MPEG-2 Error Concealment Based on Block-Matching Principles

Sofia Tsekeridou, Student Member, IEEE and Ioannis Pitas, Member, IEEE

Abstract—The MPEG-2 compression algorithm is very sensitive to channel disturbances due to the use of variable-length coding. A single bit error during transmission leads to noticeable degradation of the decoded sequence quality, in that part or an entire slice information is lost until the next resynchronization point is reached. Error concealment (EC) methods, implemented at the decoder side, present one way of dealing with this problem. An error concealment scheme that is based on block-matching principles and spatio-temporal video redundancy is presented in this paper. Spatial information (for the first frame of the sequence or the next scene) or temporal information (for the other frames) is used to reconstruct the corrupted regions. The concealment strategy is embedded in the MPEG-2 decoder model in such a way that error concealment is applied after entire frame decoding. Its performance proves to be satisfactory for packet error rates (PER) ranging from 1% to 10% and for video sequences with different content and motion and surpasses that of other EC methods under study.

Index Terms—Block matching, error concealment, MPEG-2 compression.

I. INTRODUCTION

In the prospect of digital TV (SDTV/HDTV) or multimedia applications, a wide range of video compression standards has emerged to satisfy their numerous requirements, mainly for low bit rates and good quality. The MPEG-2 compression standard is one of them, developed mainly for digital TV/HDTV broadcasting applications. Its applications further extend to video-on-demand (VOD), computer multimedia, video conferencing, electronic cinema, electronic news gathering, storage, direct broadcast satellite (DBS), or digital video broadcast (DVB). MPEG-2 [1]–[4] is a lossy algorithm that is based on motion-compensated DCT block transform and variable-length coding allowing compressed bitstreams at rates close to 5 Mb/s for NTSC/PAL quality. It exhibits a hierarchical structure: sequence, group of pictures (GOP), picture (I-, P-, or B-frames), slice, macroblock (MB), and block. Transmission of MPEG-2 compressed bitstreams through various transmission media (satellite or terrestrial) involves tasks such as packet-format specification, multiplexing, network transport, and media synchronization [3]–[5]. A complete systems approach for digital video is presented in the MPEG-2 systems standard, which, among others, defines the MPEG-2 transport stream mainly intended for use in noisy channels [4]. Each MPEG-2 transport packet consists of 188 bytes, four of which are header information and the remaining are payload (parts of variable-length PES packets). The header contains a 1-bit transport-error-indicator which indicates, if appropriately set by the system demultiplexer or an error detection/resilience mechanism [4], [6], [7], if the respective packet contains errors or not. An alternative systems approach is asynchronous transfer mode (ATM) transport, used mainly in broadband ISDN (B-ISDN) networks [4]. ATM cells with optional ATM adaptation layer (AAL) protocol functionalities are 53-bytes long with a 5-byte header and a 48-byte payload, 1 byte of which is reserved for AAL. The MPEG-2 transport packet of 188 bytes can fit into four ATM/AAL cells, thus ensuring MPEG-2 and ATM interoperability.

Transmission through physical communication channels is liable to errors, being either packet/cell losses or bit errors (isolated or in bursts), referred to as bitstream errors. They frequently result in loss of the decoder synchronization, which becomes unable to decode until the next resynchronization point. Since resynchronization points are placed at the header of each slice by the MPEG-2 coder, unless otherwise specified, transmission errors may lead to partial or entire slice information loss and to severe degradation of the currently decoded frame quality, as well as of that of the subsequent decoded frames in a GOP due to temporal error propagation. Error resilience may be achieved in three ways: 1) modifications to the encoding process to increase robustness to errors; 2) error detection and control, entropy-coding, and robust packetization prior to or during transmission; and 3) modifications to the decoding process and use of error-concealment (EC) methods at the decoder. During encoding, use of smaller or adaptive [8] slice sizes to increase the number of synchronization points in the bitstream, fractal-based [9], hierarchical pyramid-based [10], or wavelet-based [11] coding instead of the common DCT-based one, use of overlapped block motion compensation [11], a combination of VLC and FLC codes in the bitstream [12], or use of frequency scanning and MU VLC [13] instead of block scanning and VLC present methods that fall into the first error resilience category. This category further includes methods that implement a 9 × 9 with one pixel overlap block DCT [14], [15] or use layered coding (different scalability bitstreams) in combination with different prioritization transmission levels [5], [13], [16]–[21] in one or more channels. All of these methods, however, result in significant overhead in the total bit rate or increased encoder complexity. The error control and detection schemes involve use of forward error correction (FEC) or Reed-Solomon (RS).
coding/decoding [16]–[18], [22], [23] or use of the video-specific AAL protocol [4], [5], [13] in ATM networks. Such error-correction schemes, though, add redundancy to the bitstream. In order to reduce such redundancy, error-resilient entropy coding (EREC) methods have been used instead [10], [16], [21], [24], [25]. At the transmission level, error resilience can also be accomplished by robust packetization such as interleaved packetization [16], [25] or use of different prioritization levels along with alternative packing schemes [15], [16], [26]. Even when either or both of the first error-resilience methods are employed, some errors may still remain uncorrectable (commonly known as residual errors). In such cases, error-resilience methods belonging to the third category are employed. Forcing the decoder to resync [5]–[7], [18], [27]–[29] before the next resynchronization point in the bitstream might lead to correct decoding of part of otherwise lost frame area. EC methods attempt to reconstruct lost areas by exploiting spatial and/or temporal redundancy. They are distinguished to spatial, temporal, or spatio-temporal methods depending on the available information exploited for concealment. Spatial EC methods [5]–[7], [9], [14], [16], [20]–[22], [26]–[28], [30]–[42] are mainly targeted for intra-frame concealment ranging from simple bilinear interpolation ones to complex fuzzy reasoning-based ones. Temporal methods [6], [7], [12], [16], [20]–[22], [26]–[28], [31], [33], [37], [38], [40], [43]–[48] can be used for both inter- and intra-coded concealment, provided that temporal information is available. Spatio-temporal methods [5], [6], [15], [16], [22], [33], [49], [50] are based on measures of spatial and temporal activity for determining the type of concealment. The performance of these EC methods deteriorates for higher error rates. In such cases, layered coding combined with such methods [5], [20] gives better results, with a possible increase in bit rate and complexity.

The present study focuses on the problem of EC assuming the existence of appropriate error detection. The presented EC scheme consists of a spatial method, called the split-match EC method, for the first intra-frame concealment, since no temporal information is available for it, and a temporal one, called the forward–backward block-matching EC (F–B BM EC) method, for concealing the other frames of an MPEG-2 coded video sequence. The EC scheme is embedded in the MPEG-2 decoder model and is based on block-matching (BM) principles proving to be of improved performance compared to other EC’s under study, even for high packet error rates (PER’s) and sequences of varying content and motion. Care is taken to keep its complexity as low as possible for real-time capabilities. The structure of the paper is as follows. In Section II, the types of errors observable in a decoded video sequence are explained, for better problem understanding. Sections III and IV describe the two EC methods comprising the presented EC scheme. Alternative approaches to spatial EC based on anisotropic diffusion principles are also presented in Section III. A short description of some known EC methods that are compared with the presented ones is also given in Sections III and IV. Section V summarizes the simulation results. Concluding remarks are outlined in Section VI.

II. TYPES OF ERRORS

Transmission errors lead to two types of errors observable at the decoded video sequence.

1) Bitstream errors: result from direct bit errors or cell/packet loss in the transmitted bitstream and may lead to loss of the decoder synchronization.

2) Propagation errors: caused due to predictive coding. Since temporal prediction is applied in P- and B-frames, such errors are observed only in these frames. Errors in previously decoded anchor frames (I- or P-frames) propagate to subsequent in coding order frames.

Appropriate transport structure (MPEG-2 transport packets) and error-detection mechanisms (ATM/AAL, channel/source error detection) enable the decoder to decide on the spatio-temporal locations of corrupted by bitstream error blocks. Assuming erroneous packets are discarded, the errors lie at the end of a damaged slice between the erroneously received packet and the next resynchronization point in the bitstream [27], [42]. The knowledge of the locations of bitstream error corrupted blocks in anchor frames and the fact that, at the locations of propagation errors, all MB information is available at the decoder (i.e., motion vectors, MB coding modes, prediction errors) enables the localization of propagation errors in predicted frames.

In detail, if the motion vector and coding mode of an MB in a predicted frame are such that, for the decoding of this MB, information is obtained from a damaged MB in its reference frame, then the current MB is most probably partially or entirely corrupted due to propagation of the reference frame errors. It is worth noting that concealment of propagation errors is unnecessary if a good EC method is implemented in the decoder model. Even then, the knowledge of propagation error locations can prove useful in estimating and post-processing the propagation of errors due to imperfect concealment from previously reconstructed frames.

III. FIRST INTRA-FRAME EC

The best possible concealment of an I-frame is most desirable since imperfect reconstruction of I-frames result in propagation of imperfect concealment errors to all frames of the respective GOP and finally in the degradation of the video sequence visual quality. Intra-frame concealment is performed by exploiting mere spatial information from retrieved neighboring regions, since no temporal information is available. The use of bilinear interpolation between border pixels of neighboring good MB’s has been proposed in [22]. Since no left and right neighbors exist in our case, only the border pixels of the existing top \((MB_T)\), and bottom \((MB_B)\), MB’s are used to conceal the currently lost one \(MB_C\) (spatial vertical interpolation, SVI EC). Concealment is achieved by evaluating the expression

\[
MB_T(i,j) = \frac{dB}{d_{max}}MB_T(N_i,j) + \frac{dx}{d_{max}}MB_B(1,j).
\]

In the above equation, \(d_T\) and \(d_B\) represent distances in pixels of pixel \((i,j)\) from its vertically neighboring top and bottom border pixels respectively. \(d_{max}\) is the vertical size of the lost regions in pixels, which may be equal to \(N_i\), the vertical dimension of a MB, if only one slice is lost. Such an interpolation scheme performs satisfactorily when little spatial detail exists or when
vertical edges or lines are attempted to be reconstructed. In all other cases, its performance deteriorates followed by noticeable blurring in the concealed regions.

More complicated interpolation methods or spatial techniques have also been proposed in the literature [9], [14], [16], [26], [30]–[32], [34]–[39], [41], [42] to deal with these drawbacks. They are, however, usually more computationally intensive than the bilinear interpolation approach. Novel methods for the concealment of lost MB’s in the first I-frame of an MPEG-2 compressed sequence are investigated in this paper. They attempt to further reconstruct spatial details, textures, lines or edges in almost all directions in a smooth and continuous way with less artifacts and blurring. The first method, called the spatial anisotropic diffusion EC method, applies a form of anisotropic diffusion [51] in such a way that neighboring image content is diffused into the lost region when respective gradients are of great values. A second method, called the spatial split-match EC method, an initial version of which was introduced in [40], attempts to spatially match top and bottom neighboring regions and conceals the region in the direction of the match by either copying, interpolating, or diffusing the content of the matched regions. It is noted that concealment is performed on an MB basis. A more detailed description of these methods follows.

A. Spatial Anisotropic Diffusion EC Method

Motivated by the work of [51], which used anisotropic diffusion principles for edge-preservation purposes, the inverse problem is examined for concealment applications. Since the gradients between lost areas and surrounding good areas have large values, “diffusion” of the surrounding image content to the lost areas must be implementable when gradients increase and must diminish when the latter decrease. The image intensity $I_{i,j}^n$ at pixel $(i,j)$ of a lost MB is determined by iteratively employing the expression [51]

$$I_{i,j}^{n+1} = I_{i,j}^n + \lambda [c_T \nabla T I + c_B \nabla B I]_{i,j}^n$$

(2)

where $\nabla_T I_{i,j}$ and $\nabla_B I_{i,j}$ are the intensity differences between the considered pixel $(i,j)$ and the respective border one from the existing top and bottom MB’s, respectively. Initial conditions $I_{i,j}^0$ for $(i,j)$ inside the lost MB region, are set to zero (no image content is assumed available in this region). Parameter $\lambda$ controls the stability of the numerical presentation (2) of the diffusion equation and is set equal to 0.15. Based on our latter remark, the conduction coefficients $c_T$ and $c_B$ are chosen such that they are proportional to the gradient values (intensity differences) and inversely proportional to the distance of the pixel $(i,j)$ from its top or bottom estimator. Thus, the conduction coefficients are determined by

$$c_T/B = 1 - \frac{d_{T/B}}{d_{\text{max}}},$$

(3)

d_{T/B} (i.e., $d_T$ or $d_B$) and $d_{\text{max}}$ are similar to the ones already defined in (1). Constant $K$ is used to reduce the effect of large intensity differences and to ensure smooth intensity diffusion. A value of 10.0 has been employed in our simulations. Total concealment is achieved with only a few iterations (in the order of ten). Thus, no additional delays are introduced in the concealment scheme, except for a slice delay, since the next slice should be decoded before the current can be concealed.

Anisotropic diffusion reduces the blurring effect of the spatial interpolation method, although it still fails to reconstruct areas with many details or edges and lines other than vertical ones. Its use in combination with the subsequently described EC method diminishes these drawbacks, as it is explained later on.

B. Spatial Split-Match EC Method

In order to deal with the problem of reconstructing lost regions with high spatial detail, texture or edges (other than vertical ones), spatial region similarity principles have been applied in existing neighboring regions and concealment has been performed in the direction defined by the detected similar regions. These concepts have been incorporated in the split-match EC method. The flow chart of its algorithm is shown in Fig. 1. Apart from the region similarity evaluation and concealment tasks, a split-match process is embedded in the algorithm, which decreases the size of the top and bottom neighboring regions to be matched when region matching in the previous step has failed. An overview of the implementation principles of each incorporated task is given subsequently.

Two approaches are adopted for neighboring region similarity evaluation. In the first approach, the mean absolute difference (MAD) between neighboring blocks $B_T$ and $B_B$ of size $b_x \times b_y$ is estimated

$$MAD = \frac{1}{b_x \times b_y} \sum_{x=0}^{b_x} \sum_{y=0}^{b_y} |B_T(x,y) - B_B(x,y)|,$$

(4)
The combination of blocks (referred to as the “best match”) that leads to either a \( MAD \) value smaller than a predefined threshold (use of constant thresholds) or to the minimum \( MAD \) compared to respective values obtained by matching equal-sized blocks located in a predefined search region is used to recover the image content of the lost image part in the direction of the match. Constant thresholds are employed in the first steps of the split-match process (large blocks: splitting occurs), whereas search regions are defined in its last steps [smallest block: no splitting occurs, see Fig. 2(c) and (d)], as will be explained later on. In the second approach, the statistical behavior of these blocks is evaluated by means of their mean \( m_T \) (top), \( m_B \) (bottom) and variance \( \sigma_T^2, \sigma_B^2 \) values. Similarity decisions are based on the validity of the expression

\[
(|m_T - m_B| < Q \sigma_T) \quad \text{and} \quad (|m_T - m_B| < Q \sigma_B), \tag{5}
\]

In (5), \( Q \) is a constant factor, set to 0.5 in our simulations, that controls the strictness of the similarity rule. Despite its simplicity, this decision rule combined with the split-match task leads to remarkably good results. Furthermore, it alleviates the disadvantage of constant thresholds of the first approach, since region “adaptive-like” thresholds are set by this method.

The split-match process has four steps, as shown in Fig. 2, where \( MB_C \) denotes the MB to be concealed and \( MB_T, MB_{TL}, MB_{TR}, MB_B, MB_{BL}, \) and \( MB_{BR} \) represent the available top, top-left, top-right, bottom, bottom-left, and bottom-right neighboring MB’s to \( MB_C \), respectively. In the first two steps, large neighboring blocks are defined for matching, whereas, at the last two steps, the minimum sized block \( 4 \times 4 \) is used and different search regions are defined.

Such an approach is adopted to ensure smallest processing time and best possible concealment. The initialization of this process with large BM ensures that big homogeneous or textured regions will be reconstructed directly, thus avoiding processing delays or computational effort. On the other hand, the satisfactory reconstruction of smaller homogeneous or textured regions, regions with small details, lines, and edges, requires the matching of smaller blocks. A brief description of the four steps of the algorithm follows.

1. **Step 1: Initialization—Maximum Block Size**

   Initialization is performed by attempting to match the largest vertically neighboring blocks \( b_1 \) and \( b_2 \) of size \( 16 \times 8 \) pixels, as shown in Fig. 2(a). If these regions are considered similar, entire MB concealment follows by copying, as shown later on. Otherwise, the algorithm proceeds to the second step.

2. **Step 2: First Splitting—Vertical Directions Only**

   The initial blocks of the 1st step are split into two smaller ones, \( b_1 \) and \( b_2 \) with \( t_1 \), \( t_2 \), respectively, of size \( 8 \times 8 \) (horizontal splitting) and matching is performed in the vertical direction only for each pair separately, i.e. \( b_1 \) with \( t_1 \) and \( b_2 \) with \( t_2 \) [see Fig. 2(b)]. Thus, concealment of smaller flat or textured regions (or regions at the borders of high spatial activity areas) is directly performed.

3. **Step 3: Second Splitting—Definition of Initial Search Region—Smaller Set of Directions**

   The blocks of Step 2 are further split to the smallest allowable ones of size \( 4 \times 4 \), denoted by \( b_i \) and \( t_i \). Now, though, an initial horizontal search region is additionally defined as shown in Fig. 2(c). The smallest blocks to be matched \( b_i \) with \( t_i \) slide inside the search regions and all
their combinations are considered for “best match” determination based on similarity decisions introduced previously. This step is performed iteratively until the next best combination of blocks mismatches or entire MB concealment is achieved. The concept of sliding windows ensures edge or line reconstruction at any direction.

- **Step 4: No Splitting—Enlargement of Search Region—Larger Set of Directions**

The search regions of Step 3 are further enlarged, as shown in Fig. 2(d), to allow the possibility of a larger set of reconstructed directions extending to neighboring left and right top or bottom MB’s. The block size remains the same and matching is performed between $l_k$ and $t_k$ shown in Fig. 2(d). This step is performed iteratively until entire MB concealment. It is noted that reconstruction of horizontal or almost horizontal lines or edges is impossible when using information from only top and bottom available MB’s, especially in cases where several consecutive slices are lost.

As soon as a “best match” has been found, concealment is performed by one of the following three methods: 1) copying; 2) diffusion; or 3) interpolation of the image content of the best matched blocks into the lost region in the direction of the match. Copying is performed in the way shown in Fig. 2. The image content of the top/bottom neighboring block of the “best match” is copied to the top/bottom part of the lost region in the direction of the match. Copying is performed in the way shown in Fig. 2. The image content of the top/bottom neighboring block of the “best match” is copied to the top/bottom part of the lost region in the direction of the match, respectively (e.g. in case blocks $b_2$ and $t_2$ of the first step match, the image content of $b_2$ is copied into the top equally sized region of $MB_C$, represented by a light grey color, and the image content of $t_2$ is copied into the bottom part of $MB_C$ represented by a darker grey color). In the first two steps of the split-match process, concealment is directly performed without the involvement of any iterative process. In its last steps, though, an iterative procedure is necessary to conceal the regions left unconcealed in the previous steps or iteration. For this purpose, concealed areas of the lost MB are labeled to enable unconcealed area identification and they are not concealed when, in a next step or iteration, they lie in the direction of the best match. Concealment by copying leads to blocking artifacts in the concealed areas. In order to smooth this blocking effect, the post-processing technique of [52] is used. This method is performed only on the concealed areas. The algorithm of [52] decides whether the current pixel belongs to a monotone or edge MB and adjusts its behavior accordingly. The decision is made on a bigger region than the size of a MB. However, only the MB considered is affected. For the monotone MB’s, smoothing is accomplished by filtering with a separable $2D \times 3 \times 3$ FIR filter with identical responses in the horizontal and vertical dimensions equal to $\{0.227, 0.547, 0.227\}$. For the edge MB’s, first the direction of the edge is estimated and then the blockiness is reduced by $1D$ FIR filtering in the edge direction with an impulse response of $\{0.036, 0.282, 0.363, 0.282, 0.036\}$.

Concealment by anisotropic diffusion is performed in the way described in Section III-A between border pixels of the best matched blocks. The determination of the pixels to be concealed located in the direction of the best match is performed in the way illustrated in Fig. 3. Those pixels, for which the inequality $x_{\text{min}} \leq x \leq x_{\text{max}}$ is valid, are concealed by employing anisotropic diffusion between them and the border pixels of the best match. In more detail, concealment is applied to those pixels whose distance from the $(0,0)$ point in Fig. 3 when it is projected on the line perpendicular to the direction of the best match (thus producing value $x$), lies between the smallest ($x_{\text{min}}$) and largest ($x_{\text{max}}$) such distances. The latter are evaluated with respect to the location of the best matched blocks. This type of concealment is introduced to eliminate the need of employing a post-processing technique for block artifact reduction, since diffusion does not introduce many significantly visible blocking artifacts.

Concealment by bilinear interpolation is performed as in the case of anisotropic diffusion, with the difference that image content in the lost region is restored by interpolating the image content of the border pixels of the best match using a formula similar to (1). Such concealment is introduced for comparison purposes. Because blocking artifacts introduced by copying are more noticeable in the third and fourth steps of the split-match process, concealment by diffusion or interpolation is employed only in these steps, whereas concealment in the first two steps is achieved always by copying.

**IV. INTRA-/INTER-FRAME EC**

Temporal information is available at the decoder for all but the first frame in a single scene of an MPEG-2 coded sequence. This is true even for intra-coded frames (I-frames), since previously decoded P-frames already exist in the decoder frame buffer for the decoding of the subsequent in coding order B-frames. Consequently, temporal EC methods can be used for their concealment which, in contrast to spatial ones, result in nearly perfect concealment in areas of no or little motion. Temporal EC methods exploit previously decoded temporal information to conceal lost regions in current frames. Simple temporal EC methods are temporal replacement [22], motion-compensated concealment [22], and adaptive temporal concealment [5], [50]. Temporal replacement replaces the lost MB with that in the same location of the previously decoded frame, stored in the decoder buffer. No motion information is exploited, which results in noticeable shifts in strong motion areas. This method is also known as zero motion EC (ZM EC).

Better results are obtained if motion information is exploited for concealment. The motion-compensated concealment (MC EC) estimates motion for the lost MB by averaging available motion vectors from neighboring MB’s based on the
Fig. 4. Introduction of the EC method in the MPEG-2 decoder model.

Fig. 5. Temporal matching of top and/or bottom available MB’s between current and reference frame/frames. Neighboring candidates are either existing top, bottom, or combined top and bottom MB’s.

assumption of uniform motion between all neighboring MB’s. For intra-coded neighbors, however, no motion vectors are available, thus leading to similar defects as the zero motion EC. A solution to this problem is the transmission of concealment motion vectors for intra-coded MB’s. These, however, lead to an overhead of about 0.7% of the total bit rate (for bit rates of 6–7 Mb/s) [5], [12] and transmission errors might lead to their loss as well. Furthermore, the assumption of smooth motion does not always apply for all neighboring MB’s. Adaptive temporal concealment (adaptive EC) is based on local measures of spatial and temporal activity. When spatial activity is high, motion-compensated temporal concealment is applied. When temporal activity is high, spatial interpolation is used. The method is applied to I-frames, while P- and B-frames are concealed by MC EC. Its performance is usually better than the other two, judging mainly from the perceived visual quality of the concealed sequences, with a small increase in complexity.

When motion uniformity does not apply for all neighbors of the lost MB or motion information is not available due to their intra-coding, none of the above mentioned methods performs satisfactorily. The temporal F-B BM EC method attempts to alleviate these drawbacks by selecting the neighbor that leads to the best possible concealment of the lost MB based on BM principles. The concealment scheme is embedded in the decoder model, as shown in Fig. 4, in such a way that concealment is performed after the entire frame decoding has been completed. If the concealed frame is a reference one, it is stored in the previous or future picture buffer of the decoder to be used for decoding and concealing the frames which are next in the coding order, otherwise it is displayed. Thus, only the concealment of damaged by bitstream error blocks is necessary.

The temporal F-B BM EC method exploits the temporal redundancy of an image sequence. The assumption of a smooth and uniform motion for at least some adjacent blocks is adopted. Temporal BM is performed between the top and/or bottom MB’s, MB_T and MB_B, respectively, of the lost MB MB_L and equally sized blocks located in a predefined search region in the reference frame/frames, its center being as shown in Fig. 5. Neighboring candidates for BM are either the available top neighbor MB_T (case A in Fig. 5), the available bottom one MB_B (case B), or the combination of both (case C). In the latter case, their vertical distance is restricted to 16 pixels (a MB’s vertical dimension) and the movement of

<p>| PSNR Values Measured on the First I-Frame of the Three Test Sequences, Achieved by the Spatial Concealment Methods Under Study |
|---------------------------------------------------------------|--------|--------|--------|--------|
| PSNR - Y Component - Different PER values                   |        |        |        |        |</p>
<table>
<thead>
<tr>
<th>EC Method</th>
<th>Flower Garden</th>
<th>Mobile &amp; Cam.</th>
<th>Tunnel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Error Free</td>
<td>32.82</td>
<td>32.21</td>
<td>35.30</td>
</tr>
<tr>
<td>S-M EC (Adaptive Thr.)</td>
<td>28.85</td>
<td>23.73</td>
<td>31.62</td>
</tr>
<tr>
<td>S-M EC (Constant Thr.)</td>
<td>28.57</td>
<td>23.21</td>
<td>31.63</td>
</tr>
<tr>
<td>SD EC</td>
<td>26.17</td>
<td>23.75</td>
<td>31.44</td>
</tr>
<tr>
<td>SVI EC</td>
<td>26.57</td>
<td>23.62</td>
<td>32.36</td>
</tr>
<tr>
<td>Errors</td>
<td>22.76</td>
<td>12.02</td>
<td>22.68</td>
</tr>
</tbody>
</table>

<p>| Average PSNR Values for the Three Test Sequences Achieved by the Temporal Concealment Methods Under Study for Different PER’s. The First Frame is Concealed by the Split-Match EC Method, Using Adaptive Thresholds and Anisotropic Diffusion |
|---------------------------------------------------------------|--------|--------|--------|
| Average PSNR - Y Component - Different PER values            |        |        |        |</p>
<table>
<thead>
<tr>
<th>EC Method</th>
<th>Flower Garden</th>
<th>Mobile &amp; Cam.</th>
<th>Tunnel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Error Free</td>
<td>27.95</td>
<td>27.95</td>
<td>35.51</td>
</tr>
<tr>
<td>F-B BM EC</td>
<td>27.42</td>
<td>25.15</td>
<td>33.78</td>
</tr>
<tr>
<td>Adaptive EC</td>
<td>26.52</td>
<td>23.27</td>
<td>33.24</td>
</tr>
<tr>
<td>MC EC</td>
<td>26.63</td>
<td>23.29</td>
<td>33.38</td>
</tr>
<tr>
<td>ZM EC</td>
<td>26.12</td>
<td>21.51</td>
<td>32.94</td>
</tr>
<tr>
<td>Errors</td>
<td>18.76</td>
<td>11.16</td>
<td>23.52</td>
</tr>
</tbody>
</table>

TABLE I

TABLE II
Fig. 6. First frame of the Flower Garden sequence: (a) error free; (b) erroneous, PER = 0.1. Concealed by: (c) S-M EC using adaptive thresholds and copying; (d) S-M EC using adaptive thresholds and anisotropic diffusion; (e) SD EC; and (f) SVI EC.

$MB_T$ and $MB_B$ in the search region is identical. Temporal matching is performed for all neighboring candidates in a predefined search region of size $N \times M$ and the one that leads to $MAD$ minimization is used to estimate the lost MB image content. Concealment is performed by copying into the lost MB area the image content of the appropriate neighbor of the temporal “best match.” For example, if the case A candidate minimizes the respective matching error at a certain location inside the search region, the immediately bottom $16 \times 16$ neighboring region of the temporal best match is copied into the lost area of equal size. The definition and separate matching of different neighboring candidates aims at dealing with cases where the smoothness of motion applies for some but not all neighboring blocks. Temporal matching proves also useful when no motion vectors are available for adjacent MB’s due to their intra-coding. In order to reduce the computational complexity and processing time, the search region, initialized in the reference frame/frames, is set bigger if temporal matching is performed between far distant frames and decreases if they are closer in time. Furthermore, a logarithmic search rather than a full search may be employed. For I- and P-frames, the matching is performed between the current and the past reference frames, while for B-frames, matching is performed between the current and the past as well as the future reference frames. The best
match, in the latter case may be a part of either the past or the future reference frame. The reference frames are already decoded and stored in the decoder buffers, even when I-frames require concealment. Consequently, no additional memory requirements are introduced.

An alternative approach is also considered for further processing time reduction, mainly for P- and B-frames, for which adjacent motion information is available at the decoder. Specifically, much smaller search regions of size $K \times L$ are initialized around their motion compensated ones in the reference frame/frames for the candidates of cases A and B in Fig. 5. Motion compensation is performed using the available motion vectors of adjacent blocks, as shown in Fig. 5. Motion vectors are used only if they exist. Motion vectors for I-frames may
be obtained by the respective ones of the previously decoded P-frames.

V. SIMULATION RESULTS

In order to evaluate the performance of the presented concealment scheme, three different 4:2:0 CCIR 601 sequences have been used, namely the Flower Garden (125 frames, large uniform motion, movement of background), the Mobile & Calendar (40 frames, moderate irregular motion, movement of background), and the Tunnel (100 frames, large irregular motion, static background) sequence. These have been coded by an MPEG-2 encoder at 5 Mb/s at 25 frames/s (PAL) using slice sizes equal to an entire row of MB’s, $N = 12$ (number of frames in a GOP) and $M = 3$ (number of frames between successive I- and P- or P- and P-frames). Layered coding (scalable profiles) has not been used. Motion compensation has been performed on a MB basis. MPEG-2 transport packets
are considered, which are 188-bytes long composed by 4 bytes of header information and 184 bytes of payload [4]. Errors may result from either isolated bit errors or packet loss. Appropriate error detection is assumed to be performed prior to concealment. Erroneous packets are discarded by the decoder until it regains synchronization. Two different PER’s are considered: 0.01 and 0.1, corresponding to medium and high error-rate cases. For higher PER values, none of the EC’s under study performs satisfactorily without the additional use of error control, coder modifications, or layered coding.

Performance evaluation of both methods of the concealment scheme has been performed based on PSNR values and perceived visual quality of the concealed sequence. Results from luminance concealment were exploited for further chrominance concealment. The split-match EC (S-M EC) method is compared with SVI EC, proposed in [22]. Constant and adaptive thresholds for region similarity decisions have been considered. “Optimal” constant thresholds have been experimentally estimated for the Flower Garden sequence by evaluating the perceived visual quality of the concealed frame, and have been ap-

Fig. 9. Frame 12 (I-frame) of the Mobile & Calendar sequence: (a) error free; (b) erroneous, PER = 0.1. Concealed by: (c) F-B BM EC; (d) adaptive EC; and (e) MC EC.
plied to the other two sequences as well. Results are presented in Table I for all test sequences measured on the first I-frame. Table I additionally includes results obtained by the spatial diffusion EC (SD EC). Although SVI EC and SD EC result in similar PSNR values, which are generally better than the ones achieved by the S-M EC method, they also lead to significant blurring in the concealed areas, thus producing an inferior visual quality of the concealed frame. The use of constant thresholds by the S-M EC achieves better PSNR values than when adaptive ones are employed, but they have the disadvantage of their manual estimation. However, it is seen that use of “optimal” constant thresholds for the Flower Garden in the concealment of the other sequences leads to satisfactory results. Small variations in PSNR values are observed for different types of concealment employed by the S-M EC for almost all cases listed in Table I. However, diffusion does not require post-processing of block artifacts as copying does, although the latter is simpler. Judging from the perceived visual quality of the concealed frames (Fig. 6), the S-M EC method achieves to reconstruct spatial edges at various directions and results in smooth continuation of the image content in the lost areas, when spatial information is sufficient for their retrieval. It is also observed that, al-

Fig. 10. Frame 90 (P-frame) of the Tunnel sequence: (a) error free; (b) erroneous, PER = 0.1. Concealed by: (c) F-B BM EC; (d) adaptive EC; and (e) MC EC.
though the use of adaptive thresholds might not lead to the best PSNR values, the subjective quality of the concealed frames is much better than that achieved using constant ones.

Average PSNR values evaluated on the test sequences concealed by the temporal F-B BM EC method and the temporal methods presented in Section IV are tabulated in Table II. The F-B BM EC method surpasses all the other temporal EC’s under study in almost all cases examined (different PER values and test sequences). This is further supported by the PSNR plots versus frame indices shown in Fig. 7 and the achieved visual quality of the concealed frames, examples of which are illustrated in Figs. 8–10. This significant improvement results from the fact that concealment in I-frames is much better performed when taking into account temporal information as well. It is generally agreed that an MPEG-2 coded video sequence quality is greatly affected by the visual quality of its I-frames, since they serve as anchor frames for the decoding of P and B ones. The fact that they are coded independently from other frames aids in delimiting the propagation of decoding errors from frames in a previous GOP to frames in the next one. Therefore, their best possible concealment, which implies use of motion information as the F-B BM EC method does, results in a remarkable improvement of the visual quality of the decoded sequence. Furthermore, the capability of the F-B BM EC method to deal with cases for which the assumption of smooth motion for all neighbors of a lost MB is not valid, or when intra-coding of adjacent MB’s leaves the decoder with no details about their motion, enhances its performance compared to the other temporal EC’s under study, since it attempts to locate the best possible candidate from the neighboring regions for concealment. Even if the adaptive EC decides on spatial concealment for such occasions, such concealment is worse than temporal one. The control of the search region size and the use of motion information, when available, assist in reducing the processing time of the F-B BM EC method. Block artifacts caused mainly in occluded regions, where mere past information is not sufficient (observed in I- or P-frames), are less visible for smaller PER values and become noticeable if PER values become greater than 10%. The F-B BM EC proves tolerant for PER values as high as 10%, while the other temporal EC’s under study cause more artifacts, which become noticeable even at frame rates of 25 frames/s.

VI. CONCLUSION

An EC scheme based on BM principles for MPEG-2 nonlayered coded video sequences is presented. It applies the spatial split-match EC method to the first I-frame of a sequence or a scene and uses the temporal F-B BM EC method to the other frames. The concealment scheme is embedded in the decoder model in such a way that concealment is performed after entire frame decoding. It is robust for PER’s as high as 10%, when other EC methods usually fail. The split-match EC method is based on spatial block similarity principles. When spatial information is adequate for lost region retrieval, it succeeds in reconstructing edges or lines of various orientations, textures and other spatial details in a smooth and continuous way. The F-B BM EC method is based on temporal block similarity principles and performs well when motion uniformity is not valid for all adjacent MB’s or when no motion information exists for adjacent MB’s (due to their intra-coding) cases, when the other temporal EC methods under study may fail.

REFERENCES


Sofia Tsekeridou (SM’96) was born in Lagadas, Thessaloniki, Greece, in 1973. She received the Diploma in electrical engineering in 1996 from Aristotle University of Thessaloniki, Thessaloniki, Greece, and is currently working toward the Ph.D. degree at the University of Thessaloniki, Thessaloniki, Greece.

During her studies, she was trainee at the University of Saarbruecken, Saarbruecken, Germany. She has also been a Visiting Researcher at Tampere University of Technology, Tampere, Finland. She is currently a Researcher and Teaching Assistant with the Department of Informatics, University of Thessaloniki. Her current research interests lie in the areas of signal, image, and video processing, pattern recognition, and computer vision.

Ms. Tsekeridou received a scholarship from the State Scholarship Foundation of Greece during her studies. She is a student member of SPIE and ACM and is a member of the Technical Chamber of Greece.

Ioannis Pitas (SM’94) received the Diploma in 1980 and the Ph.D. degree in 1985, both in electrical engineering, from the University of Thessaloniki, Thessaloniki, Greece.

Since 1994, he has been a Professor with the Department of Informatics, University of Thessaloniki. From 1980 to 1993, he served as Scientific Assistant, Lecturer, Assistant Professor, and Associate Professor in the Department of Electrical and Computer Engineering. His current research interests include the areas digital image processing, multidimensional signal processing, and computer vision, and he has published over 250 papers and contributed to eight books in these areas.

Dr. Pitas was an Associate Editor of *IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS* and Co-Editor of *MULTIDIMENSIONAL SYSTEMS AND SIGNAL PROCESSING*. He is currently an Associate Editor of *IEEE TRANSACTIONS ON NEURAL NETWORKS*. He was the Chair of the 1995 IEEE Workshop on Nonlinear Signal and Image Processing (NSIP’95) and has been a member of the European Community ESPRIT Parallel Action Committee. He has also been an invited speaker and/or member of the program committee of several scientific conferences and workshops.


