

New Hybrid Error Concealment for Digital Compressed Video

Ofer Hadar

Communication Systems Engineering Department, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel
Email: hadar@cse.bgu.ac.il

Merav Huber

Electrical and Computer Engineering Department, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel
Email: huberm@bgu.ac.il

Revital Huber

Electrical and Computer Engineering Department, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel
Email: huberr@bgu.ac.il

Shlomo Greenberg

Communication Systems Engineering Department, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel
Email: shlomogreenberg@freescalse.com

Received 2 August 2004; Revised 26 December 2004; Recommended for Publication by Reha Civanlar

Transmission of a compressed video signal over a lossy communication network exposes the information to losses and errors, which leads to significant visible errors in the reconstructed frames at the decoder side. In this paper we present a new hybrid error concealment algorithm for compressed video sequences, based on temporal and spatial concealment methods. We describe spatial and temporal techniques for the recovery of lost blocks. In particular, we develop postprocessing techniques for the reconstruction of missing or damaged macroblocks. A new decision support tree is developed to efficiently choose the best appropriate error concealment method, according to the spatial and temporal characteristics of the sequence. The proposed algorithm is compared to three error concealment methods: spatial, temporal, and a previous hybrid approach using different noise levels. The results are evaluated using four quality measures. We show that our error concealment scheme outperforms all the other three methods for all the tested video sequences.

Keywords and phrases: error concealment, spatial/temporal/hybrid error concealment, video coding, MPEG-2, decision tree, multimedia/video communication.

1. INTRODUCTION

The demand for transmitting compressed video over data network increases as bandwidth and storage of computer networks grow. Signal loss occurring in physical communication channels is unavoidable. During data transmission of packets over the Internet, packets may be dropped or damaged, due to channel errors, congestion, and buffer limitation. Moreover, the data may arrive too late to be used in real-time applications.

These errors fall into two categories: (1) bit stream errors caused by direct signal loss of some or the whole compressed packet of a coded MB, and result in the loss of a single block, a group of blocks or macroblocks, or the whole respective slice information; (2) propagation errors

caused in *P*- and *B*-frames uniquely by the additional use of motion-compensated time information for their reconstruction at the decoder side. Errors in previously decoded reference frames propagate to their dependent frames in the decoding order.

In the case of transmission of compressed video sequences such as MPEG-2, this loss may be devastating and result in a completely damaged stream at the decoder side. Compression, which dilutes the amount of redundant information, cannot compensate for data loss, which leads to different visual artifacts [1]. Since MPEG compression standards use variable-length coding (VLC), even a bit change, which stems from imperfections over the transmission medium, may cause misinterpretation of code words. This leads to desynchronization of the following bits, until

the next synchronization word is encountered [2]. In order to deal with the problems caused by packet losses, and since retransmission is not an option for real-time application, error concealment (video resilience) techniques are required.

These techniques are divided into two major types: techniques that aim at lossless recovery, such as FEC (forward error concealment) and ECC (error-control coding), and techniques that focus on signal reconstruction and error concealment [2]. These techniques offer a close approximation of the original signal, based on natural video characteristics and on features of the human vision system [2]. The first type suggests delay, bandwidth increase, and insertion of data or codewords, and therefore requires nonstandard decoders, whereas the second type requires detection of the error positions within the image, or knowledge of motion vectors and DCT coefficients, and may result in blurred frames [3].

A different technique is implemented for *I*-frames, since the latter are coded independently from the other frames of the video sequence. This technique exploits spatial information only from available neighboring blocks of the current frame. Error concealment in *P*- and *B*-frames is performed through the use of both spatial and temporal information. The spatial information is obtained by available current frame neighboring MBs, while time information is acquired by using previously decoded frames.

Error concealment approaches can be considered as either active or passive concealment. In active concealment, both retransmission and error-control coding methods are used. Active concealment has the advantage of permitting perfect reconstruction at the decoding end if the amount of data lost is not significant [2]. Since packet loss can result in the loss of entire rows of macroblocks in an image, packetization techniques that rely on interleaving data have been proposed [4, 5].

Postprocessing techniques for error concealment at the decoder side, also referred to as “passive concealment,” utilize spatial data, temporal data, or hybrid of both [6]. Missing macroblocks can be reconstructed by estimating their DCT coefficients from the DCT coefficients of the neighboring macroblocks [7]. An alternative to spatial error concealment is to use motion compensation [8] whereby the average of the motion vectors of neighboring macroblocks is used to perform concealment.

In this paper, we describe spatial and temporal techniques for the recovery of lost macroblocks. In particular, we present a new postprocessing concealment algorithm for the reconstruction of missing and damaged blocks. The proposed error concealment algorithm combines temporal and spatial concealment methods. The type of concealment to be implemented upon a degraded block is selected according to a decision tree. Error implementation is performed by simulation, on decompressed images, and takes into account characteristics of MPEG-2, such as block coding and frame types. The performance of the suggested concealment algorithm is compared to a spatial [6], temporal, and hybrid [6] spatial/temporal concealment methods, and evaluated using several quality measures. The quality measures are based on mathematical calculations, where the last two emulate fea-

tures of the human visual system on the perception of the video sequence. The following criteria are used for quality measure: (a) mean square error (MSE), (b) the peak signal-to-noise ratio (PSNR) [9], (c) *improved* MSE, which eliminates the influence of sole peak values, (d) video perceptual distortion measure (VPDM), which performs comparisons between video sequences [10, 11], and finally (e) normalized *peak* VPDM (P_VPDM), given in dB.

The paper is organized as follows. The proposed error concealment schemes, based on post-image processing at the decoder side, are presented in Section 2. Section 3 describes error generation simulation used in this work, and consequent erroneous blocks handling. Quality criteria and performance evaluation are described in Section 4. Experiments and results are presented in Section 5, and conclusions of our work are given in Section 6.

2. ERROR CONCEALMENT SCHEME

In this section, we first describe three common types of error concealment, namely, temporal, spatial, and frequency-domain error concealment methods. Then, a previous work suggesting a hybrid error concealment method, which combines the temporal and the spatial methods, is described. Finally, the proposed hybrid decision-support-tree-based algorithm is presented.

2.1. Spatial concealment

Spatial post-error concealment is based on the fact that natural images are likely to be smooth. This means that if a pixel is lost, its value can be derived from the neighboring pixels. There are several methods for spatial reconstruction of a lost block, which differ in the amount of neighboring pixels used, in their location and distance from the lost pixel, and in their relative weight in the concealment process.

Usually spatial concealment is combined with frequency-domain concealment, since the transmitted data contains DCT values of a block of pixels. Loss of part of the block information requires spatial reconstruction of the whole block. In this paper, we suggest to use the method proposed by Dovrolis et al. [6]. The value of each missing pixel x is an average of its four neighbors, from the left (l), right (r), top (t), and bottom (b), as follows:

$$x = \frac{x_l + x_r + x_t + x_b}{4}. \quad (1)$$

Since an entire block is missing, we receive a set of 64 equations with 64 parameters, that are solved simultaneously.

For missing blocks, which do not have four neighbors within the block, we use the values of boundary pixels of the available neighboring blocks.

2.2. Temporal concealment

An important statistic characteristic of the compressed video stream is that there is no correlation between the packet loss in one encoded frame and the packet loss in the following frame. Thus, a block that suffers degradation in the previous frame is very unlikely to be degraded at the next frame.

This statistic assumption is exploited when performing temporal concealment. Replacing a lost or degraded block with the same positioned block in the previous frame is the easiest and fastest temporal concealment method. However, when there is fast motion in the block area, such a block replacement causes visible distortion. Thus, motion compensation is considered. The motion vector of the missing block is computed using a linear combination of the motion vectors of the neighboring blocks, which were correctly received or previously reconstructed. This is useful in case the motion variance in the missing block neighborhood is not very large. However, in the case of scattered motion vectors, it would be hard to achieve the correct vector, and other concealment methods should be considered. Motion vectors are reconstructed by averaging the motion vectors of the four neighboring blocks.

Another issue related to temporal concealment is determining the missing block type (*I*-, *P*-, or *B*-frame). Determining the block type can be performed in several ways. For example, when the block type cannot be identified, it is determined to be an *I*-block. Another method [2] uses the block types of the upper and lower neighboring blocks to determine the missing block type.

In this work, we assume that the positions of the lost blocks are known, and the surrounding blocks parameters are known or can be derived (i.e., motion vectors and DCT coefficients). Moreover, it is assumed that the remaining (undamaged) parameters of the damaged block are known as well. Lost motion vectors are reconstructed by averaging the four motion vectors of the neighboring blocks and rounding the result to the nearest integer, resulting in one-pixel accuracy. Since using the original difference-block coefficient does not improve the image quality, the difference block is zeroed.

2.3. Frequency-domain concealment

Frequency-domain interpolation is different from the temporal and spatial interpolations, since it uses remaining information of a damaged block, rather than ignore the whole block. The frequency domain interpolation is based on smoothness, which characterizes real video signals, and assumes high correlation between spatially adjacent blocks. Thus frequency interpolation is used especially on *I* blocks or still pictures. Frequency concealment uses DCT coefficients of neighboring blocks to reconstruct corresponding DCT coefficients of the missing block. However, the methods of reconstruction are only suitable for low-order DCT coefficients [12]. One method of error concealment based on frequency domain is given in [13]. When all the block coefficients are lost, the algorithm uses spatial interpolation for block reconstruction.

After assessing the frequency-domain concealment, it was decided not to use it within the proposed decision-tree-based algorithm, for two reasons: first, the algorithm is time consuming, and second, the results are unsatisfying. Only in few cases did this scheme result in better concealment than the temporal and spatial concealment algorithms. Using the

remaining DCT coefficients gives better results than altering them using the frequency-domain concealment.

2.4. Hybrid error concealment method

The method suggested by Dovrolis et al. [6] uses a combination of the spatial and the temporal error concealment methods mentioned above. The main assumption of this algorithm is that using the temporal estimation yields better results where no motion exists, and spatial estimation should be used where motion appears. A missing block is divided into four quarters (up, down, left, and right), and the concealment decision is performed on each quarter separately, according to the existence of motion in the section.

2.5. Decision-support-tree-based error concealment algorithm

This section describes the proposed decision support algorithm for hybrid error concealment. The decision-support mechanism considers several criteria and tries to efficiently implement the best concealment for each block in order to achieve minimal visible distortion. The following criteria are used: (a) motion level (slow/fast) in the area of the degraded block/macroblock (MB); (b) the motion variance in that area; (c) spatial smoothness; and (d) the remaining DCT coefficients. The decision is performed on each damaged block/MB separately, taking into consideration the possibility that several blocks/MBs around the degraded block are also damaged or dropped. The decision-support tree for *P*- and *B*-frames is illustrated in Figure 1.

The main steps of the decision-support algorithm are as follows.

(a) For a macroblock with valid motion vector, the amount of motion in the macroblock is evaluated, according to the motion vector size. In case of low motion level (the speed is below a predefined threshold “A” in Figure 1) the macroblock is assumed to be similar to its corresponding block/MB in the reference frame. Therefore, no concealment is performed in this case and the motion vector and the reference block are used for block decoding. The DCT coefficients of the lost difference block are set to zero.

(b) In case of high motion velocity (i.e., greater than a predefined threshold “A”), the algorithm further checks whether the number of intact DCT coefficients in the block is greater than a threshold “B.” The intact DCT coefficients are counted from the DC coefficient to the AC coefficients in increasing order, according to the zig-zag scan. If this condition is met, no error concealment is done since the lost coefficients are not dominant and zeroing them enables decoding with better quality.

(c) When the number of the intact coefficients is less than “B,” two options are considered: temporal or spatial concealment. The preferred concealment method depends on the spatial variance in the missing block/macroblock neighborhood. In the case of relatively smooth area (a spatial variance below threshold “D”), spatial concealment is preferred since spatial reconstruction should not cause large visible distortion. For nonuniform area with large variance, the temporal concealment is performed. The spatial variance is derived

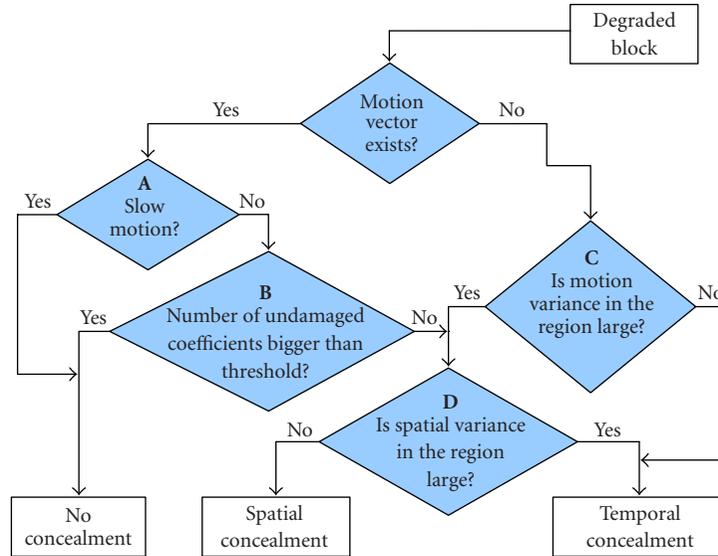


FIGURE 1: Decision-tree-based error concealment algorithm for *B*- and *P*-frames.

from the pixel values around the damaged block/MB, including only pixels belonging to intact blocks/MBs or previously reconstructed ones. The spatial variance is computed by

$$\text{Spatial_Variance} = E[(X - E[X])^2], \quad (2)$$

where X represents the pixel values around the reconstructed block, and $E[X]$ stands for the expected value of the variable X .

(d) In the case of missing motion vectors, we use the motion variance of the neighboring macroblocks to evaluate the missing motion vectors. Equation (2) can also be used for this calculation, where X would represent the motion vectors of the neighboring macroblocks of the missing block/MB. For low variance (smaller than threshold “C”), a good reconstruction of the motion vector is predicted by averaging its neighboring macroblocks motion vectors. Therefore temporal error concealment is carried out.

(e) For high variance (greater than “C”), the spatial variance in the degraded area is compared to a higher threshold “D2” in order to decide whether temporal or spatial concealment should be implemented. For high variance values for both spatial and temporal cases, the temporal concealment is preferred.

(f) For high motion variance (greater than “C”) and average (low) spatial variance (less than “D2”), spatial concealment is chosen.

A different technique is implemented for *I*-frames, since the latter are coded independently from the other frames of the video sequence. This technique exploits spatial information only, from available neighboring MBs of the current frame. For a sufficient number of correctly received DCT coefficients (greater than threshold B2), no concealment is carried out. Decision-support tree for *I*-frames is illustrated in Figure 2.

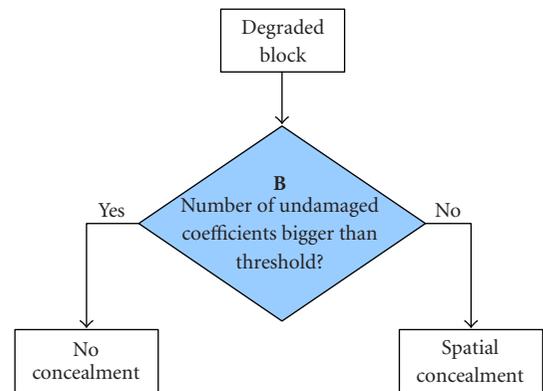


FIGURE 2: Decision-tree-based error concealment algorithm for *I*-frames.

Error concealment in *P*- and *B*-frames is performed through the use of both space and time information. The space information is obtained by the available neighboring MBs of the current frame, while time information is acquired by using previously decoded frames.

3. ERROR GENERATION AND CONSEQUENT ERRONEOUS HANDLING

We describe here the error generation simulation used in this work to simulate the various types of degradation in the compressed video. In Section 3.2, we present an algorithm for handling consequent erroneous blocks.

3.1. Error generation by simulation

We simulate the degradation of the video sequences by using an error generator which damages or drops the DCT coefficients and motion vectors according to the frame type and

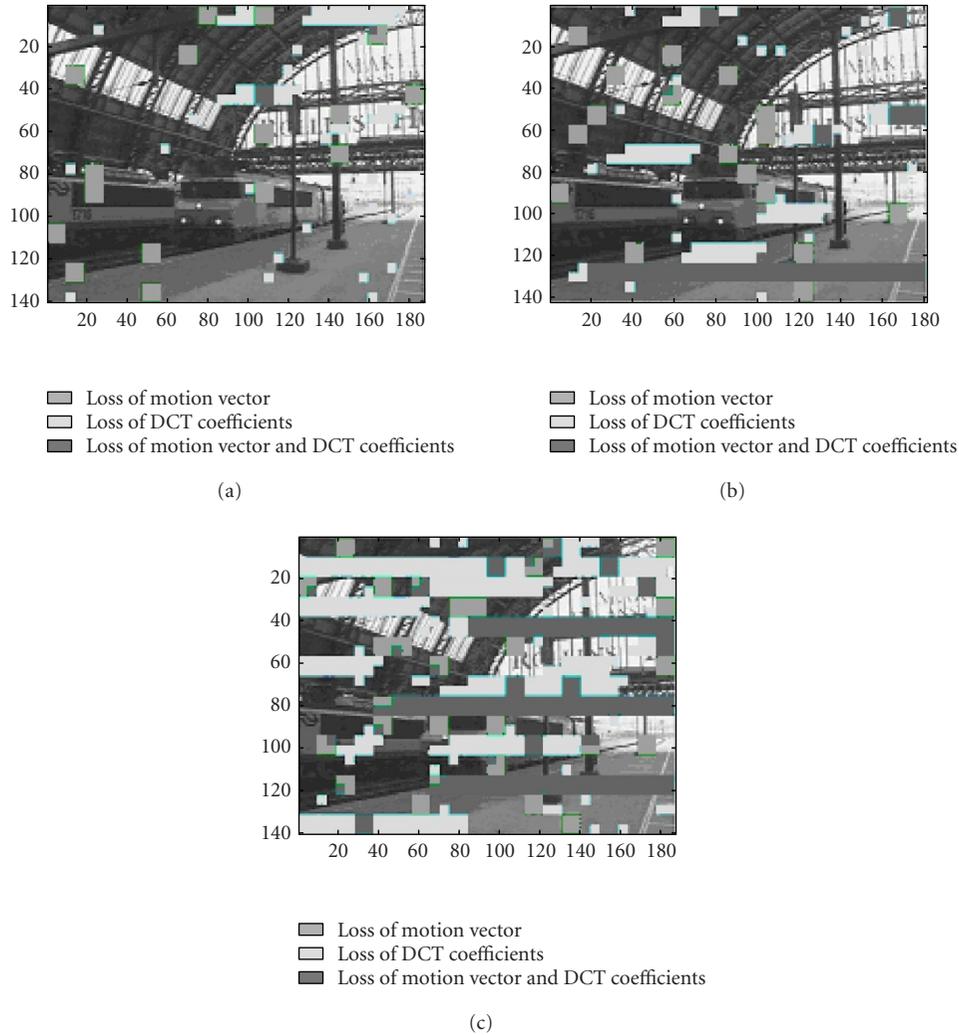


FIGURE 3: Degraded images using different noise factors: (a) noise factor of 1, (b) noise factor of 1.5, and (c) noise factor of 2.5.

amount of noise in the channel (noise level/factor). For *I*-frames, only DCT coefficients are damaged. *P*- and *B*-frames suffer from additional motion vectors loss and error propagation from the reference frames. The compression rate of each frame determines the amount of data loss in the pixel domain. For highly compressed frames, the error effect is much stronger. When a single block is damaged, a DCT coefficient within the block is chosen from which all other coefficients are damaged. Errors are applied also on several consequent blocks, and on slices from a certain block till the end of the slice. A so-called “error generation filter” or “noise-level” contains the following parameters: specified channel-noise factor, probability of a single-block loss, probability of a group of blocks, and probability of block loss until the end of a slice. Any change in one of the parameters affect the noise level and results in a different amount of degradation for the video sequence. Figure 3 shows the effect of different channel-noise factors on the quality of a video sequence. Noise factor of 1 is considered relatively low, while noise factor of 2.5 is relatively high.

3.2. Loss of consequent blocks

Signal loss occurring in physical communication channels used for transmission of compressed video sequences usually causes loss of some consequent macroblocks (and sometimes till the end of the slice). Concealment schemes use information from neighboring blocks for reconstruction: spatial concealment methods use the pixel gray level of the neighboring blocks, while temporal concealment uses the motion vectors of the adjacent macroblocks. In case of loss models which cause scattered missing blocks in the video sequence [14], the order of concealment of the damaged blocks within a frame (choosing which block is the next one to be reconstructed) is of no importance. However, loss of consequent blocks or macroblocks affects the error concealment results, thus the concealment order within a frame is important. We assume that it is preferred to use an adjacent block, which was damaged and reconstructed, rather than not to use it at all, since important information may be salvaged although it may cause spatial error propagation. This assumption applies also for temporal neighboring blocks.

The algorithm starts by reconstructing the blocks or macroblocks (MB) which have all four neighbors intact. In the next step, lost blocks or MBs with only one missing neighbor are reconstructed. This stage is repeated for blocks or macroblocks with two missing neighbors, and so on. Finally, a new search is performed to find the next missing block/macroblock to be concealed. This iterative procedure is carried out until all the damaged or dropped blocks/MBs in the frame are reconstructed.

4. PERFORMANCE EVALUATION

In order to evaluate the performance of the new proposed decision-support-tree-based error concealment algorithm, we have used four common image quality measures. Moreover, a visual interface was developed to visually assess the amount of correct decisions that were made by the proposed algorithm.

4.1. Image quality measures

The following quality measures are used in this work: (a) mean square error (MSE) [1, 9]; (b) peak signal-to-noise ratio (PSNR) measure [9]; (c) *improved* MSE measure; (d) video perceptual distortion measure (VPDM); and (e) P_VPDM. MSE eliminates the influence of the highest MSE values within a block. When using the MSE measure, a minor difference in pixel values within a block may result in large MSE values, although the human eye would not notice the difference. This elimination is done by taking only 7/8 of the overall MSE values of each block, excluding the 1/8 pixels which have the highest MSE values. VPDM uses three sequential frames in calculating the sequence quality, thus imitating temporal masking, which is a characteristic of the human viewer:

$$\begin{aligned} \text{dist}(t) = & w_1 \cdot | \text{IDM}(t-1) - \text{IDM}(t) | + w_2 \cdot | \text{IDM}(t) | \\ & + w_3 \cdot | \text{IDM}(t) - \text{IDM}(t+1) |, \end{aligned} \quad (3)$$

where $\text{dist}(t)$ is the currently received frame quality and $\text{IDM}(t)$ is the picture distortion measured between the transmitted and the received frames. In this work, we use A_{\max} [1, 15] as the image distortion measures (IDM). The weight for each element in the equation is w_i , $i = 1, 2, 3$, and it equals 1/3 if a scene cut is not detected. Otherwise, w_i of the scene transition is assigned a low value (1/30) [1, 10, 11] and P_VPDM criteria is given by

$$\text{P_VPDM} = 10 \log_{10} \left(\frac{P^2}{\text{VPDM}} \right), \quad (4)$$

where P is the maximum intensity value of the image.

4.2. Visual interface for result evaluation

This section describes a visual interface, which was built to evaluate results of the decision-tree-based concealment algorithm. The mechanism chooses between three types of

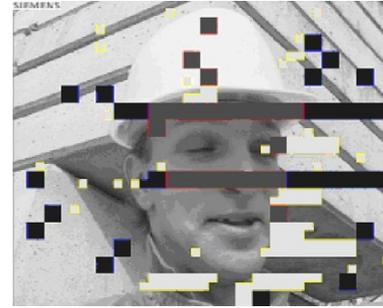


FIGURE 4: Decision map of the decision-based algorithm for each block.

concealment (temporal or spatial concealment and no concealment) for each degraded block/macroblock. For each reconstructed frame, a visual decision map is created. Along with building the decision map, a control frame is made. Each video sequence goes through both only temporal and only spatial error concealment algorithms for all the damaged blocks. The two resulting reconstructed sequences are compared to the original (undamaged) video sequence using one of the following criteria: PSNR, MSE, *improved* MSE, and VPDM. This comparison leads to a reliable decision of which concealment type is preferred for each degraded block. A block is considered to be equally concealed by more than one concealment type if the quality measurements slightly differ. Comparing the reconstructed frame and the control frame gives a performance indication for the decision-tree algorithm. This visual tool is used offline for user post-performance analysis. An example of the visual interface is given in Figure 4.

This visual interface should help the user to assess the amount of correct decisions that were made by the proposed algorithm.

Figure 5 demonstrates the performance of the proposed algorithm on a degraded “Foreman” sequence. Figure 5a indicates the different levels of degraded blocks (loss of motion vector, DCT coefficients, and both). The optimum spatial or temporal concealment is found for three different image quality criteria: MSE, *improved* MSE, and VPDM. Figure 5e shows the concealment choice of the proposed algorithm (spatial, temporal, and no concealment) for each degraded block. This yields similar results to those achieved for the optimum case with the *improved* MSE criteria. This measure often suggests concealing a degraded block in two optimal ways (appears as mixed colors), which are the combination of the concealment type that resulted from the MSE and VPDM calculations.

4.3. Computation load of the decision algorithm

The added computational load derived by the decision tree is very low. Each node contains a comparison operation,

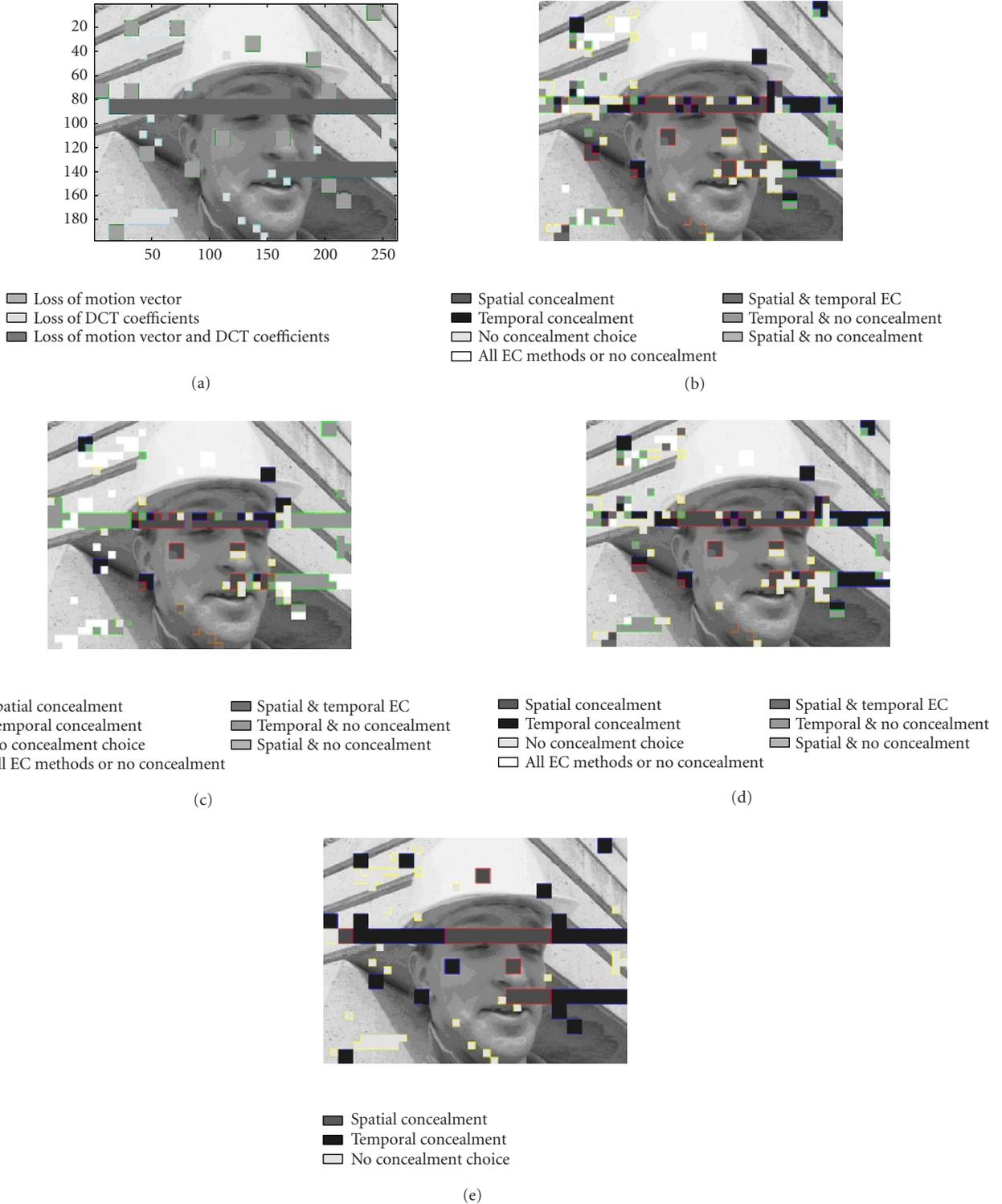


FIGURE 5: Error concealment visual representation. (a) “Foreman” sequence with different levels of degraded blocks. Optimal spatial or temporal concealment for (b) MSE, (c) improved MSE, and (d) VPDM. The concealment choice of the proposed algorithm is shown in (e).

and some nodes (“C” and “D”) contain, in addition, variance calculation. Node “C” calculates the motion variance of the available neighboring macroblocks. Assuming N avail-

able neighbors results in N vectors in each direction (X and Y). This yields $O(N)$ arithmetic operations. Node “D” calculates the variance of the pixel values along the boundaries of

the missing block. We assume M available pixels, thus node “D” results in $O(M)$ arithmetic operations. Since the longest path includes both “C” and “D” nodes, the total computation load is estimated by $O(\max(N + M))$ arithmetic operations.

The decision algorithm is performed on each missing block/macroblock in the video sequence, and the amount of the damaged data depends on the channel-noise factor. After a decision is made, the block/macroblock is concealed according to the chosen method.

5. EXPERIMENTS AND RESULTS

The video sequences used in this work are “Train,” “Foreman,” and “Ruby,” which are all originally AVI movies. The image frame size for all movies were scaled to 240×320 pixels. Decoded MPEG-2 frames are used as the original video frames, on which noise is added.

The thresholds used in the decision-support-tree-based algorithm may significantly affect the performance. Therefore, empirical determination of these thresholds is needed at the first stage. Then, the following experiments are performed using four different error concealment approaches: (a) comparing the quality of the reconstructed video for different frame types within a GOP (group of pictures); (b) evaluation of the effect of different motion levels (speed) on the resulting quality; (c) testing the threshold effect while using different thresholds per video sequence, and the same threshold set (average thresholds) for all the sequences; (d) visual inspection of the four error concealment schemes.

5.1. Determination of the decision tree thresholds

The proposed decisions-support algorithm uses four kind of thresholds: SV—spatial variance, TV—temporal variance, ML—motion level, which is the square of the correctly received motion vector value in a damaged block/MB, and DL-DCT level. Determination of these thresholds significantly affects the performance. We used a training set of several video sequences degraded with different kinds of errors in order to determine these thresholds empirically. This was done in two stages: (1) we first fixed the thresholds ML and DL and performed tests in order to find the optimum SV and TV pair (in the range of 500–3000 for SV and 2–30 for TV), (2) then using these chosen values for SV and TV, we determined the ML and DL thresholds (in the range of 0–128 for ML and 40–64 for DL). Different sets of thresholds were found for each video sequence. Once the four thresholds were determined, the performance of the decision-support algorithm was evaluated by comparing the results to both temporal and spatial error concealment methods. The test set used for evaluation was composed of 15 different degradation versions of the original video sequences. For most of the cases, the proposed algorithm yields better quality, using the VPDM. In addition to using the specific-scaled thresholds for each movie, we assess the quality achieved by using one common set of thresholds for the three sequences, for performance evaluation.

5.2. Comparing the different error concealment methods

In this section we present a comparison of four error concealment methods: spatial, temporal, hybrid [6], and the proposed decision-support-based algorithm. The comparison is carried out using different experiments with different noise levels. Two criteria for quality measurement, the PSNR and P_VPDM, are used here for performance evaluation. Both measures are normalized-log functions of the basic criteria MSE and VPDM. Since their results are given in a similar manner (proportional to the quality, and in dB) and they are both improvements of the basic criteria, we chose to present our results using these two quality measures.

5.2.1. Concealment of degraded compressed video sequences

The purpose of this experiment is to evaluate the quality of the reconstructed video sequence using different error concealment approaches. In particular we investigate here the efficiency of the error concealment for I -frames and for P - and B -frames. Along with the frame type, also the effect of the frame position within the GOP is investigated.

The test set includes nine degradation versions of each of the three original video sequences: “Ruby,” “Foreman,” and “Train.” For each video sequence, a GOP of 13 raw data frames is chosen, with order of $IBBPBBPBBPBB$, and the frame size is 320×240 .

Figure 6 shows the average quality of the three video sequences along a group of pictures (GOP) containing 13 frames. The influence of the frame type and its position within a GOP on the image quality is very clear. All the I -frames yield very high quality. The quality of the proceeding frames deteriorates until the last B -frame, with some small peaks of improvement where P -frames are found. This is a result of the compression ratio for each frame and the error propagation along a GOP until the next independent I -frame [1].

The proposed error concealment scheme outperforms all the other three methods for all the three tested video sequences. Similar results are achieved for I -frames by the temporal, spatial, and hybrid algorithms. The improvement achieved by the proposed algorithm for I -frames is due to its ability not to conceal specific degraded blocks. This gives our algorithm a head start since I -frames are used as reference frames for the entire GOP.

5.2.2. Motion speed effect

The affect of motion level is simulated here using some kind of video transrating by dropping frames. We define 6 frame-rate levels by dropping 0–5 consecutive frames from the original sequence (level 1 represents the original sequence, level 2 the sequence built from every second frame, etc.). We assume that the frame rate somehow reflects the motion level in the given sequence. However, skipping a frame does not mean accelerating the motion by a factor of 2.

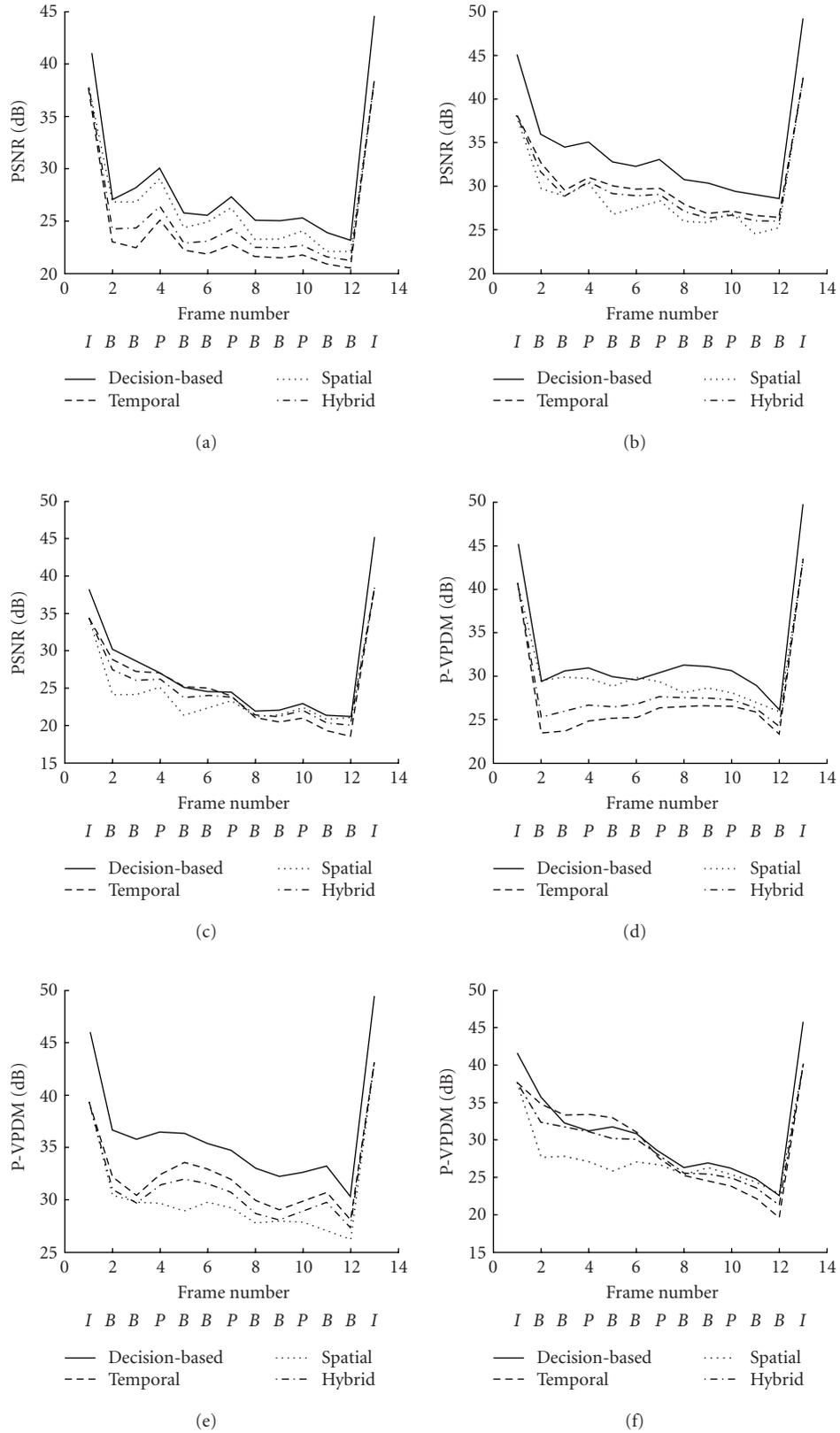


FIGURE 6: Average PSNR and P_{-VPDM} along a GOP for the different error concealment methods: (a) PSNR results of “Ruby” stream; (b) PSNR results of “Foreman” stream; (c) PSNR results of “Train” stream; (d) P_{-VPDM} results of “Ruby” stream; (e) P_{-VPDM} results of “Foreman” stream; (f) P_{-VPDM} results of “Train” stream.

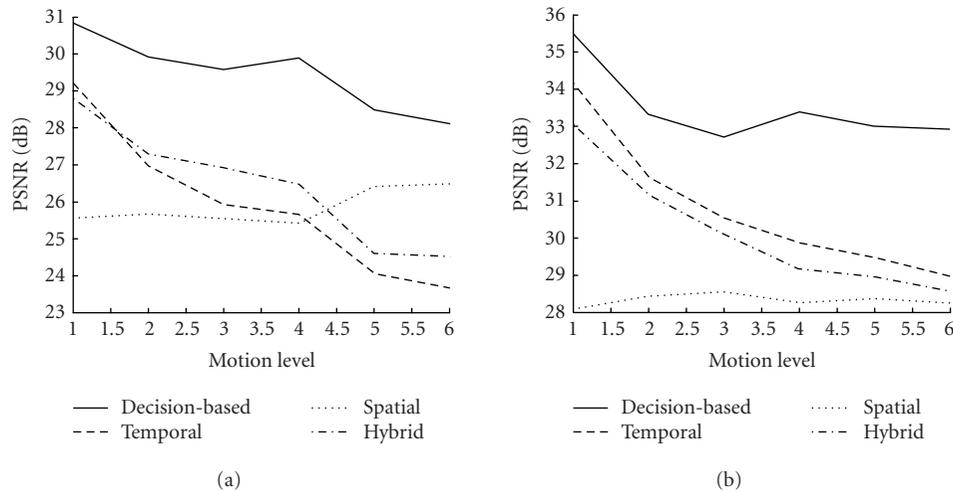


FIGURE 7: Error concealment on degraded video streams with various motion levels. Average PSNR versus motion level for (a) “Ruby” stream and (b) “Foreman” stream.

The results of error concealment for various motion levels using the PSNR measure are illustrated in Figure 7. It can be seen that the quality achieved by the temporal, hybrid, and decision-support-based error concealment methods tend to decrease as the motion level increases. This is not the case for the spatial concealment, since it does not depend on any temporal information. The proposed algorithm achieved the best quality in terms of PSNR and P_VPDM (greater than ~ 30 dB) compared to the other three concealment techniques, for all motion levels. The spatial approach yields the worst concealment quality for relatively slow motion.

5.2.3. Specific thresholds per video sequence versus average common thresholds

In order to find uniform thresholds, an attempt to average the different sets of thresholds selected for each video sequence was done. The robustness of the proposed algorithm to noise level was tested for both specific and average thresholds and the results were compared to the spatial and temporal techniques. Figure 8 demonstrates the results in term of PSNR as a function of the noise level. As expected, the video quality deteriorates as the noise level rises. This degradation is not necessarily monotonic since we simulate the noise level by the probability of block loss. In addition, for each simulated video degradation, the losses occur on different blocks, which result in different degradation, depending on the block type and content. The proposed algorithm achieves the best quality for both specific and average thresholds compared to the other two methods. Although the algorithm perform better using the specific thresholds, the average set of thresholds yields satisfactory results as well.

5.2.4. Visual comparison of the error concealment scheme

Figure 9 depicts an original and degraded image of the “Ruby” video stream. The results of applying the four de-

scribed error concealment schemes (temporal, spatial, hybrid, and the proposed algorithm) are visually presented. The temporal concealment scheme, Figure 9d, produces the worst visible result, especially for consecutive block degradation which lasts till the end of the slice. However, the entire background is well reconstructed.

The spatial concealment yields a blurred image. The hybrid concealment algorithm results in a relatively good reconstruction of the background, but the dog remains damaged.

The proposed algorithm produces the best reconstruction of the main object in the image, although the background suffers from some blurred blocks. Since usually the background is of less importance to the human viewer and is not the focus of attention, we may assume that our algorithm results in the best error concealment, for a human viewer.

6. CONCLUSIONS

In this paper, we present a new hybrid decision-support algorithm for error concealment in digital compressed video streams. We developed a hybrid error concealment algorithm for the reconstruction of missing or damaged blocks and macroblocks at the decoder side. A new decision-support tree is developed to efficiently choose the best appropriate error concealment method. Performance evaluation is carried out by comparing the proposed algorithm to three different error concealment schemes: temporal, spatial, and hybrid concealment, using various types of compressed video degradation. The proposed error concealment scheme outperforms all the other three methods for all the tested video sequences, and yields better results in terms of image quality. A unique visual interface is presented in order to visually illustrate the effect and the performance of the error concealment algorithm on each macroblock.

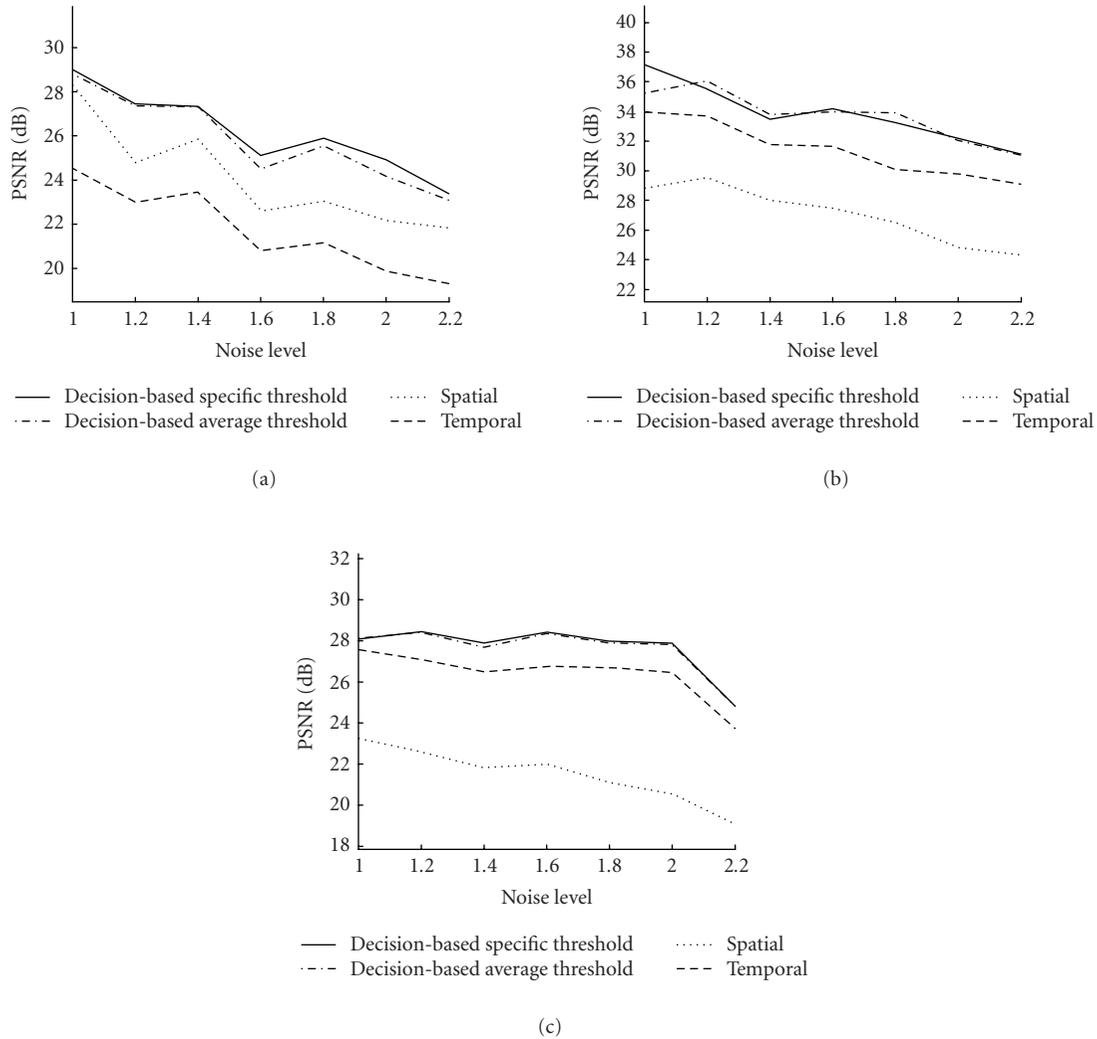


FIGURE 8: Average PSNR as a function of noise level using the proposed algorithm for (a) "Ruby" stream, (b) "Foreman" stream, and (c) "Train" stream.

Although, generally the spatial error concealment yields the worst image quality, in the presence of high motion speed it performs better than the other concealment methods. The hybrid scheme, which integrates both spatial and temporal techniques, is found to be the best approach. The additional ability not to conceal a specific degraded block contributes to the high concealment quality achieved by the proposed algorithm.

All the four different image quality criteria used for evaluation comply with optimal concealment type assigned to each macroblock by the proposed algorithm. The concealment results achieved by using the *improved* MSE criteria are very similar to the concealment achieved by our decision-support-tree hybrid algorithm. Future extension of this research will suggest employing human visual system (HVS) measures for more efficiency threshold determination as well as for performance evaluation.

In this work, we assume that error locations are known, and that the remaining information of a degraded block is usable. However, this is not always the case and therefore it will be interesting to develop error detection algorithm which works alongside the video decoder. Further improvement of the proposed decision-based hybrid algorithm can be achieved by refining the thresholds and employing adaptive intrarefresh (AIR) techniques that provide resilient coding.

ACKNOWLEDGMENTS

This work is a part of the STRIMM Consortium, sponsored by the Israeli Chief Scientist, Ministry of Trade and Industry, Israel. The authors would like to thank Mr. Yuval Kenan and Mr. Oren Peles for their valuable contribution to this research.

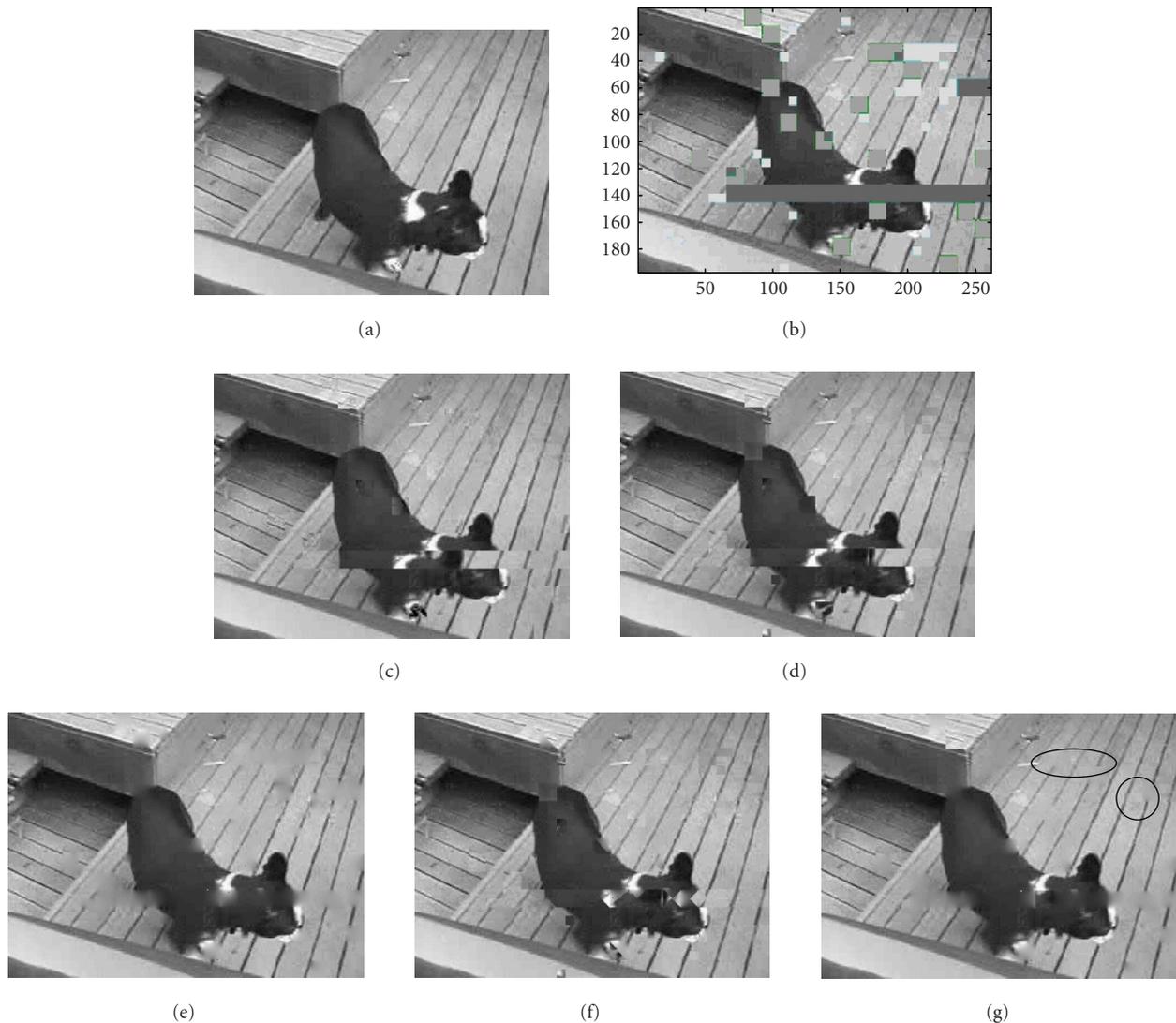


FIGURE 9: Visual error concealment for “Ruby” video stream: (a) original image; (b) degradation scheme of the original image; (c) damaged frame and reconstructed frame using (d) temporal, (e) spatial, (f) hybrid, and (g) the proposed concealment scheme.

REFERENCES

- [1] O. Hadar, R. Huber, M. Huber, and R. Shmueli, “Quality measurements for compressed video transmitted over a lossy packet network,” *Optical Engineering*, vol. 43, no. 2, pp. 506–520, 2004.
- [2] Y. Wang and Q.-F. Zhu, “Error control and concealment for video communication: a review,” *Proc. IEEE*, vol. 86, no. 5, pp. 974–997, 1998.
- [3] B. W. Wah, X. Su, and D. Lin, “A survey of error-concealment schemes for real-time audio and video transmissions over the internet,” in *Proc. International Symposium on Multimedia Software Engineering*, pp. 17–24, Taipei, Taiwan, December 2000.
- [4] W. Luo and M. El Zarki, “Analysis of error concealment schemes for MPEG-2 video transmission over ATM based networks,” in *Proc. SPIE Conference on Visual Communications and Image Processing*, vol. 1605, pp. 1358–1368, Taipei, Taiwan, May 1995.
- [5] J. Y. Park, M. H. Lee, and K. J. Lee, “A simple concealment for ATM bursty cell loss,” *IEEE Trans. Consumer Electron.*, vol. 39, no. 3, pp. 704–710, 1993.
- [6] C. Dovrolis, D. Tull, and P. Ramanathan, “Hybrid spatial/temporal loss concealment for packet video,” in *Proc. 9th International Packet Video Workshop*, New York, NY, USA, May 1999.
- [7] Q.-F. Zhu, Y. Wang, and L. Shaw, “Coding and cell-loss recovery in DCT-based packet video,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 3, no. 3, pp. 248–258, 1993.
- [8] M. Wada, “Selective recovery of video packet loss using error concealment,” *IEEE J. Select. Areas Commun.*, vol. 7, no. 5, pp. 807–814, 1989.
- [9] O. Hadar, M. Huber, R. Huber, and A. Stern, “MTF as a quality measure for compressed images transmitted over lossy packet network,” *Optical Engineering*, vol. 40, no. 10, pp. 2134–2142, 2001.
- [10] F.-H. Lin, W. Gass, and R. M. Mersereau, “Video perceptual distortion measure: two-dimensional versus

three-dimensional approaches,” in *Proc. IEEE International Conference on Image Processing (ICIP '97)*, vol. 3, pp. 460–463, Santa Barbara, Calif, USA, October 1997.

- [11] F.-H. Lin, W. Gass, and R. M. Mersereau, “Vision model based video perceptual distortion measure for video processing and applications,” in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP '97)*, vol. 4, pp. 3133–3136, Munich, Germany, April 1997.
- [12] S. Cen and P. C. Cosman, “Decision trees for error concealment in video decoding,” *IEEE Trans. Multimedia*, vol. 5, no. 1, pp. 1–7, 2003.
- [13] S. S. Hemami and T. H.-Y. Meng, “Transform coded image reconstruction exploiting interblock correlation,” *IEEE Trans. Image Processing*, vol. 4, no. 7, pp. 1023–1027, 1995.
- [14] M. Ancis, D. D. Giusto, and C. Perra, “Error concealment in the transformed domain for DCT-coded picture transmission over noisy channels,” *European Transactions on Telecommunications*, vol. 12, no. 3, pp. 197–204, 2001.
- [15] J. O. Limb, “Distortion criteria of the human viewer,” *IEEE Trans. Syst., Man, Cybern.*, vol. 9, no. 12, pp. 778–793, 1979.

Ofer Hadar received the B.S., the M.S. (cum laude), and the Ph.D. degrees from the Ben-Gurion University of the Negev, Israel, in 1990, 1992, and 1997, respectively, all in electrical and computer engineering. The prestigious Clore Fellowship supported his Ph.D. studies. His Ph.D. dissertation dealt with the effects of vibrations and motion on image quality and target acquisition. From August 1996 to February 1997, he was with CREOL at Central Florida University, Orlando, Fla, as a Research Visiting Scientist. From October 1997 to March 1999, he was a Postdoctoral Fellow in the Department of Computer Science, the Technion – Israel Institute of Technology, Haifa. Currently he is a faculty member at the Communication Systems Engineering Department, Ben-Gurion University of the Negev. His research interests include image compression, video compression, routing in ATM networks, flow control in ATM network, packet video, transmission of video over IP networks, and video rate smoothing and multiplexing. Hadar is a Member of the IEEE and SPIE.



Merav Huber received her B.S. degree in electrical engineering from the Technion–Israel Institute of Technology, Israel, in 1999, and the M.S. degree in electrical engineering from the Electrical Engineering Department, Ben-Gurion University of the Negev, Israel, in 2004. She has served as an Academic Professional Officer in the Israel Air Force, and is currently working toward her Ph.D. degree in the Communication Systems Engineering Department, Ben-Gurion University of the Negev, where she is a Teaching Assistant.



Revital Huber received her B.S. degree in electrical engineering from the Technion–Israel Institute of Technology, Israel, in 1999, and the M.S. degree in electrical engineering from the Electrical Engineering Department, Ben-Gurion University of the Negev, Israel, in 2004. She has served as an Academic Professional Officer in the Israel Air Force, and is currently working toward her Ph.D. degree in the Communication Systems Engineering Department, Ben-Gurion University of the Negev, where she is a Teaching Assistant.



Shlomo Greenberg received his B.S. degree, M.S. degree (cum laude), and his Ph.D. degree in electrical and computer engineering from the Ben-Gurion University of the Negev, Beer-Sheva, Israel, in 1976, 1984, and 1998, respectively. He has been employed with the IAEC/NRCN (Israel Atomic Energy Commission, Nuclear Research Center—Negev), Israel, from 1979 to 1999, and with Motorola Semiconductor Israel since May 2000. Currently he is a faculty member at the Communication Systems Engineering Department, Ben-Gurion University of the Negev. His research interests include computer vision, image and video compression, transmission of video over IP networks, video rate smoothing and multiplexing, signal and image processing, automatic target detection, pattern recognition, neural networks, and fuzzy logic. He is a Member of the IEEE.

