Error Concealment in MPEG Video Streams Over ATM Networks

Paul Salama, *Member, IEEE*, Ness B. Shroff, *Member, IEEE*, and Edward J. Delp, *Fellow, IEEE*

***Abstract*—When transmitting compressed video over a data network, one has to deal with how channel errors affect the decoding process. This is particularly a problem with data loss or erasures. In this paper we describe techniques to address this problem in the context of Asynchronous Transfer Mode (ATM) networks. Our techniques can be extended to other types of data networks such as wireless networks. In ATM networks channel errors or congestion cause data to be dropped, which results in the loss of entire macroblocks when MPEG video is transmitted. In order to reconstruct the missing data, the location of these macroblocks must be known. We describe a technique for packing ATM cells with compressed data, whereby the location of missing macroblocks in the encoded video stream can be found. This technique also permits the proper decoding of correctly received macroblocks, and thus prevents the loss of ATM cells from affecting the decoding process. The packing strategy can also be used for wireless or other types of data networks. We also describe spatial and temporal techniques for the recovery of lost macroblocks. In particular, we develop several optimal estimation techniques for the reconstruction of missing macroblocks that contain both spatial and temporal information using a Markov random field model. We further describe a sub-optimal estimation technique that can be implemented in real time.**

***Index Terms*—ATM, cell loss, cell packing, error concealment, motion vectors, Markov random field, spatial reconstruction, temporal reconstruction.**

I. Introduction

WHEN transmitting compressed video over a data network, one has to deal with how channel errors affect the decoding process. This is particularly a problem with data loss or erasures. In this paper we describe techniques to address this problem.

It is envisioned that one of the most important network applications will involve transmitting digital video [1]. During periods of network congestion, packets may be dropped, badly degrading the quality of the video, as a result of the missing data. Since retransmission is not a viable option for real-time multimedia applications, error concealment algorithms need to be developed. These algorithms estimate the missing data from the received video in an effort to conceal the effect of channel impairements. Furthermore, these algorithms will have to be integrated into the decoder hardware, and hence must be simple enough to be implemented in real time.

In this paper we derive error concealment techniques based on a Markov random field (MRF) model [2]–[4]. The algorithms are categorized as being either spatial (utilizing pixel data) or temporal (utilizing motion vectors) in nature. The spatial techniques rely on pixel data within a current damaged frame and an MRF model of the frame to restore damaged areas of the frame. Similarly, estimates for missing motion vectors are obtained by modeling the motion field as an MRF and finding the maximum *a posteriori* (MAP) estimate of each missing motion vector given its neighboring motion vectors. We also show that the widely used huerisitic technique based on averaging the motion vectors of neighboring macroblocks [5] is a special case of our MAP estimate. We also describe a temporal-spatial method for restoring damaged macroblocks based on the use of a ternary tree for classifying the available motion vectors and the MRF model for the corrupted frame.

The paper is organized as follows. The concealment of lost macroblocks resulting from ATM cell loss is addressed in Section III. The performance and comparison of our error concealment algorithms are provided in Section IV; and the conclusion of our work is given in Section V.

II. Previous work

Numerous schemes have been proposed in the literature to combat the effects of data loss on encoded video. Currently the approaches utilized for signal restoration/error concealment are either active concealment or passive concealment. In active concealment, error control coding techniques are used along with retransmission. Since extra data must be transmitted, it is sometimes necessary to reduce the source coder's data rate to avoid increasing network congestion. Active concealment has the advantage of permitting perfect reconstruction at the decoding end, if the amount of data lost is not significant, i.e., within the parameters of the error control coding scheme. In addition, unequal error protection can be provided by varying the number of bits used according to the priority of the data being protected [6], [7]. In passive concealment, the video stream is postprocessed to reconstruct the missing data. Although passive concealment does not result in perfect reconstruction of the lost data, it is necessary in many applications where error control coding cannot be used because of a high level of overhead, a problem with compliance with video transmission standards, or when active concealment itself fails. A general overview with an extensive bibliography of the various error concealment methods can be found in [8].

Manuscript received May 2, 1999; revised November 2, 1999. This work was supported in part by grants from the AT&T Foundation, Lucent Technologies, and by the National Science Foundation under Grant NCR-9624525.

P. Salama is with the Department of Electrical and Computer Engineering, Purdue School of Engineering and Technology, IUPUI, Indianapolis, IN 46202 USA (e-mail: salama@engr.iupui.edu).

N. B. Shroff is with the Network Engineering and Wireless Systems Research Lab, School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907-1285 USA.

E. J. Delp is with the Video and Image Processing Laboratory (VIPER), School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907-1285 USA (e-mail: ace@ecn.purdue.edu).

Publisher Item Identifier S 0733-8716(00)04336-5.

Since video sequences that have been compressed according to the current video compression standards [9]–[12] consist of DCT coefficients and motion vectors, one approach to error concealment over networks is to prioritize data as described in [13]–[20]. Encoded video data are segmented into low-priority data such as high-frequency DCT coefficients, and high-priority data such as addresses of blocks, motion vectors, and low-frequency DCT coefficients. In the event of network congestion, packets carrying low-priority data are discarded, while those carrying high-priority data are retained.

Alternative techniques for reducing the effect of packet loss were proposed in [16] and [21]. In [16] the data rate of the encoded bitstream is decreased when the packet loss rate increases. A similar approach for subband coded sequences is described in [21], wherein error correcting codes are employed to correct data lost due to packet loss. Since the use of error correcting codes increases the data rate, the source coder is throttled such that the ultimate data rate is equal to the data rate of the original unprotected video sequence. If the packet loss rate is too severe and the lost data cannot be retrieved by the error correcting code, the damaged frame areas are then filled in with data from the previous frame.

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 [17], [22]–[24]. Interleaving is performed either at the macroblock level [22], [23], or at the Slice level [17]. This has the advantage of distributing the error due to packet loss over the entire frame rather than localizing it, and hence aids passive concealment techniques that rely entirely on spatial data to restore the damaged regions. However, it increases decoder complexity.

Postprocessing (passive) techniques for the sake of error concealment utilize spatial data, or temporal data, or a hybrid of both [5], [24], [26]–[39]. Missing macroblocks can be reconstructed by estimating their low-frequency DCT coefficients from the DCT coefficients of the neighboring macroblocks [23], [29]–[31], by estimating missing edges in each block from edges in the surrounding blocks as proposed in [33], or by the method of projections onto convex sets [39] as described in [34].

An alternative to using spatial data for error concealment is to use motion compensated concealment [5] whereby the average of the motion vectors of neighboring macroblocks is used to perform concealment. We will show later that this is a special case of one of our motion compensated techniques.

Currently the periodic insertion of Resynchronization Markers (unique codes that limit the effect of data loss and establish synchronization between encoder and decoder) is being considered for the MPEG-4 and JPEG2000 standards [40], [41]. The resilience of MPEG-4 bitstreams is also increased by placing coded motion vector data prior to coded DCT coefficients. These two groups of data are separated by a field known as the Motion Boundary Marker (MBM) [40], [41]. The MBM is utilized by the decoder to discern whether coded motion vectors have been corrupted or not. Uncorrupted coded motion vectors are used to reconstruct macroblocks. However, if the motion vector data between two resynchronization markers have been corrupted, all the data between the two markers are discarded and the macroblocks constructed via temporal replacement from the previous frame. The robustness of the stream is further enhanced by using reversible variable length codes that can be uniquely forward or backward decoded, as well as by repeating important header information [40], [41].

III. Passive Error Concealment

The goal of passive error concealment is to estimate missing data. In the case of MPEG video, the objective is to estimate missing macroblocks and motion vectors. The underlying idea is that there is still enough redundancy in the sequence to be exploited by the concealment technique. In particular, in I-frames it is possible to have a lost macroblock surrounded by intact macroblocks that are used to interpolate the missing data. This is a result of the fact that macroblocks in I-frames can span across two packets. It also arises when the macroblocks in I-frames are interleaved prior to packing them into packets. In P- and B-pictures, it is possible to have entire rows of macroblocks missing. In this case, spatial interpolation will not yield acceptable reconstructions. However, the motion vectors of the surrounding regions can be used to estimate the lost vectors, and the damaged region reconstructed via motion compensated interpolation [26], [27].

Let X be an $N_1 \times N_2$ decompressed frame from an MPEG sequence, and let Y be the received version at the output of the channel. Due to channel errors, it is conceivable that Y may have missing data. Each transmitted picture consists of M macroblocks that have $N \times N$ pixels. Let $\boldsymbol{x}_i$ be the lexicographic ordering of the pixels in the ith macroblock in X. The vector $\boldsymbol{x}$ is then defined to be the concatenation of $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_M$, that is, $\boldsymbol{x} = [\boldsymbol{x}_1^T \ \boldsymbol{x}_2^T \ \cdots \ \boldsymbol{x}_M^T]^T$. The vector $\boldsymbol{y}$ is similarly defined for Y. If the jth macroblock is missing due to packet loss then $\boldsymbol{y} = \boldsymbol{D}\boldsymbol{x}$, where $\boldsymbol{D}$ is an $(N_1N_2 - N^2) \times N_1N_2$ matrix that consists of the identity matrix excluding the rows from row jN^2 to row $(j+1)N^2 - 1$. If n of the M macroblocks are missing due to packet loss, then $\boldsymbol{D}$ will be an $(N_1N_2 - nN^2) \times N_1N_2$ matrix. The goal of passive error concealment is to estimate $\boldsymbol{x}$ given the received data $\boldsymbol{y}$. A description of how this is done will be given in the following sections.

In Sections III-A–C, we use spatial restoration techniques to estimate missing pixel values from neighboring intact pixels. In particular, the original image is modeled as a Markov random field, (Section III-A), and the MAP estimate of the original image given the received image is then obtained (as described in Section III-A1). Since MAP estimation is computationally intensive, an approximation that can be implemented in real time [42] is shown in Section III-B. In addition, macroblock restoration based on estimating macroblock boundary pixels is presented in Section III-C.

We also consider the use of motion compensated concealment based on estimating missing motion vectors in Section III-D. In Section III-D1, missing motion vectors are estimated from the motion vectors belonging to surrounding intact macroblocks, while in Section III-D2, a missing motion vector is estimated based on motion vector data as well as pixel values.

A. Statistical Spatial Approach: MAP Estimation

Statistical techniques have been successfully used in image processing for edge reconstruction [4], [43]–[46]. The original image is modeled as a Markov random field (MRF) [2]–[4], and edges are reconstructed by maximum *a posteriori* (MAP) techniques. This is the approach adopted here. Each original frame X and its received version Y are modeled as discrete parameter random fields where each pixel is a continuous random variable. Assuming a prior distribution for $\boldsymbol{x}$, a maximum *a posteriori* (MAP) estimate is obtained given the received data $\boldsymbol{y}$. Denoting the estimate of $\boldsymbol{x}$ by $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}\,|\,\boldsymbol{y}=\boldsymbol{D}\boldsymbol{x}} L(\boldsymbol{x}\,|\,\boldsymbol{y})$, where $L(\boldsymbol{x}\,|\,\boldsymbol{y})$ is the log-likelihood function. In other words, $L(\boldsymbol{x}\,|\,\boldsymbol{y}) = \log f(\boldsymbol{x}\,|\,\boldsymbol{y})$, where $f(\boldsymbol{x}\,|\,\boldsymbol{y})$ is the conditional probability density function of $\boldsymbol{x}$ given $\boldsymbol{y}$. It is shown in Appendix I that

$$\hat{\boldsymbol{x}} = \arg \min_{\boldsymbol{x}\,|\,\boldsymbol{y}=\boldsymbol{D}\boldsymbol{x}} [-\log f(\boldsymbol{x})]. \tag{1}$$

Since X is modeled as a Markov Random Field (MRF), the probability density function of $\boldsymbol{x}$ is given by [2], [4], [47],

$$f(\boldsymbol{x}) = \frac{1}{Z} \exp\left(-\sum_{c \epsilon C} V_c(\boldsymbol{x})\right) \tag{2}$$

where Z is a normalizing constant known as the *partition function*, $V_c(\cdot)$ a function of a local group of points c known as cliques, and C the set of all cliques [4]. Using (1) and (2), the MAP estimate of $\boldsymbol{x}$ is then

$$\hat{\boldsymbol{x}} = \arg \min_{\boldsymbol{x}\,|\,\boldsymbol{y}=\boldsymbol{D}\boldsymbol{x}} \left[\sum_{c \epsilon C} V_c(\boldsymbol{x})\right].$$

In the following we will indicate our choice of the potential functions [4], and describe the process by which we obtain the MAP estimate.

1) The MAP Estimate: The proper choice of V_c is crucial in the reconstruction of the macroblocks. In this case, the potential functions are chosen such that

$$\sum_{c \epsilon C} V_c(\boldsymbol{x}) = \sum_{i=0}^{N_1-1} \sum_{j=0}^{N_2-1} \sum_{m=0}^{3} b_{i,j}^{(m)} \rho\left(\frac{D_m(X_{i,j})}{\sigma}\right) \tag{3}$$

where $D_0(X_{i,j}) = X_{i,j-1} - X_{i,j}, D_1(X_{i,j}) = X_{i-1,j+1} - X_{i,j}, D_2(X_{i,j}) = X_{i-1,j} - X_{i,j}$, and $D_3(X_{i,j}) = X_{i-1,j-1} - X_{i,j}$ are used to approximate the first-order derivatives at pixel i, j. $\rho(\cdot)$ is a cost function, σ a scaling factor, $b_{i,j}^{(m)}$ weighting coefficients, and the set of cliques [4] is $C = \{\{(i, j-1), (i,j)\}, \{(i-1, j+1), (i,j)\}, \{(i-1, j), (i,j)\}, \{(i-1, j-1), (i,j)\}$.

Several cost functions have been proposed [43], [45]. A convex $\rho(\cdot)$ results in the minimization of a convex functional. The cost function used here is the one introduced by Huber for obtaining robust M-estimates of location [48]. Its advantage is that it is convex, does not heavily penalize edges, and is simpler to implement than most of the convex cost functions used in the literature [45]. It is defined to be

$$\rho_\gamma(x) = \begin{cases} x^2 & |x| \le \gamma \\ \gamma^2 + 2\gamma(|x| - \gamma) & |x| > \gamma. \end{cases}$$

Hence,

$$\sum_{c \epsilon C} V_c(\boldsymbol{x}) = \sum_{i=0}^{N_1-1} \sum_{j=0}^{N_2-1} \sum_{m=0}^{3} b_{i,j}^{m} \rho_\gamma\left(\frac{D_m(X_{i,j})}{\sigma}\right).$$

Letting

$$h_\gamma(\boldsymbol{x}) = \sum_{i=0}^{N_1-1} \sum_{j=0}^{N_2-1} \sum_{m=0}^{3} b_{i,j}^{m} \rho_\gamma\left(\frac{D_m(X_{i,j})}{\sigma}\right)$$

we have

$$\hat{\boldsymbol{x}} = \arg \min_{\boldsymbol{x}\,|\,\boldsymbol{y}=\boldsymbol{D}\boldsymbol{x}} h_\gamma(\boldsymbol{x}). \tag{4}$$

The solution to (4) above can be obtained by means of the iterative conditional modes (ICM) algorithm [49]. In particular if the ith element of $\boldsymbol{x}$, denoted by x_i, corresponds to a lost pixel value, and $\boldsymbol{x}_{\partial i}$ denotes the neighborhood of x_i then

$$\hat{x}_i = \arg \max_{x_i} f(x_i \,|\, \boldsymbol{x}_{\partial i}). \tag{5}$$

Using (3) and (5), the MAP estimate of pixel (i, j), given its neighbors, is

$$\hat{X}_{i,j} = \arg \min_{X_{i,j}} \sum_{l=i}^{i+1} \sum_{k=j}^{j+1} \sum_{m=0}^{3} b_{l,k}^{m} \rho_\gamma\left(\frac{D_m(X_{l,k})}{\sigma}\right). \tag{6}$$

Let $\boldsymbol{X}_{\partial i}$ denote the neighborhood of the ith macroblock $\boldsymbol{x_i}$. The MAP estimate of $\boldsymbol{x_i}$ satisfies $\hat{\boldsymbol{x}}_{\boldsymbol{i}} = \arg\max_{\boldsymbol{x}} f(\boldsymbol{x_i} \,|\, \boldsymbol{X}_{\partial i})$. If we let $\boldsymbol{J_i}$ denote the set of indices of the pixels belonging to $\boldsymbol{x_i}$, then it can be similarly shown that

$$\hat{\boldsymbol{x}}_{\boldsymbol{i}} = \arg \min_{\boldsymbol{x_i}} \sum_{(i,j) \epsilon \boldsymbol{J_i}} \sum_{l=i}^{i+1} \sum_{k=j}^{j+1} \sum_{m=0}^{3} b_{l,k}^{m} \rho_\gamma\left(\frac{D_m(X_{l,k})}{\sigma}\right). \tag{7}$$

The solution to (7) can be obtained iteratively. This, however, is computationally intensive. Next we describe how we speed up the process of finding a solution by using median filtering techniques to obtain a suboptimal MAP estimate.

B. Median Filtering: A Suboptimal Approach

The choice of γ and σ is crucial to the reconstruction of edges. The smaller the product $\gamma\sigma$, the less the edges are penalized. Since $h_\gamma(\boldsymbol{x})$ is continuous, convex, and has continuous first partial derivatives, then by successively iterating with respect to each pixel, a global minimum is attained.[1] Using (6) we obtain

$$\begin{aligned} \frac{\partial}{\partial X_{i,j}} h_\gamma(\boldsymbol{x}) = {} & \frac{b_{i,j+1}^{0}}{\sigma} \rho'_\gamma\left(\frac{D_0(X_{i,j+1})}{\sigma}\right) \\ & + \frac{b_{i+1,j-1}^{1}}{\sigma} \rho'_\gamma\left(\frac{D_1(X_{i+1,j-1})}{\sigma}\right) \\ & + \frac{b_{i+1,j}^{2}}{\sigma} \rho'_\gamma\left(\frac{D_2(X_{i+1,j})}{\sigma}\right) \\ & + \frac{b_{i+1,j+1}^{3}}{\sigma} \rho'_\gamma\left(\frac{D_3(X_{i+1,j+1})}{\sigma}\right) \\ & - \sum_{m=0}^{3} \frac{b_{i,j}^{m}}{\sigma} \rho'_\gamma\left(\frac{D_m(X_{i,j})}{\sigma}\right) \end{aligned}$$

[1]There is, however, more than one global minimum.

where

$$\rho'_\gamma(x) = \begin{cases} 2x & |x| \le \gamma \\ 2\gamma & x > \gamma \\ -2\gamma & x < -\gamma. \end{cases} \tag{8}$$

Each pixel in the interior has eight neighbors. Let z_1, z_2, z_3, z_4, z_5, z_6, z_7, z_8 be the eight neighbors arranged in ascending order, and rename the associated weights $\{b^m_{l,k} \,|\, l = i, i+1, l = j, j+1, m = 0 \cdots 3\}$ as b_1, b_2, b_3, b_4, b_5, b_6, b_7, b_8. Defining

$$\begin{aligned} U = \{(k,l) \,|\, & k = i-1 \text{ and } l = j-1, j, j+1 \text{ or} \\ & k = i \text{ and } l = j-1\} \\ L = \{(k,l) \,|\, & k = i+1 \text{ and } l = j-1, j, j+1 \text{ or} \\ & k = i \text{ and } l = j+1\} \end{aligned}$$

and

$$\Delta_k(X_{i,j}) = \begin{cases} z_k - X_{i,j} & z_k \epsilon U \\ X_{i,j} - z_k & z_k \epsilon L \end{cases}$$

then

$$\begin{aligned} \frac{\partial}{\partial X_{i,j}} h_\gamma(\boldsymbol{x}) = & \frac{1}{\sigma} \sum_{k \epsilon L} b_k \rho'_\gamma \left(\frac{\Delta_k(X_{i,j})}{\sigma} \right) \\ & - \frac{1}{\sigma} \sum_{k \epsilon U} b_k \rho'_\gamma \left(\frac{\Delta_k(X_{i,j})}{\sigma} \right). \end{aligned} \tag{9}$$

Since, we are iterating for $X_{i,j}$, we need to solve

$$\frac{\partial}{\partial X_{i,j}} h_\gamma(\boldsymbol{x}) = 0. \tag{10}$$

When solving (10), three cases need to be considered.

Case 1: $|\boldsymbol{\Delta_k}| \le \gamma\sigma, \ \forall k$.

This occurs when $z_8 - \gamma\sigma \le z_1 + \gamma\sigma$, and the optimum value of $X_{i,j}$ satisfies the constraint: $|\Delta_k| \le \gamma\sigma \ \forall k$. Hence, $h_\gamma(\hat{X}_{i,j}) = \sum_{k=1}^{8} b_k (\Delta_k/\sigma)^2$. Using (10) and (8)

$$\hat{X}_{i,j} = \frac{\sum_{k=1}^{8} b_k z_k}{\sum_{k=1}^{8} b_k}. \tag{11}$$

Case 2: $|\boldsymbol{\Delta_k}| \le \gamma\sigma$ for some k.

We show in Appendix II, that the optimum estimate $\hat{X}_{i,j}$ satisfies

$$\hat{X}_{i,j} = \frac{\sum_{k=J_1}^{J_2} b_k z_k + \gamma\sigma \left[\sum_{k=J_2+1}^{8} b_k - \sum_{k=1}^{J_1-1} b_k \right]}{\sum_{k=J_1}^{J_2} b_k}$$

where J_1 and J_2 satisfy $z_{J_2} - z_{J_1} \le \gamma\sigma$, $J_2 \ge J_1$.

We fix σ and choose γ to be arbitrarily small and positive. Then according to Case 2A in Appendix II, if equal weights $\{b^m_{l,k} \,|\, m = 0 \cdots 3, \ l = i, i+1, \ k = j, j+1\}$ are used, one possible choice for the optimum value $\hat{X}_{i,j}$ will be the median of its neighbors, unless there are at least two neighboring pixels that are equal in value. Under such conditions, the common pixel value is used. Although this is a suboptimal strategy, the resulting reconstruction technique is faster than searching for the optimum value. The estimate of each missing pixel value is now obtained by finding the median of 8 values instead of performing line search techniques.

C. Estimation of Boundary Pixels

The above-mentioned techniques coupled with macroblock or Slice [9] interleaving are particularly useful in the restoration of intracoded macroblocks that belong to a frame that serves as the anchor frame for a new scene within the same sequence.

An alternative approach that is useful for reconstructing intracoded macroblocks when the current damaged frame and the previous reference frame belong to the same scene is next described. The underlying idea is to try to find a macroblock sized region in the previous frame, X^{-1}, that will maximize the MAP estimate of the boundary pixels of the missing macroblock given its neighbors.

Formally, suppose again that the ith macroblock $\boldsymbol{x}_i$ is missing. Let $\boldsymbol{X}_{\partial i}$ denote its neighboring macroblocks and let (m,n) denote the coordinates of the upper left corner of $\boldsymbol{x}_i$. Establish a search range $\mathcal{S}$ of $(2S+1) \times (2S+1)$ pixels in the previous frame X^{-1} centered at (m,n), that is, $\mathcal{S} = \{X^{-1}_{k,l} \,|\, k \epsilon [m-S, m+S], l \epsilon [n-S, n+S], k, l \text{ integers}\}$. Let $\boldsymbol{u}$ denote a macroblock sized region in $\mathcal{S}$, that is, $\boldsymbol{u} \subset \mathcal{S}$, and let $\boldsymbol{u}_B$ denote the boundary pixels of $\boldsymbol{u}$, then

$$\hat{\boldsymbol{x}}_{\boldsymbol{i}} = \arg\max_{\boldsymbol{u} \subset \mathcal{S}} f(\boldsymbol{u}_B \,|\, \boldsymbol{X}_{\partial i}).$$

Using the potential functions described above, this can be written as

$$\begin{aligned} \hat{\boldsymbol{x}}_{\boldsymbol{i}} = \arg\min_{\boldsymbol{u} \subset \mathcal{S}} & \sum_{(r,s) \,|\, X_{(r,s)} \epsilon \boldsymbol{u}_B} \sum_{l=r}^{r+1} \sum_{k=s}^{s+1} \sum_{m=0}^{3} b^m_{l,k} \rho_\gamma \\ & \cdot \left(\frac{D_m(X_{l,k})}{\sigma} \right). \end{aligned} \tag{12}$$

This technique is particularly useful for restoring I-frames that have been heavily damaged due to network packet loss.

In the following, we describe how we achieve motion compensated restoration by first estimating the missing motion vectors and then utilizing the estimates of the missing motion vectors to reconstruct the missing macroblocks.

D. Motion Compensated: Motion Vector Estimation

Most frames in an MPEG sequence are predicted frames that have motion vectors associated with their macroblocks. Let $\boldsymbol{v}_i$ be the motion vector associated with the ith macroblock in the current frame. In lossless transmission, $\boldsymbol{x}^{(0)}_{\boldsymbol{i}}$, the ith macroblock in the current frame, is reconstructed by the decoder as $\boldsymbol{x}^{(0)}_{\boldsymbol{i}} = \boldsymbol{x}^{(-1)}_{\boldsymbol{i}-\boldsymbol{v}_i} + \boldsymbol{n}_i$. Here $\boldsymbol{x}^{(-1)}_{\boldsymbol{i}-\boldsymbol{v}_i}$ is the macroblock in the reference frame that closely matches $\boldsymbol{x}^{(0)}_{\boldsymbol{i}}$, $\boldsymbol{n}_i$ is the error arising from having replaced $\boldsymbol{x}^{(0)}_{\boldsymbol{i}}$ by $\boldsymbol{x}^{(-1)}_{\boldsymbol{i}-\boldsymbol{v}_i}$, and $\boldsymbol{i}$ indicates the spatial coordinates of the ith macroblock. In lossy transmission, it is not possible to recover $\boldsymbol{n}_i$. The goal is to obtain an estimate for $\boldsymbol{v}_i$ that will point to $\boldsymbol{x}^{(-1)}_{\boldsymbol{i}-\boldsymbol{v}_i}$.

One approach to reconstructing a lost macroblock would be to estimate its associated missing motion vector by averaging the motion vectors of surrounding macroblocks [5], [14]. This motion vector estimate is then used to fill in the missing macroblock.

In the following we present alternative means of estimating the missing motion vector based on the use of Markov Random Field (MRF) models. In particular, in Section III-D1 we obtain the MAP estimate of the missing motion vector given its neighbors. We also describe an alternative technique for estimating missing motion vectors that incorporates spatial as well as temporal data in Section III-D2.

1) MAP Estimation of Motion Vectors: Utilizing the same MRF model described above to model each component of the motion field, we can obtain an MAP estimate of the missing motion vector given its neighboring motion vectors. In addition, using the results of Section III-B, we immediately infer that the median of the motion vectors of the surrounding macroblocks yields a suboptimal estimate of the missing motion vector. These estimates are then utilized to perform motion compensated restoration of the missing macroblock.

It is also evident from (11) that when $\sigma = 1, \gamma \to \infty$ (larger than the image dimensions suffices), $\boldsymbol{b}_k = 1$ for $k = 1, \cdots, 8$, and z_ks are motion vectors, the MAP estimate of the motion vector is the average of all surrounding motion vectors, as proposed in [5]. Thus, the median and average of the surrounding motion vectors are special cases of the MAP estimate.

2) Temporal-Spatial Approach: The above-described techniques for estimating missing motion vectors rely solely on temporal data (neighboring motion vectors). An alternative approach that uses both temporal and spatial data is described next. The use of both temporal and spatial data yields estimates for missing motion vectors that are more reliable, as will be presented in Section IV.

The alternative approach is based on using a ternary tree to classify the motion vectors neighboring the missing motion vector, and the MRF models of the image and motion field. Each neighboring motion vector is classified into one of nine classes according to whether each of its components are positive, negative, or zero. This is shown in Fig. 1, where $\boldsymbol{v}_x$ denotes the horizontal component of a motion vector and $\boldsymbol{v}_y$ its vertical component. The idea is to implicitly model the discontinuity in the motion field, and is similar to the approaches considered in [50] and [51], wherein the discontinuities of motion fields were modeled by means of binary MRF's.

After classifying all the neighboring motion vectors into their respective classes, we then determine the class or group to which the missing motion vector belongs. This is done by assigning a cost to each class and then choosing the class with the lowest cost. After the class with the lowest cost has been chosen, the motion vectors belonging to that class are modeled via the MRF and an MAP estimate of the missing motion vector obtained. If the MAP estimate is not unique, the motion vector that provides the macroblock with the "best" matching boundaries, given the neighboring macroblocks, is then chosen.

Formally, let $\mathcal{K} = \{K_i, i = 1 \cdots 9\}$ denote the set of 9 classes. Also let $\boldsymbol{v_t}, \boldsymbol{v_b}, \boldsymbol{v_l}$, and $\boldsymbol{v_r}$ be the top, bottom, left, and right neighboring motion vectors of the missing motion vector $\boldsymbol{v}_i$, respectively. Then, the cost C_i incurred by assuming that $\boldsymbol{v}_i \epsilon K_i$ is given by

$$C_i = \alpha[g_i(\boldsymbol{v}_i, \boldsymbol{v_t}) + g_i(\boldsymbol{v}_i, \boldsymbol{v_b}) + g_i(\boldsymbol{v}_i, \boldsymbol{v_l}) + g_i(\boldsymbol{v}_i, \boldsymbol{v_r})]$$

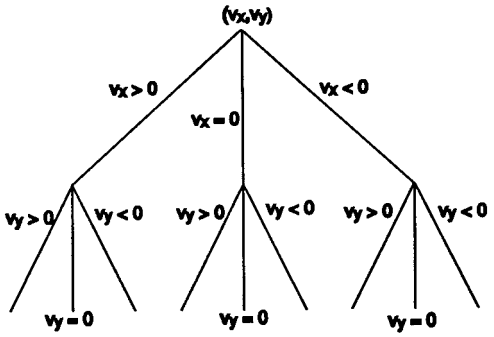

Fig. 1. Tree classification of motion vectors.

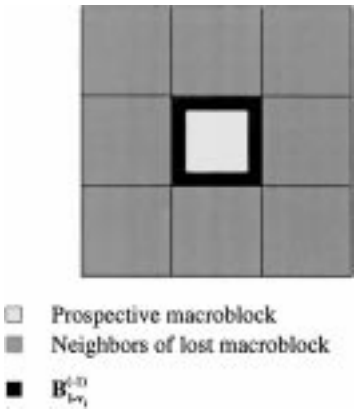

Fig. 2. Boundary pixels of prospective macroblock.

where α is some constant and

$$g_i(\boldsymbol{v}, \boldsymbol{u}) = \begin{cases} 0 & \boldsymbol{v} \epsilon K_i, \boldsymbol{u} \epsilon K_i \\ 1 & \text{otherwise} \end{cases}$$

that is, $g_i(\cdot, \cdot)$ is 0 when both arguments belong to the same class of motion vectors.

Let $\mathcal{K}^*$ denote the set of classes that have the same minimum cost. It is conceivable that two or more classes will have the same cost. In this case, we need to use spatial information to decide between which classes of vectors to choose. However, prior to that we need to obtain an MAP estimate of $\boldsymbol{v}_i$ given that it belongs to the class $K \epsilon \mathcal{K}^*$. This is done by solving the following for every $K \epsilon \mathcal{K}^*$

$$\{\hat{\boldsymbol{v}}_i\}_K = \arg\max_{\boldsymbol{v} \epsilon K} f(\boldsymbol{v} \,|\, K)$$

where $\{\hat{\boldsymbol{v}}_i\}_K$ is the class of vectors that maximize the above equation.

Choosing the motion vector with the "best" matching boundaries is performed as follows. Let $\mathcal{V}$ denote the set of MAP motion vectors, that is, $\mathcal{V} = \bigcup_{K \epsilon \mathcal{K}^*} \{\hat{\boldsymbol{v}}_i\}_K$, then $\hat{\boldsymbol{v}}_i$, the estimate of $\boldsymbol{v}_i$, is given by

$$\hat{\boldsymbol{v}}_i = \arg\max_{\boldsymbol{v} \epsilon \mathcal{V}} f\left(\boldsymbol{B}^{(-1)}_{\boldsymbol{i-v}} \,|\, \boldsymbol{X}_{\partial i}\right).$$

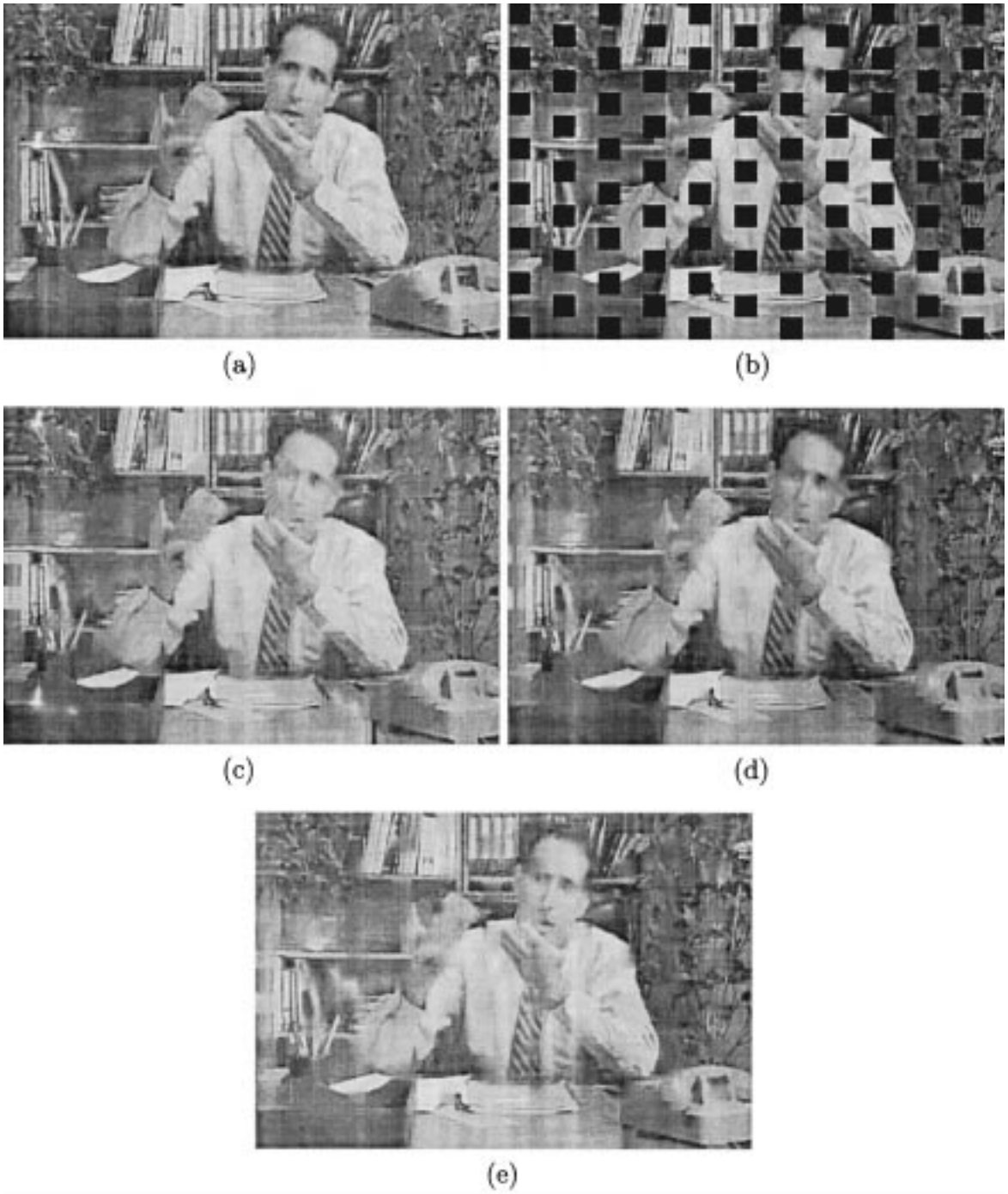

Fig. 3. Spatial reconstruction: (a) decoded frame from the *salesman* sequence, (b) same frame with missing macroblocks, (c) reconstructed using median filtering, (d) reconstructed by using line search techniques to obtain the MAP estimates with $\sigma = 100.0, \gamma = 1.0, \boldsymbol{b}_k = 1.0$ for $k = 1 \cdots 8$, and (e) reconstructed using bilinear interpolation.

Here $\boldsymbol{B}_{\boldsymbol{i}-\boldsymbol{v}}^{(-1)}$ consists of the pixels lying on the boundary of the macroblock $\boldsymbol{x}_{\boldsymbol{i}-\boldsymbol{v}}^{(-1)}$, as illustrated in Fig. 2. Using the potential functions described above, this can be rewritten as

$$\hat{\boldsymbol{v}}_i = \arg\min_{\boldsymbol{v}\in\mathcal{V}} \sum_{(r,s)\,|\,X_{(r,s)}\epsilon\boldsymbol{B}_{\boldsymbol{i}-\boldsymbol{v}}^{(-1)}} \sum_{l=r}^{r+1}\sum_{k=s}^{s+1}\sum_{m=0}^{3} b_{l,k}^{m}\rho_\gamma \cdot \left(\frac{D_m(X_{l,k})}{\sigma}\right). \tag{13}$$

It is to be noted that this approach is different from that proposed in [52] in the following sense.

1) The metric used in [52] is the L_2 norm, whereas the metric used by the temporal-spatial approach is based on the Huber cost function. Furthermore, in [52] only the difference between directly adjacent pixels is obtained, while in this case the differences between neighboring pixels that are at a 45° angle is also used.
2) When computing the metric in [52], the macroblocks above, below, and to the left of the damaged macroblock

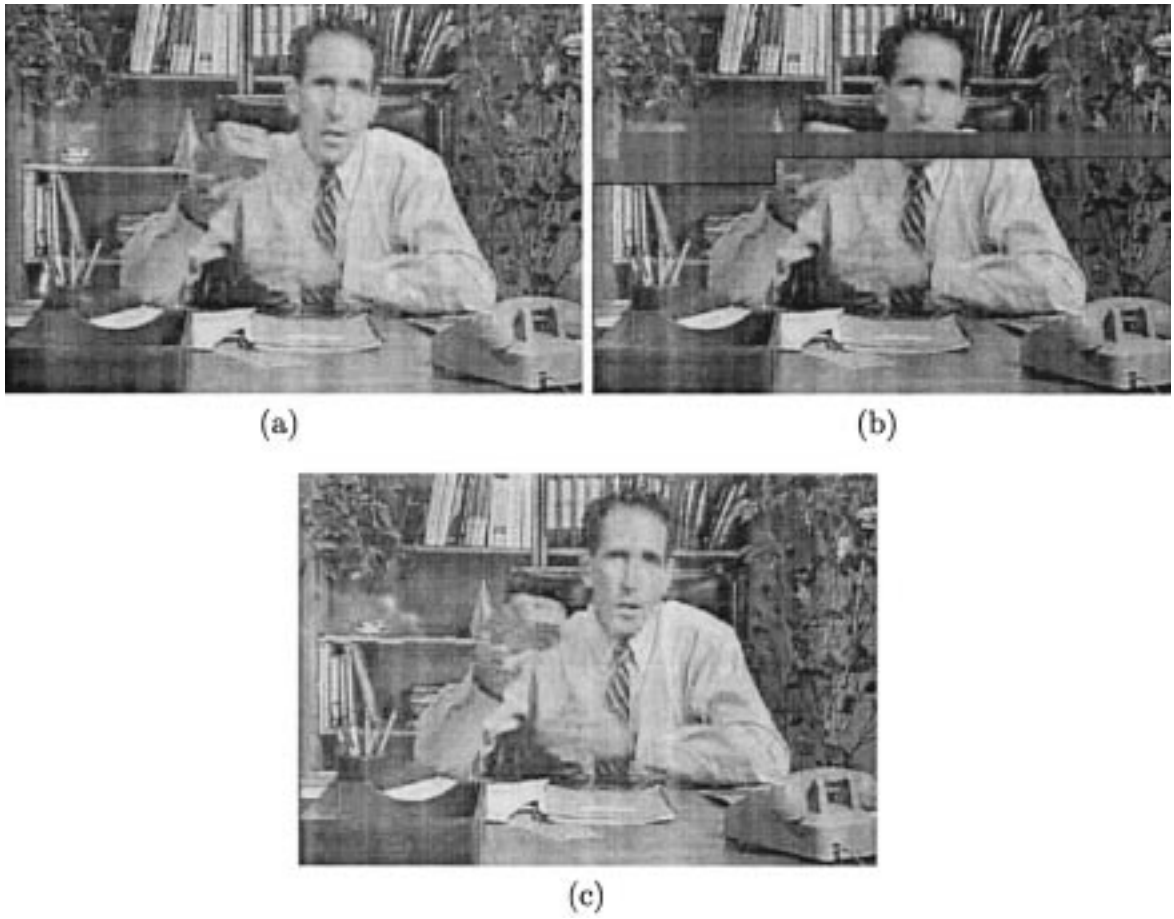

Fig. 4. Reconstruction based on finding the macroblock that has "best" matching boundaries: (a) decoded frame from the *salesman* sequence, (b) damaged version due to cell loss, and (c) the reconstructed version.

are used, whereas the temporal-spatial approach uses all available surrounding macroblocks.

3) In [52], candidate motion vectors are chosen from the following:
   a) the zero motion vector,
   b) one of the neighboring motion vectors,
   c) the median of the neighboring motion vectors,
   d) the average of the neighboring motion vectors,
   e) the motion vector of the macroblock in the previous frame that has the same spatial location as the current damaged macroblock.

   On the other hand, the temporal-spatial approach attempts to estimate the missing motion vector given the neighboring motion vectors.

In Section IV we will compare the performance of the various motion compensated error concealment techniques.

IV. Experimental Results

To test the reconstruction algorithms, the *salesman*, *football*, *flowergarden*, and hockey sequences, encoded at data rates of 0.3 Mbits/s, 1.15 Mbits/s, 1.5 Mbits/s, and 1.5 Mbits/s, respectively,[2] were multiplexed with MPEG1 Layer II audio streams into MPEG1 System Layer Streams. For our experiments, we assumed that the protocol for transmitting video over networks is the ATM protocol. In ATM, information is transmitted in fixed size packets of data called "cells." The respective system layer streams were then packed into ATM cells. This is done by packing important header information such as MPEG1 system layer or MPEG2 Program Stream pack and packet headers, and any header data necessary for the proper decoding of the compressed sequences, into high-priority cells.[3] All other data, such as motion vectors and DCT coefficients, are packed into low-priority cells. An extra 16 bits are inserted at the start of each cell, nine of which provide the location of the first macroblock being packed into the cell, and the rest are used as an ATM cell counter. The extra nine bits are also used to indicate when a macroblock spans across more than one cell. In addition, for every first macroblock being packed into a cell, the following is performed.

- The address of the macroblock is coded relative to the position of the Slice [9], [10] to which it belongs.
- Any motion vectors associated with the macroblock are coded as is, not differentially.

[2]The GOP's in each sequence were 15 frames, and the frame pattern of each GOP was IBBPBBPBBPBBPBB.

[3]Our goal is to protect the least amount of data and yet be able to reconstruct damaged frames unlike other approaches [13]–[16], [20]. Hence, we only protect the headers, without which proper decoding of the sequence would not be possible.

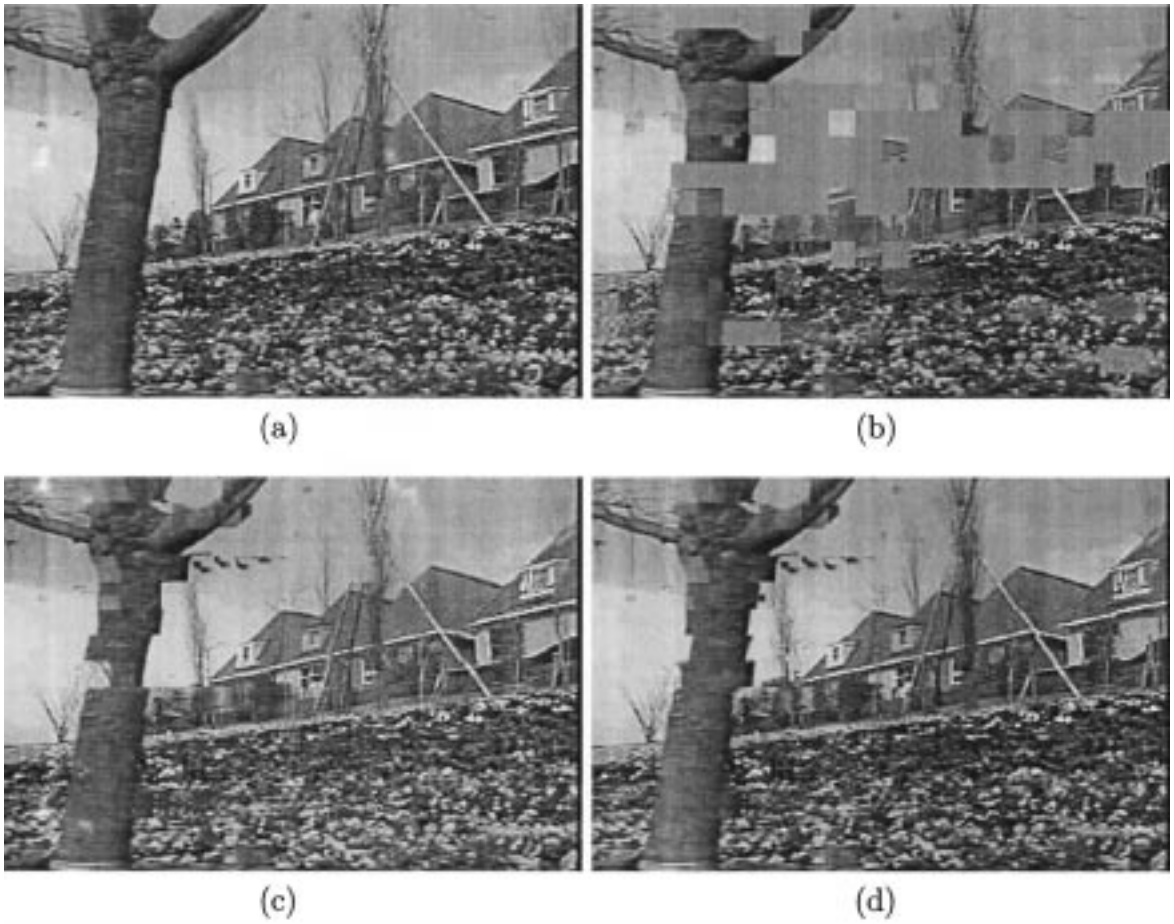

Fig. 5. (a) Decoded frame from the *flowergarden* sequence, (b) frame is damaged due to 5% ATM cell loss, (c) the frame was restored by using temporal replacement, and (d) the frame was reconstructed by finding the average of the neighboring motion vectors. The PSNR values are 26.75 dB and 27.50 dB, respectively.

- If the macroblock is intracoded, then the DC coefficients of its first Y block and chrominance blocks are coded relative to a value of 1024.

The above measures are performed for every first macroblock being packed into an ATM cell to prevent the loss of macroblocks in prior cells from affecting the decoding of succeeding macroblocks. The packed System Layer Streams were then subjected to 2% and 5% random ATM cell loss. Fifty trials with different random number generator seeds were conducted for each error rate.

It is to be noted that coding the address of a macroblock relative to the position of the Slice to which it belongs, absolute coding of its motion vectors, and coding its DC coefficients relative to a value of 1024, are described in the MPEG video standards [9], [10] as the means for encoding the first macroblock of every Slice. However, this does not mean that we start a new Slice at the beginning of every ATM cell, as this would result in a larger increase in the overhead than required. Rather, it means that we only utilize the concepts for coding the first macroblock in a Slice. The codes used were those specified by the Huffman coding tables provided in the MPEG standards [9], [10].

To determine which macroblocks were lost due to cell loss, the difference between the addresses of the two most recent correctly decoded macroblocks is obtained. This difference between both addresses is taken to be the number of macroblocks (between both macroblocks) that were lost. This can lead to, in the case of MPEG-1 and MPEG-2, an error in determining which macroblocks in a P- or B-frame were actually lost or damaged. This is a consequence of the fact that in MPEG video, predicted macroblocks that have zero motion vectors and negligible difference DCT coefficients are not coded but skipped. Under lossless conditions, a macroblock address difference that is greater than 1 is interpreted by the MPEG decoder to mean that the intervening macroblocks are to be duplicated from the reference frame. In particular, those macroblocks in the reference frame that have the same locations as those that are being decoded are used. Thus, our techniques for estimating which macroblocks are missing can misinterpret skipped macroblocks as being lost. This may result in skipped macroblocks being improperly constructed, particularly if all the surrounding macroblocks have nonzero motion vectors.

Having assigned the missing macroblocks, reconstruction proceeds as follows.

- If the damaged frame is an I-frame and all the neighbors of a missing macroblock are available, then reconstruction is performed by means of one of the spatial recon-

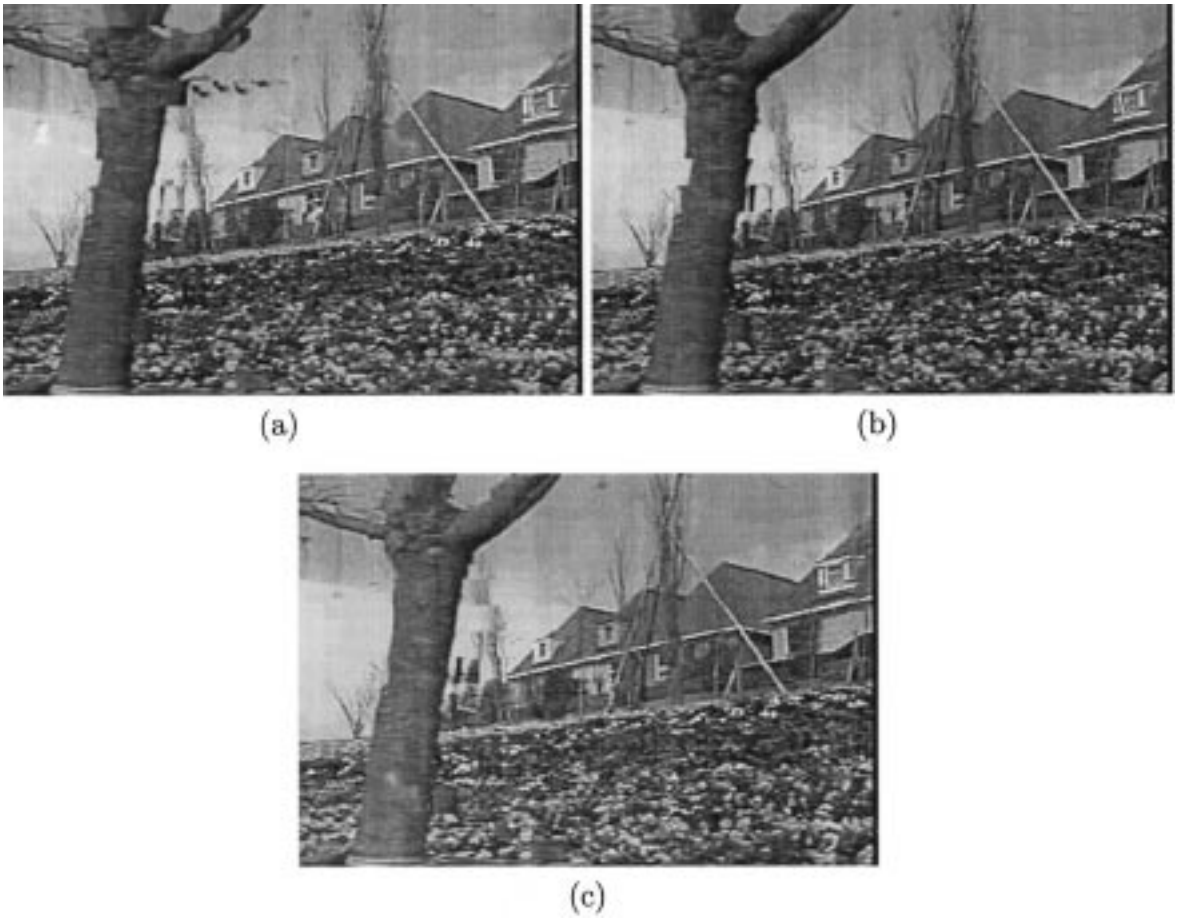

Fig. 6. (Continuation of previous figure.) (a) The frame was restored by finding the median of the neighboring motion vectors, (b) the frame was reconstructed by finding the MAP estimate of the missing motion vector, and (c) the frame was restored by using the temporal-spatial approach. The PSNR values are 28.77 dB, 30.18 dB, and 30.30 dB, respectively.

struction techniques discussed above. In particular, the approach based on median filtering is used.

- If the damaged frame is an I-frame and some of the neighbors of a missing macroblock are not available, then reconstruction is performed by searching for the macroblock in the most recent I- or P-frame that optimizes the boundary pixels [equation (12)]. The search space used is 21×21 pixels in size.
- If the damaged frame is a P- or B-frame, reconstruction proceeds as follows.
  - If the missing macroblock is surrounded by intracoded macroblocks, it is then reconstructed by the same method used for restoring macroblocks missing from an I-frame,
  - Otherwise, the missing macroblock is assumed to be intercoded and it is reconstructed by first estimating its associated motion vector. This is done by a number of ways enumerated below.
    1) Temporal replacement, that is, a motion vector with zero components is used.
    2) The average of the surrounding motion vectors is obtained.
    3) The median of the surrounding motion vectors is obtained.
    4) The MAP estimate of the missing motion vector given its neighboring motion vectors is obtained.
    5) The temporal-spatial approach is used to estimate the missing motion vector.

    The estimate of the missing motion vector is then used to provide error concealment. This is achieved by replacing the missing macroblock by the region in the past I- or P-reference frame to which the estimated motion vector is pointing. We will compare the performance of all the above enumerated motion vector estimation techniques later on.

For all of the motion compensated techniques, the parameters $\sigma, \gamma, \{\boldsymbol{b}_k,\ k = 1, \ldots, 8\}$ were set to unit value. It was observed that changing the values of these parameters did not significantly impact the quality of the reconstruction.

To compare the performance of the different spatial reconstruction techniques, 25% of the macroblocks belonging to a frame from the *salesman* sequence were dropped. In Fig. 3(a) and (b) we show the original and damaged frames (due to missing macroblocks), from the *salesman* sequence. Reconstruction was performed spatially in Fig. 3(c) via median filtering, in Fig. 3(d) by iteratively solving for the MAP estimate of each missing pixel within the damaged macroblock, and in Fig. 3(e) by means of bilinear interpolation [25]. The

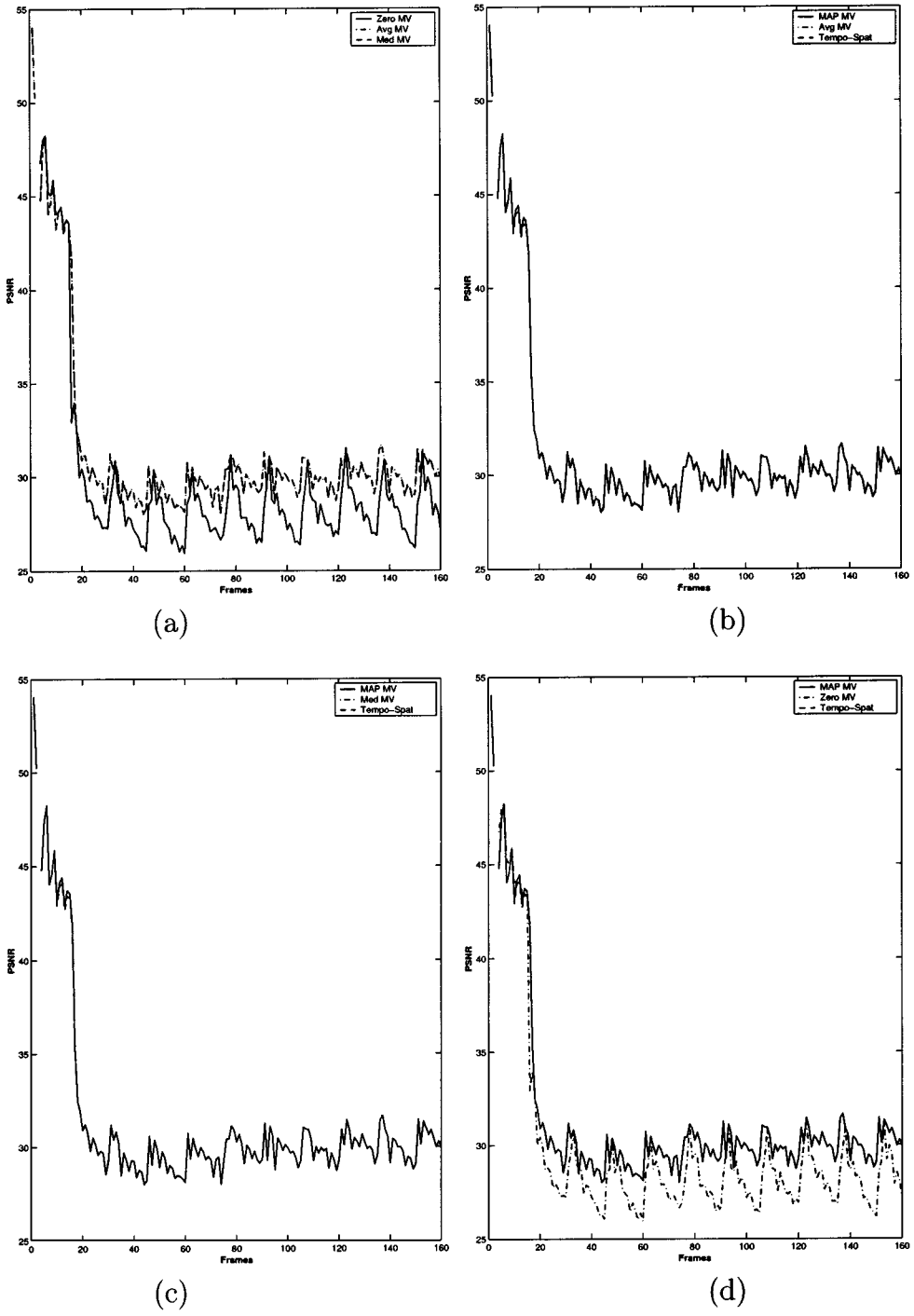

Fig. 7. Average reconstruction PSNR values for the *flowergarden* sequence when 5% of the cells were dropped. (a) Zero motion vector (temporal replacement), the average of the neighboring motion vectors, and the median of the neighboring motion vectors used. (b) The MAP estimate of the missing motion vector, the temporal spatial approach, and the average of the neighboring motion vectors are used. (c) The MAP estimate of the missing motion vector, the temporal spatial approach, and the median of the neighboring motion vectors are used. (d) The MAP estimate of the missing motion vector, the temporal-spatial approach, and temporal replacement are used.

reconstruction PSNR's were found to be 29.16 dB, 29.85 dB, and 29.86 dB, respectively. From a PSNR point of view, the performance of all techniques is comparable. It was observed that the reconstructions due to the first two techniques have sharper edges. Furthermore, our approach based on median filtering is attractive since it can be implemented in realtime [42], does not require a search for dominant edges, and operates on spatial data rather than DCT coefficients which may not be available.

In our experiments it was observed that using values for σ that are greater than 1 and $\gamma = 1.0$ provided the best MAP restoration when the weights $\{\boldsymbol{b}_k,\ k = 1 \cdots 8\}$ were of unit value. This can be seen since for $\gamma = 1.0, \sigma > 1$ values will not heavily penalize edges. The image in Fig. 3(d) was reconstructed with

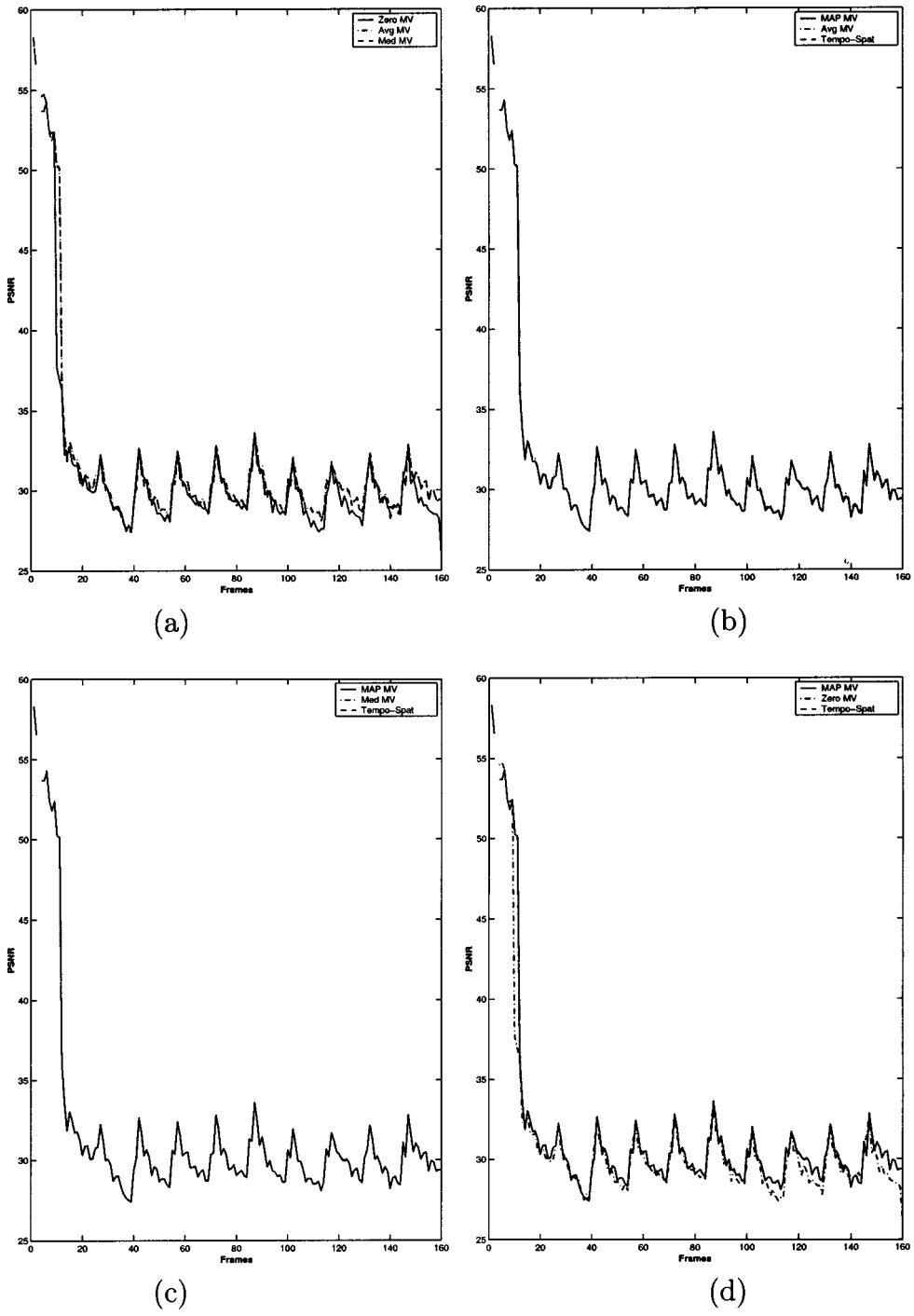

Fig. 8. Average reconstruction PSNR values for the *football* sequence when 5% of the cells were dropped. (a) Zero motion vector (temporal replacement), the average of the neighboring motion vectors, and the median of the neighboring motion vectors used. (b) The MAP estimate of the missing motion vector, the temporal spatial approach, and the average of the neighboring motion vectors are used. (c) The MAP estimate of the missing motion vector, the temporal spatial approach, and the median of the neighboring motion vectors are used. (d) The MAP estimate of the missing motion vector, the temporal spatial approach, and temporal replacement are used.

$\sigma = 100$. It was also observed that using the median values of the border pixels as initial values led to rapid convergence to the optimal MAP estimate.

Fig. 4(a) is a decoded frame from the *salesman* sequence. Due to random cell loss, major portions of the frame were lost as shown in Fig. 4(b). Reconstruction is performed by minimizing (12). The search space for the motion vectors did not exceed an area of 7×7 pixels, and hence an exhaustive search for the motion vector was implemented. A small search region was used in this case since there is little motion in this sequence. In the case of sequences where a substantial amount of motion exists, an exhaustive search may be too costly, and thus other searches such as the logarithmic search may be implemented but at a cost of lower fidelity. As seen, the reconstructed version in Fig. 4(c)

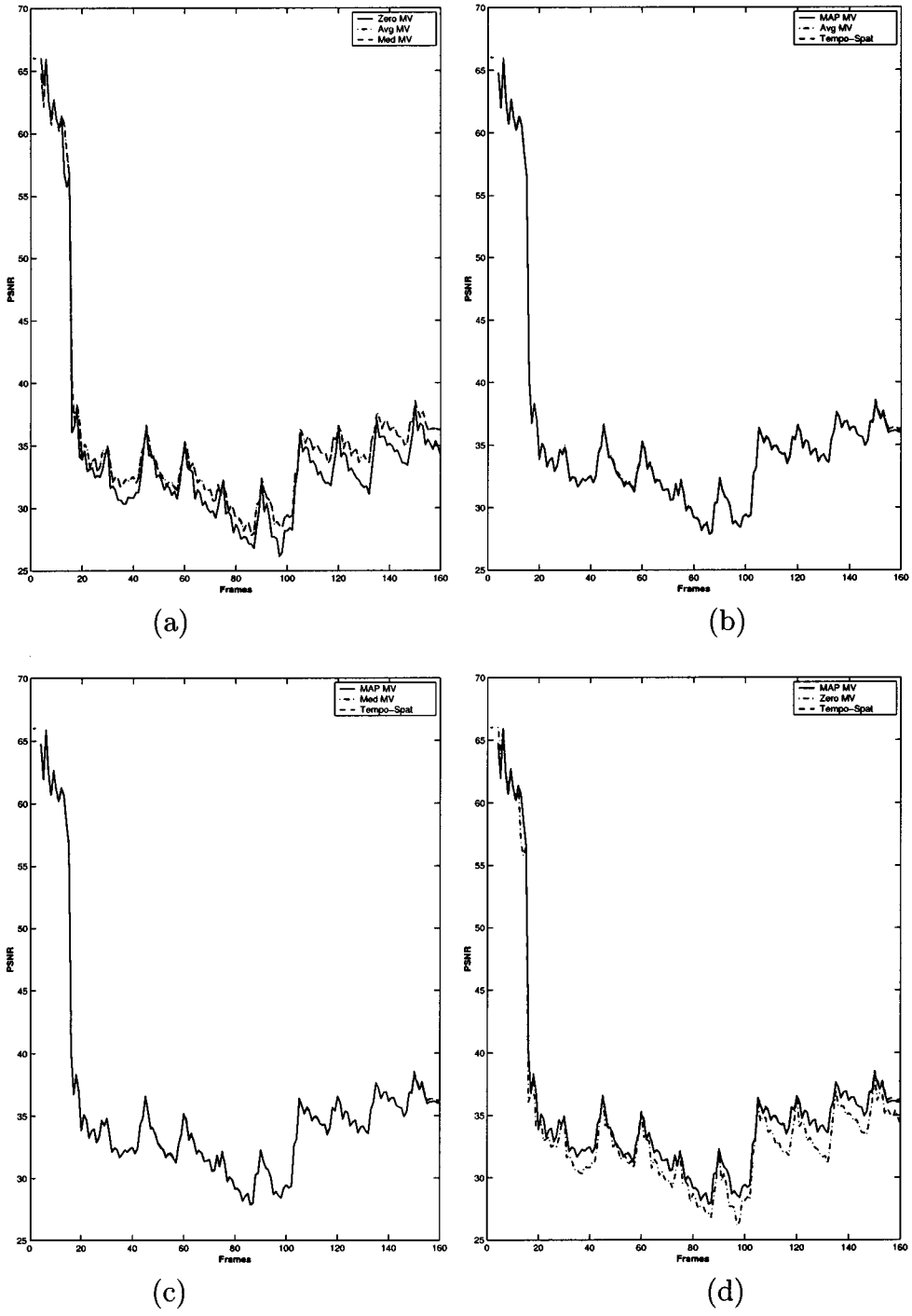

Fig. 9. Average reconstruction PSNR values of all trials for the *hockey* sequence when 5% of the cells were dropped. (a) Zero motion vector (temporal replacement), the average of the neighboring motion vectors, and the median of the neighboring motion vectors used. (b) The MAP estimate of the missing motion vector, the temporal spatial approach, and the average of the neighboring motion vectors are used. (c) The MAP estimate of the missing motion vector, the temporal spatial approach, and the median of the neighboring motion vectors are used. (d) The MAP estimate of the missing motion vector, the temporal spatial approach, and temporal replacement are used.

closely matches the original in Fig. 4(a) with a reconstruction PSNR value of 35.72 dB. The PSNR value was obtained via

$$\text{PSNR} = 10 \log \frac{255^2}{\dfrac{\text{MSE}(Y) + \text{MSE}(U) + \text{MSE}(V)}{3}}$$

where $\text{MSE}(\cdot)$ denotes the mean square error of the reconstructed color component.

Fig. 5(a) is a frame from the *flowergarden* sequence, and Fig. 5(b) is the same frame with missing macroblocks from one of the trials of that involved 5% ATM cell loss. The frame is restored via temporal replacement, obtaining the average of the neighboring motion vectors, finding the median of the neighboring motion vectors, obtaining the MAP estimate of the missing motion vector, and via the temporal-spatial approach in Figs. 5(c), (d), and 6(a)–(c), respectively. The reconstruction

TABLE I
AVERAGE PSNR VALUES IN DECIBELS FOR THE DIFFERENT ERROR CONCEALMENT SCHEMES FOR THE *FLOWER GARDEN*, *FOOTBALL*, AND *HOCKEY* SEQUENCES AT A 2% ATM CELL LOSS RATE

Error Concealment Technique	*flowergarden*	*football*	*hockey*
Temporal Replacement	31.37	33.07	34.92
Average Motion Vector	33.32	33.58	36.07
Median Motion Vector	34.13	34.13	36.78
MAP estimation of Motion Vector	34.86	34.15	36.74
Temporal-Spatial Approach	35.33	34.96	38.73

TABLE II
MINIMUM PSNR VALUES IN DECIBELS FOR THE DIFFERENT ERROR CONCEALMENT SCHEMES FOR THE *FLOWER GARDEN*, *FOOTBALL*, AND *HOCKEY* SEQUENCES AT A 2% ATM CELL LOSS RATE

Error Concealment Technique	*flowergarden*	*football*	*hockey*
Temporal Replacement	30.51	31.95	33.31
Average Motion Vector	32.51	32.58	34.41
Median Motion Vector	33.00	32.80	34.94
MAP estimation of Motion Vector	33.64	33.08	34.92
Temporal-Spatial Approach	34.46	33.75	36.73

TABLE III
MAXIMUM PSNR VALUES IN DECIBELS FOR THE DIFFERENT ERROR CONCEALMENT SCHEMES FOR THE *FLOWER GARDEN*, *FOOTBALL*, AND *HOCKEY* SEQUENCES AT A 2% ATM CELL LOSS RATE

Error Concealment Technique	*flowergarden*	*football*	*hockey*
Temporal Replacement	32.13	34.48	36.60
Average Motion Vector	34.11	34.93	38.01
Median Motion Vector	35.22	35.69	38.61
MAP estimation of Motion Vector	35.91	35.51	38.80
Temporal-Spatial Approach	36.51	36.56	40.80

TABLE IV
AVERAGE PSNR VALUES IN DECIBELS FOR THE DIFFERENT ERROR CONCEALMENT SCHEMES FOR THE *FLOWER GARDEN*, *FOOTBALL*, AND *HOCKEY* SEQUENCES AT A 5% ATM CELL LOSS RATE

Error Concealment Technique	*flowergarden*	*football*	*hockey*
Temporal Replacement	28.45	29.83	31.94
Average Motion Vector	30.27	30.26	33.19
Median Motion Vector	31.16	30.88	33.92
MAP estimation of Motion Vector	31.64	30.62	33.53
Temporal-Spatial Approach	32.10	31.57	35.64

TABLE V
MINIMUM PSNR VALUES IN DECIBELS FOR THE DIFFERENT ERROR CONCEALMENT SCHEMES FOR THE *FLOWER GARDEN*, *FOOTBALL*, AND *HOCKEY* SEQUENCES AT A 5% ATM CELL LOSS RATE

Error Concealment Technique	*flowergarden*	*football*	*hockey*
Temporal Replacement	27.58	28.81	30.65
Average Motion Vector	29.51	29.20	32.00
Median Motion Vector	30.34	29.88	32.57
MAP estimation of Motion Vector	30.84	29.63	32.20
Temporal-Spatial Approach	31.21	30.63	34.55

TABLE VI
MAXIMUM PSNR VALUES IN DECIBELS FOR THE DIFFERENT ERROR CONCEALMENT SCHEMES FOR THE *FLOWER GARDEN*, *FOOTBALL*, AND *HOCKEY* SEQUENCES AT A 5% ATM CELL LOSS RATE

Error Concealment Technique	*flowergarden*	*football*	*hockey*
Temporal Replacement	29.14	30.92	33.37
Average Motion Vector	31.17	31.05	34.42
Median Motion Vector	32.12	31.71	35.91
MAP estimation of Motion Vector	32.62	31.55	34.51
Temporal-Spatial Approach	32.79	32.55	36.72

PSNR values are 26.75 dB, 27.50 dB, 28.77 dB, 30.18 dB, and 30.30 dB, respectively.

As is evident, the use of an MRF field that does not penalize the discontinuities in the motion field outperforms using the average or the median of the surrounding motion vectors. It is also observed that using spatial information, as is done in the temporal-spatial approach, results in better restoration. This can be seen in Fig. 6(c) where it is evident that the damaged portions of the tree trunk are almost perfectly restored. This is attributed to the fact that the temporal-spatial approach attempts to preserve the discontinuities in the motion field while matching the macroblock boundaries. In this case, the "best" matching boundaries were those that preserved the discontinuity between the tree trunk and the background. Also shown in Figs. 5(c), (d), and 6(a)–(c) are the effect of error propagation due to inaccurate restoration.

For comparison purposes, we provide the average PSNR curves of all trials for the various error concealment schemes when 5% of the ATM cells were lost, for the *flowergarden*, *football*, and *hockey* sequences, respectively, in Figs. 7–9. We also provide the average, minimum, and maximum PSNR values for the different error concealment strategies at 2% and 5% ATM cell loss rate for the same three sequences in Tables I–VI, respectively.

In general, restoration based on using the temporal-spatial approach was better than that attained by the other techniques by at least 1 dB. Furthermore, the gap in performance between the temporal-spatial approach and the MAP estimation of the missing motion vector widened in the case of both the *football* and *hockey* sequences. This is also seen in Figs. 8 and 9, and can be attributed to the fact that the motion field of *flowergarden* was more uniform than those of *football* and *hockey*. In the case of the latter, the motion vectors do not point in one general direction. In addition, it is possible for a macroblock to lie on the boundary between two objects moving in opposite directions. In such a case, the use of spatial data was needed to determine which object the missing macroblock belonged to. It is also observed that using the median of the neighboring motion vectors results in better performance than using the average of the neighboring motion vectors.

V. CONCLUSION

Due to the nature of how channel errors and congestion are exhibited on data networks, error concealment must be used when video is transmitted. We have presented both spatial and temporal error concealment techniques for compressed video. Our spatial technique based on median filtering, although suboptimal, can be implemented in real time. We have also shown that using temporal-spatial data yields better reconstructions than only using temporal data. We are currently investigating the use of DSP's for the deployment of error concealment in video codecs used in ATM networks. Our methods can

be extended to wireless networks and streaming video on packet-switched networks, for example, the Internet.

APPENDIX I

We show here that $\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}\,|\,\boldsymbol{y}=\boldsymbol{D}\boldsymbol{x}}[-\ln f(\boldsymbol{x})]$. Let $\mathcal{A}$ be the event that $\boldsymbol{y} = \boldsymbol{D}x$. In general [53], $f(\boldsymbol{x},\boldsymbol{y})P(\mathcal{A}|\boldsymbol{x},\boldsymbol{y}) = f(\boldsymbol{x},\boldsymbol{y}|\mathcal{A})P(\mathcal{A})$. Using Bayes' rule, then,

$$f(\boldsymbol{x}\,|\,\boldsymbol{y})P(\mathcal{A}\,|\,\boldsymbol{x},\boldsymbol{y}) = \frac{f(\boldsymbol{x},\boldsymbol{y}\,|\,\mathcal{A})P(\mathcal{A})}{f(\boldsymbol{y})}.$$

But

$$P(\mathcal{A}\,|\,\boldsymbol{x},\boldsymbol{y}) = \begin{cases} 0 & \boldsymbol{y} \neq \boldsymbol{D}x \\ 1 & \boldsymbol{y} = \boldsymbol{D}x. \end{cases}$$

Thus,

$$f(\boldsymbol{x}\,|\,\boldsymbol{y}) = \frac{f(\boldsymbol{x},\boldsymbol{y}\,|\,\mathcal{A})P(\mathcal{A})}{f(\boldsymbol{y})}.$$

Let $I_{\delta x} = \{\boldsymbol{z}\,|\,\|\boldsymbol{z}-\boldsymbol{x}\| \leq \delta x\}$, and $I_{\delta y} = \{\boldsymbol{u}\,|\,\|\boldsymbol{u}-\boldsymbol{y}\| \leq \delta y\}$. Now

$$f(\boldsymbol{x},\boldsymbol{y}\,|\,\mathcal{A}) = \lim_{\delta x\to 0}\lim_{\delta y\to 0} P(\boldsymbol{x}\epsilon I_{\delta x}, \boldsymbol{y}\epsilon I_{\delta y}\,|\,\mathcal{A}).$$

Rewriting $P(\boldsymbol{x}\epsilon I_{\delta x}, \boldsymbol{y}\epsilon I_{\delta y}\,|\,\mathcal{A})$ as $P(\boldsymbol{x}\epsilon I_{\delta x}, \boldsymbol{y}\epsilon I_{\delta y}\,|\,\mathcal{A}) = P(\boldsymbol{y}\epsilon I_{\delta y}\,|\,\boldsymbol{x}\epsilon I_{\delta x}, \mathcal{A})P(\boldsymbol{x}\epsilon I_{\delta x}\,|\,\mathcal{A})$, where

$$P(\boldsymbol{y}\epsilon I_{\delta y}\,|\,\boldsymbol{x}\epsilon I_{\delta x}, \mathcal{A}) = \begin{cases} 0 & \boldsymbol{D}I_{\delta x} \cap I_{\delta y} = \phi \\ 1 & \boldsymbol{D}I_{\delta x} \cap I_{\delta y} \neq \phi \end{cases}$$

and $\boldsymbol{D}I_{\delta x} = \{\boldsymbol{D}\boldsymbol{z}\,|\,\boldsymbol{z}\epsilon I_{\delta x}\}$, then $f(\boldsymbol{x},\boldsymbol{y}\,|\,\mathcal{A})P(\mathcal{A}) = f(\boldsymbol{x}\,|\,\mathcal{A})P(\mathcal{A})$. Since

$$f(\boldsymbol{x}\,|\,\mathcal{A})P(\mathcal{A}) = f(\boldsymbol{x})P(\mathcal{A}\,|\,\boldsymbol{x})$$

then $L(\boldsymbol{x}\,|\,\boldsymbol{y}) = \ln f(\boldsymbol{x}) + \ln P(\mathcal{A}\,|\,\boldsymbol{x}) - \ln f(\boldsymbol{y})$. For a given observed image, the third term in the preceding equation is independent of $\boldsymbol{x}$, hence the MAP estimate is obtained as

$$\begin{aligned}\hat{\boldsymbol{x}} &= \arg\max_{\boldsymbol{x}\,|\,\boldsymbol{y}=\boldsymbol{D}\boldsymbol{x}}[\ln f(\boldsymbol{x}) + \ln P(\mathcal{A}\,|\,\boldsymbol{x})] \\ &= \arg\min_{\boldsymbol{x}\,|\,\boldsymbol{y}=\boldsymbol{D}\boldsymbol{x}}[-\ln f(\boldsymbol{x}) - \ln P(\mathcal{A}\,|\,\boldsymbol{x})].\end{aligned}$$

Since

$$P(\mathcal{A}\,|\,\boldsymbol{x}) = \begin{cases} 0 & \boldsymbol{y} \neq \boldsymbol{D}x \\ 1 & \boldsymbol{y} = \boldsymbol{D}x \end{cases}$$

then $\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}\,|\,\boldsymbol{y}=\boldsymbol{D}\boldsymbol{x}}[-\ln f(\boldsymbol{x})]$.

APPENDIX II

We show here that

$$\hat{X}_{i,j} = \frac{\displaystyle\sum_{k=J_1}^{J_2} b_k z_k + \gamma\sigma\left[\sum_{k=J_2+1}^{8} b_k - \sum_{k=1}^{J_1-1} b_k\right]}{\displaystyle\sum_{k=J_1}^{J_2} b_k}$$

where J_1 and J_2 satisfy $z_{J_2} - z_{J_1} \leq \gamma\sigma,\ J_2 \geq J_1$. We assume positive weights as well as positive values for $z_1 \cdots z_8$.

Consider the following special case.

Case 2A: $|\boldsymbol{\Delta_k}| > \gamma\sigma\ \forall\, k$

In this case, the optimal estimate $\hat{X}_{i,j}$ will lie between z_1 and z_8. It will either be within a $\gamma\sigma$ of some z_n or not. If the former is true, then $h_\gamma(\hat{X}_{i,j}) = b_n(\hat{\Delta}_n/\sigma)^2 + \sum_{k=1,k\neq n}^{8} b_k[(2\gamma|\hat{\Delta}_k|/\sigma) - \gamma^2]$; otherwise,

$$\begin{aligned}h_\gamma(\hat{X}_{i,j}) &= \sum_{k=1}^{8} b_k\left[\frac{2\gamma|\hat{\Delta}_k|}{\sigma} - \gamma^2\right] \\ &= \sum_{k=1}^{8} b_k\,\frac{2\gamma|\hat{\Delta}_k|}{\sigma} - \gamma^2\sum_{k=1}^{8} b_k\end{aligned}$$

where

$$\hat{\Delta}_k = \begin{cases} z_k - \hat{X}_{i,j} & z_k \epsilon U \\ \hat{X}_{i,j} - z_k & z_k \epsilon L. \end{cases}$$

Considering the latter, that is, $\hat{X}_{i,j} \notin [z_k - \gamma\sigma, z_k + \gamma\sigma]\ \forall\, k$, let $\tilde{h}_\gamma(X_{i,j}) = \sum_{k=1}^{8} b_k(2\gamma|\Delta_k|/\sigma)$ For the following intervals and for the variable $X_{i,j}, \tilde{h}_\gamma(X_{i,j})$ is a straight line with negative, zero, or positive slope.

1) $X_{i,j} < z_1 - \gamma\sigma$:

$$\tilde{h}_\gamma(X_{i,j}) = \frac{2\gamma}{\sigma}\left[-\left(\sum_{k=1}^{8} b_k\right)X_{i,j} + \sum_{k=1}^{8} b_k z_k\right]$$

2) $z_l - \gamma\sigma < X_{i,j} < z_{l+1} - \gamma\sigma$ for $l = 1 \cdots 7$

$$\begin{aligned}\tilde{h}_\gamma(X_{i,j}) = \frac{2\gamma}{\sigma}\Bigg[&\left(\sum_{k=1}^{l} b_k - \sum_{k=l+1}^{8} b_k\right)X_{i,j} \\ &+ \sum_{k=l+1}^{8} b_k z_k - \sum_{k=1}^{l} b_k z_k\Bigg]\end{aligned}$$

3) $z_8 - \gamma\sigma < X_{i,j}$:

$$\tilde{h}_\gamma(X_{i,j}) = \frac{2\gamma}{\sigma}\left[\left(\sum_{k=1}^{8} b_k\right)X_{i,j} - \sum_{k=1}^{8} b_k z_k\right].$$

Since $\tilde{h}_\gamma(X_{i,j})$ is continuous, then there exists an integer J such that $\sum_{k=1}^{J-1} b_k - \sum_{k=J}^{8} b_k < 0$, $\sum_{k=1}^{J} b_k - \sum_{k=J+1}^{8} b_k = 0$, and $\sum_{k=1}^{J+1} b_k - \sum_{k=J+2}^{8} b_k > 0$, that is, $\tilde{h}_\gamma(X_{i,j})$ has a flat bottom, or $\sum_{k=1}^{J-1} b_k - \sum_{k=J}^{8} b_k < 0$ and $\sum_{k=1}^{J} b_k - \sum_{k=J+1}^{8} b_k > 0$. If $\tilde{h}_\gamma(X_{i,j})$ does have a flat bottom, then $z_J \leq \hat{X}_{i,j} \leq z_{J+1}$, otherwise $\hat{X}_{i,j} = z_J$.

Case 2B: $|\boldsymbol{\Delta_k}| \leq \gamma\sigma$ for some $\boldsymbol{k}$

Suppose now that there exist J_1 and J_2 such that $J_1 < J_2$ and $z_{J_2} - z_{J_1} \leq \gamma\sigma$, then for $z_{J_1} \leq X_{i,j} \leq z_{J_2}$,

$$\begin{aligned}h_\gamma(X_{i,j}) = &\sum_{k=1}^{J_1-1} b_k\left[\frac{2\gamma}{\sigma}\,|X_{i,j} - z_k| - \gamma^2\right] \\ &+ \sum_{k=J_2+1}^{8} b_k\left[\frac{2\gamma}{\sigma}\,|X_{i,j} - z_k| - \gamma^2\right] \\ &+ \sum_{k=J_1}^{J_2} b_k\left(\frac{X_{i,j} - z_k}{\sigma}\right)^2.\end{aligned}$$

In light of Case 2A above, if $\sum_{k=1}^{J_1-1} b_k - \sum_{k=J_1}^{8} b_k > 0$, then there exists an integer J_3 such that $J_3 \leq J_1$ and $z_{J_3} \leq \hat{X}_{i,j} \leq z_{J_3+1}$, or $\hat{X}_{i,j} = z_{J_3}$. Similarly, if $\sum_{k=1}^{J_2} b_k - \sum_{k=J_2+1}^{8} b_k < 0$,

then there exists an integer J_4 such that $J_4 \geq J_2$ and $z_{J_4} \leq \hat{X}_{i,j} \leq z_{J_4+1}$ or $\hat{X}_{i,j} = z_{J_4}$. Otherwise, there exists an integer J_5 such that $J_1 \leq J_5 \leq J_2$ and $z_{J_5} - \gamma\sigma \leq \hat{X}_{i,j} \leq z_{J_5} + \gamma\sigma$. Since

$$\begin{aligned} h_\gamma(X_{i,j}) = &\sum_{k=1}^{J_1-1} b_k \left[\frac{2\gamma}{\sigma}\left|X_{i,j} - z_k\right| - \gamma^2\right] \\ &+ \sum_{k=J_2+1}^{8} b_k \left[\frac{2\gamma}{\sigma}\left|X_{i,j} - z_k\right| - \gamma^2\right] \\ &+ \sum_{k=J_1}^{J_2} b_k \left(\frac{X_{i,j} - z_k}{\sigma}\right)^2 \\ &\cdot \sum_{k=1}^{J_1-1} b_k \left[\frac{2\gamma}{\sigma}\left[X_{i,j} - z_k\right] - \gamma^2\right] \\ &+ \sum_{k=J_2+1}^{8} b_k \left[\frac{2\gamma}{\sigma}\left[z_k - X_{i,j}\right] - \gamma^2\right] \\ &+ \sum_{k=J_1}^{J_2} b_k \left(\frac{X_{i,j} - z_k}{\sigma}\right)^2 \end{aligned}$$

$$\frac{\partial}{\partial X_{i,j}} h_\gamma(\boldsymbol{x}) = 0, \Rightarrow$$

$$\hat{X}_{i,j} = \frac{\displaystyle\sum_{k=J_1}^{J_2} b_k z_k + \gamma\sigma\left[\sum_{k=J_2+1}^{8} b_k - \sum_{k=1}^{J_1-1} b_k\right]}{\displaystyle\sum_{k=J_1}^{J_2} b_k}$$

where J_1 and J_2 satisfy $z_{J_2} - z_{J_1} \leq \gamma\sigma, J_2 \geq J_1$. As $\gamma\sigma \to \infty$, Case 2A above is satisfied, that is, there exists an integer J such that $z_J \leq \hat{X}_{i,j} \leq z_{J+1}$ when $\sum_{k=1}^{J} b_k = \sum_{k=J+1}^{8} b_k$, or $\hat{X}_{i,j} = z_J$ when $\sum_{k=1}^{J} b_k \geq \sum_{k=J+1}^{8} b_k$. If $J = 3$, then a possible value for $\hat{X}_{i,j}$, would be the median of $z_1 \cdots z_8$. This is in fact the case when all the weights $b_1 \cdots b_8$ are equal.

REFERENCES

[1] M. Tomordy, "Airline with the personal touch," *IEE Rev.*, vol. 44, no. 6, pp. 261–264, Nov. 1998.

[2] J. Besag, "Spatial interaction and the statistical analysis of lattice systems," *J. Royal Stat. Soc.*, ser. B, vol. 36, pp. 192–326, 1974.

[3] R. Kinderman and J. L. Snell, *Markov Random Fields and their Applications.*: Amer. Math. Soc., 1980, vol. 1, Contemporary Mathematics.

[4] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-6, pp. 721–741, Nov. 1984.

[5] M. Wada, "Selective recovery of video packet loss using error concealment," *IEEE J. Select. Areas Commun.*, vol. 7, pp. 807–814, June 1989.

[6] H. Man, F. Kossentini, and M. J. T. Smith, "A class of EZW image coders for noisy channels," in *Proc. Int. Conf. Image Processing*, vol. III, Santa Barbara, CA, Oct. 26–29, 1997, pp. 90–93.

[7] J. Hagenauer, "Rate compatible punctured convolutional codes (RCPC Codes) and their applications," *IEEE Trans. Commun.*, vol. 36, pp. 389–400, Apr. 1988.

[8] Y. Wang and Q. Zhu, "Error control and concealment for video communication: A review," *Proc. IEEE*, vol. 86, pp. 974–996, May 1998.

[9] ISO/IEC 11172–2 MPEG-1 Video Coding Standard, "ISO/IEC 11172–2, Information technology-coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbits/s—Part 2: Video,", 1993. ISO.

[10] ISO/IEC 13818-2 MPEG-2 Video Coding Standard, "ISO/IEC 13818–2, Generic coding of moving pictures and associated audio information—Part 2: Video,", 1995. ISO.

[11] ITU-T, "CCIR recommendation H.261: Codec for audiovisual services at $p \times 64$ kbits/sec,", 1990.

[12] ITU-T, "Draft ITU-T recommendation H.263 version 2: Video coding for low bitrate communication,", Sept. 1997.

[13] F. Kishino, K. Manabe, Y. Hayashi, and H. Yasuda, "Variable bit rate coding of video signals for ATM networks," *IEEE J. Select. Areas Commun.*, vol. 7, pp. 801–806, June 1989.

[14] M. Ghanbari and V. Seferidis, "Cell-loss concealment in ATM video codecs," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 3, pp. 238–247, June 1993.

[15] M. Ghanbari and C. J. Hughes, "Packing coded video signals into ATM cells," *IEEE/ACM Trans. Networking*, vol. 1, pp. 505–508, Oct. 1993.

[16] P. Pancha and M. El Zarki, "MPEG coding for variable bit rate video transmission," *IEEE Commun. Mag.*, vol. 32, pp. 54–66, May 1994.

[17] W. Luo and M. El Zarki, "Analysis of error concealment schemes for MPEG-2 video transmission over ATM based networks," in *Proc. SPIE Conf. Vis. Commun. Image Processing*, vol. 1605, Taipei, Taiwan, May 1995, pp. 1358–1368.

[18] L. H. Kieu and K. N. Ngan, "Cell loss concealment techniques for layered video codec in an ATM network," *IEEE Trans. Image Processing*, vol. 3, pp. 666–677, Sept. 1994.

[19] D. Raychaudhuri, H. Sun, and R. S. Girons, "ATM transport and cell-loss concealment techniques for MPEG video," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, Minneapolis, MN, Nov. 1993, pp. 117–120.

[20] C. Hahm and J. Kim, "An adaptive error concealment in SNR scalable system," in *Proc. SPIE Conf. Vis. Commun. Image Processing*, vol. 2501/3, Taipei, Taiwan, May 24–26, 1995, pp. 1380–1387.

[21] V. Parthasarathy, J. Modestino, and K. S. Vastola, "Design of a transport coding scheme for high quality video over ATM networks," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, pp. 358–376, Apr. 1997.

[22] A. S. Tom, C. L. Yeh, and F. Chu, "Packet video for cell loss protection using deinterleaving and scrambling," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, Toronto, Canada, May 1991, pp. 2857–2860.

[23] Q. Zhu, Y. Wang, and L. Shaw, "Coding and cell loss recovery in DCT based packet video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 3, pp. 248–258, June 1993.

[24] J. Y. Park, M. H. Lee, and K. J. Lee, "A simple concealment for ATM bursty cell loss," *IEEE Trans. Consumer Electron.*, vol. 39, pp. 704–710, Aug. 1993.

[25] P. Salama, N. Shroff, E. J. Coyle, and E. J. Delp, "Error concealment techniques for encoded video streams," in *Proc. Int. Conf. Image Processing*, vol. I, Washington, DC, Oct. 23–26, 1995, pp. 9–12.

[26] P. Salama, N. Shroff, and E. J. Delp, "A Bayesian approach to error concealment in encoded video streams," in *Proc. Int. Conf. Image Processing*, vol. II, Lausanne, Switzerland, Sept. 16–19, 1996, pp. 49–52.

[27] ——, "A fast suboptimal approach to error concealment in encoded video streams," in *Proc. Int. Conf. Image Processing*, vol. II, Santa Barbara, CA, Oct. 26–29, 1997, pp. 101–104.

[28] ——, "Error concealment in encoded video streams," in *Image Recovery Techniques for Image Compression Applications*, N. P. Galatsanos and A. K. Katsaggelos, Eds. Norwell, MA: Kluwer, 1998.

[29] Y. Wang, Q. Zhu, and L. Shaw, "Maximally smooth image recovery in transform coding," *IEEE Trans. Commun.*, vol. 41, pp. 1544–1551, Oct. 1993.

[30] Y. Wang and Q. Zhu, "Signal loss recovery in DCT-based image and video codecs," in *Proc. SPIE Conf. Vis. Commun. Image Processing*, vol. 2501/3, Boston, MA, Nov. 1991, pp. 667–678.

[31] L. T. Chia, D. J. Parish, and J. W. R. Griffiths, "On the treatment of video cell loss in in the transmission of motion-JPEG and JPEG images," *Computers Graphics: Image Commun.*, vol. 18, no. 1, pp. 11–19, Jan.–Feb. 1994.

[32] H. Sun and J. Zdepski, "Adaptive error concealment algorithm for MPEG compressed video," in *Proc. SPIE Conf. Vis. Commun. Image Processing*, vol. 1818, Boston, MA, Nov. 1992, pp. 814–824.

[33] W. Kwok and H. Sun, "Multidirectional interpolation for spatial error concealment," *IEEE Trans. Consumer Electron.*, vol. 3, pp. 455–460, Aug. 1993.

[34] H. Sun and W. Kwok, "Concealment of damaged block transform coded images using projections onto convex sets," *IEEE Trans. Image Processing*, vol. 4, pp. 470–477, Apr. 1995.

[35] S. Aign and K. Fazel, "Error detection and concealment measures in MPEG-2 video decoder," in *Proc. Int. Workshop HDTV*, Torino, Italy, Oct. 1994.

[36] ——, "Temporal and spatial error concealment techniques for hierarchical MPEG-2 video codec," in *Proc. Int. Workshop Commun.*, Seattle, WA, June 18–20, 1995, pp. 1778–1783.
[37] S. Aign, "Error concealment enhancement by using the reliability outputs of a SOVA in MPEG-2 video decoder," in *Proc. URSI Int. Symp. Signal, Syst., Electron.*, San Francisco, CA, Oct. 25–27, 1995, pp. 59–62.
[38] J. F. Shen and H. M. Hang, "Compressed image concealment and postprocessing for digital video recording," in *Proc. IEEE Asia-Pacific Conf. Circuits Syst.*, Taipei, Taiwan, Dec. 5–8, 1994, pp. 636–641.
[39] D. C. Youla and H. Webb, "Image restoration by the method of convex projections: Part 1—Theory," *IEEE Trans. Med. Imag.*, vol. MI-1, pp. 81–94, Oct. 1982.
[40] R. Talluri, "Error resilient video coding in the ISO MPEG-4 standard," *IEEE Commun. Mag.*, vol. 36, pp. 112–119, June 1998.
[41] J. Liang and R. Talluri, "Tools for robust image and video coding in JPEG2000 and MPEG4 standards," in *Proc. SPIE Conf. Vis. Commun. Image Processing*, vol. 3653, San Jose, CA, Jan. 23–29, 1999, pp. 40–51.
[42] E. Asbun and E. J. Delp, "Real-time error concealment in compressed digital video streams," in *Proc. Picture Coding Symp.*, Portland, OR, Apr. 21–23, 1999.
[43] R. L. Stevenson, B. E. Schmitz, and E. J. Delp, "Discontinuity preserving regularization of inverse visual problems," *IEEE Trans. Syst. Man Cybern.*, vol. 24, pp. 455–469, Mar. 1994.
[44] D. Geman and G. Reynolds, "Constrained restoration and the recovery of discontinuities," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, pp. 367–382, Mar. 1992.
[45] C. Bouman and K. Sauer, "A generalized Gaussian image model for edge-preserving MAP estimation," *IEEE Trans. Image Processing*, vol. 2, pp. 296–310, July 1993.
[46] J. Marroquin, S. Mitter, and T. Poggio, "Probabilistic solution of ill-posed problems in computational vision," *J. Amer. Stat. Assoc.*, vol. 82, no. 397, pp. 76–89, Mar. 1987.
[47] R. Schultz and R. L. Stevenson, "A Bayesian approach to image expansion for improved definition," *IEEE Trans. Image Processing*, vol. 3, pp. 233–241, May 1994.
[48] P. J. Huber, *Robust Statistics*. New York: Wiley, 1981.
[49] J. Besag, "On the statistical analysis of dirty pictures," *J. Royal Stat. Soc.*, ser. B, vol. 48, no. 3, pp. 259–302, 1986.
[50] J. Konrad and E. Dubois, "Bayesian estimation of motion vector fields," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, pp. 910–926, Sept. 1992.
[51] J. Li, X. Lin, and C. C. J. Kuo, "Boundary control vector motion field representation and estimation by using a markov random field model," *J. Vis. Commun. Image Represent.*, vol. 7, no. 3, pp. 230–243, Sept. 1996.
[52] W. M. Lam, A. R. Reibman, and B. Liu, "Recovery of lost or erroneously received motion vectors," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, Minneapolis, MN, Apr. 27–30, 1993, pp. V417–V420.
[53] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd ed. New York: McGraw-Hill, 1984.

Paul Salama (S'94–M'00) received the B.S. degree (first class honors) from the University of Khartoum, and the M.S.E.E. and Ph.D. degrees from Purdue University, West Lafayette, IN.

He is currently an Assistant Professor in the Department of Electrical and Computer Engineering, Purdue School of Engineering and Technology, IUPUI, Indianapolis. His research interests include image and video compression, image processing, and medical imaging.

Dr. Salama is a member of SPIE, Tau Beta Pi, and Eta Kappa Nu.

Ness B. Shroff (S'90–M'94) received the B.S. degree from the University of Southern California, Los Angeles, the M.S.E. degree from the University of Pennsylvania, Philadelphia, and the M.Phil. and Ph.D. degrees from Columbia University, New York, NY.

He is currently an Assistant Professor in the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN. His current research interests are in broadband networks and wireless communication networks.

Dr. Shroff was the Conference Chair for the 14th Annual IEEE Computer Communications Workshop (CCW) and is the program co-chair for the symposium on High-Speed Networks at the IEEE Globecomm'2000. He is the secretary of the IEEE Communications Society Technical Committee on Computer Communications (TCCC). He is also on the Editorial Board for Computer Networks and IEEE COMMUNICATION LETTERS. He received the NSF CAREER award from the National Science Foundation in 1996.

Edward J. Delp (S'70–M'79–SM'86–F'97) was born in Cincinnati, OH. He received the B.S.E.E. (cum laude) and M.S. degrees from the University of Cincinnati, Cincinnati, OH, and the Ph.D. degree from Purdue University, West Lafayette, IN.

From 1980 to 1984, he was with the Department of Electrical and Computer Engineering, University of Michigan, Ann Arbor. Since August 1984, he has been with the School of Electrical and Computer Engineering at Purdue University, where he is a Professor of electrical engineering. His research interests include image and video compression, image processing, multimedia security, medical imaging, parallel processing, multimedia systems, nonlinear filtering, and communication and information theory. He has also consulted for various companies and government agencies in the areas of signal and image processing, robot vision, pattern recognition, and secure communications.

Dr. Delp is a member of Tau Beta Pi, Eta Kappa Nu, Phi Kappa Phi, Sigma Xi, ACM, and the Pattern Recognition Society. He is a Fellow of the SPIE, and a Fellow of the Society for Imaging Science and Technology (IS&T). In 1997, he was elected Chair of the Image and Multidimensional Signal Processing (IMDSP) Technical Committee of the IEEE Signal Processing Society. From 1994 to 1998, he was Vice-President for Publications of IS&T. He was Co-Chair of the SPIE/IS&T Conference on Security and Watermarking of Multimedia Contents held in San Jose in January 1999. He was the General Co-Chair of the 1997 Visual Communications and Image Processing Conference (VCIP) held in San Jose. He was Program Chair of the IEEE Signal Processing Society's Ninth IMDSP Workshop held in Belize in March 1996. He was General Co-Chairman of the 1993 SPIE/IS&T Symposium on Electronic Imaging.