QUALITATIVE ANALYSIS OF VIDEO PACKET LOSS CONCEALMENT WITH GAUSSIAN MIXTURES

Daniel Persson, Thomas Eriksson and Per Hedelin

Chalmers University of Technology
Department of Signals and Systems
412 96 Göteborg
Sweden

ABSTRACT

We have developed a Gaussian mixture model-based technique for the compensation of lost pixel blocks during real time transmission of video. In this paper, we pursue an argument in order to better understand how the Gaussian mixture model-based estimator works. The discussion is supported with subjective evaluations and examples. Naive viewers preferred the result of our proposed technique 72 percent of the time in comparison with the linear minimum mean square error estimator. Our scheme increases performance measured in peak signal-to-noise ratio for all the 11 standard evaluation movie clips that were used.

1 Introduction

We are today gradually moving toward a situation where competitive network-solutions offer the possibility for users to communicate via video. Video streaming is already established as a popular application and peer-2-peer real-time video communications are on the increase. However, because of heavy global traffic load and local traffic bursts, the Internet is not reliable when it comes to transmission performance [1]. Especially in real-time 2-way video communication, these problems make themselves felt. In this scenario, we have to face both high compression requirements and demands for error concealment.

We address the video error concealment problem (see [2] for an overview), by proposing a new estimator for lost blocks. Our estimator is derived from a Gaussian mixture model (GMM) for video data and it uses information in the surrounding to the lost block to yield a replacement. In investigations preliminary to this paper [3], our method was shown to increase performance in peak signal-to-noise ratio (PSNR) compared to the best linear method when both training and evaluation data was taken from the same movie archive [4]. The models used in this paper are also trained on data from this source but evaluated on the standard movie clips.

In this paper, we provide a qualitative analysis of our method. We place the error-concealment scheme as a post-processor after the decoder in order to avoid coder details. The motion vectors are considered as lost for lost blocks at the decoder and are estimated by assuming zero motion. This is known to be a good approximation in scenes with relatively little motion, see [2]. As we will see, GMM-based estimators may understand the current situation of image motion, even if motion vector estimation is excluded in the analysis. This is one of the benefits of the model compared to the linear MMSE (LMMSE) estimator.

The rest of this paper is organized as follows. In Section 2, our Gaussian mixture-based method is explained and a qualitative analysis is performed. Thereafter, in Section 3, simulation details are described. Results of the subjective tests are presented in Section 4. The paper is concluded in Section 5.

2 Gaussian mixture model and estimation

Let the elements of a vector stochastic variable W represent the pixel luminance values in a context of the video containing a number of neighboring pixels in space-time. Each pixel in our context is labeled by spatial indexes x and y and a time index t. In this paper, we consider the motion vectors lost and estimate them by the zero motion vector. Further presume that the pdf of W, $f_W(w)$, can be described by a Gaussian mixture model,

$$f_W(w) = \sum_{m=1}^{M} \rho^m f_W^m(w) \tag{1}$$

where the distributions $f_W^m(w), m = 1...M$ are Gaussian with mean μ_W^m and covariance C_{WW}^m. The weights ρ^m are all positive and sum to one.

We divide the vector W into two parts $W^T = (U^T, V^T)$, where the values of U are assumed to be lost and we wish to use V to estimate the lost values. The GMM-based MMSE estimator of U from V is

$$\hat{u}(v) = \sum_{m=1}^{M} \pi^m(v) \mu_{U|V}^m(v) \tag{2}$$

where

$$\pi^{(m)}(v) = \frac{\rho^{(m)} f_V^{(m)}(v)}{\sum_{k=1}^{M} \rho^{(k)} f_V^{(k)}(v)} \tag{3}$$

and

$$\mu_{U|V}^{(m)}(v) = C_{UV}^{(m)} (C_{VV}^{(m)})^{-1} (v - \mu_V^{(m)}) + \mu_U^{(m)}. \tag{4}$$

The weights $\pi^m(v)$ sum to one. The LMMSE estimator of U,

$$\hat{u}_{\text{LMMSE}}(v) = C_{UV} (C_{VV})^{-1} (v - \mu_V) + \mu_U \tag{5}$$

is achieved by making a Gaussian assumption of the joint distribution of U and V. When comparing (2) and (5), one sees that a benefit of the GMM is that it allows several different modes of operation depending on the value v of V, i.e. given different values of V, the GMM-based estimator provides $\hat{u}(v)$ as different affine transformations of v.

If we assume that each of the matrices C_{WW}^m is stationary

in the sense that each element may be expressed as a function $C^m_{WW}(\Delta x, \Delta y, \Delta t)$, where Δx and Δy are spatial position differences and Δt is the temporal position difference for the pixels in question ($C^m_{WW}(\Delta x, \Delta y, \Delta t) \neq C^m_{WW}(-\Delta x, \Delta y, \Delta t)$ and similarly for Δy), we may average the elements of C^m_{WW} and achieve an estimate of a correlation-like function $R^m_{WW}(\Delta x, \Delta y, \Delta t)$ in space-time. It is obvious that $R^m_{WW}(\Delta x, \Delta y, \Delta t)$ has the property that $R^m_{WW}(\Delta x, \Delta y, \Delta t) \neq R^m_{WW}(-\Delta x, \Delta y, \Delta t)$ (and similarly for Δy). In Figure 1, we see $R^m_{WW}(\Delta x, \Delta y, \Delta t)$ for a mixture with $M = 64$. Each two-square row represents one Gaussian and the squares show $R^m_{WW}(\Delta x, \Delta y, \Delta t)$ for time differences $\Delta t = 0, 1$, from left to right. In each square, Δx and Δy range from -3 to 3.

It is seen in the figure that each Gaussian has specialized for estimating a special situation. Some Gaussians motion-compensate by using information that has moved in a special direction while others ignore temporal information for example. In this way, it is

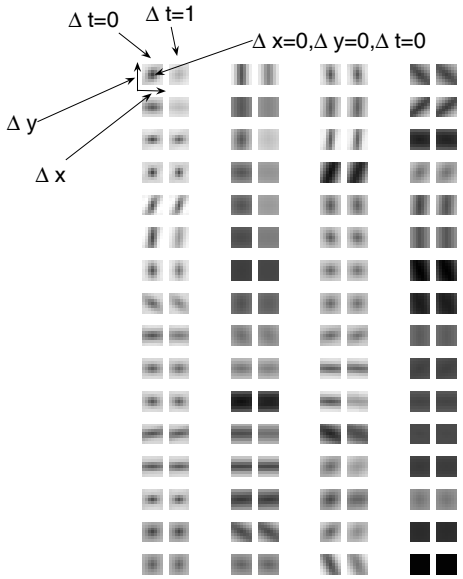

Fig. 1. Correlation-like functions of Δx, Δy and Δt for the Gaussians in a mixture with 64 Gaussians. Each two-square row represents one Gaussian and the squares show the spatial correlation measure for time differences $\Delta t = 0, 1$, from left to right. In each square, Δx and Δy range from -3 to 3.

reasonable to expect that the GMM can understand the current situation from v. For this to be possible, v should provide enough information. Moreover, the method should be more stable if the lost blocks are on a scale that is small compared to the level of detail in the frames.

3 Simulation prerequisites

The estimator is evaluated for error concealment of lost 8×8-blocks distributed as in Figure 2. This error distribution is repeated in two consecutive frames that are followed by one error-free frame and the in this way generated loss-pattern is in turn repeated through the whole of the movie clip. For each lost 8×8-block, one 4×4-block is estimated at a time, see Figure 3. If the future 4×4-block is missing or if spatial information is missing on the sides of the frame, the estimator has to be reformulated. This is done by assuming rotational invariance of the problem and storing six special cases of estimators (2) that may all be achieved in one model training because of the division of W into arbitrary U and V. In preliminary simulations, the assumption of rotational invariance was shown not to affect performance. Already estimated temporal information is reused for estimation in consecutive frames. On the contrary, already estimated spatial information is not reused. An example of a situation in which our method could perform better than the best linear method is seen in Figure 4.

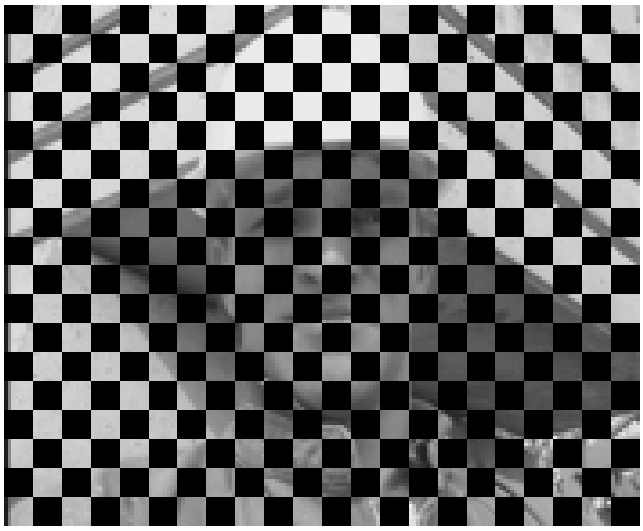

Fig. 2. Distribution of lost 8×8-blocks. This error distribution is repeated in two consecutive frames that are followed by an error-free frame. The in this way obtained three-frame loss-pattern is in turn repeated through the whole movie clip.

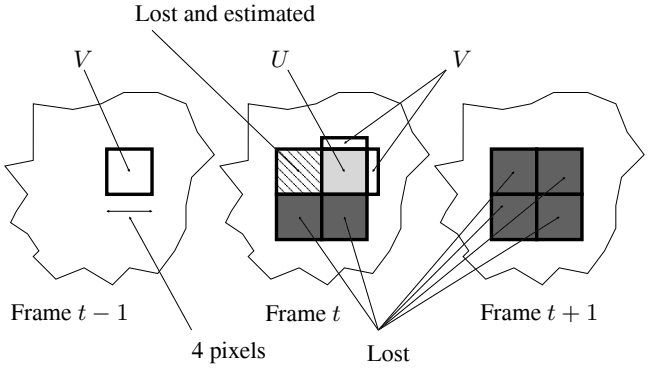

Fig. 3. A lost 8×8-block is concealed, one 4×4-block U from it's context V at a time. In this example, the future information is lost.

4 Subjective evaluation

The subjective evaluation was conducted by showing the same movie clip where error concealment was performed by the LMMSE-based estimator and the GMM-based estimator to 11 naive view-

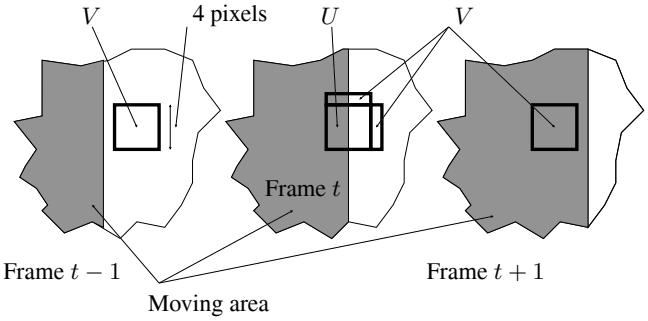

Fig. 4. Example of situation where the GMM is able to understand the motion field. A 4×4-block U is estimated from it's context V.

	Viewers' choices		PSNR	
Clip	LMMSE	GMM	LMMSE	GMM
Miss America	8	3	38.1	39.3
Football	0	11	20.1	20.7
Foreman	5	6	26.9	27.9
Mobile	3	8	18.5	18.7
Container	0	11	29.6	35.8
Carphone	6	5	28.4	29.2
Hall	4	7	30.7	33.9
Silent	1	10	32.1	34.3
Suzie	2	9	32.2	33.4
Tennis	0	11	21.6	23.3
Trevor	5	6	30.8	32.3

Table 1. Results for LMMSE- and GMM-based estimators in terms of viewers' preference and PSNR.

ers and letting the viewers choose the best clip. This was repeated for 11 standard movie clips. The results are presented in Table 1. The viewers' choices were in favor of GMM 72 percent of the time. Only the LMMSE treatment of Miss America and Carphone were preferred over the GMM treatment. Assuming for each algorithm that the number of votes for each movie is the outcome of a Gaussian variable with parameters that can be measured with good precision, the hypothesis that the GMM method gives better subjective quality is 95 percent confident. Performance in PSNR was always better with GMM. In Figure 5 to 7, we observe how the GMM manages to understand the motion and in this way improve performance in comparison to LMMSE. Another benefit of GMM is to lower flickering noise. The GMM may select a bad mode of operation at rare occasions, see Figure 8 to 10, where the square marks the place of a lost 8×8-block in frame 28. One sees how the GMM chooses a bad mode that still can be justified from the information in v seen in the figures. The future information in frame 29 was also lost in this case. Error concealment by means of LMMSE in Figure 9 yields a better result than that of the GMM-estimator seen in Figure 10 in this case.

5 Conclusion

The GMM scheme clearly improves performance in subjective tests and in PSNR compared to the LMMSE estimator in general. In some rare cases, with little information and high detail, the GMM scheme may produce worse results than LMMSE. Future work could be to improve the GMM method without increasing the computational complexity, for example by introducing a maximum correlation length beyond which the elements of the covariance matrices in the mixture are set to zero, and in this way be able to include more pixels in the model.

Fig. 5. Frame 51 of Suzie without losses.

Fig. 6. Frame 51 of Suzie with losses and after error concealment by means of LMMSE.

6 References

[1] S. Kalidindi and M. J. Zekauskas, "Surveyor: An infrastructure for Internet performance measurements," in *Proc. INET*, June 1999.

[2] Y. Wang and Q.-F. Zhu, "Error control and concealment for video communication: a review," *Proc. IEEE*, vol. 86, pp. 974–997, May 1998.

[3] D. Persson and P. Hedelin, "A statistical approach to packet loss concealment for video," in *Proc. ICASSP*, Mar. 2005, pp. 293–296.

[4] "Prelinger archives," http://www.archive.org/details/prelinger, Online resource.

Fig. 7. Frame 51 of Suzie with losses and after error concealment by means of GMM.

Fig. 9. Part of frame 28 of Miss America with losses and after error concealment by means of LMMSE. The square marks the place of a lost 8×8-block in the current frame.

Fig. 8. Part of frame 27 of Miss America without losses. The square marks the place of a lost 8×8-block in frame 28.

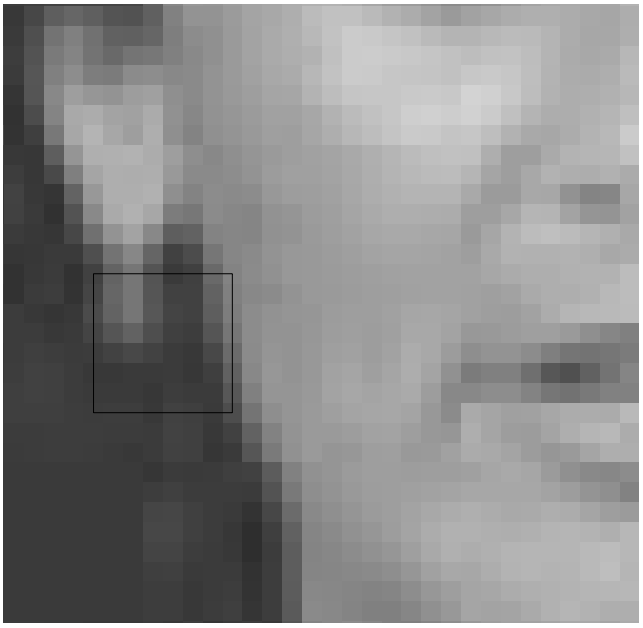

Fig. 10. Part of frame 28 of Miss America with losses and after error concealment by means of GMM. The square marks the place of a lost 8×8-block in the current frame. An estimation error is visible in the square.