END-TO-END RATE-DISTORTION OPTIMIZED MOTION ESTIMATION

Shuai Wan[1]*, Ebroul Izquierdo*[1]*, Fuzheng Yang*[2] *and Yilin Chang*[2]

[1] Multimedia and Vision Research Lab, Queen Mary, University of London
{shuai.wan, ebroul.izquierdo}@elec.qmul.ac.uk
[2] State Key Lab of Integrated Service Networks, Xidian University
{fzhyang, ylchang}@mail.xidian.edu.cn

ABSTRACT

An end-to-end rate-distortion optimized motion estimation method for robust video coding in lossy networks is proposed. In this method the expected reconstructed distortion after transmission and the total bit rate for displaced frame difference are estimated at the encoder. The results are fed into the Lagrangian optimization at the encoder to perform motion estimation. Here the encoder automatically finds an optimized motion compensated prediction by estimating the best trade off between coding efficiency and end-to-end distortion. Computer simulations in lossy channel environments were conducted to assess the performance of the proposed method. A comparative evaluation using other conventional techniques from the literature was also conducted.

Index Terms— ***Motion estimation, rate distortion optimization, packet loss resilience, video coding***

1. INTRODUCTION

Video communication over packet-switched networks such as IP faces the critical problem of packet loss. This impairs the quality of the reconstructed video at the decoder, and causes error propagation in both spatial and temporal domains. Various kinds of packet-loss resilient techniques have been proposed to enhance the robustness of the video communication system. In addition to the conventional error control techniques such as forward error correction (FEC) and automatic repeat request (ARQ), a variety of robust video coding techniques have been proposed for pack-loss resilience, e.g., periodic intra-mode selection [1], or rate-distortion (RD) optimized mode decision methods [2], [3].

Although RD optimized mode decision methods have contributed to a significant improvement on the error-resilience performance of video coders, error propagation remains a critical issue. That is, if a block is predicted from a corrupt image area, the channel errors introduced in previous frames may still propagate to the current frame along the motion-compensation path. To deal with this problem, a few robust motion estimation techniques have been proposed so far. For instance, modeling the error propagation process over multiple reference frames and tackling its effects on the decoded video after transmission [4]. An error-resilient motion prediction criterion has also been proposed, with the aim to minimize the expected decoder prediction error [5].

Though these techniques use different strategies and models, they feature a crucial common aspect: During motion estimation the displaced frame difference is estimated between the current block and corresponding error-prone reference area at decoder. Furthermore, when Lagrangian optimization is used, only the bit rate for motion information is employed. However, the displaced frame difference does not reflect the real end-to-end reconstructed distortion when packet loss occurs. Thus, the end-to-end RD performance should be optimized considering the total bit rate for motion compensation.

In this paper an RD optimized motion-compensated prediction (MCP) method for robust video coding in lossy transmission channels is presented. Contrasting methods from the conventional literature, the proposed technique uses the final expected reconstructed distortion after transmission, instead of an error-free or “end-to-end” displaced frame difference. Furthermore, the total bit rate for MCP coding is also estimated during motion estimation using a different rate distortion model. This model includes both the bit rate for coding motion information and residual texture. A recursive algorithm is utilized for distortion estimation, which is close in spirit to the recursive optimal per-pixel estimate (ROPE) algorithm introduced in [2]. However, critical changes are made to approximate unknown information in motion estimation at the encoder. Computer simulations in lossy channel environments were conducted to assess the performance of the proposed method. A comparative evaluation using other conventional techniques from the literature was also conducted.

The remainder of this paper is organized as follows. In Section 2, the description of the proposed Lagrangian optimization framework and the detailed implementation are given. To validate the performance of the proposed method, simulation results are reported in Section 3. Section 4 presents conclusion and future research directions.

2. END-TO-END RATE-DISTORTION OPTIMIZED MOTION ESTIMATION

When inter-frame prediction is employed, errors may propagate from previous corrupt frames to the current one via the prediction. In order to reduce error propagation, robust mode decision methods introduce more intra coding blocks, which also facilities robust motion estimation by providing more reliable reference areas. Error-resilient motion estimation method aims to find a good trade off between prediction accuracy and error resilience, which can be accomplished through Lagrangian optimization.

Let B_n be the block for motion estimation in frame n. The process of end-to-end RD optimized motion estimation for B_n aims at minimizing the following Lagrangian cost function J:

$$J = D_{dec} + \lambda \cdot R_{mc}\,, \tag{1}$$

where the end-to-end reconstructed distortion D_{dec} of B_n is gauged against the number of bits R_{mc} for motion compensation coding using the Lagrangian multiplier λ. In this paper λ is taken as the one used for the conventional error-free case. However, neither the end-to-end reconstructed distortion nor the amount of bits for motion compensation coding is available at the encoder during motion estimation. Therefore, D_{dec} and R_{mc} need to be estimated.

2.1. End-to-end distortion estimation

Suppose $d_n^{(i)}$ is the reconstructed distortion of $p_n^{(i)}$, where $p_n^{(i)}$ is the i^{th} pixel in frame n and $p_n^{(i)} \in B_n$. Then D_{dec} can be expressed as:

$$D_{dec} = \sum_{p_n^{(i)} \in B_n} d_n^{(i)}. \tag{2}$$

Moreover, let $\hat{p}_n^{(i)}$ and $\tilde{p}_n^{(i)}$ denote the reconstructed value of $p_n^{(i)}$ at the encoder and at the decoder, respectively. Since the encoder has no access to $\tilde{p}_n^{(i)}$ and the occurrence of packet loss is assumed to be an independent random event, $\tilde{p}_n^{(i)}$ can be regarded as a random variable at the encoder side. Using the mean squared error to measure the reconstructed distortion at the decoder, the end-to-end distortion of the concerned pixel can be formulated as:

$$\begin{aligned} d_n^{(i)} &= E\left\{ \left(p_n^{(i)} - \tilde{p}_n^{(i)} \right)^2 \right\} \\ &= \left(p_n^{(i)} \right)^2 - 2p_n^{(i)} E\left\{ \tilde{p}_n^{(i)} \right\} + E\left\{ \left(\tilde{p}_n^{(i)} \right)^2 \right\}. \end{aligned} \tag{3}$$

Let $\hat{p}_k^{(j)}$ be the pixel from which $p_n^{(i)}$ predicted, and $\hat{p}_k^{(j)} \in B_k$ where B_k is a reconstructed reference block in frame k. Then the prediction error $e_n^{(i)}$ and the quantized prediction error $\hat{e}_n^{(i)}$ are given by: $e_n^{(i)} = p_n^{(i)} - \hat{p}_k^{(j)}$, and $\hat{e}_n^{(i)} = \hat{p}_n^{(i)} - \hat{p}_k^{(j)}$. Then the first and second moments of $\tilde{p}_n^{(i)}$ can be computed as:

$$E\left\{ \tilde{p}_n^{(i)} \right\}_P = (1-\wp)\left(\hat{e}_n^{(i)} + E\left\{ \tilde{p}_k^{(j)} \right\} \right) + \wp E\left\{ \tilde{p}_{n-1}^{(i)} \right\}, \tag{4}$$

$$\begin{aligned} E\left\{ \left(\tilde{p}_n^{(i)} \right)^2 \right\}_P &= (1-\wp) E\left\{ \left(\hat{e}_n^{(i)} + \tilde{p}_k^{(j)} \right)^2 \right\} + \wp E\left\{ \left(\tilde{p}_{n-1}^{(i)} \right)^2 \right\} \\ &= (1-\wp)\left(\left(\hat{e}_n^{(i)} \right)^2 + 2\hat{e}_n^{(i)} E\left\{ \tilde{p}_k^{(j)} \right\} + E\left\{ \left(\tilde{p}_k^{(j)} \right)^2 \right\} \right) \\ &\quad + \wp E\left\{ \left(\tilde{p}_{n-1}^{(i)} \right)^2 \right\}, \end{aligned} \tag{5}$$

where $\wp$ represents the packet-loss rate.

In a real coding situation the quantized prediction error $\hat{e}_n^{(i)}$ is not available at the time of motion estimation. In our method, the most straightforward approximation is used by ignoring the effects of quantization errors, i.e., by setting

$$\hat{e}_n^{(i)} = e_n^{(i)}. \tag{6}$$

Though this choice may appear crude and simplistic, it is effective since quantization errors are negligible when compared with the error caused by packet loss or error propagation in an error-prone environment. The validity of this argument has been empirically demonstrated through extensive experiments. Inserting (6) into (4) and (5), the first and second moments of $\tilde{p}_n^{(i)}$ are obtained. Using the first and second moments of $\tilde{p}_n^{(i)}$, $d_n^{(i)}$ and D_{dec} can be estimated.

2.2. Bit rate estimation

The total bit rate for motion compensated coding R_{mc} in (1) includes the bits needed to code motion vectors, reference parameters and residual error. Thus, R_{mc} can be expressed as:

$$R_{mc} = R_{mv} + R_r\,, \tag{7}$$

where R_{mv} is the bit budget allocated for both motion vector and reference parameter while R_r is the bit budget allocated for the residual error or displace frame difference. The value of R_{mv} can be extracted from a lookup table as in conventional RD optimized motion estimation. However, the real value of R_r is only available after the actual coding. Since computing R_r for each single MV would renders the approach unfeasible in practical applications, an accurate approximation for R_r is needed in which the actual coding can be avoided. RD models are useful for estimating the value of R_r. In order to simplify the estimation, the following quadratic RD model [6] is used. In this model only quantization step Q and the SAD are involved.

$$\frac{R_r}{SAD} = \frac{b_1}{SAD} + b_2 Q^{-1} + b_3 Q^{-2}\,, \qquad (8)$$

where $b_m,\ m = 1, 2, 3$ are the model parameters. Therefore R_r can be estimated as:

$$R_r = c_1 + c_2 SAD\,, \qquad (9)$$

where $c_1 = b_1$, $c_2 = b_2 Q^{-1} + b_3 Q^{-2}$ are the model parameters, and the values for $b_m,\ m = 1, 2, 3$ are updated for each frame using linear regression.

3. EXPERIMENTAL RESULTS

The proposed method has been extensively evaluated using the test model JM7.6 of the H.264 video coder [7]. Without loss of generality, the first frame was intra coded and the following frames were all inter coded. It was assumed that each row of MBs, i.e., each group of blocks (GOB), was transmitted in a separate packet. To simulate packet loss, the common conditions for wire-line, low delay IP/UDP/RTP packet loss resilient testing defined in [8] were used. For the sake of conciseness the results reported in this paper include three test sequences only: FOREMAN sequence in QCIF format, IRENE and PARIS in CIF format. Similar results can be observed for other sequences.

In all experiments, the ConstrainedIntraPrediction Flag of the video coder was switched on, i.e., inter pixels were not used in intra prediction. The rate control was used during coding to bring the resulting coding bit rate as close as possible to the channel bit rate. The selected results reported in this section consider 3%, 5%, 10%, and 20% average packet-loss rate. According to the used error patterns all frames, including the first one, were subject to transmission errors. These packet-loss rates were simulated using the respective error pattern files defined in [9].

For overall performance comparison, 30 simulation runs were performed under each one of the error rate conditions, each run using a different packet-loss pattern. For the objective video quality assessment the luminance PSNR (Y-PSNR) was averaged over all decoded frames and all the implemented channel conditions. Furthermore, to compare the consistency of the performance, PSNR of every frame was also collected for certain packet rates with certain packet loss patterns.

The following three methods were utilized for the comparative evaluation. Observe that the RD optimized mode decision technique (OM) achieves significant gains over the traditional robust intra update methods (random, regular) [3], this option was selected in the H.264 video coder to generate the anchor. When multiple reference frames are used, the frame restriction method can be applied together with OM to improve the performance [10]. So the optimal mode decision with frame restriction (OMR) was also used in comparison in the cases of using multiple reference frames. To enable overall comparison, the proposed end-to-end RD optimized motion estimation was integrated with the OM, denoted OME. As suggested in [3], K=30 was used in the RD optimized mode decision methods for the comparative assessment.

Table 1: Performance comparison on average PSNR for the IRENE sequence: CIF, 30fps, 384 kbit/s.

Packet loss rate	OM	OMR	OME
3%	34.61 dB	34.63 dB	36.82 dB
5%	34.13 dB	34.12 dB	36.54 dB
10%	33.69 dB	33.68 dB	35.45 dB
20%	31.03 dB	31.07 dB	32.93 dB

Table 2: Performance comparison on average PSNR for the FOREMAN sequence: QCIF, 7.5 fps, 144 kbit/s.

Packet loss rate	OM	OMR	OME
3%	32.58 dB	32.68 dB	34.39 dB
5%	31.61 dB	31.49 dB	32.51 dB
10%	29.61 dB	29.61 dB	30.17 dB
20%	26.74 dB	26.74 dB	27.05 dB

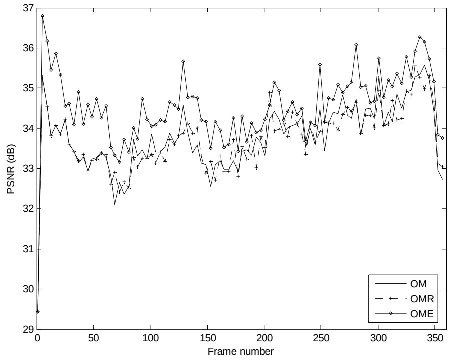

Fig. 1 PSNR versus frame number: IRENE sequence in CIF format, 30fps, 384 kbit/s, packet loss rate 10% .

Table 1 and Table 2 summarize the performance results under different packet loss rate, when five reference frames were used. The results support the claim that the overall end-to-end RD performance is significantly improved by introducing the proposed end-to-end RD optimized motion estimation. Specifically, for the FOREMAN sequence gains up to 1.81 dB in average can be achieved over the OM, and 1.71 dB over the OMR. For the IRENE sequence, gains up to 2.21 dB over OM and 2.19 dB over OMR can be observed. Fig. 1 and Fig. 2 depict the PSNR curves over frame number using different packet loss rates and patterns. In these experiments five reference frames were used. The proposed OME technique clearly outperforms the OM and OMR.

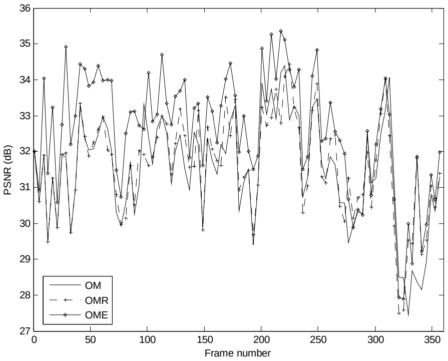

Fig. 2 PSNR versus frame number: FOREMAN sequence in QCIF format, 7.5fps, 144 kbit/s, packet loss rate 5% .

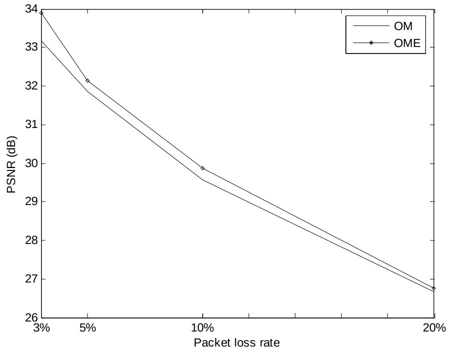

Fig. 3 Performance comparison for single reference frame under different packet loss rates: FOREMAN sequence in QCIF format, 7.5fps, 144 kbit/s.

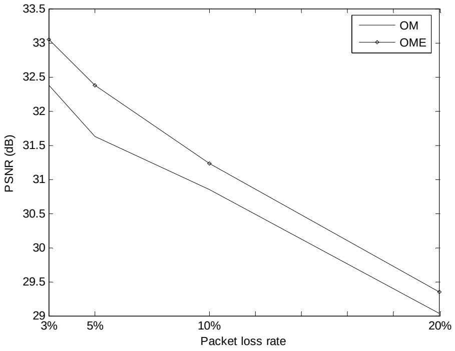

Fig. 3 Performance comparison for single reference frame under different packet loss rates: PARIS sequence in CIF format, 15fps, 384 kbit/s.

Results using single reference frame obtained for various packet loss rates are shown in Fig. 3 and Fig. 4. It can be observed that up to 0.73 dB in average are gained over the OM for the FOREMAN sequence in QCIF format, and up to 0.75 dB can be gained for the PARIS sequence also in CIF format. Similar conclusions for the multiple reference frames case can be drawn.

4. CONCULUSION

In this paper an end-to-end RD optimized motion estimation method for packet loss resilient video coding has been proposed. Given the packet loss rate, the reconstructed distortion at the decoder and the total bit rate for motion compensation coding are estimated at the encoder during motion estimation. Then the estimated values are incorporated into the Lagrangian optimization framework. Simulation results show that the end-to-end performance can be improved using the proposed method. Since the proposed method introduces additional encoder complexity, future work will consider a solution to estimate the reconstructed distortion at the encoder with reduced complexity.

ACKNOWLEDGEMENT

We wish to acknowledge support provided by the European Commission under contract FP6-001765 aceMedia.

REFERENCES

[1] G. Cote and F. Kossentini, “Optimal intra coding of blocks for robust video communication over the Internet,” Signal Processing: Image Communication, Vol. 15, pp 25-34, Sept. 1999.

[2] R. Zhang, S.L. Regunathan, K. Rose, “Video coding with optimal Inter/Intra-mode switching for packet loss resilience,” IEEE Journal of Selected Areas in Communication, vol. 18, no. 6, pp. 966-976, June 2000.

[3] T. Stockhammer, D. Kontopodis, and T. Wiegand, “Rate-distortion optimization for H.26L video coding in packet loss environment,” 12th International Packet Video Workshop, Pittsburg, PY, May 2002.

[4] T. Wiegand, N. Farber, K. Stuhlmuller, B. Girod, “Error-resilient video transmission using long-term memory motion-compensated prediction,” IEEE Journal of Selected Areas in Communication, vol. 18, no. 6, pp.1050-1062, June 2000.

[5] H. Yang, K. Rose. “Source-channel prediction in error resilient video coding,” IEEE ICME 2003, vol. 2, pp. 233–236, Baltimore, USA, July 2003.

[6] S.Wan, Y. Chang and F. Yang, “Frame-layer rate control for JVT video coding using improved quadratic rate distortion model”, VCIP 2005, Beijing, China, Jul.2005.

[7] ITU-T Rec. H.264 | ISO/IEC 14496-10 AVC Joint Model, version JM7.6, Apr. 2002. http://bs.hhi.de/~suehring/tml/.

[8] Stephan Wenger, “Common conditions for wire-line, low delay IP/UDP/RTP packet loss resilient testing,” ITU- T VCEG document VCEG-N79r1, Sep. 2001.

[9] Stephan Wenger “Proposed Error Patterns for Internet Experiments” ITU- T VCEG document Q15-I-16r1, October, 1999.

[10] T. Stockhammer and D. Kontopodis, “Error robust macroblock mode and reference frame restriction”, Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, JVT-B102, Jan. 2002.