Hindawi Publishing Corporation
EURASIP Journal on Image and Video Processing
Volume 2009, Article ID 683510, 15 pages
doi:10.1155/2009/683510

Research Article

Improved Side Information Generation for Distributed Video Coding by Exploiting Spatial and Temporal Correlations

Shuiming Ye, Mourad Ouaret, Frederic Dufaux, and Touradj Ebrahimi (EURASIP Member)

Institute of Electrical Engineering, Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland

Correspondence should be addressed to Shuiming Ye, shuiming@gmail.com

Received 22 May 2008; Revised 15 October 2008; Accepted 14 December 2008

Recommended by Stefano Tubaro

Distributed video coding (DVC) is a video coding paradigm allowing low complexity encoding for emerging applications such as wireless video surveillance. Side information (SI) generation is a key function in the DVC decoder, and plays a key-role in determining the performance of the codec. This paper proposes an improved SI generation for DVC, which exploits both spatial and temporal correlations in the sequences. Partially decoded Wyner-Ziv (WZ) frames, based on initial SI by motion compensated temporal interpolation, are exploited to improve the performance of the whole SI generation. More specifically, an enhanced temporal frame interpolation is proposed, including motion vector refinement and smoothing, optimal compensation mode selection, and a new matching criterion for motion estimation. The improved SI technique is also applied to a new hybrid spatial and temporal error concealment scheme to conceal errors in WZ frames. Simulation results show that the proposed scheme can achieve up to 1.0 dB improvement in rate distortion performance in WZ frames for video with high motion, when compared to state-of-the-art DVC. In addition, both the objective and perceptual qualities of the corrupted sequences are significantly improved by the proposed hybrid error concealment scheme, outperforming both spatial and temporal concealments alone.

1. Introduction

Nowadays, the most popular digital video coding solutions are represented by the ISO/IEC MPEG and ITU-T H.26x standards [1], which rely on a highly complex encoder. However, in some emerging applications, such as wireless low-power video surveillance, multimedia sensor networks, wireless PC cameras, and mobile camera phone, low complexity encoding is required. Distributed video coding (DVC) [2], a new paradigm in coding which allows for very low complexity encoding, is well suited for these applications.

In DVC, the complex task of exploiting the source statistics, that is, the motion estimation can be moved from the encoder to the decoder. The Slepian-Wolf theorem on lossless distributed source coding states that the optimal rate of joint encoding and decoding of two statistically dependent discrete signals can be achieved by using two independent encoders and a joint decoder [3]. Wyner-Ziv coding extends this result to lossy coding with side information (SI) in the case of Gaussian memoryless sources and mean-squared error distortion [4]. DVC generally divides a video sequence into key frames and WZ frames. The key task to exploit source statistics is carried out in SI generation process to produce an estimation of the WZ frame being decoded. SI has a significant influence on the rate distortion (RD) performance of DVC. Indeed, more accurate SI at the decoder implies that fewer bits are requested from the encoder through a feedback channel, so that the bitrate is reduced for the same quality. In common DVC codecs, the SI is obtained by motion compensated temporal interpolation (MCTI) from the previous and next key frames and utilizes the block matching algorithm (BMA) for motion estimation. However, motion vectors from BMA are often not faithful to true object motions. Unlike classical video compression, it is more important to find true motion vectors for SI generation in DVC. Therefore, it is important to improve the SI generation in DVC in order to achieve better RD performance.

Another appealing property of DVC is its good resilience to transmission errors due to its intrinsic joint source-channel coding framework. A thorough analysis of

its performance in the presence of transmission errors has been presented in [5], showing its good error resilience properties. This results from the fact that DVC is based on a statistical framework rather than the closed-loop prediction used in conventional video coding. Recently, the rapid growth of Internet and wireless communications has led to increased interest for robust transmission of compressed video. However, transmission errors may severely impact video quality as compressed data is very sensitive to these errors [6]. Thus, error control techniques are necessary for efficient video transmission over error prone channels.

This paper proposes a new SI generation scheme by exploiting spatio-temporal correlations at the decoder. It uses partially decoded WZ frames generated by the WZ decoder to improve SI generation. In other words, the proposed scheme is not only based on the key frames, but also on the WZ bits already decoded. Furthermore, enhanced temporal frame interpolation is applied, including motion vector refinement and smoothing to re-estimate and filter the motion vectors, and optimal compensation mode selection to select the mode with minimum matching distortion. Based on these techniques, we also propose a new hybrid spatial and temporal error concealment (EC) scheme for WZ frames in DVC. It uses the error-concealed results from spatial EC to improve the performance of the temporal EC, instead of simply switching between spatial and temporal EC. Spatial EC based on the edge-directed filter [7] is firstly applied to the corrupted blocks, and the results are used as partially decoded WZ frames to improve the performance of temporal EC. In other words, the temporal EC is not only based on the key frames, but also on the WZ bits already decoded. Experimental results show that the proposed scheme significantly improves the quality of the SI and RD performance of DVC, and the performance of the proposed hybrid scheme is superior to spatial EC and temporal EC alone.

This paper is organized as follows. First, the DVC architecture and other related work are introduced in Section 2. Then, the proposed SI generation scheme is presented in Section 3. Section 4 introduces a new hybrid spatio-temporal EC based on the improved SI generation technique. Simulation results are presented in Section 5. Finally, Section 6 concludes the paper.

2. Related Work

2.1. DVC Architecture. Without loss of generality, in this paper, we consider the transform domain Wyner-Ziv (TDWZ) DVC architecture from [8], as shown in Figure 1. A video sequence is divided into key frames (Y) and WZ frames (X). Hereafter, we consider a Group of Pictures (GOPs) size of 2, namely, the odd and even frames are key frames and WZ frames, respectively. Key frames Y are conventionally encoded using H.264/AVC Intra coding [1]. Conversely, for WZ frames X, a DCT transform is firstly applied to the input stream, and the resulting transform coefficients undergo quantization. The quantized coefficients are then split into bitplanes which are turbo encoded. At the decoder, SI approximating the WZ frames is generated by MCTI of the decoded key frames. The SI is used in the turbo decoder, along with WZ parity bits requested from feedback channel, in order to reconstruct the decoded WZ frames X'. In this paper, the turbo decoder stops requesting more bits if the bitplane bit error rate is below a given threshold equal to 10^{-3}.

2.2. Motion-Compensated Temporal Interpolation. Motion-compensated temporal interpolation (MCTI) has been used in almost the all DVC codecs to generate SI by interpolating the current frame from key frames. The purpose of MCTI is to create an interpolation of a particular frame by using blocks from previous and next reference frames. This problem is similar to video frame rate upconversion interpolation to improve temporal resolution at the decoder [9, 10]. In contrast to the motion compensation (MC) technique used in conventional codecs, MCTI has no knowledge about the frame being decoded. To estimate frame k, bidirectional motion estimation is generally used in MCTI using a bidirectional motion estimation scheme similar to the B-frames coding mode used in current video standards. For every block in frame $k-1$, the most similar block in frame $k+1$ is found, and its motion vector is calculated. Once the motion vector is obtained, the interpolated frame can be filled by simply using bidirectional motion compensation. Due to block-matching techniques not being ideal, forward and backward searches usually do not produce the same results, and they need to be averaged. This scheme holds as long as the block has constant velocity. However, when there is large or asymmetric motion, MCTI fails to generate a good SI estimate.

Spatial motion vector smoothing was proposed to improve the performance of bidirectional MCTI in [8, 11]. It is observed that the motion vectors have sometimes low spatial coherence [11]. Therefore, spatial smoothing filter was proposed to improve motion estimation by reducing the number of false motion vectors, that is, incorrect motion vectors when compared to the true motion field. This scheme uses weighted vector median filters, which maintains the motion field spatial coherence by looking, at each block, for candidate motion vectors at neighboring blocks. This filter is also adjusted by a set of weights controlling the filter smoothing strength depending on the prediction mean square error of the block for each candidate motion vector. However, spatial motion smoothing is only effective at removing false vectors that occur as isolated pulse spikes.

Subpixel interpolation is also proposed to improve motion estimation for SI generation [12]. The subpixel interpolation method of H.264 is used to generate the pixel value at subpixel position. At the decoder, the side information is motion-compensated refined according to the chosen multiestimation mode from backward, forward, and bidirectional modes. This motion refinement procedure uses subpixel interpolation to improve the precision of search. The subpixel interpolation is effective at improving the generated SI, but it also fails in the presence of large or asymmetric motion. Moreover, it increases the decoder complexity.

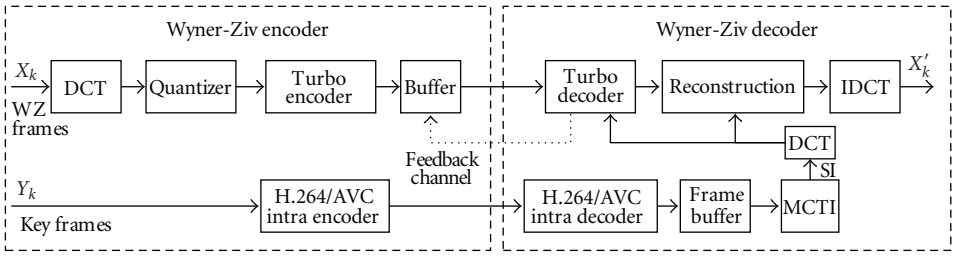


Figure 1: DVC architecture.

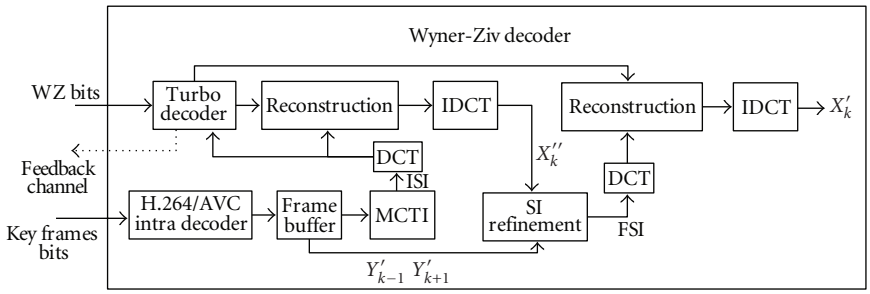


Figure 2: DVC decoder architecture with proposed SI generation.

2.3. Encoder Aided Motion Estimation for SI Generation. Encoder aided motion estimation to improve SI generation was proposed to conduct more accurate motion estimation at the decoder with the help of auxiliary information sent by the encoder, such as the cyclic redundancy check (CRC) [13] and hash bits [14]. In [13], CRC bits for every block are calculated at the encoder and transmitted for the decoder to perform motion search and choose the candidate block that produces the same CRC. The encoder transmits a CRC check of the quantized sequence. Motion estimation is carried out at the decoder by searching over the space of candidate predictors one-by-one to decode a sequence from the set labeled by the syndrome. When the decoded sequence matches the CRC check, decoding is declared to be successful. However, the way to generate and exploit CRC is complicated, and it increases the complexity not only at the decoder, but also at the encoder. In [14], it is proposed to send robust hash codewords from the encoder, in addition to the Wyner-Ziv bits, to aid the decoder in estimating the motion and generating the SI. These hash bits carry the motion information to the decoder without actually estimating the motion at the encoder. The robust hash code for a block simply consists of a very coarsely subsampled and quantized version of the block. The decoder performs a motion search based on the hash to generate the best SI block from the previous frame. In this scheme, the encoder is no longer an intraframe coder because of the hash store. The hash bits do help in motion estimation, but they increase the encoder complexity and transmission payload. In addition, in [14], the SI is generated only based on the previous key frame, which is not as good as bidirectional motion estimation.

It was also proposed to split the Wyner-Ziv frame into two subsets at the encoder based on a checkerboard pattern, in order to exploit spatial correlations between these subsets at the decoder [15]. Each subset is encoded independently. At the decoder, the first subset is decoded using the SI obtained by MCTI, thus exploiting only temporal correlation. Then, the second subset is decoded by MCTI based on key frames, or by interpolating the first decoded subset. When the estimated temporal correlation is high, the temporal SI is used. Otherwise, the spatial SI is used. However, this approach can only achieve a modest improvement, and the encoder must be modified accordingly.

2.4. Iterative Decoding and Motion Estimation for SI Generation. The idea of iterative decoding and motion estimation has also been proposed to improve SI, such as motion vector refinement via bitplane refinement [16] and iterative MCTI techniques [17, 18], but with a high cost of several iterations of motion estimation and decoding. In [16], the reconstructed image and the adjacent key frames are used in order to refine the motion vectors and, thus, obtain a new and improved version of the decoded and SI frames, including matching criteria function to perform motion estimation and three decoding interpolation modes to select the best reference frame. This scheme is based on bitplane refinement for pixel-domain DVC, and only minor improvements have been achieved. In [17], the first outcome of the distributed decoder is called partially decoded picture. A second motion-compensated interpolation is applied, which uses the partially decoded picture as well as the previous and next key frames. For each aligned block in the partially decoded picture, the most similar block is searched in the previous frame, the next key frames, the motion-compensated average of the previous and the next frames, and the result from the MCTI previously performed. However, only minor improvement has been achieved. In

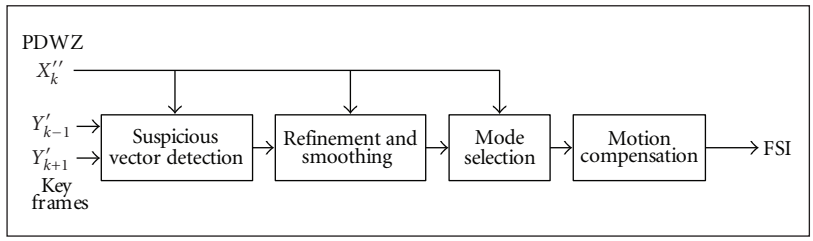


Figure 3: Proposed SI refinement procedure.

this paper, we use the same idea of using partially decoded picture, but it is further augmented with suspicious vector detection, new matching criteria for motion estimation, and motion vector filtering, resulting in much better improvements. An iterative approach based on multiple SI with motion refinement has also been proposed in [18]. Multiple SI streams are used at the decoder: the first SI stream is predicted by motion extrapolating the previous two closest key frames and the second SI stream is predicted using the immediate key frame and the closest Wyner-Ziv frame. Based on the error probability, the turbo decoder decides which SI stream is used for decoding a given block.

2.5. Error Concealment. EC consists in estimating or interpolating corrupted data at the decoder from the correctly received information. It can improve the quality of decoded video corrupted by transmission errors, without any additional payload on the encoder or channel. EC can be classified into three categories: spatial concealment [19–21], temporal concealment [22–24], and hybrid spatial and temporal concealments [25–27].

Spatial EC is used to interpolate the lost block from its spatially neighboring available blocks or coefficients in the current frame. It relies on the inherent spatial smoothness of the data. For example, the technique proposed in [19] exploits the smoothness property of image signals and recovers the damaged blocks by a smoothness measure based on second-order derivatives. However, this smoothness measure would lead to blurred edges in the recovered frame due to the simple second-order derivative-based measure to represent the edges. In [28], through benchmarking results on the existing error concealment approaches, it was observed that none of the existing approaches is an all time champion. A classification-based concealment approach was proposed which can combine the better performance of different spatial approaches [29].

Temporal EC techniques use the temporally neighboring frames to estimate the lost blocks in the current frame, based on the assumption that the video content is smooth and continuous in the temporal domain. A very simple scheme for temporal error concealment is used to just copy the block at the same spatial location in the previous frame to conceal the lost block. A bidirectional temporal error concealment algorithm that can recover loss of a whole frame was proposed in [23]. However, the accuracy of the motion estimation may affect the results significantly.

Temporal EC usually leads to better results when compared to spatial concealment, given the typically high temporal correlation in video. However, for video with scene changes or with very large or irregular motion, spatial EC is preferred. Some attempts have been made to combine both spatial and temporal ECs to improve performance [26, 27]. These schemes use some mode selection methods to decide whether to use spatial or temporal EC. For example, temporal activity (measured as the prediction error in the surrounding blocks) and spatial activity (measured as the variance of the same surrounding blocks) are used to decide which concealment mode to use [26]. In general, however, these methods have achieved very limited success, mostly due to the simple mode selection mechanisms at the decoder to merge the results from both spatial and temporal ECs.

In [30], a forward error correcting coding scheme is proposed for traditional video coding, where an auxiliary redundant bitstream generated at the encoder using Wyner-Ziv coding is sent to the decoder for error concealment. However, in the literature, few error concealment schemes for DVC can be found.

3. Proposed SI Generation Scheme

The DVC decoder architecture including the proposed SI generation scheme is illustrated in Figure 2. Firstly, the MCTI with spatial motion smoothing from [8] is used to compute motion vectors and to estimate the initial SI (ISI) for the frame being decoded. Based on the ISI, the WZ decoder is first applied to generate a partially decoded WZ (PDWZ) frame denoted by X_k''. The PDWZ frame, that is, the decoded result after the first run of WZ decoding, is then exploited to generate an improved SI as detailed in Figure 3. More specifically, the SI refinement procedure first detects suspicious motion vectors based on the matching errors between the PDWZ frame and the reference key frames. These motion vectors are then refined using a new matching criterion and a spatial smoothing filter. Furthermore, optimal motion compensation mode selection is conducted. Namely, based on the spatio-temporal correlations between the PDWZ frame and the reference key frames, the interpolated block can be selected from a number of sources: the previous frame, the next frame, and the bidirectional motion-compensated average of the previous and the next frame. The final SI (FSI) is constructed using motion compensation based on the refined motion vectors and the optimal compensation mode. Finally, based on the

MV_1	MV_2	MV_3
MV_4	MV_c	MV_5
MV_6	MV_7	MV_8

Figure 4: Neighboring motion vectors for weighted median vector filter.

FSI, the reconstruction step is performed again to get the final decoded WZ frame.

Common MCTI techniques use only the previous and next key frames to generate the SI. In comparison, the proposed SI generation scheme appears to perform much better than common MCTI, since it has additional information (from WZ bits) about the frame it is trying to estimate. Moreover, the spatio-temporal correlations are exploited based on the PDWZ frame using the SI refinement procedure. The decoded frame obtained here could then be used again as PDWZ frame for a subsequent iteration. However, our experiments show that extra iterations do not provide a significant performance improvement. In other words, the additional information carried by parity bits is fully exploited in a single run of our SI generation scheme. Therefore, only one iteration is used in the proposed scheme, avoiding additional complexity at the decoder.

3.1. Matching Criterion. To exploit the spatio-temporal correlations between the PDWZ frame and reference key frames, a new matching criterion is used to evaluate the errors in motion estimation. Generally, the goal of motion estimation is to minimize a cost function that measures the prediction error, that is, how similar the original blocks and the estimated block are. For example, the popular mean absolute difference (*MAD*) for the estimated motion vector MV of the block B_1 is defined as

$$\begin{aligned} &MAD(P_0, F_1, F_2, MV) \\ &\quad = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \left| F_1(i + x_0, j + y_0) \right. \\ &\qquad\qquad \left. - F_2(i + x_0 + MV_x, j + y_0 + MY_y) \right|, \end{aligned} \tag{1}$$

where (x_0, y_0) is the coordinate of the top left point P_0 of the original block in the current frame F_1, F_2 is the reference frame, (MV_x, MY_y) is the candidate motion vector MV, and (M, N) are the dimensions of the block. However, when there are changes in pixel intensity and noises, minimizing *MAD* often leads to false motion vectors.

On the other hand, boundary absolute difference (*BAD*) is proposed in the error concealment literature [9, 24] to measure the accuracy of motion compensation to enforce the spatial smoothness property by minimizing the side matching distortion between the internal and external borders of the recovered block. It is defined as

$$\begin{aligned} &BAD(P_0, F_1, F_2, MV) \\ &\quad = \frac{1}{M} \sum_{i=0}^{M-1} \left| F_1(i + x_0, y_0) \right. \\ &\qquad\qquad \left. - F_2(i + x_0 + MV_x, y_0 + MV_y - 1) \right| \\ &\qquad + \frac{1}{M} \sum_{i=0}^{M-1} \left| F_1(i + x_0, y_0 + N - 1) \right. \\ &\qquad\qquad \left. - F_2(i + x_0 + MV_x, y_0 + MV_y + N) \right| \\ &\qquad + \frac{1}{N} \sum_{j=0}^{N-1} \left| F_1(x_0, j + y_0) \right. \\ &\qquad\qquad \left. - F_2(x_0 + MV_x - 1, j + y_0 + MV_y) \right| \\ &\qquad + \frac{1}{N} \sum_{j=0}^{N-1} \left| F_1(x_0 + M - 1, j + y_0) \right. \\ &\qquad\qquad \left. - F_2(x_0 + MV_x + M, j + y_0 + MV_y) \right|. \end{aligned} \tag{2}$$

Unfortunately, *BAD* is not efficient at picking out bad motion vectors when local variation is large [9].

In this paper, we propose a new matching criterion based on *MAD* and *BAD*. The matching distortion (D_{ST}) for the motion vector (MV) of the current block with upper-left point P_0 is defined as

$$\begin{aligned} D_{ST}(P_0, F_1, F_2, MV) &= \alpha\, BAD\,(P_0, F_1, F_2, MV) \\ &\quad + (1 - \alpha) MAD\,(P_0, F_1, F_2, MV), \end{aligned} \tag{3}$$

where α is a weighting factor, and MV is the candidate motion vector. *MAD* is utilized to measure how well the candidate MV can keep temporal continuity. The smaller *MAD* is, the better the candidate MV keeps temporal continuity. On the other hand, *BAD* is used to measure how well the candidate MV can keep spatial continuity. The smaller *BAD* is, the better the candidate MV keeps spatial continuity.

This matching criterion is exploited in suspicious vector detection, motion vector refinement and smoothing, and optimal motion compensation mode selection in the proposed SI generation pipeline.

3.2. Suspicious Vector Detection. Generally, for most sequences with low and smooth motion, the majority of motion vectors estimated by MCTI are close to the true motion. However, erroneous vectors may result in serious block artifacts if they are directly used in frame interpolation. In this paper, a threshold T is established to define the candidate blocks for further refinement based on the matching criterion D_{ST}. If an estimated MV satisfies the criteria defined in (4), it is considered to be a good estimation; otherwise, it is identified as a suspicious vector and will be further processed as follows:

$$D_{ST}(P_0, X_k'', Y_{k-1}', MV) + D_{ST}(P_0, X_k'', Y_{k+1}', MV) < T, \tag{4}$$

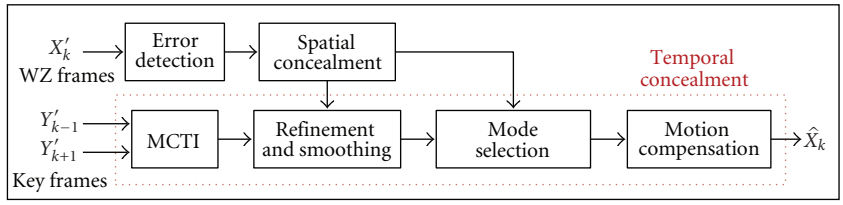


Figure 5: Proposed spatio-temporal error concealment.

(a) Original frame (b) Errors in WZ frame (c) Errors in H.264 intra-coded frame

Figure 6: Errors in WZ frame (*Foreman*, frame 54).

where Y'_{k-1} and Y'_{k+1} are the previous and next decoded key frames, respectively, and X''_k is the PDWZ frame.

3.3. Motion Vector Refinement and Smoothing. The spatio-temporal correlations between the PDWZ frame and the reference key frames are exploited to refine and smooth the estimated motion vectors. More specifically, the motion vectors are re-estimated by bidirectional motion estimation using the matching criterion defined in (3) and the PDWZ frame. They are then filtered using a spatial smoothing filter. This process generates a new estimation of the motion vector for the block to be interpolated.

It is observed that motion vectors have sometimes low spatial coherence. A spatial motion smoothing filter is therefore used, similar to [11], but with the matching criterion defined in (3) and the PDWZ frame. More precisely, a weighted vector median filter is used to maintain the motion field spatial coherence. This filter is adjusted by a set of weights controlling the smoothing strength. The weighted vector median filter is defined as

$$MV_F = \arg\min_{MV_i} \sum_{j=1}^{Num} w_j ||MV_i - MV_j||, \quad i \in [1, Num], \tag{5}$$

where $MV_1, \ldots, MV_{Num}$ are the motion vectors of the corresponding nearest neighboring blocks. MV_F is the motion vector output of the weighted vector median filter, which is chosen in order to minimize the sum of distances (L^2-norm used in this paper) to the other $Num - 1$ vectors. 8-neighborhood is used in this paper ($Num = 8$), as shown in Figure 4. The weights $w_1, \ldots, w_{Num}$ are calculated based on the new matching criterion and the PDWZ frame as follows:

$$w_j = \frac{D_{ST}(P_0, X''_k, Y'_{k-1}, MV_c) + D_{ST}(P_0, X''_k, Y'_{k+1}, MV_c)}{D_{ST}(P_0, X''_k, Y'_{k-1}, MV_j) + D_{ST}(P_0, X''_k, Y'_{k+1}, MV_j)}, \tag{6}$$

where MV_c is the current estimated vector for the block to be smoothed. The weight is small if there is a high prediction error using MV_j, that is, the median filter is to substitute the previously estimated motion vector with a neighboring vector which has the smallest prediction error.

3.4. Optimal Motion Compensation Mode Selection. The objective of this step is to generate an optimal motion-compensated estimate. In most DVC schemes, while bidirectional prediction is shown to be effective, it is limited to motion-compensated average of the previous and the next key frames.

Based on the PDWZ frame, the most similar block to the current block can be selected from three sources: the previous frame, the next frame, and the bidirectional motion-compensated average of the previous and the next frames. More specifically, the block is estimated by selecting the mode with minimum matching error from the following three modes of motion compensation.

(i) Backward mode: the block in the SI is interpolated using only one block from the previous key frame.

(ii) Forward mode: the block in the SI is interpolated using only one block from the next key frame.

(iii) Bidirectional mode: the block in the SI is interpolated using the average of one block in the next key frame

(a) Original (b) SI (TDWZ, 20.3 dB) (c) SI (proposed, 26.6 dB)

(d) Decoded (TDWZ, 27.5 dB) (e) Decoded (proposed, 29.8 dB)

Figure 7: Visual result comparisons.

and another block in the previous key frame, at arbitrary positions.

Among these modes, the decision is performed according to the matching criterion defined in (3), and the one with the minimum matching error is retained.

Based on the refined motion vectors and the selected interpolation mode, motion compensation is applied to generate the final SI. Based on this SI, the final decoded frame X'_k is got after running the WZ decoder again.

4. Application of Improved SI Generation Technique to EC

The techniques used to improve SI generation for DVC not only can improve the performance of DVC, but also are useful to improve the error resilience of DVC when applied to a hybrid error concealment scheme. A hybrid error concealment scheme is proposed based on the improved SI generation techniques, as illustrated in Figure 5. The error location is firstly detected. In this paper, we assume that the error locations are known at the decoder, as often presumed in error concealment literature, which can be done at transport level or based on syntax and watermarking [6]. For example, the UDP protocol generally used for video streaming provides the parity check information. If an error is detected, the entire packet is discarded and an error is reported. Spatial EC is then applied to obtain a partially error-concealed frame. This frame is much closer to the error-free frame than the corrupted one. The partially error-concealed frame is used for motion vector refinement,

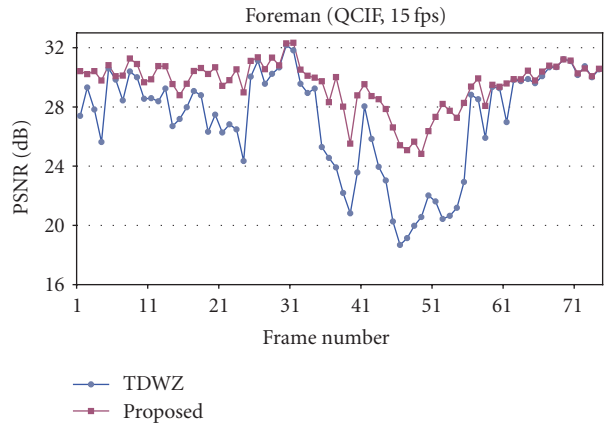


Figure 8: PSNR of SI for *Foreman* frames.

smoothing, and optimal compensation mode selection to obtain an estimate of the motion vector of the corrupted block. Motion compensation is finally used to obtain the final error-concealed frame $\hat{X}_k$.

4.1. Spatial Concealment Based on Edge Directed Filter. In DVC, the decoded WZ frames are based on the SI generated by MCTI of the key frames. WZ bits are then used to improve the quality of the approximate estimation of SI, and to obtain the decoded WZ frame. Motion estimation for SI would be more correct for smooth areas than edges, that is, less WZ bits are used for smooth areas. Therefore, the transmission errors in WZ bits tend to cause noises around edges in the

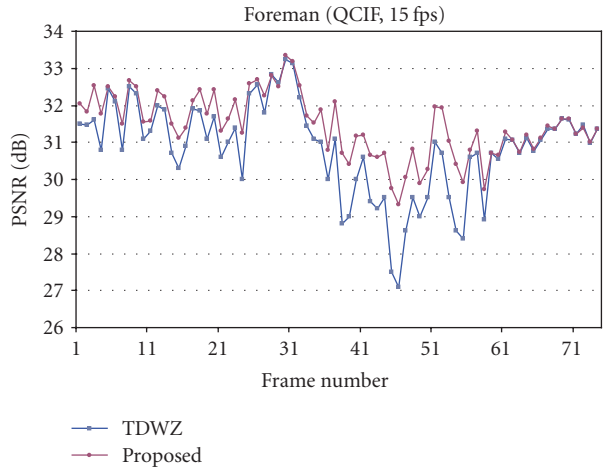


Figure 9: PSNR of decoded *Foreman* frames.

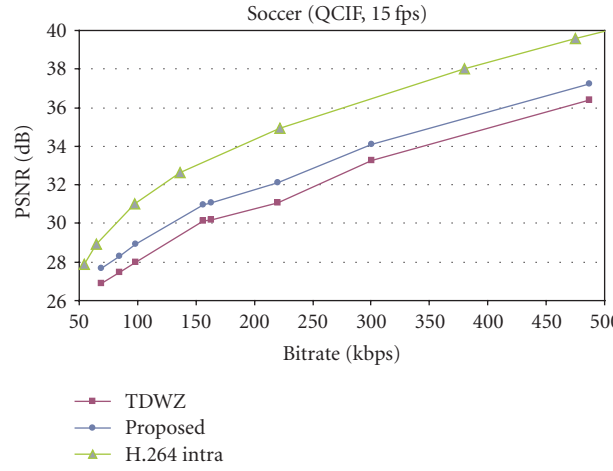


Figure 11: RD performance for sequence *Soccer*.

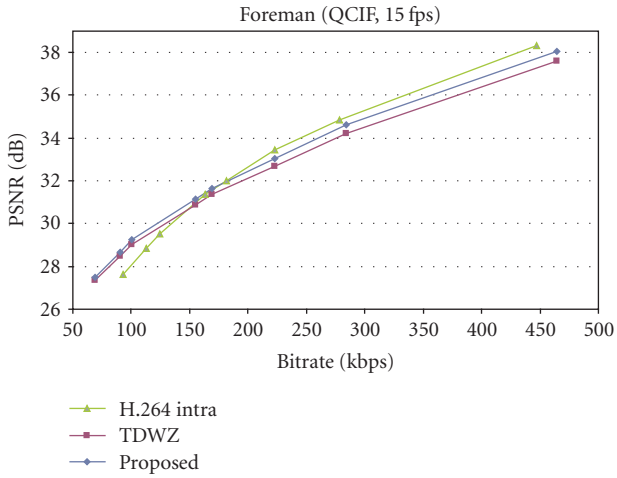


Figure 10: RD performance for sequence *Foreman*.

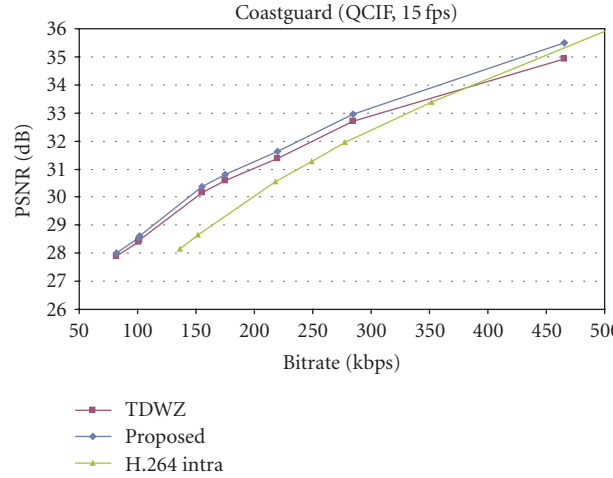


Figure 12: RD performance for sequence *Coastguard*.

corrupted WZ frames. For example, when there are errors in WZ bits, the error pattern of the damaged WZ frames, as shown in Figure 6(b), is different from that of traditional video coding schemes (Figure 6(c)). Therefore, the error concealment schemes proposed for traditional video coding schemes cannot be directly applied to conceal errors in WZ frames. In this paper, since errors in WZ bits tend to cause artifacts around edges, an edge-directed filter is constructed to remove the noises without serious blurriness.

Anisotropic diffusion techniques have been widely used in image processing for their efficiency at smoothing noisy images while preserving sharp edges. We adopt the anisotropic diffusion as a direction diffusion operation and use the diffusion function for spatial error concealment as in [7]. The error concealment method in [7] is designed for wavelet-based images and contains wavelet domain constraints and rectifications. In this paper, we only use the edged directed filter without any constraint or rectification. Based on the error patterns generated by WZ corrupted frames, an edge-directed filter is constructed to remove the noises around edges caused by errors in the WZ bits by adopting the anisotropic diffusion as a direction diffusion operation and the diffusion function for spatial error concealment proposed in [7]

$$f(\nabla I) = \frac{\exp\left(-|\nabla I|/M\right)}{\max\left(\exp(\Delta I), 1 + |\nabla I|\right)},$$
$$M = \max_{P \in \Gamma}\left(|\nabla I_P|\right), \tag{7}$$

where Γ is the 16×16 pixels blocks where the corrupted pixel belongs to, ∇ is the gradient operator, and $|\nabla I|$ is the magnitude of ∇I.ΔI is the Laplacian of the frame I, a second-order derivative of I.

The edge-directed filter is applied iteratively as follows:

$$I^{n+1} = I^n + \frac{\Delta t}{N}\sum_{i=1}^{N} f\left(\nabla I_i^n\right) \cdot \nabla I_i^n, \tag{8}$$

where I^{n+1} is the recovered frame after $n+1$ iterations, I^0 is the corrupted frame, and Δt is the anisotropic diffusion step. For each pixel (I_i^n), the filtering is carried out on the neighboring N (16×16) pixels.

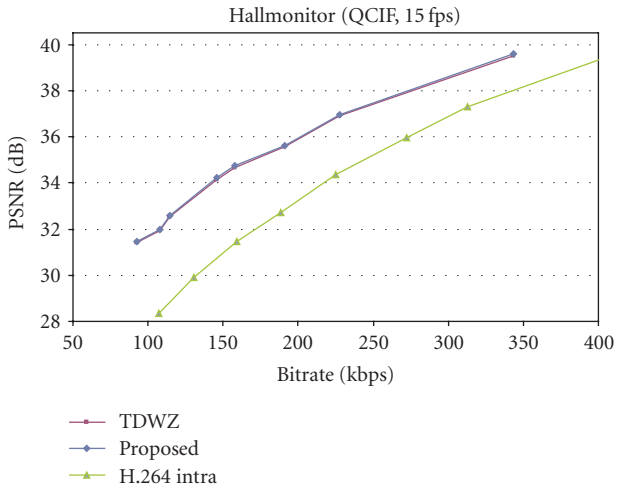


Figure 13: RD performance for sequence *Hallmonitor*.

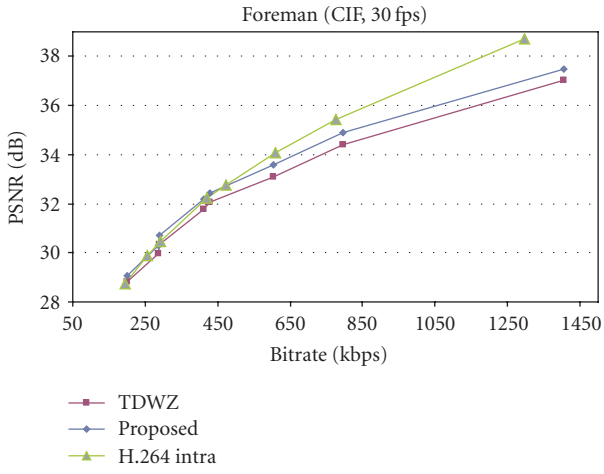


Figure 14: RD performance for sequence *Foreman* (CIF).

4.2. Enhanced Temporal Error Concealment. The techniques used in the improved SI generation scheme, as described in Section 3, are exploited to improve the performance of the temporal EC. The approach is based on MCTI and motion vector filtering as proposed in [11]. One of the key novelties is that the partially error-concealed frame is used to improve the temporal EC, unlike [11], where MCTI is based on the previous and next key frames. Indeed, the reconstructed frame by spatial concealment contains additional information about the current frame carried by the correctly received WZ bits. Therefore, by using the partially error-concealed frame resulting from spatial EC, the spatio-temporal correlations between this frame and the reference key frames can be better exploited. Hence, the performance of the temporal EC is improved.

The matching criterion in (3) is used to evaluate the error in motion estimation based on the partially error-concealed frame and the reference key frames. The spatio-temporal correlations between the partially error-concealed frames and the key frames are then exploited to refine and smooth the estimated motion vectors using the technique presented in Section 3.3. The block most similar to the corrupted block is selected from a number of sources: the previous frame, the next frame, the bidirectional motion-compensated average of the previous and the next frames, as presented in Section 3.4. Based on the estimated motion vectors and the interpolation modes, motion compensation is applied to generate the reconstructed blocks as the result of temporal concealment.

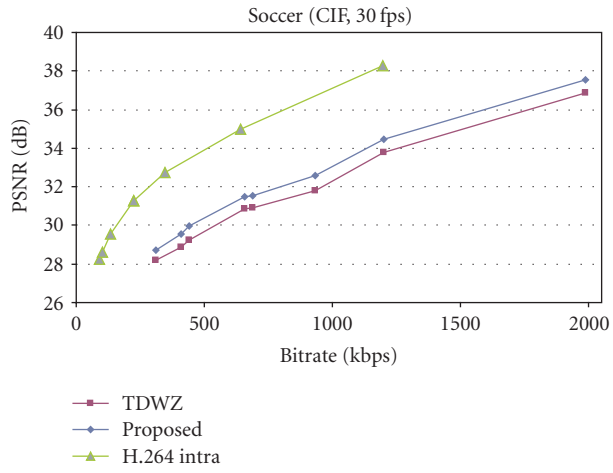


Figure 15: RD performance for sequence *Soccer* (CIF).

5. Results and Discussions

The TDWZ DVC codec proposed in [8] is used in our experiments, and only luminance data is coded. The video sequences *Foreman, Soccer, Coastguard*, and *Hallmonitor* are used in QCIF format and at 15 fps. The DVC codec is run for the first 149 frames. Eight RD points are computed per sequence. The results are compared to the TDWZ codec [8]. The quantization parameters of the key frames in our experiments were selected in such a way that the qualities of the output key frames are kept similar with those of the WZ frames. The weight α in (3) and the threshold T in (4) are empirically set to 0.3 and 10, respectively.

5.1. Performance Improvement by the Proposed SI Generation Method. Figure 7 shows the visual results of SI and decoded frames for *Foreman*. The face and the building in the SI generated by the TDWZ contain block artifacts (Figure 7(b)). On the contrary, the SI generated by the proposed method (Figure 7(c)) is much better, with the same requested bits (20.6 Kbits). The improvement in the SI also results in a better quality of the decoded WZ frame. There are much fewer block artifacts on the face and building in the decoded frame (Figure 7(e)) when compared to the proposed method by the TDWZ (Figure 7(d)).

Figure 8 shows the SI quality for *Foreman*. The proposed algorithm achieves up to 6.7 dB and an average of 2.4 dB improvement, when compared to the SI in the TDWZ. The PSNR values of the decoded WZ frames are shown in

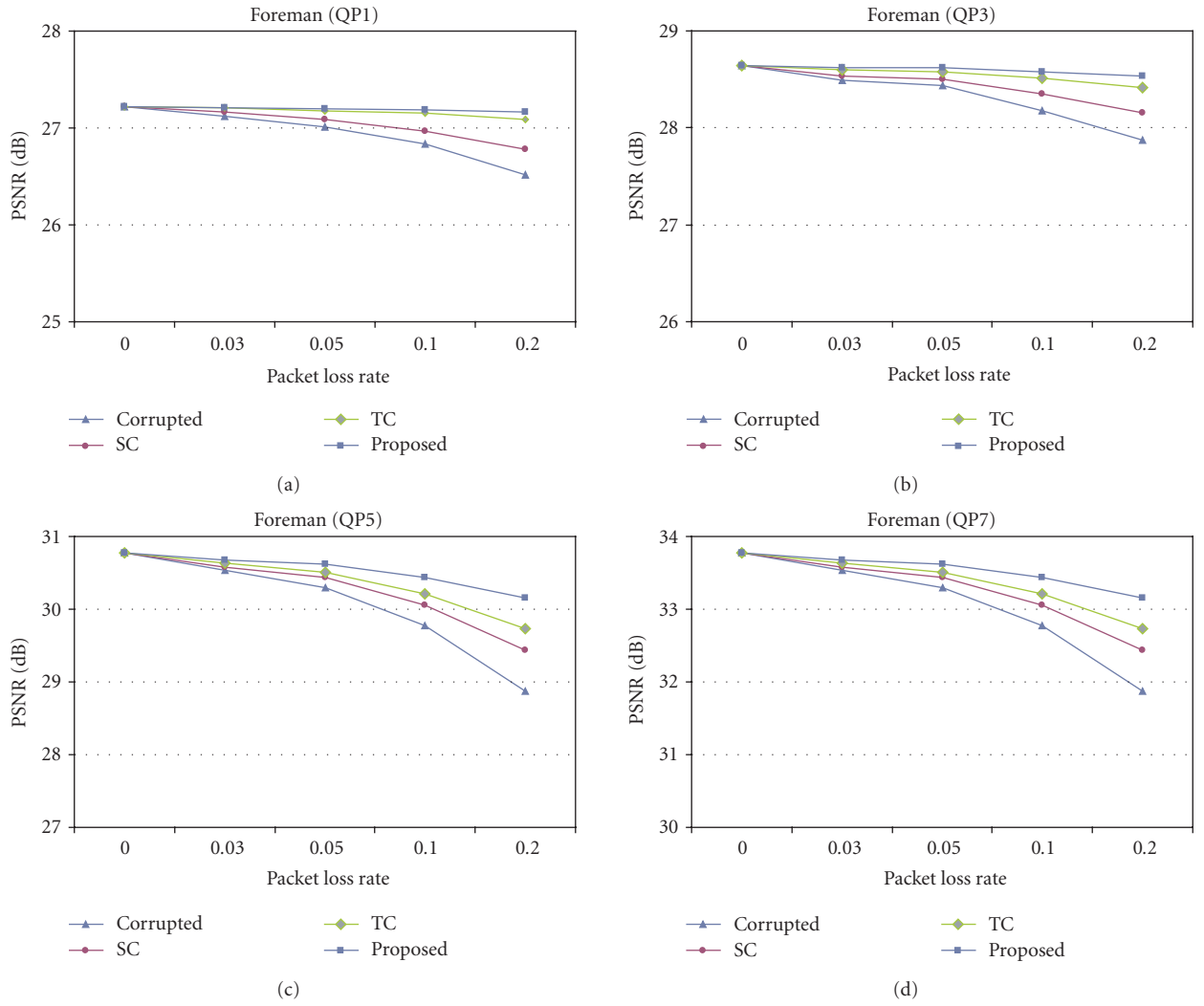


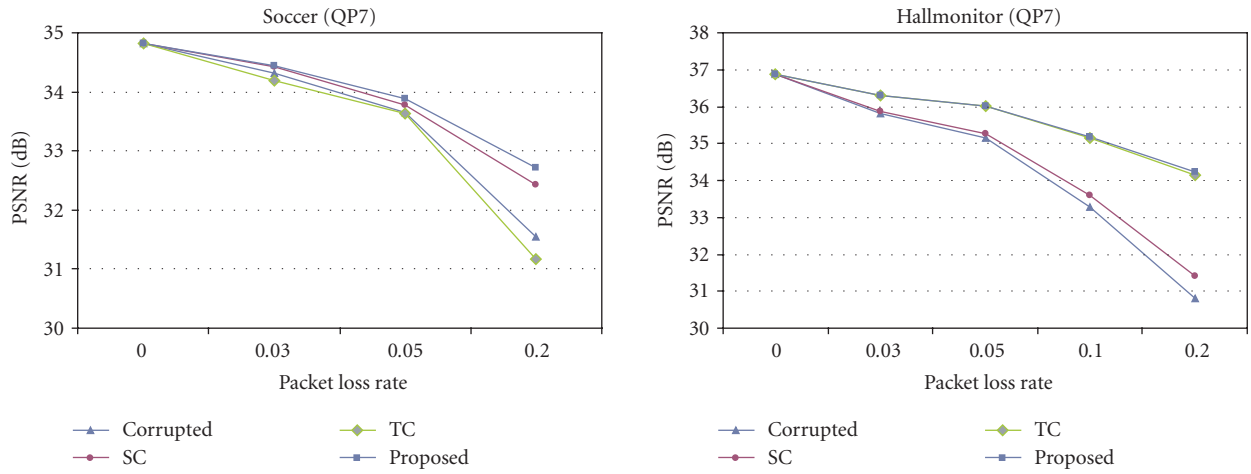
Figure 16: PSNR performance for *Foreman* (only WZ frames are corrupted).

Table 1: Effect of the weight α (*Foreman*, QCIF, 15 fps, $T = 10$).

	PSNR (dB)			
	$\alpha = 0.0$	$\alpha = 0.3$	$\alpha = 0.7$	$\alpha = 1.0$
Bitrate (kbps)				
72.86	28.31	28.36	28.10	27.85
99.41	29.62	29.66	29.26	28.94
170.06	32.15	32.19	31.65	31.26
297.16	35.51	35.53	34.88	34.38
Average	31.40	31.44	30.97	30.61

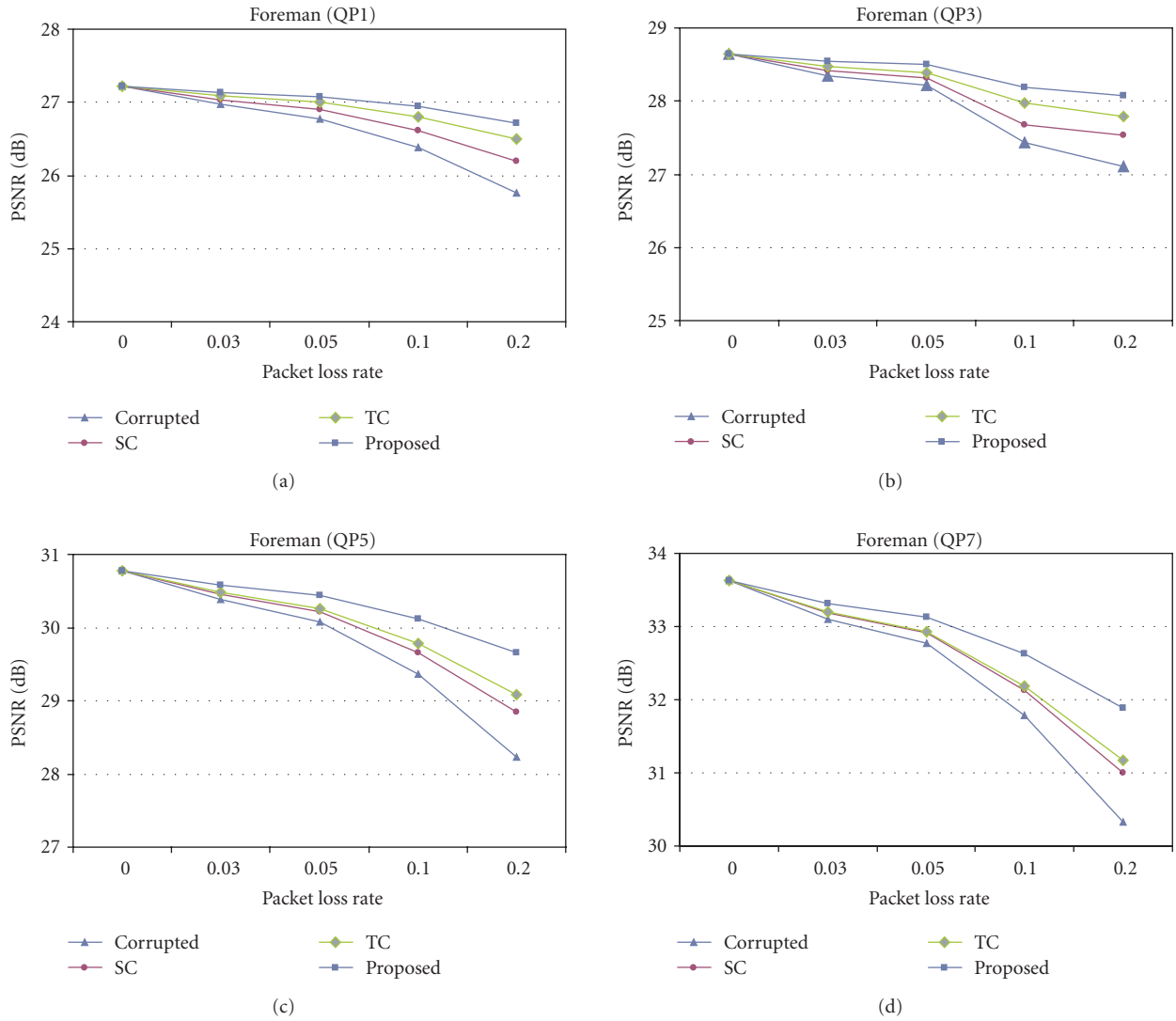
Table 2: Effect of the threshold T (*Foreman*, QCIF, 15 fps, $\alpha = 0.3$).

Bitrate (kbps)	$T = 10$		$T = 40$		$T = 70$		$T = 100$	
	PSNR (dB)	Percentage	PSNR (dB)	Percentage	PSNR (dB)	Percentage	PSNR (dB)	Percentage
72.86	28.36	84.13%	28.34	18.73%	28.3	9.45%	28.06	0%
99.41	29.66	91.24%	29.63	19.47%	29.55	10.83%	29.26	0%
170.06	32.19	89.82%	32.12	21.73%	32	11.82%	31.65	0%
297.16	35.53	94.24%	35.41	24.21%	35.23	13.05%	34.84	0%
Average	31.44	89.86%	31.38	21.03%	31.27	11.29%	30.95	0%



(a) *Soccer* (b) *Hallmonitor*

Figure 17: PSNR performance of different sequences (only WZ frames are corrupted).



(a) (b) (c) (d)

Figure 18: PSNR performance for *Foreman* (both key and WZ frames are corrupted).

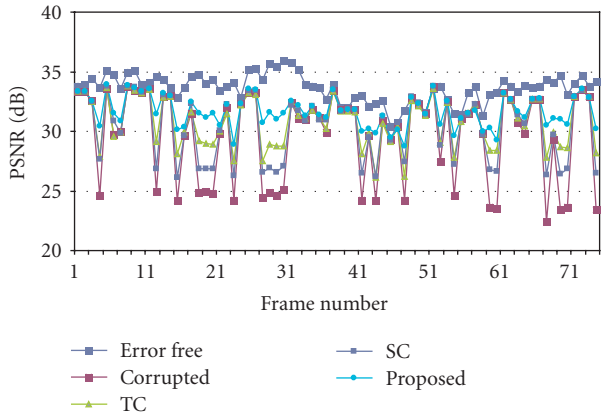


Figure 19: PSNR of error-concealed frames for *Foreman*.

Figure 9. Compared to the TDWZ, the proposed method achieves up to 2.2 dB and an average of 0.6 dB improvement. Both TDWZ and the proposed algorithm have the same bitrate, 12.0 kbps. From these figures, we can also find that the proposed algorithm gets larger improvement over TDWZ when there is large or asymmetric motion, for example, from frame 35 to 55. The reason is that for small or symmetric motions, the motion estimation and compensation are correct enough by TDWZ.

The RD performance of the decoded WZ frames for all the sequences is shown in Figures 10, 11, 12, and 13, respectively. RD performance improvements over the TDWZ are observed for all sequences. For *Soccer* (Figure 11), the proposed method significantly improves the objective quality (up to 1.0 dB at QP6) with respect to TDWZ. When compared to H.264/AVC intra, the performance is still inferior but the gap is brought down to around 2 dB, while it was around 3 dB for the TDWZ. For video with simple motion such as *Hallmonitor* (Figure 13), the performance of the proposed method is only slightly better than the TDWZ, which is already around 3 dB superior to H.264/AVC Intra. For *Foreman*, the introduced SI improves the performance over H.264/AVC intra for low bitrates (Figure 10). For Coastguard, the introduced SI outperforms H.264/AVC intra for all different bitrates (Figure 12).

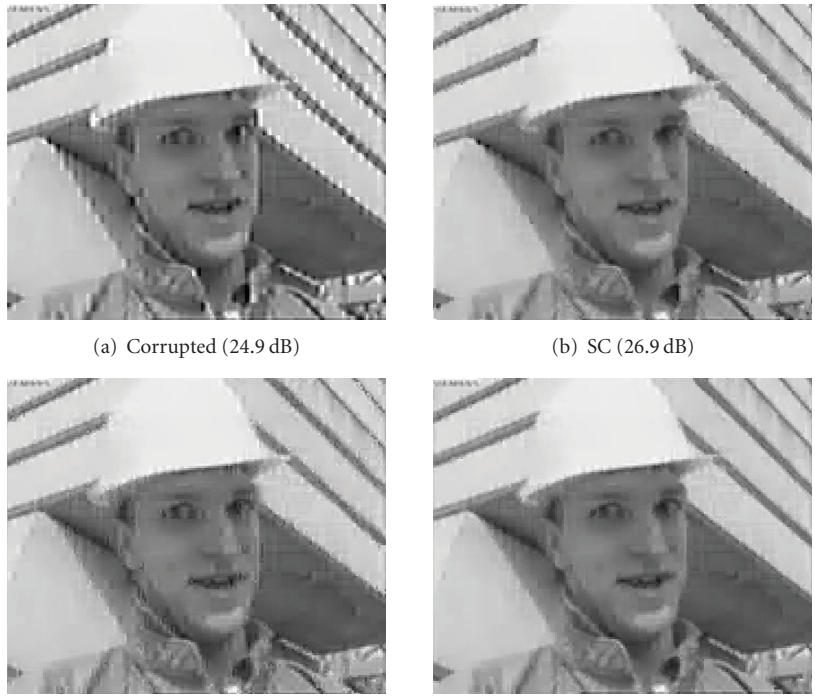
We also test the proposed SI generation with sequences in CIF format at 30 fps. The RD performance results are show in Figure 14 for *Foreman* and in Figure 15 for *Soccer*, respectively. Compared to the QCIF results (Figures 10 and 11), the improvement of the proposed method over TDWZ for CIF format is a bit smaller than that for QCIF format. The reason is that the better quality of CIF format and the larger number of frame per second, which means that the motion between two subsequent frames is smaller, improve motion estimation, which makes the improvement of the proposed method smaller.

The effect of the weight α and the threshold T on the finally decoded frames has been also studied for the sequence *Foreman*. The average PSNR values of the decoded frames at different bitrates with different α are shown in Table 1. We can observe that the α value of 0.3 always generates the best PSNR. The average PSNR values and the corresponding percentage of detected suspicious motion vectors with different T are shown in Table 2. Naturally, for larger T, the percentage of detected suspicious vectors gets smaller, and then the complexity of out method decreases. However, as a good number motion vectors estimated by MCTI are close to the true motion, the false motion vectors are always filtered. Hence, the coding performance remains essentially constant for small enough T.

5.2. Error Concealment Simulations. In simulations on error concealment for DVC, a communication channel, characterized by the error pattern files provided in [31] with different Packet Loss Ratios (PLRs), is used. The test sequences are corrupted with packet loss rates of 3%, 5%, 10%, and 20%. For the various test conditions in terms of PLR and Quantization Parameters (QPs), the average PSNR is measured. Results are obtained by averaging over ten runs using different error patterns. In video streaming, due to the tight delay constraint, a feedback channel is not preferred since it will cause additional delay, and the received retransmitted bits beyond time constraint will be useless. In this case, no feedback channel is used when errors occurred during the transmission. However, the rate control of the DVC architecture (Figure 1) used in this paper uses a feedback channel driven rate control. Therefore, in this section, we suppose that the encoder performs ideal rate control, that is, the number of requested bits for each bitplane for the error-free case is determined a priori and used for decoding the corresponding corrupted bitstream. Furthermore, as used in [8], if the bit error probability of the decoded bitplane in WZ decoder is higher than 10^{-3}, the decoder uses the corresponding bitplane of the SI. The header of the WZ bitstream, which contains critical information such as frame size, quantization parameters, and intraperiod, is assumed as correctly received.

To evaluate the performance of the proposed error concealment method, first only errors in WZ frames are simulated. Figure 16 shows the PSNR values of error-concealed results at several quantization indexes for sequence *Foreman*. The results of the proposed concealment method are compared to those of spatial concealment (SC, based on edge directed filter [7]), and temporal concealment (TC, based on MCTI [11]). As shown in Figure 16, the proposed error concealment method achieves better objective qualities than both SC and TC, for all packet loss ratios and quantization indexes. The improvements of the proposed method can be considered as quite good when compared to the-state-of-the-art. From the results, we can also observe that, although TC generally performs better than SC, the performance of SC can be very close to TC when the video quality is good (for high QP or low PLR).

Figure 17 shows the results of sequence *Soccer* (Figure 17(a)) and *Hallmonitor* (Figure 17(b)). For sequences with large and complex motion such as *Soccer*, the temporal EC performance can become worse than that of spatial EC (Figure 17(a)). For sequences with simple motion such as *Hallmonitor*, the performance of the proposed EC becomes close to TC (Figure 17(b)).



(a) Corrupted (24.9 dB) (b) SC (26.9 dB)

(c) TC (28.9 dB) (d) Proposed method (31.6 dB)

Figure 20: Error-concealed results for *Foreman*.

To evaluate the performance of the proposed method in a more realistic scenario, errors in both key and WZ frames are also simulated. The spatial EC defined in H.264/AVC JM 11.0 is used to conceal errors in key frames. Figure 18 shows the PSNR values of error-concealed results at several quantization indexes for sequence *Foreman*. As shown in Figure 18, errors in key frames further decrease PSNR values for all EC methods, but the proposed error concealment method leads to even larger improvements when compared to the-state-of-the-art. The reason may be that the errors in key frames cause distortions in WZ frame, and these distortions can also be better concealed by the proposed method.

Figure 19 shows the PSNR values of all frames of sequence *Foreman* (QP7, PLR 20%). The improvement by the proposed method is quite important, with gains up to 8.0 dB. The average improvement for the whole sequence is 2.0 dB, significantly outperforming SC (0.82 dB) or TC (1.13 dB).

Figure 20 shows the visual results of frame 56 (QP7, PLR 20%). Distortions are still visible near the edges in the frames concealed by SC (Figure 20(b)) and TC (Figure 20(c)). The visual quality obtained with the proposed method (Figure 20(d)) is better.

6. Conclusions

A new side information generation scheme is proposed to improve the performance of DVC. The use of the partially decoded WZ frame, which is exploited in motion vector refinement, smoothing, and optimal compensation mode selection improves the performance of SI generation. Simulation results show that a large improvement of performance has been achieved by the proposed approach, compared to the-state-of-the-art DVC. The proposed approach achieves better RD performances than TDWZ for all test sequences. For sequences with large motion, such as *Soccer*, the proposed method improves the objective quality (up to 1.0 dB), compared with TDWZ. These results were obtained with no additional complexity or modification to the DVC encoder.

The new SI generation technique is also very helpful in the proposed hybrid EC scheme for WZ frames. Simulation results on the proposed error concealment method also show that the proposed method can bring significant improvements in terms of both objective and perceptual qualities for corrupted sequences. The proposed hybrid EC scheme is only applied at the decoder, without any modification to the DVC encoder or transmission channels.

Acknowledgments

This work was partially supported by the European Network of Excellence VISNET II (http://www.visnet-noe.org) IST Contract 1-038398, funded under the European Commission IST 6th Framework Program.

References

[1] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, 2003.

[2] B. Girod, A. M. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding," *Proceedings of the IEEE*, vol. 93, no. 1, pp. 71–83, 2005.

[3] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions on Information Theory*, vol. 19, no. 4, pp. 471–480, 1973.

[4] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Transactions on Information Theory*, vol. 22, no. 1, pp. 1–10, 1976.

[5] J. Pedro, L. Soares, C. Brites, et al., "Studying error resilience performance for a feedback channel based transform domain Wyner-Ziv video codec," in *Proceedings of the Picture Coding Symposium (PCS '07)*, Lisbon, Portugal, November 2007.

[6] Y. Wang and Q.-F. Zhu, "Error control and concealment for video communication: a review," *Proceedings of the IEEE*, vol. 86, no. 5, pp. 974–997, 1998.

[7] S. Ye, Q. Sun, and E.-C. Chang, "Edge directed filter based error concealment for wavelet-based images," in *Proceedings of IEEE International Conference on Image Processing (ICIP '04)*, vol. 2, pp. 809–812, Singapore, October 2004.

[8] C. Brites, J. Ascenso, and F. Pereira, "Improving transform domain Wyner-Ziv video coding performance," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '06)*, vol. 2, pp. 525–528, Toulouse, France, May 2006.

[9] J. Zhai, K. Yu, J. Li, and S. Li, "A low complexity motion compensated frame interpolation method," in *Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS '05)*, vol. 5, pp. 4927–4930, Kobe, Japan, May 2005.

[10] B.-D. Choi, J.-W. Han, C.-S. Kim, and S.-J. Ko, "Motion-compensated frame interpolation using bilateral motion estimation and adaptive overlapped block motion compensation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 4, pp. 407–415, 2007.

[11] J. Ascenso, C. Brites, and F. Pereira, "Improving frame interpolation with spatial motion smoothing for pixel domain distributed video coding," in *Proceedings of the 5th EURASIP Conference on Speech and Image Processing, Multimedia Communications and Services*, Smolenice, Slovakia, July 2005.

[12] L. Wei, Y. Zhao, and A. Wang, "Improved side-information in distributed video coding," in *Proceedings of the 1st International Conference on Innovative Computing, Information and Control (ICICIC '06)*, vol. 2, pp. 209–212, Beijing, China, August 2006.

[13] R. Puri and K. Ramchandran, "PRISM: a "reversed" multimedia coding paradigm," in *Proceedings of IEEE International Conference on Image Processing (ICIP '03)*, vol. 1, pp. 617–620, Barcelona, Spain, September 2003.

[14] A. Aaron, S. Rane, and B. Girod, "Wyner-Ziv video coding with hash-based motion compensation at the receiver," in *Proceedings of IEEE International Conference on Image Processing (ICIP '04)*, vol. 5, pp. 3097–3100, Singapore, October 2004.

[15] M. Tagliasacchi, A. Trapanese, S. Tubaro, J. Ascenso, C. Brites, and F. Pereira, "Exploiting spatial redundancy in pixel domain Wyner-Ziv video coding," in *Proceedings of IEEE International Conference on Image Processing (ICIP '06)*, pp. 253–256, Atlanta, Ga, USA, October 2006.

[16] J. Ascenso, C. Brites, and F. Pereira, "Motion compensated refinement for low complexity pixel based distributed video coding," in *Proceedings of IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS '05)*, pp. 593–598, Como, Italy, September 2005.

[17] X. Artigas and L. Torres, "Iterative generation of motion-compensated side information for distributed video coding," in *Proceedings of IEEE International Conference on Image Processing (ICIP '05)*, vol. 1, pp. 833–836, Genova, Italy, September 2005.

[18] W. A. R. J. Weerakkody, W. A. C. Fernando, J. L. Martínez, P. Cuenca, and F. Quiles, "An iterative refinement technique for side information generation in DVC," in *Proceedings of IEEE International Conference on Multimedia and Expo (ICME '07)*, pp. 164–167, Beijing, China, July 2007.

[19] W. Zhu, Y. Wang, and Q.-F. Zhu, "Second-order derivative-based smoothness measure for error concealment in DCT-based codecs," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 8, no. 6, pp. 713–718, 1998.

[20] W. Zeng and B. Liu, "Geometric-structure-based error concealment with novel applications in block-based low-bit-rate coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, no. 4, pp. 648–665, 1999.

[21] X. Li and M. T. Orchard, "Novel sequential error-concealment techniques using orientation adaptive interpolation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 10, pp. 857–864, 2002.

[22] P.-J. Lee, H. H. Chen, and L.-G. Chen, "A new error concealment algorithm for H.264 video transmission," in *Proceedings of the International Symposium on Intelligent Multimedia, Video and Speech Processing (ISIMP '04)*, pp. 619–622, Hong Kong, October 2004.

[23] Y. Chen, K. Yu, J. Li, and S. Li, "An error concealment algorithm for entire frame loss in video transmission," in *Proceedings of the Picture Coding Symposium (PCS '04)*, pp. 389–392, San Francisco, Calif, USA, December 2004.

[24] Y. Chen, O. Au, C. Ho, and J. Zhou, "Spatio-temporal boundary matching algorithm for temporal error concealment," in *Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS '06)*, pp. 686–689, Island of Kos, Greece, May 2006.

[25] O. Hadar, M. Huber, R. Huber, and S. Greenberg, "New hybrid error concealment for digital compressed video," *EURASIP Journal on Applied Signal Processing*, vol. 2005, no. 12, pp. 1821–1833, 2005.

[26] D. Agrafiotis, D. R. Bull, T. K. Chiew, P. Ferre, and A. Nix, "Enhanced error concealment for video transmission over WLANS," in *Proceedings of the International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS '05)*, Montreux, Switzerland, April 2005.

[27] C. Yim, W. Kim, and H. Lim, "Hybrid error concealment method for H.264 video transmission over wireless networks," in *Proceedings of the International Wireless Communications and Mobile Computing Conference (IWCMC '07)*, pp. 606–611, Honolulu, Hawaii, USA, August 2007.

[28] M. Chen and Y. Zheng, "Classification-based spatial error concealment for visual communications," *EURASIP Journal on Applied Signal Processing*, vol. 2006, Article ID 13438, 17 pages, 2006.

[29] S. Ye, X. Lin, and Q. Sun, "Content based error detection and concealment for image transmission over wireless channel," in *Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS '03)*, vol. 2, pp. 368–371, Bangkok, Thailand, May 2003.

[30] R. Bernardini, M. Fumagalli, M. Naccari, et al., "Error concealment using a DVC approach for video streaming applications," in *Proceedings of the EURASIP European Signal Processing Conference (EUSIPCO '07)*, Poznan, Poland, September 2007.

[31] S. Wenger, "Proposed error patterns for Internet experiments," in *Proceedings of the 9th Meeting for Video Coding Experts Group (VCEG '99)*, Red Bank, NJ, USA, October 1999, Doc. VCEG Q15-I-16R1.