DISTORTION ESTIMATION FOR TEMPORAL LAYERED VIDEO CODING

Sila Ekmekci, Pascal Frossard

Swiss Federal Institute of Technology
Signal Processing Institute
CH-1015 Lausanne / Switzerland

Thomas Sikora

Technical University Berlin
Institute for Telecommunications
D-10587 Berlin / Germany

ABSTRACT

We present a recursive block based decoder distortion estimation model for temporal layered video transmission, based on a DPCM structure. Each block in a video frame is modeled as a sample from an AR(1) source. The correlation coef£cient of this source depends on the loop £ltering effects, whereas the additional noise term on the motion compensated block difference and the quantization distortion of the block. Distortion estimations are compared to simulation results, and the model is shown to accurately capture the video distortion in various lossy streaming scenarios. The low implementation complexity, and high estimation accuracy of the proposed technique makes it particularly attractive for adaptive video communication applications, that try to optimize the streaming policy.

1. INTRODUCTION

Hybrid coders employ motion compensation for increasing coding gain, with the penalty of decreasing error resilience. In case of losses, errors propagate both spatially and temporally, since the pixels in the current MB may have been motion compensated from pixels in different MB's in the previous frame, each with a different error propagation history. In this work, we are interested in errors occurring due to packet losses. There are mainly two sources of packet losses in packet switched networks: buffer over¤ow at intermediate nodes of the network, and long queuing delays. The packet loss rate in internet communications may reach 20%. Decoder distortion estimation can be used to design rate distortion optimized mode selection which considers both the error concealment as well as quantization distortion [2]. Transcoding is another application area which can make use of distortion estimation.

The work presented in [7] computes the total decoder distortion recursively at pixel level precision to accurately account for spatial and temporal error propagation. The £rst and second moments of random variables (pixel luminance values) are needed, which increase the computation complexity. Moreover the computation is applicable to integer pixel accuracy becoming impractical for extension to half pixel accuracy or higher. Similarly the work in [6] introduces an analytical model to capture the effects of error concealment and interframe error propagation at the video decoder. The drawback of the system is the imprecise estimation of the distortion. Reference [5] gives a multidimensional bitrate control approach based on distortion measures.

Our approach is based on the block based recursive estimation of the distortion. The blocks in each frame are modeled as samples from an AR(1) source with their corresponding blocks along the sequence, where the correlation coef£cient of the source depends on the loop £ltering effects and the additional noise term on the motion compensated block difference as well as on the quantization distortion of the block [1], [4]. The presented technique is an extension of our previous work on decoder distortion estimation for single layer video transmission [3] towards layered coding and transmission. Section 2 describes the estimation algorithm, the experiments and the results are presented in section 3, £nally section 4 concludes the paper.

This work was supported in part by Technical University Berlin, and the Swiss National Science Foundation.

2. RECURSIVE DECODER DISTORTION ESTIMATION: EXTENDED TO B-FRAMES

We assume that each frame is transmitted in a single and separate packet, i.e. I, P or B-frames, though having different sizes, are transmitted in a single packet each.

Distortion estimation for single layer coding is described in [3] which we apply for the base layer. Section 2.1 gives a brief summary of the algorithm. Distortion estimation for the temporal enhancement layer is discussed in section 2.2.

2.1. Distortion Estimation for Base Layer

The base layer distortion estimation works the same way as for single layer coding. The only difference is the increased temporal difference between the adjacently coded frames. Each frame is divided into 4x4 pixel blocks and the decoder distortion estimate for each block of the frame is calculated depending on the packet loss parameter p_l. Block b in current frame f is supposed to constitute a sequence with its corresponding blocks on the other frames along the video sequence (see Figure 1). We suppose that the error due to channel impairments and error concealment propagates only among the corresponding blocks along time. We also assume indirectly that the corresponding blocks along time can be represented by an AR(1) source where a factor depending on the loop £ltering drives the correlation coef£cient between blocks. The motion compensated block difference as well as the quantization distortion of the block constitute the additional noise term. For each frame f and for each block b given by coordinates (b_v, b_h) in f, two distortion estimates are considered: $D_l(f, b_v, b_h)$ denotes the decoder distortion if f and thus b is lost, whereas $D_r(f, b_v, b_h)$ denotes the decoder distortion if the packet containing the data for f and b is received. b_v and b_h are the vertical and horizontal block coordinates respectively. $D(f, b_v, b_h)$, on the other hand, is the overall decoder distortion estimation calculated as the weighted average of $D_l(f, b_v, b_h)$ and $D_r(f, b_v, b_h)$. To develop the recursive formula to calculate the decoder distortion estimate $D(f, b_v, b_h)$ for block b we take the recursion depth 1, i.e., we consider only the distortion of the corresponding block in the reference frame (the previous frame in the formulas) and we differentiate also between the cases where the latter is received or not. To incorporate the block intra updates into the recursive formula we differentiate also between intra-updated blocks and the rest. We have four different cases: $D_{rr}(f, b_v, b_h)$ gives the distortion of block b if the previous and the current frames are both received and $D_{ll}(f, b_v, b_h)$ gives the distortion if both are lost. Similarly $D_{lr}(f, b_v, b_h)$ gives the distortion when the previous one is lost but the current one is received , contrary $D_{rl}(f, b_v, b_h)$, the distortion when the previous one is received and the current one is lost. The details of the algorithm are given in [3].

2.2. Distortion Estimation for Temporal Enhancement Layer

We assume that the motion information is always received, or can be accurately inferred, even when the corresponding frame difference image is lost (using FEC or other method). For B-frames we not only have forward motion vectors but also backward motion vectors depending on the coding mode. Let $m_{f_h}(f, b_v, b_h)$ and $m_{f_v}(f, b_v, b_h)$ be the horizontal and vertical components of the forward motion vector at frame f and block b whose horizontal and vertical coordinates are b_h and b_v respectively. Similarly $m_{b_h}(f, b_v, b_h)$ and $m_{b_v}(f, b_v, b_h)$ are the horizontal and vertical components of the backward motion vector. Motion compensated difference frame $mcdf$ is calculated as the difference between the current frame and the weighted average of the motion compensated forward and backward reference images. Let F be the current frame (B-frame) and F_{-1} and F_{+1} are the previous (forward reference) and the next frames (backward reference) in the sequence. $mcdb(f, b_v, b_h)$, the motion compensated difference block with block coordinates (b_v, b_h) can be given as follows if bidirectional prediction is used:

$$
\begin{aligned}
mcbd(f, b_v, b_h) &= F(b_v, b_h) - \frac{F_{+1}(b''_v, b''_h) + F_{-1}(b'_v, b'_h)}{2} \\
b'_v &= b_v + mv_{f_v}(f, b_v, b_h); \\
b'_h &= b_h + mv_{f_h}(f, b_v, b_h); \\
b''_v &= b_v + mv_{b_v}(f, b_v, b_h); \\
b''_h &= b_h + mv_{b_h}(f, b_v, b_h);
\end{aligned}
$$

Five modes to predict blocks in B-frames are: 1-intra prediction, 2-forward prediction, 3-backward prediction, 4-bidirectional prediction 5-direct prediction. The algorithm for intra and forward prediction is given in [3]. Direct prediction is a special case of bidirectional prediction where the forward motion vector is half of the motion vector of the corresponding block in the P Frame. The backward motion vector is then just the opposite of the forward motion vector.

We will investigate only the bidirectional prediction for B frames here. We had to consider only the loss states of the reference and current frames in distortion estimation for forward prediction. For bidirectional distortion estimation, however, we consider the loss states of the current, forward and backward reference frames. In other words, there are eight cases to be investigated to calculate the distortion of a block in the current frame F. In D_{xyz}, x gives the state of the forward reference frame, y the state of the current frame and z of the backward reference frame respectively, e.g. D_{lrr} denotes the distortion when the forward reference frame is lost, but the current frame and the backward reference frame are received. Similarly in D_x, x gives the state of the current frame regardless of the states of the reference frames. D_q is the quantization distortion.

<u>Case 1:</u> Reference, current and next frames received
The block b in frame f can be expressed as:

$$
\begin{aligned}
b(f, b_v, b_h) &= \frac{1}{2}\sqrt{\alpha}(b(f-1, b'_v, b'_h) + q(f-1, b'_v, b'_h) \\
&\quad + b(f+1, b''_v, b''_h) + q(f+1, b''_v, b''_h)) \\
&\quad + mcbd(f, b_v, b_h),
\end{aligned} \tag{1}
$$

where $b(f-1, b'_v, b'_h)$ and $b(f+1, b''_v, b''_h)$ are the corresponding blocks of block $b(f, b_v, b_h)$ in the forward and backward reference frames and $q(f-1, b'_v, b'_h)$ and $q(f+1, b''_v, b''_h)$ their quantization errors respectively. The factor $\sqrt{\alpha}$ stands for the loop £ltering effect. Denoting the accumulated error terms on $b(f, b_v, b_h)$, $b(f-1, b'_v, b'_h)$ and $b(f+1, b''_v, b''_h)$ as $d(f, b_v, b_h)$, $d(f-1, b'_v, b'_h)$ and $d(f+1, b''_v, b''_h)$ respectively, we have:

$$
\begin{aligned}
& b(f, b_v, b_h) + d(f, b_v, b_h) = \\
&\quad \frac{1}{2}\sqrt{\alpha}(b(f-1, b'_v, b'_h) + d(f-1, b'_v, b'_h) \\
&\quad + b(f+1, b''_v, b''_h) + d(f+1, b''_v, b''_h)) \\
&\quad + mcbd(f, b_v, b_h) + q(f, b_v, b_h),
\end{aligned} \tag{2}
$$

yielding:

$$
\begin{aligned}
d(f, b_v, b_h) &= \frac{1}{2}\sqrt{\alpha}(d(f-1, b'_v, b'_h) - q(f-1, b'_v, b'_h) \\
&\quad + d(f+1, b''_v, b''_h) - q(f+1, b''_v, b''_h)) \\
&\quad + q(f, b_v, b_h).
\end{aligned} \tag{3}
$$

If the current and the previous frames are both received:

$$
\begin{aligned}
d(f-1, b'_v, b'_h) &= q(f-1, b'_v, b'_h) + d'_{rem} \\
d(f+1, b''_v, b''_h) &= q(f+1, b''_v, b''_h) + d''_{rem},
\end{aligned} \tag{4}
$$

where d'_{rem} and d''_{rem} are the components of $d(f-1, b'_v, b'_h)$ and $d(f+1, b''_v, b''_h)$ which are independent of $q(f-1, b'_v, b'_h)$ and $q(f+1, b''_v, b''_h)$. The correlation between $d(f-1, b'_v, b'_h)$ and $q(f-1, b'_v, b'_h)$, similarly between $d(f+1, b''_v, b''_h)$ and $q(f+1, b''_v, b''_h)$ is given as:

$$
\begin{aligned}
E[d(f-1, b'_v, b'_h)q(f-1, b'_v, b'_h)] &= D_q(f-1, b'_v, b'_h) \\
E[d(f+1, b''_v, b''_h)q(f+1, b''_v, b''_h)] &= D_q(f+1, b''_v, b''_h),
\end{aligned} \tag{5}
$$

where $D_q(f-1, b'_v, b'_h)$ and $D_q(f+1, b''_v, b''_h)$ are the quantization distortions of the blocks on the forward and reference frames respectively. Moreover $d(f+1, b''_v, b''_h)$ is given in terms of $d(f-1, b'_v, b'_h)$ as:

$$
\begin{aligned}
d(f+1, b''_v, b''_h) &= \sqrt{\alpha}(d(f-1, b'_v, b'_h) - q(f-1, b'_v, b'_h)) \\
&\quad + q(f+1, b''_v, b''_h) + mcbd(f+1, b''_v, b''_h),
\end{aligned} \tag{6}
$$

$$d(f+1, b_v'', b_h'') - q(f+1, b_v'', b_h'') = \quad (7)$$
$$d(f-1, b_v', b_h') - q(f-1, b_v', b_h') + mcbd(f+1, b_v'', b_h''),$$

$$E[d'_{rem} d''_{rem}] = \sqrt{\alpha}(D_r(f-1, b_v', b_h') - D_q(f-1, b_v', b_h')). \quad (8)$$

$D_{rrr}(f, b_v, b_h)$ is then calculated by combining (3), (4), (5), (6), (7) and (8):

$$\begin{aligned} D_{rrr}(f, b_v, b_h) = & \quad (9) \\ & \frac{\alpha}{4}(D_r(f-1, b_v', b_h') - D_q(f-1, b_v', b_h') \\ & +(D_r(f+1, b_v'', b_h'') - D_q(f+1, b_v'', b_h'')) \\ & +\frac{\alpha^{\frac{3}{2}}}{2}(D_r(f-1, b_v', b_h') - D_q(f-1, b_v', b_h')) \\ & +D_q(f, b_v, b_h) \end{aligned} \quad (10)$$

Case 2: Reference and current frames received,next frame lost
$D_{rrl}(f, b_v, b_h)$ is calculated in a similar way as case 1 considering that the backward reference frame is lost. The only difference with case 1 is that:

$$E[(d(f+1, b_v'', b_h'') - q(f+1, b_v'', b_h''))^2] = \quad (11)$$
$$D_l(f+1, b_v'', b_h'') + D_q(f+1, b_v'', b_h''),$$

since $d(f+1, b_v'', b_h'')$ and $q(f+1, b_v'', b_h'')$ are uncorrelated with each other.

Case 3: Previous frame received, current and next frames lost
The calculation is the same as for case 2 with the difference that the additional quantization distortion $D_q(f, b_v, b_h)$ is replaced by $pow\_mcbd(f, b_v, b_h)$, the averaged sum of the squared pixel intensities of the motion compensated block difference.

Case 4: Previous and next frames received, current frame lost
When the reference frames are received and the current one is lost, the formula we obtain is the same as in case 1 with the difference that the additional quantization distortion $D_q(f, b_v, b_h)$ is replaced by $pow\_mcbd(f, b_v, b_h)$, the averaged sum of the squared pixel intensities of the motion compensated block difference.
The distortion terms for the remaining cases are calculated similarly, details are omitted here due to lack of space. To sum up, the formula of distortion for each case consists of 4 parts (corresponding to the four lines of the equation): The £rst term and the second terms have a $'-'$ if the forward reference frame and backward reference frames are received respectively and a $'+'$ otherwise. The third term has always the same sign as the £rst term (since it depends on the reception of the forward reference frame). Additionally, The last term is $pow\_mcbd(f, b_v, b_h)$ if the current frame is lost and $D_q(f, b_v, b_h)$ if it is received.
$D_l(f, b_v, b_h)$ and $D_r(f, b_v, b_h)$ are calculated as weighted averages of the estimated distortion terms corresponding to the cases where current frame was lost or received respectively.
PSNR values are calculated directly from the distortion terms as:

$$PSNR_{xyz}(f) = 10log10 \frac{255^2}{\frac{\sum_{i=1}^{max_i} \sum_{j=1}^{max_j} D_{xyz}(f,i,j)}{total\_block\_number}}$$

	$PSNR_{avg}$ **[dB]**	R **[kbits/s]**	**i.-GOB per.**
F.	37.02	160.86	0
A.	37.67	18.15	0
F.+i.-GOB	36.99	197.55	1
A.+i.-GOB	37.38	47.89	1

Table 1. TLC, no-intra-coding Parameters

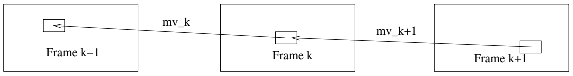
where xyz is the index of the corresponding distortion term. The overall PSNR is the average of all PSNR terms weighted by their respective probabilities (to simulate the way how average PSNR is calculated over all frames by the video codec).

3. EXPERIMENTS AND RESULTS

We assume that each frame whether I, P, or B frame is transmitted in a separate and single packet. The packets from base and enhancement layer are lost with a given uniform loss rate of p_l. If a frame gets lost, assuming that the exact motion information is received (using FEC or other method), we use motion compensated error concealment, i.e. the motion compensated difference value is set to zero for reconstruction. We simulated the lossy channel with a random uniform loss generator and sequences are coded and decoded with H.264 Codec (TML Version 9.0). For each loss rate, 100 different random loss patterns are considered. We used £ve QCIF video sequences coded at 30 fps: Foreman, Akiyo, Claire, Mother & Daughter and Salesman. Because of the place limitations the results are shown only for the high motion sequence Foreman and low motion sequence Akiyo. We consider two coding options: 1- without any intra-updates, 2- with updates of GOB's. To generate the estimation values we needed four parameters: 1- quantization distortion, 2- motion vector £eld, 3- energy of motion compensated block differences (pow_mcbd) for each block of each frame and 4- α. All of these parameters are calculated off-line (at the sender side). Quantization distortion should be adapted to the coding mode (intra updates) whereas motion vector £eld and motion compensated frame differences can be chosen identical independent of the coding mode to simplify the calculations. α is determined manually for each sequence. Coding parameters for with and without GOB updates are given in Table 1. A total of 100 Frames are considered for each sequence. Odd frames of the sequence are coded in the base layer, whereas even frames in the temporal enhancement layer. Experiments show the good accordance of the simulation results to the estimation ones. Although the simulation results are generated using uniformly distributed loss patterns, the model is applicable to all kinds of loss models. Figures 2 and 3 show the model & simulation accordance for the sequences Foreman and Akiyo respectively when no-intra coding is used. The results with intra-GOB-coding are depicted in £gures 4 and 5. Frame PSNR, $PSNR(f)$ is shown on the y-axis and frame number N on the x-axis.

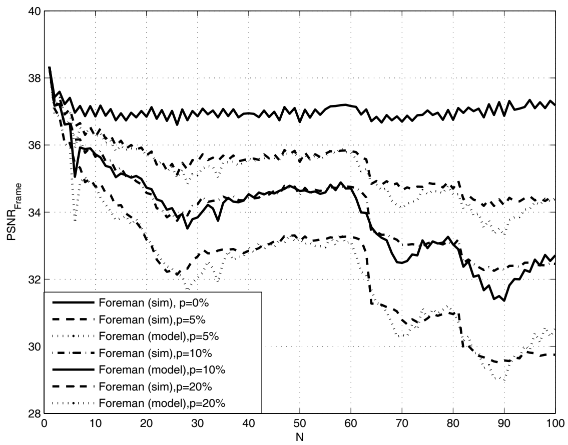
4. CONCLUSIONS

We introduced a block based recursive decoder distortion estimation technique for temporal layered coded video and veri£ed that the estimated frame PSNR values using the technique show good accordance with the simulation results. The low implementation complexity and high estimation accuracy are the advantages of the technique.



,

Fig. 1. AR(1) Model for Sequence of Blocks



,

Fig. 2. TLC, PSNR over Frame Number, Model-Simulation Accordance, Foreman, no-intra-updates

5. REFERENCES

[1] C. Chen. *Motion-Compensated Hybrid Coders in Video Communications*. Ph.D. Thesis, Monash University, Australia, 1992.

[2] G. Cote and F. Kossentini. Optimal intra coding of blocks for robust video communication over the internet. *Image Communication, pp. 25-34*, September 1999.

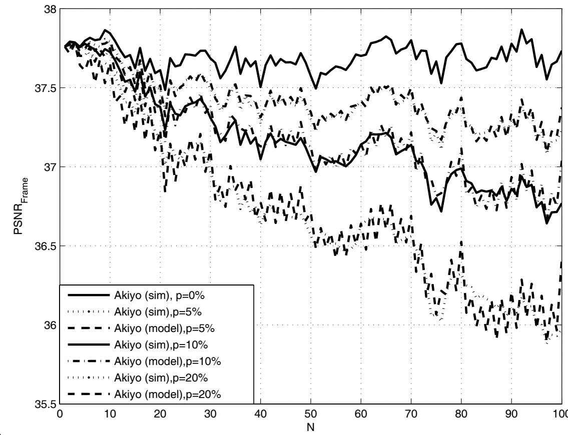
[3] S. Ekmekci and T. Sikora. Recursive decoder distortion estimation based on AR(1) modeling for video. *IEEE Int. Conf. on Image Processing (ICIP'04)*, October 2004.

[4] A. Ingle and V.A. Vaishampayan. DPCM system design for diversity system with applications to packetized speech. *IEEE Trans. Speech Audio Processing, vol. 3, pp. 48-57*, Januar 1995.

[5] E. Reed and J.S. Lim. Optimal multidimensional bit-rate control for video communication. *IEEE Transactions on Image Processing, Vol. 11, No. 8*, August 2002.

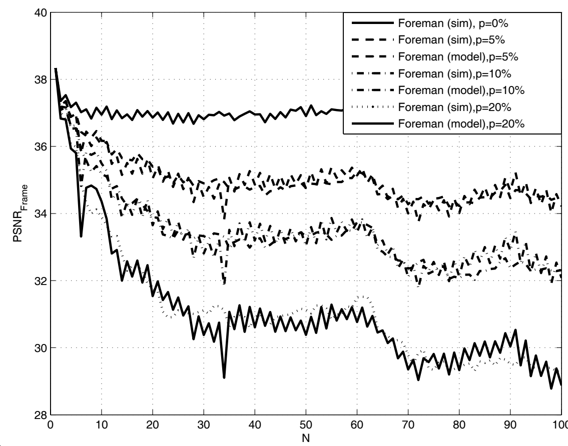
[6] K. Stuhlmueller, N. Faerber, M. Link, and B. Girod. Analysis of video transmission over lossy channels. *IEEE Journal on Selected Areas in Communications, Special Issue on Error Resilient Image and Video Transmission, pp. 1012-1032, 18(6)*, June 2000.

[7] R. Zhang, S. L. Regunathan, and K. Rose. Video coding with optimal inter/intra-mode switching for packet loss resilience. *IEEE Journal on Selected Areas in Communications, Vol. 18, No. 6*, June 2000.



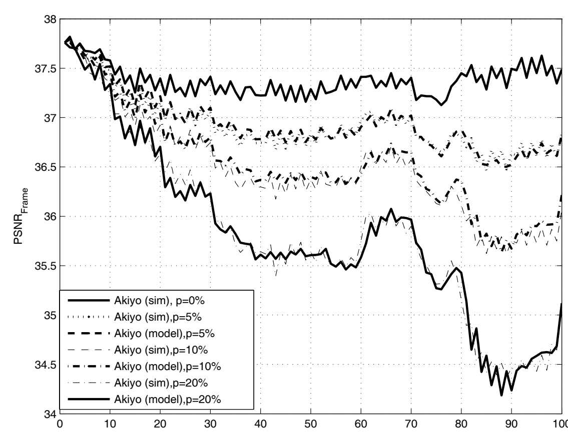
,

Fig. 3. TLC, PSNR over Frame Number, Model-Simulation Accordance, Akiyo, no-intra-updates



,

Fig. 4. TLC, PSNR over Frame Number, Model-Simulation Accordance, Foreman, intra-GOB-updates



,

Fig. 5. TLC, PSNR over Frame Number, Model-Simulation Accordance, Akiyo, intra-GOB-updates