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 maximum 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. Our method was shown to increase performance in peak signal-to-noise ratio (PSNR) compared to the best linear method. 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,

$$f_W(w) = \sum_{m=1}^{M} \rho^m f_{W | V}^{m}(w)$$

where the distributions $f_{W | V}^{m}(w)$, $m = 1...M$ are Gaussian with mean $\mu_{W}^{m}$ and covariance $C_{W | V}^{m}$. The weights $\rho^m$ 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

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

where

$$\pi^m(v) = \frac{\rho_{U}^{(m)} f_{U}^{(m)}(v)}{\sum_{k=1}^{M} \rho_{U}^{(k)} f_{V}^{(k)}(v)}$$

and

$$\mu_{U | V}^{(m)}(v) = C_{U | V}^{(m)} (C_{V | V}^{(m)})^{-1} (v - \mu_{V}^{(m)}) + \mu_{U}^{(m)}$$

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

$$\hat{u}_{LMMSE}(v) = C_{U | V} (C_{V | V})^{-1} (v - \mu_{V}) + \mu_{U}$$

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 $C_{W | V}^{m}$ is stationary...
in the sense that each element may be expressed as a function
\( C_{WW}^m(\Delta x, \Delta y, \Delta t) \), where \( \Delta x \) and \( \Delta y \) are spatial position differ-
ences and \( \Delta t \) is the temporal position difference for the pixels in
question (\( C_{WW}^m(\Delta x, \Delta y, \Delta t) \neq C_{WW}^m(-\Delta x, \Delta y, \Delta t) \) and sim-
ilarly for \( \Delta y \)). We may average the elements of \( C_{WW}^m \) and achieve
an estimate of a correlation-like function \( R_{WW}^m(\Delta x, \Delta y, \Delta t) \) in
space-time. It is obvious that \( R_{WW}^m(\Delta x, \Delta y, \Delta t) \) has the prop-
erty that \( R_{WW}^m(\Delta x, \Delta y, \Delta t) \neq R_{WW}^m(-\Delta x, \Delta y, \Delta t) \) (and sim-
ilarly for \( \Delta y \)). In Figure 1, we see \( R_{WW}^m(\Delta x, \Delta y, \Delta t) \) for a mix-
ture with \( M = 64 \). Each two-square row represents one Gaussian
and the squares show \( R_{WW}^m(\Delta x, \Delta y, \Delta t) \) for time differences
\( \Delta t = 0, 1 \), from left to right. In each square, \( \Delta x \) and \( \Delta y \) 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
reasonable to expect that the GMM can understand the current sit-
tuation 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 \( 8 \times 8 \)-
blocks distributed as in Figure 2. This error distribution is re-
peated 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 \( 8 \times 8 \)-
block, one \( 4 \times 4 \)-block is estimated at a time, see Figure 3. If the
future \( 4 \times 4 \)-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 stor-
ing 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 esti-
imated 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 x 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.](image)

![Fig. 3. A lost 8 x 8-block is concealed, one 4 x 4-block U from it’s context V at a time. In this example, the future information is lost.](image)
Fig. 4. Example of situation where the GMM is able to understand the motion field. A $4 \times 4$-block $U$ is estimated from it’s context $V$.

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

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 algo-


Fig. 5. Frame 51 of Suzie without losses.


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

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

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 com-
putational 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.

6 References

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

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

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 $8 \times 8$-block in the current frame.

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 $8 \times 8$-block in the current frame. An estimation error is visible in the square.