Linköping University Post Print

Packet Video Error Concealment With Gaussian Mixture Models

Daniel Persson, Thomas Eriksson and Per Hedelin

N.B.: When citing this work, cite the original article.

©2009 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

Daniel Persson, Thomas Eriksson and Per Hedelin, Packet Video Error Concealment With Gaussian Mixture Models, 2008, IEEE Transactions on Image Processing, (17), 2, 145-154.
<http://dx.doi.org/10.1109/TIP.2007.914151>

Postprint available at: Linköping University Electronic Press
<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-53662>

Packet Video Error Concealment With Gaussian Mixture Models

Daniel Persson, Thomas Eriksson, and Per Hedelin

Abstract—In this paper, Gaussian mixture modeling is applied to error concealment for block-based packet video. A Gaussian mixture model for video data is obtained offline and is thereafter utilized online in order to restore lost blocks from spatial and temporal surrounding information. We propose estimators on closed form for missing data in the case of varying available neighboring contexts. Our error concealment strategy increases peak signal-to-noise ratio compared to previously proposed schemes. Examples of improved subjective visual quality by means of the proposed method are also supplied.

Index Terms—Error concealment, Gaussian mixture model (GMM), packet video, video modeling.

I. INTRODUCTION

BLOCK-BASED video coders such as MPEG-1, MPEG-2, MPEG-4, H.261, and H.263 [1] are frequently used for digital video compression. The bandwidth requirements are met in this way, but the sensitivity to transmission channel impairments increases. Packet errors, where much information is lost at the same time, are caused by noisy channels and error propagation in the decoder.

Error concealment is a postprocessing technique for recreating the original video stream from redundancy in the stream with errors at the decoder. Efforts are usually categorized into spatial approaches that use spatially surrounding pixels for estimation of lost blocks, and temporal approaches, that replaces lost pixels with pixels in previous frames by means of motion vectors.

A. Previous Efforts

In order to show how our contribution fits into the history of the problem, we shortly revise a few famous spatial and temporal methods, and also some spatiotemporal methods that combine both approaches.

Spatial methods may yield better performance than temporal methods in scenes with high motion, or after a scene change. Lost transform coefficients are linearly interpolated from the same coefficients in adjacent blocks in [2]. Minimization of a first-order derivative-based smoothness measure was proposed for spatial error concealment in [3]. In order to reduce the blurring of edges, second-order derivatives are considered in [4]. A

replacement block is formed by iterative projections of the lost block and its surrounding onto two convex sets that guarantee that the replacement block has in-range color values and frequency content matching the surrounding in [5]. Further, in [6], recovery vectors, containing both known and unknown pixels, are alternately projected on the best-matched surrounding, and on convex sets guaranteeing in-range color values and a maximum difference between adjacent color values. Details inside lost blocks cannot be recreated by spatial approaches. In this case, information from the past frame may improve the result.

For temporal error concealment, rather than using the block at the same position as the lost block in the previous frame for replacement, the motion-compensated block should be used [7]. If the motion vector (MV) is available at the decoder side, it can be utilized for motion-compensated error concealment. When the MV is also lost, it has to be estimated. This is the major challenge in temporal error concealment. MV estimation is often performed by using the median of the MVs of the surrounding blocks, or the MV of the corresponding block in the previous frame [8]. The MV that yields the minimum difference between a replacement block and its spatial surrounding is chosen as an estimate in [9]. In [10], the missing MVs are estimated in a two-stage maximum *a posteriori* (MAP) process first considering a Markov random field (MRF) model for MVs, and then a MRF model for pixels. The spatial and temporal contexts are considered at the same time in order to find the MVs in [11], using a multiscale adaptive Huber MRF-MAP scheme.

From an information theoretic perspective, replacing a lost block with both spatial and temporal context should be superior to only using one of the two types of information. A first derivative-based smoothness measure yields a spatiotemporal replacement in [12]. More specifically, an objective function imposing smooth transitions in space and time is minimized offline, and yields a replacement for the lost block combining transform coefficients, pixels on the border of the lost block, and pixels from a previous frame. A constant that is as well defined offline sets the level of spatial and temporal smoothing.

An adaptive Gaussian MRF model for the prediction error field yields a MAP estimate of missing pixel values based on spatial and temporal information in [13]. In a first stage of [13], MVs are estimated if not present. Thereafter, the prediction error field is modeled as a Gaussian MRF, and a MAP estimate of the prediction error field for the lost block is formed. The weight corresponding to the difference between a pixel and one of the pixels in its clique is set adaptively, depending on edges in the blocks surrounding of the loss whose directions imply that they pass through the missing block.

A mixture of principal components for spatiotemporal error concealment of tracked objects is proposed in [14].

Manuscript received August 11, 2005; revised October 19, 2007. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Yucel Altunbasak.

The authors are with the Department of Signals and Systems at Chalmers University of Technology, S-412 96 Göteborg, Sweden (e-mail: f97danp@chalmers.se; thomase@chalmers.se; per.hedelin@chalmers.se).

Digital Object Identifier 10.1109/TIP.2007.914151

B. Our Contribution

In this paper, we propose an error concealment method that combines spatial information and motion-compensated pixels from a previous frame, given MVs. Our scheme may be employed with correctly received MVs, possibly delivered to the receiver in a base layer, or with any of the techniques [8]–[10], and [11] for estimating MVs in the case where they are lost.

The approach is based on Gaussian mixture modeling (GMM)¹ of adjacent pixel color values. It is known that a GMM may describe distributions arbitrarily well by increasing the number of mixture components; see, for example, [15]. GMM has been used for a variety of tasks in image processing, e.g., object detection in images in [16] and noise reduction, image compression, and texture classification in [15]. Our GMM-based estimator can be seen as a soft classifier that combines different Gaussian solutions with weights that depend on the current video behavior. Previous work [17] has showed that an *ad-hoc* classification of pixels increases performance when interpolating skipped frames.

In our formulation, the problem of estimation of lost pixel blocks is split into an offline model parameter estimation problem solved by means of the expectation maximization (EM) algorithm, and an online minimum mean square error (MMSE)-based estimation of lost pixels from the surrounding context using the previously obtained model parameters. Introduced model assumptions are carefully stated.

When several neighboring macroblocks are assigned to the same packet, and variable-length coding is employed between packets, a packet loss may lead to a big loss locally in the video stream [18]. The error robustness of this scheme may be substantially enhanced by the simple block interleaving strategy proposed in [12]. In this way, in [12], the error concealment algorithm usually has access to surrounding spatial information. Since the block interleaving is performed frame by frame, it does not increase the algorithmic delay. Also, it was shown in [12] that this interleaving scheme did not give rise to any important decrease in compression gain. In this paper, we employ an interleaving scheme similar to [12] in order to achieve robust coding.

Some introductory work for this paper was presented in [19] and [20].

The rest of the paper is organized as follows. In Section II, modeling by means of GMM is investigated and estimates of lost pixel information are derived for various situations. The estimators are thereafter experimentally evaluated for error concealment in Section III. Section IV concludes the paper.

II. ESTIMATION OF LOST PIXEL AREAS

In this section, we will derive MMSE estimates of lost pixel areas. The MVs are considered to be available at the decoder or previously estimated on the decoder side. To keep the treatment general, we avoid specifying the spatial and temporal location of the modeled pixels for now. When a part of the video data is missing, we make an MMSE estimate of it from its context by means of a GMM model. However, there are cases when a

part or all of the surrounding context is also missing. Under such conditions, we resort to special extensions of the theory in order to conceal the loss. Section II-A introduces our stochastic notation and the GMM model. We consider estimation in the specific situation of fully available modeled context in Section II-B. Thereafter, an investigation of the case of partially missing modeled context follows in Section II-C.

A. GMM

Parts of the video are represented by multivariate stochastic variables. The lost pixels are represented by a vector X and its surrounding pixels are represented by a vector Y . An MMSE estimate of X from Y may be formed by considering a model for $Z^T = [X^T, Y^T]$ and the values y of Y . We will from now on refer to Y as the *modeled context* to X . A GMM for the probability density function (pdf) of Z is

$$f_Z(z) = \sum_{m=1}^M \rho^{(m)} f_Z^{(m)}(z) \quad (1)$$

where $f_Z^{(m)}(z)$ are Gaussian densities with means $\mu_Z^{(m)}$ and covariances $C_{ZZ}^{(m)}$. The weights $\rho^{(m)}$ are all positive and sum to one. In all that follows, we will assume that our models describe the modeled parts of the source perfectly.²

B. Modeled Context Available

If all values y of the modeled context Y are available, we may form an MMSE estimator of X

$$\hat{x}(y) = E[X | Y = y] = \int x f_{X|Y}(x|y) dx. \quad (2)$$

In order to derive an expression for this estimator, we first have to evaluate

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}. \quad (3)$$

The pdf $f_{X,Y}(x,y) = f_Z(z)$ is known in (1). The marginal pdf of Y can be computed from $f_{X,Y}(x,y)$ as

$$f_Y(y) = \int f_{X,Y}(x,y) dx \quad (4)$$

$$= \int \sum_{m=1}^M \rho^{(m)} f_{X,Y}^{(m)}(x,y) dx \quad (5)$$

$$= \sum_{m=1}^M \rho^{(m)} \int f_{X,Y}^{(m)}(x,y) dx \quad (6)$$

$$= \sum_{m=1}^M \rho^{(m)} f_Y^{(m)}(y). \quad (7)$$

The functions $f_Y^{(m)}(y)$ are Gaussian densities with means $\mu_Y^{(m)}$ and covariances $C_{YY}^{(m)}$ where

$$\mu_Z^{(m)} = \begin{bmatrix} \mu_X^{(m)} \\ \mu_Y^{(m)} \end{bmatrix} \quad C_{ZZ}^{(m)} = \begin{bmatrix} C_{XX}^{(m)} & C_{XY}^{(m)} \\ C_{YX}^{(m)} & C_{YY}^{(m)} \end{bmatrix}. \quad (8)$$

¹It will be clear from the context whether the acronym GMM refers to Gaussian mixture model or Gaussian mixture modeling.

²While this assumption is not true in general, GMM has been successfully used in many previous applications.

Inserting (7) and (1) in (3), we get

$$f_{X|Y}(x|y) = \frac{\sum_{m=1}^M \rho^{(m)} f_{X,Y}^{(m)}(x,y)}{\sum_{k=1}^M \rho^{(k)} f_Y^{(k)}(y)} \quad (9)$$

$$= \sum_{m=1}^M \underbrace{\frac{\rho^{(m)} f_Y^{(m)}(y)}{\sum_{k=1}^M \rho^{(k)} f_Y^{(k)}(y)}}_{\pi^{(m)}(y)} f_{X|Y}^{(m)}(x|y) \quad (10)$$

$$= \sum_{m=1}^M \pi^{(m)}(y) f_{X|Y}^{(m)}(x|y). \quad (11)$$

For a fixed value y , the functions $f_{X|Y}^{(m)}(x|y)$ are Gaussian densities with means $\mu_{X|Y}^{(m)}(y)$ and covariances $C_{X|Y}^{(m)}$ where

$$\mu_{X|Y}^{(m)}(y) = C_{XY}^{(m)} \left(C_{YY}^{(m)} \right)^{-1} \left(y - \mu_Y^{(m)} \right) + \mu_X^{(m)} \quad (12)$$

$$C_{X|Y}^{(m)} = C_{XX}^{(m)} - C_{XY}^{(m)} \left(C_{YY}^{(m)} \right)^{-1} C_{YX}^{(m)}. \quad (13)$$

The function $\pi^{(m)}(y)$ is the *a posteriori* probability for mixture component density m given y . The *a posteriori* probabilities sum to one

$$\sum_{m=1}^M \pi^{(m)}(y) = \sum_{m=1}^M \frac{\rho^{(m)} f_Y^{(m)}(y)}{\sum_{k=1}^M \rho^{(k)} f_Y^{(k)}(y)} = 1. \quad (14)$$

This implies that (11) is a GMM for a fixed value y . By means of (2) and (11), we may now compute our MMSE estimator as

$$\hat{x}(y) = \int x f_{X|Y}(x|y) dx \quad (15)$$

$$= \int x \sum_{m=1}^M \pi^{(m)}(y) f_{X|Y}^{(m)}(x|y) dx \quad (16)$$

$$= \sum_{m=1}^M \pi^{(m)}(y) \int x f_{X|Y}^{(m)}(x|y) dx \quad (17)$$

$$= \sum_{m=1}^M \pi^{(m)}(y) \mu_{X|Y}^{(m)}(y). \quad (18)$$

As expected, the estimator is a function of the known values y .

C. Missing Modeled Context

In this section, we still want to estimate X from Y but some of the values of the vector Y are now missing. We divide the vector Z into three vector parts, $Z^T = [X^T, Y_1^T, Y_2^T]$, where the values of X are to be estimated, the values y_1 of Y_1 are known, and the values of Y_2 are missing. Similarly to (8), the means and covariances of the components of the GMM (1) are

$$\mu_Z^{(m)} = \begin{bmatrix} \mu_X^{(m)} \\ \mu_{Y_1}^{(m)} \\ \mu_{Y_2}^{(m)} \end{bmatrix} \quad C_{ZZ}^{(m)} = \begin{bmatrix} C_{XX}^{(m)} & C_{XY_1}^{(m)} & C_{XY_2}^{(m)} \\ C_{Y_1X}^{(m)} & C_{Y_1Y_1}^{(m)} & C_{Y_1Y_2}^{(m)} \\ C_{Y_2X}^{(m)} & C_{Y_2Y_1}^{(m)} & C_{Y_2Y_2}^{(m)} \end{bmatrix}. \quad (19)$$

We will study three possible solutions: marginalization, estimation based on data that are external to the model, and repeated estimation.

- *Marginalization.* We may choose to estimate X from Y_1 alone. We then have to get rid of the missing part Y_2 of the modeled context in (3) by marginalization. By applying the treatment in Section II-B, we arrive at an MMSE estimator

$$\hat{x}(y_1) = \sum_{m=1}^M \pi^{(m)}(y_1) \mu_{X|Y_1}^{(m)}(y_1) \quad (20)$$

where the weights and means are given by

$$\pi^{(m)}(y_1) = \frac{\rho^{(m)} f_{Y_1}^{(m)}(y_1)}{\sum_{k=1}^M \rho^{(k)} f_{Y_1}^{(k)}(y_1)} \quad (21)$$

$$\mu_{X|Y_1}^{(m)}(y_1) = C_{XY_1}^{(m)} \left(C_{Y_1Y_1}^{(m)} \right)^{-1} \left(y_1 - \mu_{Y_1}^{(m)} \right) + \mu_X^{(m)}. \quad (22)$$

- *Estimation based on unmodeled context.* Assume that the values of Y_2 are missing, but that we have access to the values u of the vector U that represents a neighborhood that is external to the model. Suppose further that we have a model $f_{Y,U}(y, u)$ and that U is conditionally independent of X given Y , i.e., we have a Markov model

$$f_{U|X,Y}(u|x,y) = f_{U|Y}(u|y). \quad (23)$$

An MMSE estimate of X may then be computed from Y_1 and U

$$\hat{x}(y_1, u) = \int \int x f_{X|Y_1, Y_2}(x|y_1, y_2) \times f_{Y_2|Y_1, U}(y_2|y_1, u) dx dy_2. \quad (24)$$

We consider all models to have the same number of Gaussian component densities. In this case, the MMSE estimator (24) is only obtainable on closed form when $M = 1$.

- *Repeated estimation.* If the value of the modeled context Y_2 is unavailable but we have access to a previous estimate $\hat{y}_2$ of Y_2 , we might form an estimate of X using (18)

$$\hat{x}(y_1, \hat{y}_2) = \sum_{m=1}^M \pi^{(m)} \left(\begin{bmatrix} y_1 \\ \hat{y}_2 \end{bmatrix} \right) \mu_{X|Y}^{(m)} \left(\begin{bmatrix} y_1 \\ \hat{y}_2 \end{bmatrix} \right) \quad (25)$$

where $\pi^{(m)}(y)$ and $\mu_{X|Y}^{(m)}(y)$ are computed as in (10) and (12), respectively. It is shown in the Appendix that when $M = 1$, the repeated estimation and the estimation based on unmodeled context are the same. This means that in the case when $M = 1$, repeated estimation is MMSE optimal. For a general M , there is no MMSE optimality measure for repeated estimation. The advantage of repeated estimation lies in its ease of implementation.

III. EXPERIMENTS

The derived estimators from Section II will now be applied for concealment of lost packets in transmitted video sequences. Our error concealment scheme is integrated into a generic block-based coder with block size 8×8 pixels. For error concealment, the lost 8×8 blocks are divided into blocks X of size 4×4 pixels that are concealed one by one, cf. Section II-A and Fig. 1. The lost block X has a modeled context Y

Fig. 1. Typical error concealment situation. An 8×8 -block in frame t is lost. Error concealment is performed by estimation of one 4×4 -block at a time. The 4×4 -block X is currently being estimated.

Fig. 2. For the estimation of the lost 4×4 -block X , a modeled context, containing spatially and temporally surrounding pixels Y , is being used. The vector $Z^T = [X^T, Y^T]$.

containing spatially and temporally adjacent pixels, see Fig. 2. Estimates of X from Y are formed by means of a model (1) for $Z^T = [X^T, Y^T]$. Some or all of the values of Y may also be lost at the receiver. In this case, we have to resort to the treatments in Section II-C for the concealment of X . For reasons of computational complexity, we choose to work with estimators on closed form, i.e., we choose to combine marginalization (20) and repeated estimation (25) in cases when parts of the modeled context Y are lost. Simulation details are given in Section III-A. Section III-B presents the results.

A. Prerequisites

The prerequisites are chosen to comply with state-of-the-art block-based video coders, and are impartial to all the compared schemes.

Coder: The frames are predictively coded (P-frames) (An application of our method to restoration of intracoded frames (I-frames) is completely analogous) and the corresponding prediction errors are sent. MVs are calculated for 8×8 -blocks. A search for a MV is performed by checking every integer displacement vector $(\Delta x, \Delta y)$ where $-16 \leq \Delta x, \Delta y \leq 16$. The coder works in the limit of perfect quantization.

Motion Vectors for Error Concealment: The error concealment scheme is evaluated in the case of correctly received MVs

Fig. 3. Block interleaving. One row of 16×16 -blocks is separated into two packets.

Fig. 4. Four situations when a 4×4 -block X in a lost 8×8 -block is estimated from an available surrounding Y_1 . These cases can all be handled by prestoring one estimator and mirroring X and Y_1 .

that are protected in a high priority layer, and in the case of lost MVs that are estimated by the median of the MVs of the available neighboring blocks [8]. Separate GMMs are trained for these two cases.

Benchmarking: The GMM-based estimator is compared to two other schemes that mix spatial and temporal information given the MVs: namely the methods in [12] and [13]. Also, motion-compensated copying [8] is used as a reference method. Two versions of our scheme are compared to the previously proposed methods: A GMM with $M = 64$ and a GMM with only one Gaussian component. It is easy to show that (18) with $M = 1$ is identical to the solution of the linear MMSE estimation problem [21]. In every experiment, all methods use the same motion-compensated previous pixels.

Mirror Invariance: Estimators based on marginalization according to (20) for different cases of missing surrounding pixels, are precomputed offline. The MVs are calculated for 8×8 -blocks whereas the models are trained for 4×4 -blocks. By means of mirroring the realizations of Z , see Fig. 4, an estimator can be utilized in four different situations. Using mirroring, 16 instead of 64 estimators need to be prestored.

GMM Parameter Estimation: The EM algorithm [22] for training of mixture densities is treated in [23]. It is shown in [23] that the EM algorithm guarantees an increasing log-likelihood from iteration to iteration. For the case interesting in this paper, the standard EM algorithm performs well and is, thus, used to obtain models of the form (1).

Numerical problems may arise if the covariance matrices become close to singular [24]. This occurs in the limits of many mixture components, small number of realizations in the database, and many dimensions. In order to avoid singularities, the covariance matrices are monitored and the eigenvalues were not allowed to decrease below a threshold. Since open tests are run, the results would be better if more data were used in the training.

The means of the mixture components are initialized by an estimate of the source mean. For the initialization of the covariances of the components, individual covariance matrices for the components are created by adding different small positive numbers to the eigenvalues of the estimated source covariance matrix. In the EM algorithm, 20 iterations are run to achieve convergence.

Data: We use the luminance component of 124 MPEG-1 movies from [25] that have a frame rate of 29.97 frames per second and an image size of 352×240 pixels. The movies are divided into two sets, one for GMM parameter estimation and another for evaluation. In order to show the robustness of our scheme, we use more movies for the evaluation than for the training. The sets used for parameter estimation and evaluation contain 35 and 89 randomly selected movies respectively. Also, for subjective visual evaluation, an MP4 movie from [26] was used.

Evaluation Criterion: The peak signal-to-noise ratio (PSNR), calculated for the lost pixel blocks, is used for evaluation.

B. Results

The experiments are divided into four groups. First, the offline GMM parameter estimation is investigated. Then spatial, temporal, and spatiotemporal error concealment by means of GMM are compared. Further, the measures in case of missing modeled context discussed in Section II-C are addressed. Finally, we compare our scheme to previous state-of-the-art error concealment methods.

- *GMM parameter estimation.* In this experiment, offline GMM parameter estimation by means of the EM algorithm is considered. Models are obtained for Z in Fig. 2. The MV is lost, and estimated by the median of the MVs of the neighboring blocks. For GMM training, 1 470 000 realizations of Z in Fig. 2 are drawn from the training set in a uniformly random manner and in such a way that no two vectors coincide. For the evaluation, 480 000 realizations of Z are drawn from the evaluation set in the same way. The log-likelihood for the realizations of Z in the evaluation set is shown in Fig. 5 for models with different numbers of mixture components M . As we can see, the log-likelihood increases as a function of the number of mixture components. In Fig. 6, the PSNR for the estimation of X from Y according to (18) is shown for models with different numbers of mixture components M . Confidence intervals have been calculated, assuming that the square Euclidean norm of the difference between X and its estimate is distributed according to a normal distribution. The 0.95-confidence intervals are marked by dashed lines in Fig. 6. By augmenting M from 1 to 64, we increase PSNR

Fig. 5. Log-likelihood for 480 000 realizations of Z as in Fig. 2 in the evaluation set, as a function of the number of mixture components M . MVs are estimated by the median of the MVs of the neighboring blocks.

Fig. 6. PSNR for the estimate of X from Y as in Fig. 2 for 480 000 realizations of Z in the evaluation set, as a function of the number of mixture components M . MVs are estimated by the median of the MVs of the neighboring blocks. The 0.95-confidence bounds are marked by dashed lines.

by 2.6 dB while the computational complexity for the estimation increases linearly. Copying motion-compensated past information gives a PSNR of 29.4 dB if the MV is estimated by the median of the MVs of the surrounding blocks. We conclude that augmenting the number of mixture components in the GMM-based estimator is beneficial when there is access to spatial and temporal information. By comparing Figs. 5 and 6, we see that increasing the log-likelihood does not necessarily yield a corresponding increase in PSNR. In the case when the MV is correctly received, the PSNR values increase, but the conclusions remain the same.

- *Spatial, temporal, and spatiotemporal error concealment by means of GMM.* Fig. 7 shows estimation of X from different modeled contexts in the case when the MV is estimated by the median of the MVs of the neighboring blocks. The PSNR given by an estimator with $M = 64$ is shown in the figure. By comparing $\mathcal{A}$ and $\mathcal{B}$, we see that temporal data are valuable for the creation of an estimate of the lost part. From a comparison of $\mathcal{A}$ and $\mathcal{C}$, it is noticed that spatial data are also important. The PSNR in $\mathcal{C}$ is almost

Fig. 7. Estimation of X from different modeled contexts. Frame numbers $t-1$ and t are seen in $\mathcal{A}$ and remain the same in the other problems. The number of mixture components $M = 64$. MVs are estimated by the median of the MVs of the neighboring blocks. Performance in PSNR, for 480 000 realizations of Z in the evaluation set, is shown for each experiment.

Fig. 8. Spatial marginalization and spatial repeated estimation for varying loss rates. A few tens of randomly chosen consecutive frames from each of the evaluation movies are coded. Packet errors are distributed in an independently random manner. The MV is estimated by the median of the MVs of the neighboring blocks. In case of lost pixels in a previous frame, repeated estimation is used in all experiments.

as low as the PSNR obtained by copying motion-compensated previous pixels. This means that GMM does not improve performance compared to trivial error concealment if it only has access to temporal information. Through comparison of Figs. 6 and 7, we observe that a GMM with $M = 1$ and access to both spatial and temporal context performs almost 3 dB better than a GMM with $M = 64$ and access to temporal context only. We conclude that a combination of spatial and temporal information is beneficial for GMM-based estimation of the lost pixels. In the case when the MV is correctly received, the PSNR values increase, but the conclusions about the behavior of the GMM-based estimator remain the same.

- *Measures in case of missing modeled context.* In the case of temporally adjacent lost blocks, we utilize repeated estimation from previously corrected information. This strategy is

Fig. 9. Performance of the different error concealment methods for varying loss rates in the case when the MVs are estimated by the median of the MVs of the surrounding blocks. A few tens of randomly chosen consecutive frames from each of the evaluation movies are coded. Packet errors are distributed in an independently random manner.

Fig. 10. Performance of the different error concealment methods for varying loss rates in the case when the MVs are correctly received. A few tens of randomly chosen consecutive frames from each of the evaluation movies are coded. Packet errors are distributed in an independently random manner.

applied by many others, e.g., in [12] and [13]. For spatially adjacent lost blocks, a comparison between marginalization according to (20) and repeated estimation according to (25) for different loss rates is presented in Fig. 8. Each row of 16×16 -blocks is separated into two packets according to Fig. 3. A few tens of randomly chosen consecutive frames from each of the evaluation movies are coded. Packet errors are distributed in an independently random manner. The MV is estimated by the median of the MVs of the neighboring blocks. In the case when the MV is correctly received, the PSNR values increase, but the conclusions about the behavior of the GMM-based estimator remain the same.

Since the performances of marginalization and repeated estimation are almost the same, marginalization should be chosen because it has lower computational complexity. *If some spatially neighboring pixels are missing and previously estimated, the corresponding variables are marginalized according to (20) in the following experiments. Also in*

Fig. 11. Restoration of a coded frame with fast motion, by means of the different error concealment methods, in the case of a previous frame without errors, and lost MVs that are estimated by the median of the MVs of the neighboring blocks. (a) Original frame; (b) previous frame; (c) error pattern; (d) motion-compensated copying; (e) method in [12]; (f) method in [13]; (g) GMM $M = 1$; (h) GMM $M = 64$. The used movie clip was originally encoded as MPEG-1 and taken from [25].

the following, in case of temporally adjacent lost blocks, we utilize repeated estimation from previously corrected pixels according to (25).

- *Comparison to previous state-of-the-art error concealment schemes.* Table I presents the performance of the different error concealment methods in the case of temporally and spatially isolated lost 8×8 -blocks. If the MVs are lost, they are estimated by the median of the MVs of the surrounding blocks. A few tens of randomly chosen frames from each of the evaluation movies are used for evaluation. Fig. 9 presents the performance of the different error concealment methods for different loss rates. Each row of 16×16 -blocks is separated into two packets according to Fig. 3. A few tens of randomly chosen consecutive frames from each of the evaluation movies are coded. Packet er-

rors are distributed in an independently random manner. The MVs are estimated by the median of the MVs of the surrounding blocks. Fig. 10 presents the performance of the different methods under the same conditions but with available MVs on the decoder side.

Figs. 11 and 12 present restorations of coded frames with fast and slow motions, respectively, by means of the different error concealment methods, in the case of a previous frame without errors, and lost MVs that are estimated by the median of the MVs of the neighboring blocks.

In Figs. 11 and 12, (a) shows the original frame, (b) shows the previous frame, where motion-compensated pixels are extracted for error concealment, (c) shows the error pattern, and (d)–(h) show the results obtained with the different error concealment methods.

Fig. 12. Restoration of a coded frame with slow motion, by means of the different error concealment methods, in the case of a previous frame without errors, and lost MVs that are estimated by the median of the MVs of the neighboring blocks. (a) Original frame; (b) previous frame; (c) error pattern; (d) motion-compensated copying; (e) method in [12]; (f) method in [13]; (g) GMM $M = 1$; (h) GMM $M = 64$. The used movie clip was originally encoded as MP4 and was taken from [26].

TABLE I

PERFORMANCE OF DIFFERENT ERROR CONCEALMENT METHODS IN THE CASE OF TEMPORALLY AND SPATIALLY ISOLATED LOST 8×8 -BLOCKS. IF THE MVs ARE LOST, THEY ARE ESTIMATED BY THE MEDIAN OF THE MVs OF THE SURROUNDING BLOCKS. A FEW TENS OF RANDOMLY CHOSEN FRAMES FROM EACH OF THE EVALUATION MOVIES ARE USED FOR EVALUATION

Method	PSNR (dB)	
	MV not received	MV received
GMM $M = 64$	30.4	33.1
GMM $M = 1$	29.4	32.7
Method in [12]	28.9	32.3
Method in [13]	28.4	31.8
Motion-compensated past	28.1	31.9

IV. CONCLUSION

We present a GMM-based method for solving the packet video error concealment problem. An estimator on closed form, that can be modified depending on the available neighborhood, is derived. The only introduced modeling assumptions are the order of the GMM, and the validity of repeated estimation in case of missing temporal information surrounding the loss.

GMM increases performance in PSNR compared to previously proposed methods for spatiotemporal error concealment. The results are valid for a wide range of stationary loss probabilities. It is verified that augmenting the number of mixture components increases performance compared to the usage of only

one Gaussian, and also that a spatiotemporal context is beneficial for GMM-based estimation. Examples of improved subjective visual quality by means of the proposed method are also supplied.

A further increase in performance is expected if more neighboring data of the lost blocks would be incorporated into the model. The stochastic theory is general in the sense that data that are represented in different ways, for example in the pixel and transform domains, may be combined for error concealment without special arrangements. To what extent the two last claims may contribute to improvement of the method remains to be experimentally investigated.

Whereas the GMM is a well-accepted scheme that can describe densities asymptotically, it is possible that there exist other mixtures that work better for small numbers of mixture components, and give a better trade-off between performance and computational complexity. This issue is currently under investigation.

APPENDIX

PROOF OF THE EQUIVALENCE BETWEEN REPEATED ESTIMATION AND ESTIMATES BASED ON UNMODELED CONTEXT IN THE CASE WHEN THE NUMBER OF MIXTURE COMPONENT DENSITIES $M = 1$

Assume that X is estimated from Y_1 , and an estimate of Y_2 that is, in turn, estimated from Y_1 and U . We always consider all involved GMM models to have the same order and so if $M = 1$, $f_{Y_2|Y_1,U}(y_2|y_1,u)$ is a Gaussian pdf. The repeated estimator (25) then is

$$\hat{x}(y_1, \hat{y}_2(y_1, u)) = \mu_{X|Y}^{(1)} \left(\begin{bmatrix} y_1 \\ \hat{y}_2(y_1, u) \end{bmatrix} \right) \quad (26)$$

$$= \mu_{X|Y}^{(1)} \left(\begin{bmatrix} y_1 \\ \mu_{Y_2|Y_1,U}^{(1)} \left(\begin{bmatrix} y_1 \\ u \end{bmatrix} \right) \end{bmatrix} \right) \quad (27)$$

where

$$\begin{aligned} \mu_{Y_2|Y_1,U}^{(1)} \left(\begin{bmatrix} y_1 \\ u \end{bmatrix} \right) &= \begin{bmatrix} C_{Y_2 Y_1}^{(1)} & C_{Y_2 U}^{(1)} \\ C_{Y_1 Y_1}^{(1)} & C_{Y_1 U}^{(1)} \end{bmatrix}^{-1} \left(\begin{bmatrix} y_1 \\ u \end{bmatrix} - \begin{bmatrix} \mu_{Y_1} \\ \mu_U \end{bmatrix} \right) + \mu_{Y_2}. \end{aligned} \quad (28)$$

If (23) holds, by (24), the estimate based on unmodeled context is

$$\begin{aligned} \hat{x}(y_1, u) &= \iint x f_{X|Y_1,Y_2}(x|y_1,y_2) f_{Y_2|Y_1,U}(y_2|y_1,u) dx dy_2 \end{aligned} \quad (29)$$

$$= \iint x f_{X|Y_1,Y_2}(x|y_1,y_2) dx f_{Y_2|Y_1,U}(y_2|y_1,u) dy_2 \quad (30)$$

$$= \int \mu_{X|Y}^{(1)} \left(\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \right) f_{Y_2|Y_1,U}(y_2|y_1,u) dy_2 \quad (31)$$

$$= \mu_{X|Y}^{(1)} \left(\begin{bmatrix} y_1 \\ \mu_{Y_2|Y_1,U}^{(1)} \left(\begin{bmatrix} y_1 \\ u \end{bmatrix} \right) \end{bmatrix} \right) \quad (32)$$

that is the same expression as (27).

REFERENCES

- [1] B. G. Haskell, P. G. Howard, Y. A. LeCun, A. Puri, J. Ostermann, M. R. Civanlar, L. Rabiner, L. Bottou, and P. Haffner, "Image and video coding-emerging standards and beyond," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, no. 7, pp. 814–837, Nov. 1998.
- [2] S. S. Hemami and T. H.-Y. Meng, "Transform coded image reconstruction exploiting interblock correlation," *IEEE Trans. Image Process.*, vol. 4, no. 7, pp. 1023–1027, Jul. 1995.
- [3] Y. Wang, Q.-F. Zhu, and L. Shaw, "Maximally smooth image recovery in transform coding," *IEEE Trans. Commun.*, vol. 41, no. 10, pp. 1544–1551, Oct. 1993.
- [4] W. Zhu, Y. Wang, and Q.-F. Zhu, "Second-order derivative-based smoothness measure for error concealment in DCT-based codecs," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, no. 6, pp. 713–718, Oct. 1998.
- [5] H. Sun and W. Kwok, "Concealment of damaged block transform coded images using projections onto convex sets," *IEEE Trans. Image Process.*, vol. 4, no. 4, pp. 470–477, Apr. 1995.
- [6] J. Park, D. C. Park, R. J. Marks, and M. A. El-Sharkawi, "Recovery of image blocks using the method of alternating projections," *IEEE Trans. Image Process.*, vol. 14, no. 4, pp. 461–474, Apr. 2005.
- [7] Y. Wang and Q.-F. Zhu, "Error control and concealment for video communication: A review," *Proc. IEEE*, vol. 86, no. 5, pp. 974–997, May 1998.
- [8] P. Haskell and D. Messerschmitt, "Resynchronization of motion compensated video affected by ATM cell loss," in *Proc. ICASSP*, Mar. 1992, pp. 545–548.
- [9] W. M. Lam, A. R. Reibman, and B. Liu, "Recovery of lost or erroneously received motion vectors," in *Proc. ICASSP*, Apr. 1993, pp. 417–420.
- [10] P. Salama, N. B. Shroff, and E. J. Delp, "Error concealment in MPEG video streams over ATM networks," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 6, pp. 1129–1144, Jun. 2000.
- [11] Y. Zhang and K.-K. Ma, "Error concealment for video transmission with dual multiscale Markov random field modeling," *IEEE Trans. Image Process.*, vol. 12, no. 2, pp. 236–242, Feb. 2003.
- [12] Q.-F. Zhu, Y. Wang, and L. Shaw, "Coding and cell-loss recovery in DCT-based packet video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 3, no. 3, pp. 248–258, Jun. 1993.
- [13] S. Shirani, F. Kossentini, and R. Ward, "A concealment method for video communications in an error-prone environment," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 6, pp. 1122–1128, Jun. 2000.
- [14] D. S. Turaga and T. Chen, "Model-based error concealment for wireless video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 6, pp. 483–495, Jun. 2002.
- [15] K. Popat and R. W. Picard, "Cluster-based probability model and its application to image and texture processing," *IEEE Trans. Image Process.*, vol. 6, no. 2, pp. 268–284, Feb. 1997.
- [16] J. Zhang and D. Ma, "Nonlinear prediction for Gaussian mixture image models," *IEEE Trans. Image Process.*, vol. 13, no. 6, pp. 836–847, Jun. 2004.
- [17] C.-K. Wong and O. C. Au, "Modified motion compensated temporal frame interpolation for very low bit rate video," in *Proc. ICASSP*, May 1996, vol. 4, pp. 2327–2330.
- [18] M. Ghanbari and V. Seferidis, "Cell-loss concealment in ATM video codecs," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 3, no. 3, pp. 238–247, Jun. 1993.
- [19] D. Persson and P. Hedelin, "A statistical approach to packet loss concealment for video," in *Proc. ICASSP*, Mar. 2005, pp. II-293–II-296.
- [20] D. Persson, T. Eriksson, and P. Hedelin, "Qualitative analysis of video packet loss concealment with Gaussian mixtures," in *Proc. ICASSP*, May 2006, pp. II-961–II-964.
- [21] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
- [22] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Statist. Soc. B*, vol. 39, pp. 1–38, 1977.
- [23] R. A. Redner and H. F. Walker, "Mixture densities, maximum likelihood and the EM algorithm," *SIAM Rev.*, vol. 26, pp. 195–239, 1984.

- [24] D. A. Reynolds and R. C. Rose, "Robust text-independent speaker identification using Gaussian mixture speaker models," *IEEE Trans. Speech Audio Process.*, vol. 3, no. 1, pp. 72–83, Jan. 1995.
- [25] *Prelinger Archives*, [Online]. Available: <http://www.archive.org/details/prelinger>
- [26] *Internet Archive*, [Online]. Available: <http://www.archive.org/index.php>

Daniel Persson was born in Halmstad, Sweden, in 1977. He graduated from Ecole Polytechnique, Paris, France, and received the M.Sc. degree in engineering physics from Chalmers University of Technology, Göteborg, Sweden, in 2002. He is currently pursuing the Ph.D. degree at the Department of Signals and Systems, Chalmers University of Technology.

His research interests are source coding and image processing.

Thomas Eriksson was born in Skövde, Sweden, on April 7, 1964. He received the M.Sc. degree in electrical engineering and the Ph.D. degree in information theory from the Chalmers University of Technology, Göteborg, Sweden, in 1990 and 1996, respectively.

He was with AT&T Labs-Research from 1997 to 1998, and in 1998 and 1999, he was working on a joint research project with the Royal Institute of Technology and Ericsson Radio Systems AB. Since 1999, he has been an Associate Professor at the Chalmers University of Technology, and his research interests include vector quantization, speaker recognition, and system modeling of nonideal hardware.

Per Hedelin was born in Karlskoga, Sweden, in 1948. He received the M.S. and Ph.D. degrees in electrical engineering from the School of Electrical Engineering, Chalmers University of Technology, Göteborg, Sweden, in 1971 and 1976, respectively.

He was appointed Professor of information theory with data communications at Chalmers University of Technology in 1988. His research interests cover several branches of information theory, signal processing, and related subjects. Four basic fields can be distinguished in his work, namely source and channel coding, estimation and optimal filtering, adaptive signal processing and, finally, modeling and speech processing. Speech coding is also often the subject of study. He has been working with a number of different schemes for speech compression such as sinusoidal coding, glottal-pulse coding, and CELP. He has also been active in language processing.