RECURSIVE DECODER DISTORTION ESTIMATION BASED ON AR(1) SOURCE MODELING FOR VIDEO

Sila Ekmekci and Thomas Sikora

Technical University Berlin

ABSTRACT

We introduce a recursive block based decoder distortion estimation technique for video and present results showing the accordance of the estimation results with the simulation results. Each block in a frame with all its corresponding blocks along the video sequence are modeled as an AR(1) source where the correlation coefficient of the source depends on the loop filtering effects and the additional noise term on the motion compensated block difference as well as on the quantization distortion of the block. The distortion term for each block of each frame in the video sequence is calculated recursively depending on the packet loss rate of the channel.

1. INTRODUCTION

Hybrid coders employ inter-frame prediction to remove temporal redundancies and transform coding to remove spatial redundancies. Each video frame is divided into macroblocks which can be coded in intra or inter mode. In inter-mode the MB is first predicted from the previously decoded frame via motion compensation then the prediction error is transform coded. In intra mode, on the other hand, the original MB data is directly transform coded. Intra coding is used as an important tool to reduce the effects of packet loss. To stop error propagation, for certain blocks the inter-frame prediction loop is switched off so that the reproduced blocks are not dependent on past frames. However intra coding requires a higher bitrate than inter-coding. Clearly motion compensation leads to spatial propagation beyond MB boundaries 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 overflow at intermediate nodes of the network, and long queuing delays. The packet loss rate in internet communications may reach 20%.

One application area for decoder distortion estimation is designing a rate distortion optimized mode selection which considers both the error concealment as well as quantization distortion. The loss probability can be deducted from the network conditions. Transcoding is another application area for distortion estimation.

The work presented in [1] computes recursively the total decoder distortion at pixel level precision to accurately account for spatial and temporal error propagation. The first 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 only becoming impractical for extension to half pixel accuracy or higher. Similarly the work in [2] introduces an analytical model to capture the effects of error concealment and interframe error propagation at the video decoder. The drawback of this system is the inadequate estimation accuracy of the overall distortion.

Our recursive block based algorithm to optimally estimate the overall distortion of decoder frame reconstruction due to quantization error propagation and concealment is presented in section 2. The estimation accuracy is shown via simulation results in section 3. Section 4 concludes the paper.

2. RECURSIVE DECODER DISTORTION ESTIMATION

We assume that each frame whether I or P is transmitted in a separate and single packet. 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. We suppose that the error due to channel impairments and error concealment propagates only among the corresponding blocks along time. Although this assumption does not hold for sequences with occlusions and exposures and also for high motion sequences, the results achieved with the technique are quite successful as presented in Section 3. Here we also assume indirectly that the corresponding blocks along time corresponds to an AR(1) source where a factor depending on the loop filtering corresponds to the correlation coefficient between blocks and the motion compensated block difference as well as the quantization distortion of the block

0-7803-8554-3/04/$20.00 ©2004 IEEE.

constitute the additional noise term, which will be shown below.

For each frame f and for each block b given by coordinates (b_v, b_h) in f two distortion estimates are considered: $D_l(f, b_v, b_h)$ denoting the decoder distortion if f and this way b is lost whereas $D_r(f, b_v, b_h)$ denoting the decoder distortion if the packet containing the data for f and so for 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 $D_r(f, b_v, b_h)$:

$$D(f, b_v, b_h) = p_l D_l(f, b_v, b_h) + (1 - p_l) D_r(f, b_v, b_h)$$

We assume that the first frame is never lost. According to this, all blocks of the first frame have just the quantization distortion at the decoder:

$$D(1, b_v, b_h) = D_r(1, b_v, b_h) = D_q(1, b_v, b_h)$$

The blocks in the following frames can be intra or inter coded. If the block is intra coded using intra prediction modes we estimate its distortion using the average of the distortion estimates of its left and upper neighbours $D_{left}(f, b_v, b_h)$ and $D_{up}(f, b_v, b_h)$ respectively:

$$D_{left}(f, b_v, b_h) = \begin{cases} 0 & \text{if } (b_h - 1) < 1 \\ D(f, b_v, b_h - 1) & \text{else} \end{cases}$$

and

$$D_{up}(f, b_v, b_h) = \begin{cases} 0 & \text{if } (b_v - 1) < 1 \\ D(f, b_v - 1, b_h) & \text{else} \end{cases}$$

resulting in

$$D_{old}(f, b_v, b_h) = \frac{D_{up}(f, b_v, b_h) + D_{left}(f, b_v, b_h)}{2}$$

$D_{old}(f, b_v, b_h)$ represents here distortion component coming from the reference block. The component of $D_{old}(f, b_v, b_h)$ due to quantization is calculated as:

$$D_{old_q}(f, b_v, b_h) = \frac{D_{up_q}(f, b_v, b_h) + D_{left_q}(f, b_v, b_h)}{2}$$

.

On the other hand, If block b is inter coded we have to consider the distortion of the corresponding block b' in the reference frame $f' = f - 1$ to calculate $D_{old}(f, b_v, b_h)$. b' has the coordinates (b'_v, b'_h) in f':

$$b'_v = b_v + round(m_v(f, b_v, b_h)/16)$$
$$b'_h = b_h + round(m_h(f, b_v, b_h)/16)$$

where $m_v(f, b_v, b_h)$ and $m_h(f, b_v, b_h)$ are the vertical and horizontal components of the motion vector field at block position (b_h, b_v). The vector components are divided to 16, first because we consider 4x4 pixel blocks and second because of quarter pixel accuracy. If b'_v or b'_h lies out of the frame it is clipped into the frame.

For inter coded blocks we differentiate between $D_{old_l}(f, b_v, b_h)$ and $D_{old_r}(f, b_v, b_h)$, for the cases that the reference frame is lost and received respectively. $D_{old_l}(f, b_v, b_h)$, $D_{old_r}(f, b_v, b_h)$ and similarly $D_{old_q}(f, b_v, b_h)$, the quantization distortion of b' are given as:

$$D_{old_l}(f, b_v, b_h) = D_l(f - 1, b'_v, b'_h)$$
$$D_{old_r}(f, b_v, b_h) = D_r(f - 1, b'_v, b'_h)$$
$$D_{old_q}(f, b_v, b_h) = D_q(f - 1, b'_v, b'_h)$$

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 whether this one is received or not. To incorporate the block intra updates into the recursive formula we differentiate also between intra-updated blocks and the rest. $D_{rr}(f, b_v, b_h)$ 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.

If block b is intra-updated and also received, the decoder distortion is just the quantization distortion whether the previous frame is received or not:

$$D_{rr}(f, b_v, b_h) = D_q(f, b_v, b_h)$$
$$D_{lr}(f, b_v, b_h) = D_q(f, b_v, b_h)$$

On the other hand, if b is received but not updated, D_{rr} and D_{lr} are given as:

$$\begin{aligned} D_{rr}(f, b_v, b_h) &= \alpha(D_{old_r}(f, b_v, b_h) - D_{old_q}(f, b_v, b_h)) \\ &\quad + D_q(f, b_v, b_h) \end{aligned}$$

$$\begin{aligned} D_{lr}(f, b_v, b_h) &= \alpha(D_{old_l}(f, b_v, b_h) + D_{old_q}(f, b_v, b_h)) \\ &\quad + D_q(f, b_v, b_h) \end{aligned}$$

where α is a constant depending on the sequence, particularly on the scene activity representing the effect of the loop filter in hybrid coders. Note that block b is given as:

$$b = \sqrt{\alpha}(b^{'} + q^{'}) + mcbd(b)$$

where $b^{'}$ is the corresponding block to block b in the reference frame, $mcbd(b)$ is the motion compensated block difference and q and $q^{'}$ are the quantization distortions on blocks b and $b^{'}$ respectively. According to this formula, b can be considered as a AR(1) source where $\sqrt{\alpha}$ corresponds to the correlation coefficient and $\sqrt{\alpha}q^{'} + mcbd(b)$ is the additional noise term. The noise term is approximated as uncorrelated with b, which is only the case if motion compensation is ideal. Denoting the accumulated distortion on block $b^{'}$ as $d^{'}$ and on block b as d, we have:

$$b + d = \sqrt{\alpha}(b^{'} + d^{'}) + mcbd(b) + q$$

$$d = \sqrt{\alpha}(d^{'} - q^{'}) + q$$

If the current and the previous frames are both received:

$$d^{'} = q^{'} + d_{rem}$$

where d_{rem} is the component of $d^{'}$ which is independent of $q^{'}$. The correlation between $d^{'}$ and $q^{'}$ is given as:

$$E[d^{'}q^{'}] = D_{old_q}$$

resulting in the formula for D_{rr}:

$$\begin{aligned} D_{rr}(f, b_v, b_h) &= E[d^2] \\ &= \alpha(D_{old_r}(f, b_v, b_h) + D_{old_q}(f, b_v, b_h) \\ &\quad -2D_{old_q}(f, b_v, b_h)) + D_q(f, b_v, b_h) \\ &= \alpha(D_{old_r}(f, b_v, b_h) - D_{old_q}(f, b_v, b_h)) \\ &\quad +D_q(f, b_v, b_h) \end{aligned}$$

On the other hand, if the previous frame is lost, $d^{'}$ and $q^{'}$ are uncorrelated to each other so that $D_{lr}(f, b_v, b_h)$ is calculated as:

$$\begin{aligned} D_{lr}(f, b_v, b_h) &= E[d^2] \\ &= \alpha(D_{old_r}(f, b_v, b_h) + D_{old_q}(f, b_v, b_h)) \\ &\quad +D_q(f, b_v, b_h) \end{aligned}$$

If the current frame is lost, the distortions $D_{rl}(f, b_v, b_h)$ and $D_{ll}(f, b_v, b_h)$ are calculated as follows:

$$\begin{aligned} D_{rl}(f, b_v, b_h) &= \alpha(D_{old_r}(f, b_v, b_h) - D_{old_q}(f, b_v, b_h)) \\ &\quad +pow\_mcbd(f, b_v, b_h) \end{aligned}$$

$$\begin{aligned} D_{ll}(f, b_v, b_h) &= \alpha(D_{old_l}(f, b_v, b_h) + D_{old_q}(f, b_v, b_h)) \\ &\quad +pow\_mcbd(f, b_v, b_h) \end{aligned}$$

where $pow\_mcbd(f, b_v, b_h)$ is the averaged sum of the squared pixel intensities of the motion compensated block difference.

Again considering $b^{'}$ when the current frame is lost:

$$\begin{aligned} b + d &= \sqrt{\alpha}(b^{'} + d^{'}) \\ d &= \sqrt{\alpha}(b^{'} + d^{'}) - \sqrt{\alpha}(b^{'} + q^{'}) - mcdf(b) \\ &= \sqrt{\alpha}(d^{'} - q^{'}) - mcdf(b) \end{aligned}$$

If the previous frame is received, $q^{'}$ and $d^{'}$ are correlated and otherwise uncorrelated to each other. The distortion is again calculated as $E[d^2]$ resulting in the equations given for $D_{rl}(f, b_v, b_h)$ and $D_{ll}(f, b_v, b_h)$ above.

$D_l(f, b_v, b_h)$ and $D_r(f, b_v, b_h)$, on the other hand, are calculated as the weighted averages of $D_{ll}(f, b_v, b_h)$ and $D_{rl}(f, b_v, b_h)$ and similarly of $D_{lr}(f, b_v, b_h)$ and $D_{rr}(f, b_v, b_h)$ respectively:

$$\begin{aligned} D_l(f, b_v, b_h) &= p_l D_{ll}(f, b_v, b_h) + (1 - p_l) D_{rl}(f, b_v, b_h) \\ D_r(f, b_v, b_h) &= p_l D_{lr}(f, b_v, b_h) + (1 - p_l) D_{rr}(f, b_v, b_h) \end{aligned}$$

The estimated PSNR for each frame PSNR(f) is calculated as:

$$\begin{aligned} PSNR(f) &= {p_l}^2 PSNR_{ll}(f) + p_l(1 - p_l) PSNR_{lr}(f) \\ &\quad +(1 - p_l)^2 PSNR_{rr}(f) \\ &\quad +(1 - p_l) p_l PSNR_{rl}(f) \end{aligned}$$

where the individual PSNR values are given as:

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

where $total\_block\_number$ is the total number of blocks in a frame.

The remaining PSNR values are calculated by exchanging the subscripts of the variables respectively.

3. EXPERIMENTS AND RESULTS

We assume that each frame whether I or P frame is transmitted in a separate and single packet. The packets are lost with a given uniform loss rate of p_l. If a frame gets lost, assuming that the exact motion information is known, we use motion compensated error concealment, particularly the motion compensated difference value for these blocks is set to zero. 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 loss patterns are considered. We used five QCIF video sequences coded at 30 fps: Foreman, Akiyo, Claire, Mother & Daughter and Salesman for the simulations and the estimations. Because of the place limitations the results, i.e. comparison of the simulation to estimated decoded frame PSNR's over the frame number N are shown only for the sequence Foreman. We consider three cases: 1- without any intra-updates, 2- with updates of GOB's, 3- with intra frame updates. To generate the estimation values we needed four parameters: 1- quantization distortion, 2- motion vector field, 3- energy of motion compensated block differences (pow_mcbd) for each block of each frame and 4- α. All of these parameters are calculated offline (at the sender side). Quantization distortion should be adapted to the coding mode (intra updates) whereas motion vector field and motion compensated frame differences can be taken the same independent of the coding mode to simplify the calculations. α is determined manually for each sequence. Figure Fig.1. (Foreman: R=189.13 kbit/s, $PSNR_{avg}$=37.14 dB, 6.30 kbits/P-frame) shows the comparison of simulation results to the estimation results for the non-intra update case. Figure Fig.2. (Foreman: R=264.59 kbit/s, $PSNR_{avg}$=37.05 dB, 8.82 kbits/P-frame) on the other depicts the same comparison for the intra updates of GOB's (group of blocks) and Figure Fig.3. (Foreman: R=261.76 kbit/s, $PSNR_{avg}$=37.53 dB, 8.82 kbits/P-frame) for the intra updates of frames. Intra frame update period is 10 (Figure 3) and in each frame one GOB is intra updated periodically along the frame (Figure 2). A total of 100 Frames are considered for each sequence. 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.

4. CONCLUSION

We introduced a block based recursive decoder distortion estimation technique for video and verified that the estimated frame PSNR values using the technique show good accordance with the simulation results. The low implementation complexity and high estimation accuracy of the technique makes it attractive compared to other decoder distortion estimation methods in our knowledge so far.

5. REFERENCES

[1] Rui Zhang, Shankar L. Regunathan, Kenneth 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.

[2] K. Stuhlmueller, N. Faerber, M. Link, 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

[3] G. Cote, F Kossentini, "Optimal intra coding of blocks for robust video communication over the Internet " *Image Communication*, Sept. 1999, pp. 25-34.

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

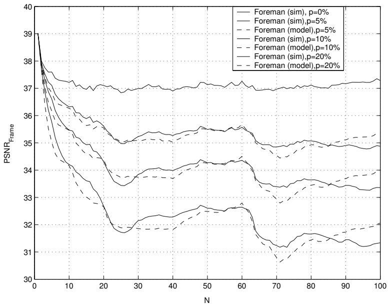

Fig.1. PSNR over Frame Number, Foreman, no intra updates

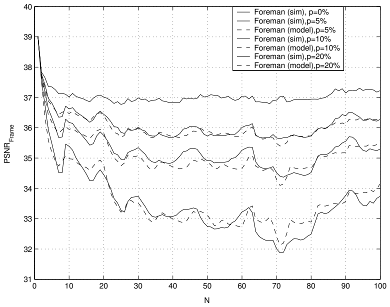

Fig.2. PSNR over Frame Number, Foreman, GOB intra updates

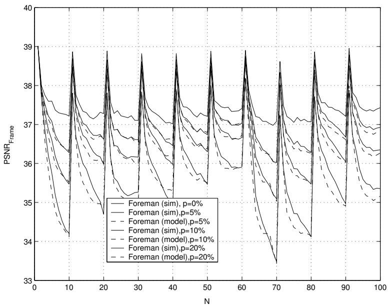

Fig.3. PSNR over Frame Number, Foreman, frame intra updates