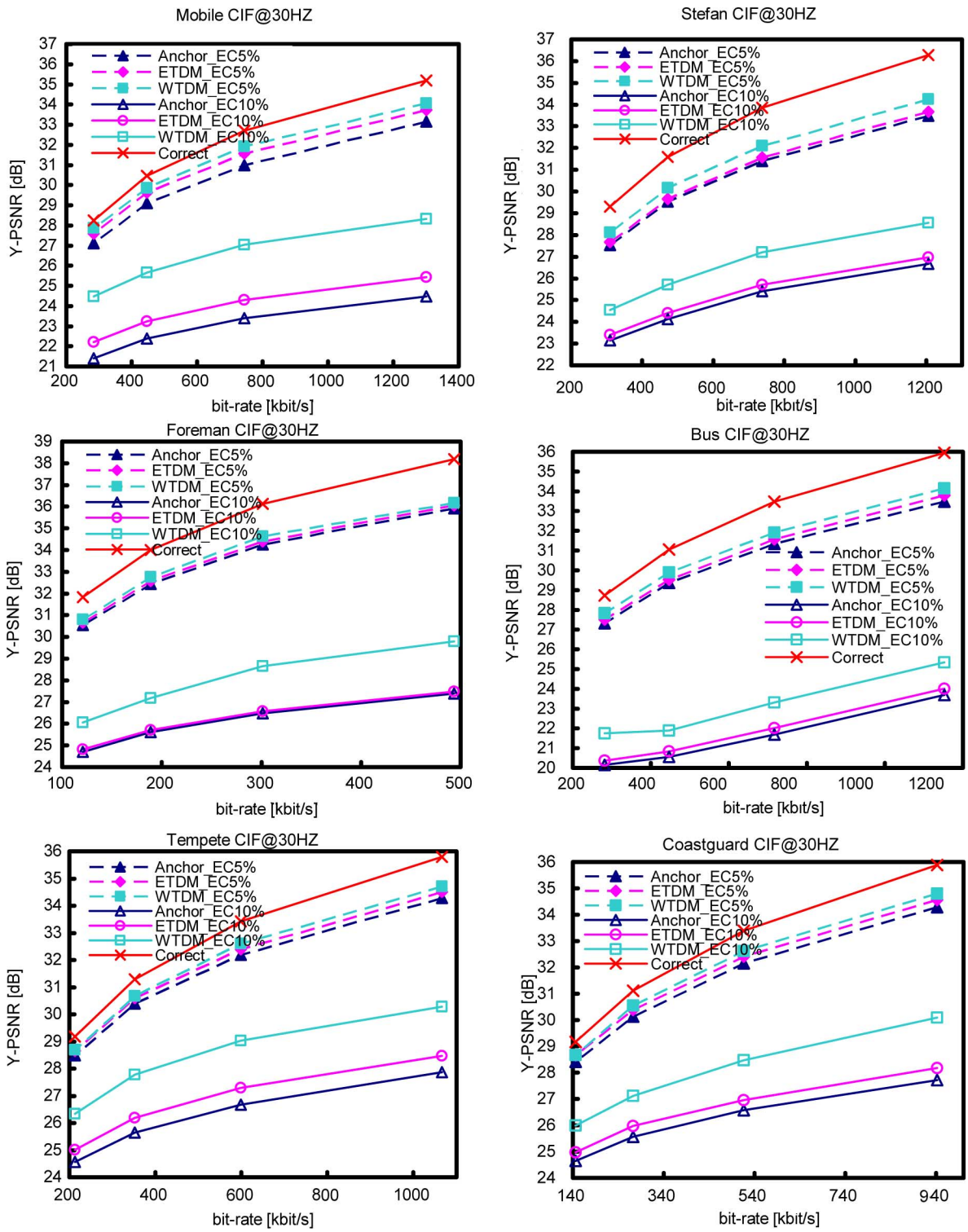


Fig. 7. Rate distortion curves comparison for no error case *correct* and different error concealment strategies Anchor_EC, ETDM_EC and WTDM_EC with %5 and %10 PLRs.

a B-picture allows two prediction blocks from *List 0* and *List 1* reference buffers which contain an arbitrary set of reference pictures in forward and/or backward directions [16]. Thus, $\tilde{F}_{t_0}$ and $\tilde{F}_{t_1}$ refers to the List_0 and List_1 reference pictures at time instants t_0 and t_1 for the current block S with the List_0 and List_1 motion vectors $\vec{v}'_0$ and $\vec{v}'_1$, accordingly.

At decoder, for a block S in the lost picture $\hat{F}_t^m$, which is concealed as a B-picture, its prediction signal can be achieved by

$$\hat{P}(\vec{x}) = w_0 \hat{F}_{t_0}(\vec{x} + \vec{v}_0) + w_1 \hat{F}_{t_1}(\vec{x} + \vec{v}_1). \quad (3)$$

TABLE I
ΔPSNR(DB) FOR ΔETDM_EC AND ΔWTDM_EC WITH DIFFERENT PLRS FOR LUMINANCE COMPONENT

PLR	QP	Stefan		Foreman		Bus		Tempete		Mobile		Coastguard	
		ΔETDM_EC	*ΔWTDM_EC*	*ΔETDM_EC*	*ΔWTDM_EC*	*ΔETDM_EC*	*ΔWTDM_EC*	*ΔETDM_EC*	*ΔWTDM_EC*	*ΔETDM_EC*	*ΔWTDM_EC*	*ΔETDM_EC*	*ΔWTDM_EC*
3%	28	0.06	0.13	0.10	0.16	0.05	0.05	0.09	0.06	0.17	0.25	0.11	0.10
	32	0.06	0.14	0.08	0.13	0.06	0.06	0.08	0.06	0.17	0.25	0.11	0.09
	36	0.06	0.14	0.07	0.13	0.05	0.05	0.08	0.04	0.16	0.22	0.09	0.08
	40	0.05	0.12	0.06	0.09	0.05	0.05	0.07	-0.02	0.15	0.21	0.07	0.05
	Avg.	**0.05**	**0.13**	**0.08**	**0.13**	**0.05**	**0.05**	**0.08**	**0.03**	**0.16**	**0.23**	**0.09**	**0.08**
5%	28	0.17	0.76	0.13	0.25	0.31	0.66	0.23	0.42	0.58	0.91	0.27	0.51
	32	0.17	0.69	0.15	0.37	0.23	0.57	0.22	0.42	0.57	0.92	0.25	0.46
	36	0.14	0.63	0.14	0.33	0.17	0.54	0.21	0.29	0.54	0.77	0.26	0.43
	40	0.12	0.56	0.13	0.26	0.19	0.51	0.17	0.20	0.49	0.75	0.24	0.25
	Avg.	**0.15**	**0.66**	**0.14**	**0.30**	**0.22**	**0.57**	**0.21**	**0.33**	**0.54**	**0.84**	**0.26**	**0.41**
10%	28	0.30	1.87	0.09	2.40	0.32	1.65	0.61	2.41	0.97	3.85	0.46	2.37
	32	0.30	1.79	0.11	2.18	0.33	1.63	0.61	2.34	0.91	3.65	0.38	1.90
	36	0.28	1.59	0.09	1.57	0.28	1.33	0.54	2.12	0.85	3.29	0.41	1.55
	40	0.25	1.39	0.10	1.35	0.22	1.61	0.46	1.78	0.81	3.06	0.32	1.34
	Avg.	**0.28**	**1.66**	**0.10**	**1.87**	**0.29**	**1.55**	**0.55**	**2.16**	**0.89**	**3.46**	**0.39**	**1.79**
20%	28	0.11	1.28	0.20	1.26	0.32	2.43	0.58	2.12	1.02	3.06	0.39	1.68
	32	0.13	1.12	0.21	1.05	0.34	2.45	0.55	2.03	0.98	2.89	0.37	1.33
	36	0.14	1.18	0.18	0.99	0.25	1.68	0.48	1.59	0.90	2.85	0.34	1.18
	40	0.12	1.08	0.19	0.39	0.26	1.65	0.45	1.46	0.85	3.49	0.30	0.90
	Avg.	**0.12**	**1.17**	**0.19**	**0.92**	**0.29**	**2.05**	**0.51**	**1.80**	**0.94**	**3.07**	**0.35**	**1.27**

(a)

(b)

(c)

Fig. 8. Subjective quality comparison of the 195th concealed picture at temporal level 4 for *Mobile*. (a) Anchor_EC (21.56 dB), (b) WTDM_EC and ETDM_EC (26.30 dB), (c) *Correct* (27.87 dB).

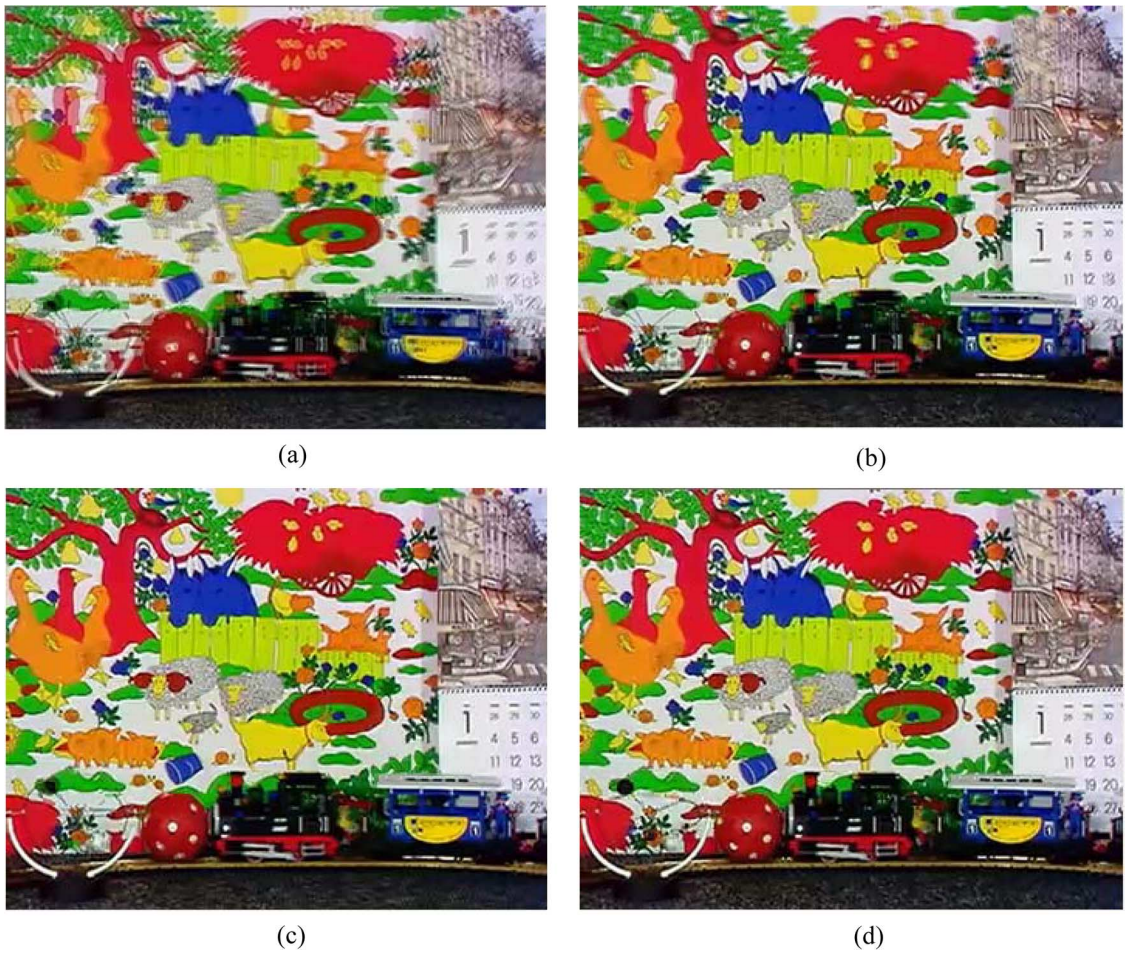


Fig. 9. Subjective quality comparison of the 230th concealed picture at temporal level 3 for *Mobile*: (a) Anchor_EC (18.29 dB), (b) ETDM_EC (23.49 dB), (c) WTDM_EC (27.03 dB), (d) *Correct* (27.70 dB).

Here, $\vec{v}_0$ and $\vec{v}_1$ correspond to the recovered List_0 and List_1 motion vectors of $\vec{v}'_0$ and $\vec{v}'_1$. The weights of w0 and w1 are both simply set to 1/2. Of course, implicit weighted prediction also can be used. $\hat{F}_{t_0}$ and $\hat{F}_{t_1}$ are the decoded List_0 and List_1 reference pictures. This prediction signal will be directly taken as the decoded signal for the block S since no reconstructed residual is received. Correspondingly, for a block S in the lost key picture to be concealed as a P-picture, only List_0 motion needs to be recovered.

B. Motion Parameters Recovery Based on the Enhanced TDM

In the case of the whole-picture loss, the motion vectors of each block in the lost picture have to be recovered based on the temporal motion correlation since all the neighboring blocks are also lost. Assume motion among the adjacent pictures is translational, a simple and efficient method to estimate the motion vector of the block in the lost B-picture can be implemented based on temporal direct mode (TDM), as proposed in [17]. In this subsection, to recover more accurate motion vectors, the traditional TDM is enhanced in the case of hierarchical B-picture coding, in which the co-located block to derive TDM motion vector can be flexibly selected in different cases.

In H.264/AVC SVC, TDM is able to take advantage of bi-prediction and does not need to transmit the motion parameters including the reference index and motion vectors. $\hat{F}_{t1\_f}$ denotes the picture referred to the first picture in the List_1 reference buffer. It can be taken as the List_1 reference picture for TDM. And the picture pointed by the List_0 motion vector $\vec{v}_{C,0}$ of the co-located block in $\hat{F}_{t1\_f}$, denoted as $\hat{F}_{t0\_r}$, is chosen as the List_0 reference picture. As illustrated in Fig. 2, the List_0 and List_1 motion vectors $\vec{v}_0$ and $\vec{v}_1$ for TDM block are derived from of the co-located block in $\hat{F}_{t1\_f}$ as follows:

$$\begin{cases} \vec{v}_0 = \frac{\mathrm{TRb}}{\mathrm{TRp}_0} \times \vec{v}_{C,0} \\ \vec{v}_1 = \frac{\mathrm{TRd}}{\mathrm{TRp}_0} \times \vec{v}_{C,0} \end{cases} \tag{4}$$

where TRb is the temporal distance, or more precisely picture order count (POC) distance [2] of the current B-picture $\hat{F}_t^m$ relative to $\hat{F}_{t0\_r}$, namely $\mathrm{POC}(\hat{F}_t^m) - \mathrm{POC}(\hat{F}_{t0\_r})$. TRp_0 is the temporal distance of $\hat{F}_{t1\_f}$ relative to $\hat{F}_{t0\_r}$, namely $\mathrm{POC}(\hat{F}_{t1\_f}) - \mathrm{POC}(\hat{F}_{t0\_r})$. TRd is the temporal distance of the current B-picture relative to $\hat{F}_{t1\_f}$, namely $\mathrm{POC}(\hat{F}_t^m) - \mathrm{POC}(\hat{F}_{t1\_f})$ and it should be equal to $\mathrm{TRb} - \mathrm{TRp}_0$. Note that actually, in H.264/AVC SVC, motion vector scalings in (4) and later equations are implemented with shift and multiplication operations.

In the aforementioned motion vector derivation for the TDM block according to (4), it is possible that $\vec{v}_{C,0}$ of the co-located block in $\hat{F}_{t1\_f}$ does not point to the temporally most recent List_0 picture. In this case, the derived motion vectors usually

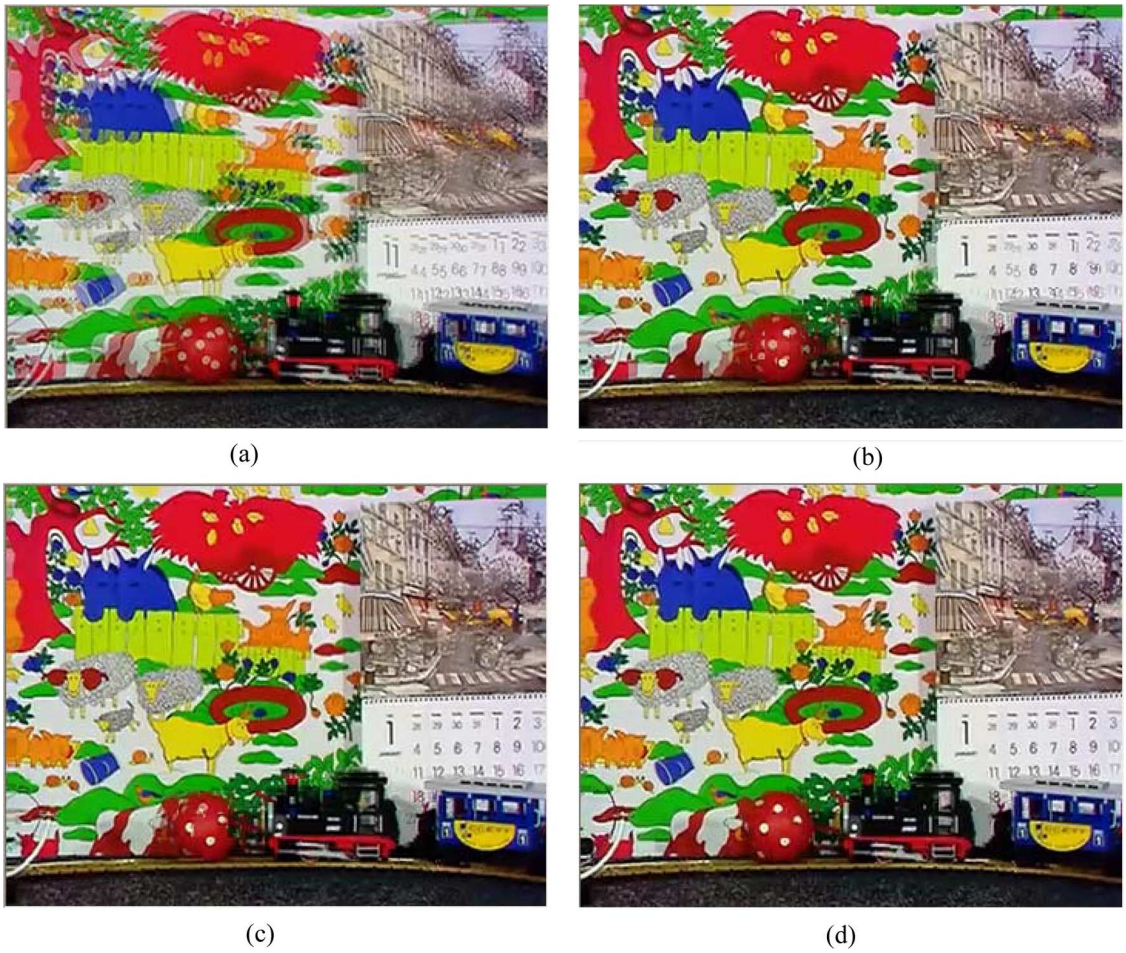


Fig. 10. Subjective quality comparison of the 156th concealed picture at temporal level 2 for *Mobile*: (a) Anchor_EC (16.36 dB), (b) ETDM_EC (22.06 dB), (c) WTDM_EC (24.82 dB), (d) *Correct* (27.89 dB).

are not precise enough to reflect the true motion since the temporal distances of the current B-picture picture relative to the reference pictures involved are too long. To tackle this situation, we can derive the List_0 and List_1 motion vectors of the TDM block based on the first picture in the List_0 reference (referred as $\hat{F}_{t0\_f}$), if it has a short temporal distance relative to the picture pointed by the List_1 motion vector $\vec{v}_{C,1}$ of the co-located block in $\hat{F}_{t0\_f}$. In such a case, $\hat{F}_{t0\_f}$ is taken as the List_0 reference picture for TDM block and the picture pointed by $\vec{v}_{C,1}$ of the co-located block in $\hat{F}_{t0\_f}$, denoted as $\hat{F}_{t1\_r}$, is chosen as the List_1 reference picture. As illustrated in Fig. 3, the List_0 and List_1 motion vectors $\vec{v}_0$ and $\vec{v}_1$ of TDM block can be derived from $\vec{v}_{C,1}$ of the co-located block in $\hat{F}_{t0\_f}$ by

$$\begin{cases} \vec{v}_0 = \frac{\mathrm{TRb}}{\mathrm{TRp}_1} \times \vec{v}_{C,1} \\ \vec{v}_1 = \frac{\mathrm{TRd}}{\mathrm{TRp}_1} \times \vec{v}_{C,1} \end{cases} \tag{5}$$

where $\mathrm{TRb} = \mathrm{POC}(\hat{F}_t^m) - \mathrm{POC}(\hat{F}_{t0\_f}), \mathrm{TRp}_1 = \mathrm{POC}(\hat{F}_{t0\_f}) - \mathrm{POC}(\hat{F}_{t1\_r}), \mathrm{TRd} = \mathrm{POC}(\hat{F}_t^m) - \mathrm{POC}(\hat{F}_{t1\_r})$ and thus, $\mathrm{TRd} = \mathrm{TRP}_1 + \mathrm{TRb}$.

However, if the picture pointed by $\vec{v}_{C,1}$ or $\vec{v}_{C,0}$ of the co-located block in the corresponding $\hat{F}_{t0\_f}$ or $\hat{F}_{t1\_f}$ picture, can not be accessed by the current lost B-picture $\hat{F}_t^m$ as they may not reside in the current reference buffer, a new available reference needs to be assigned. As a result, the motion vectors derivation according to (4) or (5) for TDM block is not applicable any more. For an example as shown in Fig. 4(a), when the picture $\hat{F}_{t0\_r}$ pointed by $\vec{v}_{C,0}$ of the co-located block in $\hat{F}_{t1\_f}$ can not be accessed by the current lost B-picture $\hat{F}_t^m$, the first picture $\hat{F}_{t0\_f}$ in List_0 reference buffer can be taken as the desired List_0 reference picture instead. In this case, the motion vectors of the TDM block in the lost B-picture $\hat{F}_t^m$ can be obtained by (4) with the new motion vector scaling, in which TRb is the temporal distance of the current B-picture relative to the picture $\hat{F}_{t0\_f}$ not $\hat{F}_{t0\_r}$ and thus, TRd is no longer equal to $\mathrm{TRb} - \mathrm{TRp}_0$. Similarly, as shown in Fig. 4(b), if the picture $\hat{F}_{t1\_r}$ pointed by $\vec{v}_{C,1}$ of the co-located block in the List_0 reference picture $\hat{F}_{t0\_f}$ can not be accessed by the current lost B-picture $\hat{F}_t^m, \hat{F}_{t1\_f}$ in List_1 reference buffer will be taken as the desired List_1 reference picture. And the motion vectors for TDM block can be obtained by (5) with the new scaling technique, in which TRd is the temporal distance of the current B-picture $\hat{F}_t^m$ relative to the picture $\hat{F}_{t1\_f}$ not $\hat{F}_{t1\_r}$ and thus, TRd is no longer equal to $\mathrm{TRP}_1 + \mathrm{TRb}$.

The detailed error concealment algorithm based on the enhanced TDM (ETDM_EC) for a lost B-picture can be described as follows.

ETDM_EC **algorithm:**

For each block in the lost B-picture:

if ($\mathrm{abs}(\mathrm{TRp}_0) <= \mathrm{abs}(\mathrm{TRp}_1)$){

  if ($\hat{F}_{t0\_r}$ is available)

    $\hat{F}_{t0} = \hat{F}_{t0\_r}, \hat{F}_{t1} = \hat{F}_{t1\_f}$ and (4) is used to derive $\vec{v}_0, \vec{v}_1$;

  else{

    $\hat{F}_{t0} = \hat{F}_{t0\_f}, \hat{F}_{t1} = \hat{F}_{t1\_f}$;

    if ($\mathrm{TRp}_0 != \mathrm{MAXVALUE}$)

      (4) is used to derive $\vec{v}_0$ and $\vec{v}_1$ with new scaling;

    else

      $\vec{v}_0$ and $\vec{v}_1$ are both set to zero;

    }

}

else {

  if($\hat{F}_{t1\_r}$ is available)

    $\hat{F}_{t0} = \hat{F}_{t0\_f}, \hat{F}_{t1} = \hat{F}_{t1\_r}$ and (5) is used to derive $\vec{v}_0, \vec{v}_1$;

  else

    $\hat{F}_{t0} = \hat{F}_{t0\_f}, \hat{F}_{t1} = \hat{F}_{t1\_f}$ and (5) is used to derive $\vec{v}_0$ and $\vec{v}_1$ with new scaling;

}

Notes: The abs ( ) function returns the absolute value of a number. TRp_0 is set to the maximum value (MAXVALUE) if the List_0 motion of the co-located block in $\hat{F}_{t1\_f}$ is NA and TRp_1 is also set to MAXVALUE if the List_1 motion of the co-located block in $\hat{F}_{t0\_f}$ is NA.

C. Further Improvement on Motion Parameters Recovery

The aforementioned approach is able to efficiently derive the motion vectors of each block in the lost B-picture because it is reasonable to assume that the motion among the adjacent pictures is translational if the temporal distances of the pictures involved are short. Therefore, it works well to recover the lost B-picture $\hat{F}_t^{M-1}$ at the highest temporal level $M-1$ since it usually has short temporal distances relative to their reference pictures. However, for a lost B-picture $\hat{F}_t^m$ which is not at the highest temporal level $M-1$, the temporal motion relationship among the pictures involved tends to considerably weaken as the temporal distances among these pictures become longer [16]. On the other hand, the aforementioned ETDM_EC algorithm only focuses on the concealment of the lost B-picture. If the key picture $\hat{F}_t^0$ is lost and recovered as a P-picture, the motion vectors of its block usually can not be derived based on TDM. As a result, the motion vector of each block in the lost key picture will be set to zero and its replacement is yielded by directly copying from the temporally previous key picture. To tackle this problem, an improved motion vector derivation approach is further employed.

In this method, the List_0 and List_1 reference pictures for the lost B-picture are the temporal most recent pictures $\hat{F}_{t_{m0}}$ and $\hat{F}_{t_{m1}}$ at lower temporal levels with time instants

$$t_{m0} = t - 2^{(M-m-1)} \quad \text{and} \quad t_{m1} = t + 2^{(M-m-1)}.$$

As shown in Fig. 5, it calculates the motion vectors $\vec{v}_0$ and $\vec{v}_1$ for a block in the lost B-picture according to the motion vectors of the co-located blocks $S_c^{t_L}$ and $S_c^{t_R}$ in the left and right neighboring B-pictures $\hat{F}_{t_L}^{m+1}$ and $\hat{F}_{t_R}^{m+1}$ at temporal level $m+1$ as follows:

$$\vec{v}_0 = \begin{cases} \vec{v}_{c,0}^{t_L} - \vec{v}_{c,1}^{t_L} & S_c^{t_L} \text{ refers to } \hat{F}_{t_{m0}} \text{ and } \hat{F}_t^m \\ 2 \times \vec{v}_{c,0}^{t_L} & S_c^{t_L} \text{ refers to } \hat{F}_{t_{m0}} \text{ but not } \hat{F}_t^m \\ -2 \times \vec{v}_{c,1}^{t_L} & S_c^{t_L} \text{ refers to } \hat{F}_t^m \text{ but not } \hat{F}_{t_{m0}} \\ \mathrm{NA} & \text{Otherwise} \end{cases} \quad (6)$$

and

$$\vec{v}_1 = \begin{cases} \vec{v}_{c,1}^{t_R} - \vec{v}_{c,0}^{t_R} & S_c^{t_R} \text{ refers to } \hat{F}_t^m \text{ and } \hat{F}_{t_{m1}} \\ -2 \times \vec{v}_{c,0}^{t_R} & S_c^{t_R} \text{ refers to } \hat{F}_t^m \text{ but not } \hat{F}_{t_{m1}} \\ 2 \times \vec{v}_{c,1}^{t_R} & S_c^{t_R} \text{ refers to } \hat{F}_{t_{m1}} \text{ but not } \hat{F}_t^m \\ \mathrm{NA} & \text{otherwise} \end{cases} \quad (7)$$

with

$$t_L = t - 2^{(M-m-2)} \text{ and } t_R = t + 2^{(M-m-2)}.$$

Here, $\vec{v}_{c,0}^{t_L}$ and $\vec{v}_{c,1}^{t_L}$ are the List_0 and List_1 motion vectors of the co-located block $S_c^{t_L}$ in $\hat{F}_{t_L}^{m+1}$. $\vec{v}_{c,0}^{t_R}$ and $\vec{v}_{c,1}^{t_R}$ are the List_0 and List_1 motion vectors of the co-located block $S_c^{t_R}$ in $\hat{F}_{t_R}^{m+1}$. Note that $S_c^{t_L}$ and $S_c^{t_R}$ will be enforced to perform at the minimum same block size. For example, if the macroblocks having $S_c^{t_L}$ and $S_c^{t_R}$ have partitions of 16×8 and 8×16, respectively, the recovery of motion vectors of the lost block will be performed at 8×8 level. On the other hand, due to intra coding or the further loss of the left and/or right neighboring B-picture, if one of motion vector $\vec{v}_0$ and $\vec{v}_1, \vec{v}_i$, is NA (not available), the other motion vector $\vec{v}_{1-i}$ will be further derived by

$$\vec{v}_i = \begin{cases} \mathrm{NA} & \vec{v}_{1-i} \text{ is NA} \\ -\vec{v}_{1-i} & \text{otherwise.} \end{cases} \quad (8)$$

According to (8), if $\vec{v}_0$ and $\vec{v}_1$ for a block in the lost B-picture are both taken as NA, the ETDM_EC algorithm described in Section II-B will be used instead.

It should be noted that this further improvement on motion parameters recovery perhaps introduce an extra delay for video decoding within one GOP since the motion vectors recovery of the lost picture at low temporal level demands the motion information decoding of the B-picture at higher temporal level. This can be solved by using a receiver-side buffer. The requirement of such a buffer is common in today's video streaming systems like Microsoft Windows Media Player 9, which has a default buffer setting to five seconds [22].

In terms of concealing the lost key picture as a P-picture, the proposed improved motion vectors recovery approach can also work very well to tackle this situation because the forward motion vector of the block in the lost key picture can be derived from the decoded motion vector of the co-located block in the left neighboring B-picture $\hat{F}_{t_L}^1$ by (6), as shown in Fig. 6. If $\vec{v}_0$ is not available due to intra coding, the motion vector of TDM block will be set to be zero. However, when $\hat{F}_{t_L}^1$ is also lost, the ETDM_EC algorithm described in Section II.B can not be further used since it only works for B-picture. To address this problem, the motion vectors of the co-located block in $\hat{F}_{t_L}^1$ should be first recovered and then, (6) is used to further derive

(a) (b) (c)

Fig. 11. Subjective quality comparison of the 104th concealed picture at temporal level 1 for *Mobile*: (a) Anchor_EC and ETDM_EC (12.80 dB)}, (b) WTDM_EC (16.93 dB), (c) *Correct* (28.14 dB).

the forward motion vector of the block in the lost key picture $\hat{F}_t^0$ based on $\hat{F}_{t_L}^1$.

Finally, the whole proposed motion parameters recovery strategy (WTDM_EC) can be depicted as follows.

WTDM_EC **algorithm:**

For each block in the lost picture:

if (the lost picture $\hat{F}_t^m$ is a key picture with $m = 0$){

if ($\hat{F}_{t_L}^1$ is also lost)

$\hat{F}_{t_L}^1$ should be first concealed as a B-picture;

$\hat{F}_{t0} = \hat{F}_{t+2M-1}$ and (6) is used to derive $\vec{v}_0$;

}

else {

If (both $\vec{v}_0$ and $\vec{v}_1$ are NA)

ETDM_EC algorithm is used;

else

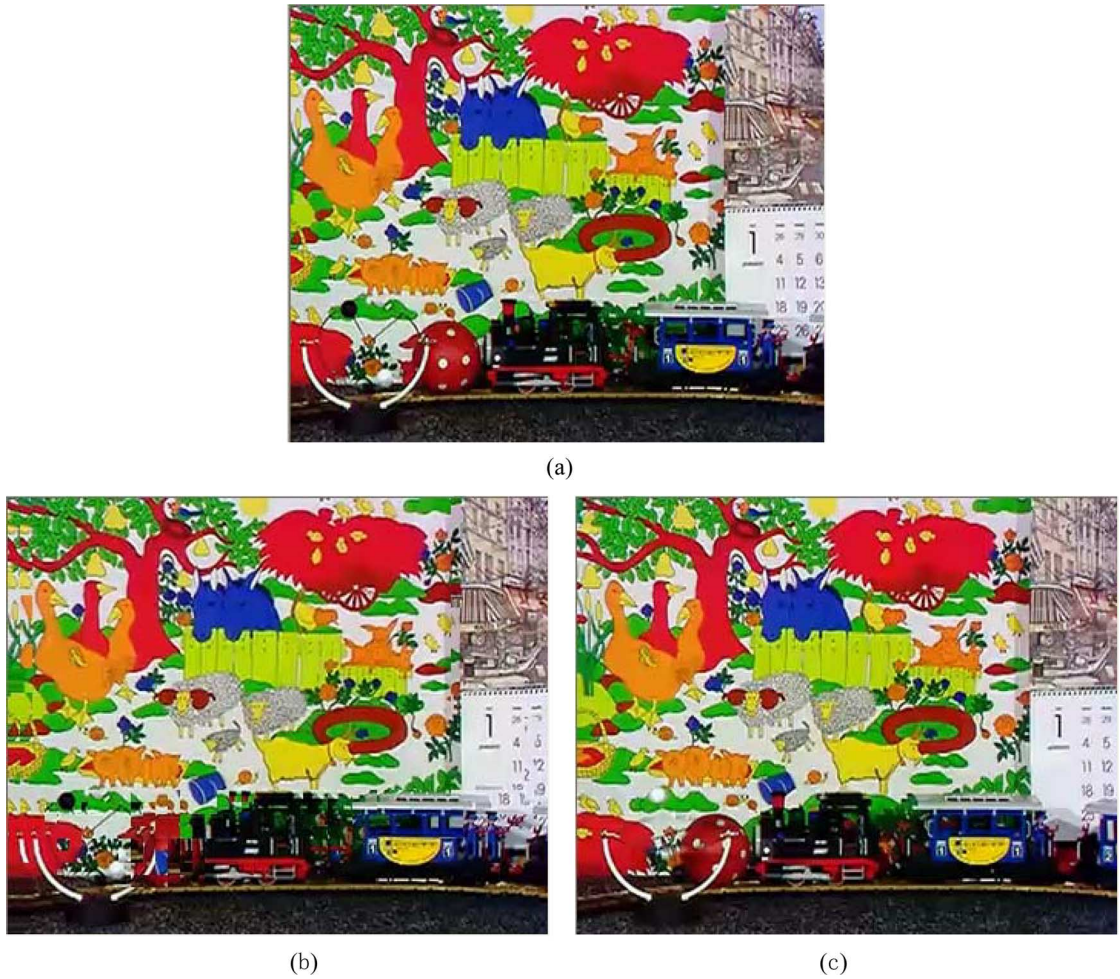
$\hat{F}_{t0} = \hat{F}_{t_{m0}}, \hat{F}_{t1} = \hat{F}_{t_{m1}}$ and (6), (7) and (8) are used to derive $\vec{v}_0$ and $\vec{v}_1$;

}

III. Experimental Results

To verify the proposed method in terms of both the objective and subjective qualities, it was integrated into JSVM_8_6, which is the H.264/AVC SVC reference software. In the experiments, only the whole-picture loss is considered although the proposed algorithms can be simply extended to the slice loss case. Test sequences are comprised of Mobile, Stefan, Foreman, Bus, Tempete and Coastguard in CIF@30 Hz. For each test sequence, the temporal hierarchical B-picture decomposition level is set to 4 ($N_{GOP} = 16$) and one I-picture is inserted for every 32 pictures. Quantization parameters for highest temporal level are composed of 28, 32, 36 and 40 and the quantization parameters for other lower temporal levels are set with the cascading quantization parameter [23]. Note that in real implementation, rate control tool is usually used to generate the desired bitstream at a given bitrate and it will also have impact on the performance of error concealment.

In the testing, the bitstream is formatted as a series of NAL units (NALUs) and then packetized. Four packet loss patterns with average packet loss rates of 3%, 5%, 10%, and 20% used in ITU-T VCEG were employed. These error patterns were generated from the experiments on the Internet backbone between one sender and three reflector sites. Detailed descriptions of the



(a) (b) (c)

Fig. 12. Subjective quality comparison of the 256th concealed key picture for *Mobile*: (a) A*nchor_EC and ETDM*_EC (12.99 dB)}, (b) WTDM_EC (19.38 dB), (c) *Correct* (30.06 dB).

generation, file format, and usage of the error patterns are available in [24].

A. Objective Quality Comparison

In terms of objective quality comparison, the rate distortion curves of the luminance component for no error case and different error concealment strategies are shown in Fig. 7 with 5% and 10% packet loss rates (PLRs), which are the very representatives of the wireless environment [25]. In Fig. 7, Anchor_EC indicates the existing temporal error concealment method in H.264/AVC SVC and *Correct* indicates no error case. Since in the proposed method WTDM_EC, ETDM_EC can be taken as an independent error concealment algorithm without an extra decoding delay requirement, its performance was also evaluated. Table I gives the detailed Δ PSNR results on ETDM_EC vs. Anchor_EC (ΔETDM_EC) and WTDM_EC vs. Anchor_EC (ΔWTDM_EC) for luminance component with 3%, 5%, 10%, and 20% PLRs. It can be observed from Fig. 7 and Table I that ETDM_EC outperforms Anchor_EC and WTDM_EC is able to further improve the PSNR of the reconstructed video especially at high PLRs. For example, WTDM_EC is able to gain 0.84 dB and 3.46 dB against Anchor_EC on average in the cases of 5% and 10% PLRs for *Mobile* sequence, respectively. On the other hand, although in Table I, overall coding performance improvement in terms of PSNR is marginal on average for low PLRs by the proposed methods since the lost pictures account for low percentage, the improvement on only the concealed picture is still notable.

B. Subjective Quality Comparison

In this subsection, the visual qualities of the concealed pictures using Anchor_EC, ETDM_EC and WTDM_EC and the correctly decoded picture are compared for *Mobile* with QP $=$ 40 and 10% PLRs. Figs. 8–11 give subjective quality comparison of the concealed B-picture at different temporal levels, respectively. As shown in Fig. 8, for the 195th picture at temporal level 4, the concealed picture by Anchor_EC is obviously blurred due to inaccurate motion vectors derivation. By the proposed ETDM_EC approach, its visual quality can be significantly improved. It should be noted that the concealed picture by WTDM_EC is the same as the one by ETDM_EC because at the highest temporal level, no B-picture at higher temporal can be utilized in WTDM_EC. In Fig. 9, for the 230th picture at temporal level 3, the concealed picture by Anchor_EC is also obviously blurred. By the proposed ETDM_EC approach, its visual quality is improved but some parts of the concealed picture are still blurred. Compared to ETDM_EC, WTDM_EC is

able to efficiently avoid blurring effect and thus, offers an excellent visual quality for the concealed picture. Compared with ETDM_EC, the further improvement by WTDM_EC is very important since in hierarchical B-picture coding, if the lost picture is not at highest temporal level, its concealment also affects the neighboring pictures with reference to it. The same phenomenon also can be observed in Fig. 10 for the 156th picture at temporal level 2.

In Fig. 11 for the 104th picture at temporal level 1, compared with the Anchor_EC, the proposed methods also can improve the quality of the concealed picture. However, due to weak temporal motion correlation between the picture being concealed and the reference picture, the quality of the concealed picture is obviously degraded in comparison with the correctly decoded one. It should be noted that the concealed picture by ETDM_EC is the same as the one by Anchor_EC because at the lowest temporal level, only the co-located block in the List_1 reference picture can be used to derive TDM motion vector of each block in the lost B-picture. In addition, in Fig. 12, subjective quality comparison of the 256th concealed key picture is also given. To conceal the lost key picture, Anchor_EC and ETDM_EC only directly copy the whole picture from the previous key picture with zero motion. In WTDM_EC, the improved motion vector recovery is used and thus, is able to significantly improve the quality of the concealed picture although its visual quality is still obviously worse than the correctly decoded one due to weak temporal motion correlation.

To sum up, compared to Anchor_EC, both ETDM_EC and WTDM_EC are able to obviously improve the objective and subjective qualities for the concealed pictures, especially for WTDM_EC. As mentioned in Section II.C, WTDM_EC is improved from ETDM_EC and perhaps introduces an extra delay for video decoding within one GOP since the motion vector recovery of the lost picture at low temporal level demands the motion information decoding of the B-pictures at higher temporal level. This can be solved by using a receiver-side buffer and such a buffer requirement is common in today's video streaming systems. If no extra receiver-side buffer is employed, certainly, only ETDM_EC can be used for error concealment. In addition, all these error concealment algorithms do not require the high-computational estimation process like in BMA and thus, are very applicable for practical applications.

IV. Conclusion

This paper proposes a novel error concealment strategy on combating the whole-picture loss for the hierarchical B-picture coding. Compared with the existing temporal error concealment approach in H.264/AVC SVC reference software, the proposed algorithm is able to more efficiently derive the motion vector of the damaged block in the lost picture by utilizing the motion information of the co-located blocks in the temporally neighboring previous and/or subsequent pictures. The experimental results demonstrate that the proposed algorithm is able to yield the high-quality recovered picture. In fact, the proposed method also can be easily extended to the slice loss case with the error resilient tool like FMO. An exemplified approach is that the recovery motion vector by the proposed method can be taken as an extra candidate motion vector to speed up the motion vector search process for BMA.

References

[1] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1103–1120, Sept. 2007.

[2] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, July 2003.

[3] J. W. Suh and Y. S. Hu, "Error concealment based on directional interpolation," *IEEE Trans. Consumer Electron.*, vol. 43, no. 3, pp. 295–302, Aug. 1997.

[4] W. M. Lam, A. R. Reibman, and B. Liu, "Recovery of lost or erroneously received motion vectors," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Process.*, Apr. 1993, vol. 5, pp. 417–420.

[5] J. Feng, K. Lo, H. Mehropour, and A. E. Karbowiak, "Error concealment for MPEG video transmissions," *IEEE Trans. Consumer Electron.*, vol. 43, no. 2, pp. 183–187, May 1997.

[6] C. S. Park, J. Ye, and S. U. Lee, "Lost motion vector recovery algorithm," in *Proc. IEEE Int. Symp. Circuits and Systems*, May 1994, vol. 3, pp. 229–232.

[7] Z. Alkachouch and M. G. Bellanger, "Fast DCT-based spatial domain interpolation of blocks in images," *IEEE Trans. Image Process.*, vol. 9, no. 4, pp. 729–732, Apr. 2000.

[8] L. Atzori, F. G. B. De Natale, and C. Perra, "A spatio-temporal concealment technique using boundary matching algorithm and mesh-based warping (BMA-MBW)," *IEEE Trans. Multimedia*, vol. 3, no. 3, pp. 326–338, Sep. 2001.

[9] K. Meisinger and A. Kaup, "Spatial error concealment of corrupted image data using frequency selective extrapolation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, May 2004, vol. 3, pp. 209–212.

[10] Y. Wang, Q. F. Zhu, and L. Shaw, "Maximally smooth image recovery in transform coding," *IEEE Trans. Commun.*, vol. 41, no. 10, pp. 1544–1551, Oct. 1993.

[11] S. Belfiore, M. Grangetto, E. Magli, and G. Olmo, "Spatio-temporal error concealment with optimized mode selection and application to H.264," *Signal Process.: Image Commun.*, vol. 18, pp. 907–923, Nov. 2003.

[12] S. Wenger, "H.264/AVC over IP," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 645–656, July 2003.

[13] S. Belfiore, M. Grangetto, E. Magli, and G. Olmo, "Concealment of whole-frame losses for wireless low bit-rate video based on multiframe optical flow estimation," *IEEE Trans. Multimedia*, vol. 7, no. 2, pp. 316–329, Apr. 2003.

[14] Q. Peng, T. W. Yang, and C. Q. Zhu, "Block-based temporal error concealment for video packet using motion vector extrapolation," in *Proc. IEEE Commun., Circuits and Syst. and West Sino Expositions*, July 2002, vol. 1, pp. 10–14.

[15] M. Flierl and B. Girod, "Generalized B pictures and the draft H.264/AVC video compression standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 587–597, July 2003.

[16] A. M. Tourapis, F. Wu, and S. Li, "Direct mode coding for bipredictive slices in the H.264 standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 1, pp. 119–126, Jan. 2005.

[17] Y. Chen, J. Boyce, and K. Xie, "Frame loss error concealment for SVC," *Joint Video Team(JVT), Doc. JVT-Q046*, Oct. 2005.

[18] Y. Chen, K. Xie, F. Zhang, P. Purvin, and J. Boyce, "Frame loss error concealment for SVC," *J. Zhejiang Univ.*, vol. 7, pp. 677–683, Apr. 2006.

[19] J. Zheng, X. Ji, G. Ni, W. Gao, and F. Wu, "Extended direct mode for hierarchical B picture coding," in *Proc. IEEE Int. Conf. Image Process.*, Sept. 2005, vol. 2, pp. 265–268.

[20] H. Schwarz, D. Marpe, and T. Wiegand, "Hierarchical B pictures," *Joint Video Team(JVT), Doc. JVT-P014*, Jul. 2005.

[21] H. Schwarz, D. Marpe, and T. Wiegand, "Analysis of hierarchical B-pictures and MCTF," in *Proc. IEEE Int. Conf. Multimedia and Expo.*, Jul. 2006.

[22] I. V. Bajic and J. W. Woods, "Error concealment for scalable motion-compensated subband/wavelet video coders," *IEEE Trans. Circuits Syst. Video Technol.*, no. 4, pp. 508–514, Apr. 2007.

[23] Joint Scalable Video Model. Klagenfurt, Austria, July 2006, ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, Document JVT-T202.

[24] S. Wenger, "Error patterns for internet experiments," in *ITU Telecommunications Standardization Sector, Doc. Q15-1-16rl*, Oct. 1999.

[25] Z. He, J. Cai, and C. W. Chen, "Joint source channel rate-distortion analysis for adaptive mode selection and rate control in wireless video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 6, pp. 511–523, Jun. 2002.

Xiangyang Ji received the B.S. and M.S. degrees in computer science from Harbin Institute of Technology, Harbin, China, in 1999, and 2001, respectively, and the Ph.D. degree in computer science from Institute of Computing Technology, and the Graduate School of Chinese Academy of Science, Beijing, in 2008. He has authored or co-authored over 40 conference and journal papers. His research interests include video/image coding, video streaming, and multimedia processing.

Debin Zhao received the B.S., M.S. and Ph.D. degrees in computer science, all from Harbin Institute of Technology (HIT), Harbin, China, in 1985, 1988 and 1998, respectively.

He joined the Department of Computer Science of HIT as an associate professor in 1993. He is now Professor at HIT and the Institute of Computing Technology, Chinese Academy of Sciences. He has been a Research Fellow in Department of Computer Science, City University of HK. His research interests include multimedia compression and its related applications. He has authored or co-authored over 100 publications.

Dr. Zhao has received the National Science and Technology Progress Award of China (Second Prize) three times. He also received an Excellent Teaching Award from Baogang Foundation.

Wen Gao (M'99–SM'05) received the M.S. and the Ph.D. degrees in computer science from Harbin Institute of Technology, Harbin, China, in 1985 and in 1988, respectively, and the Ph.D. degree in electronics engineering from University of Tokyo, Tokyo, Japan, in 1991.

He was a Research Fellow with the Institute of Medical Electronics Engineering, University of Tokyo, in 1992, and a Visiting Professor at Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, in 1993. From 1994 to 1995, he was a Visiting Professor with the Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge. Currently, he is a Professor with the School of Electronic Engineering and Computer Science, Peking University, Peking, China, and a Professor of computer science at Harbin Institute of Technology. He is also the Honor Professor in computer science at City University of Hong Kong, and the External Fellow of International Computer Science Institute, University of California, Berkeley. He has published seven books and over 200 scientific papers. His research interests are in the areas of signal processing, image and video communication, computer vision, and artificial intelligence.

Dr. Gao is Editor-in-Chief of the *Chinese Journal of Computers*. He chairs the Audio Video coding Standard (AVS) workgroup of China. He is the head of Chinese National Delegation to MPEG working group (ISO/SC29/WG11).