COMPLEXITY CONSTRAINED ROBUST VIDEO TRANSMISSION FOR HAND-HELD DEVICES

Waqar Zia, Klaus Diepold

Dept. of Elect. Engineering and Information
Technology, Technische Universität München,
Arcisstraße 21, 80333, Munich, Germany
{waqar.zia, kldi}@tum.de

Thomas Stockhammer

Nomor Research,
Tannenweg 25, 83346, Bergen, Germany
stockhammer@nomor.de

ABSTRACT

Robust video conversational applications for hand-held devices come with numerous challenges, e.g. real-time processing, complexity constrained devices and small end-to-end delays, etc. Transmission losses of compressed video data result in spatio-temporal error propagation in the decoded video sequence. To ensure some QoS, the video codec has to be well tuned to combat the degradation resulting from losses. Several feedback based error mitigation technique are assessed in this work. The proposed error robustness technique based on reference picture selection (RPS) and error tracking enhances the overall performance of the target system by more than 4 dB for moderate radio link control (RLC) PDU loss rates of 1.5%. This enhancement is achieved without any additional computational complexity.

Index Terms— Mobile video, interactive error control, complexity constrained coding, feedback, error tracking

1. INTRODUCTION

Mobile communication networks and devices are rapidly progressing in capacity and service quality. At the same time the requirement of high-end services like high quality video is increasing. Mobile channels suffer from frequent fading that results in bursty losses. For video communication a single loss can cause spatio-temporal error propagation due to the temporal and spatial dependencies in the compressed data. Channel coding is an expensive option, since it requires even more bandwidth. Regardless of channel protection, losses are bound to happen in mobile environments and only robust, *loss aware* coding can terminate the error propagation.

H.264/AVC is the state-of-the-art codec of choice for robust video communication. Several of its error resilience tools e.g. slice structure coding and flexible macroblock (MB) ordering (FMO) [1] etc. provide some robustness in an open-loop video communication system. For a more robust system, long-term memory (LTM) motion compensated prediction (MCP) along with average statistical information of channel knowledge has been employed in selecting optimal mode decisions, for example in [2, 3]. Further feedback based techniques in conjunction with accelerated retroactive decoding (ARD) have been investigated in [4, 5, 6]. In [7], proxy-based RPS is employed along with temporal sub-sequences for conversational applications. However, these techniques are unsuitable to be directly applied to the target system because of the complexity and delay constraints, as discussed in the following section. We propose a complexity constrained interactive error control (IEC) technique based on simple and efficient interactive error tracking (IET). A realistic simulation of the system is done and several objective and subjective assessments will be provided that will help identify the most robust system configuration.

The target system is introduced in Section 2 along with its constraints. This is followed by the description of the proposed IEC techniques in Section 3. The performance evaluation criteria, and results are presented in Section 4, before the conclusion in Section 5.

2. CONVERSATIONAL PACKET SWITCHED VIDEO SERVICES

3GPP mobile video telephony services are enabled by conversational packet switched multimedia services [8] which are based on IP multimedia subsystem (IMS). In this service, a hand-held transceiver is connected via bidirectional high speed packet access (HSPA) link to base station, which connects to the core network. The remote terminal may also be a hand-held device. Bidirectional transmission of compressed H.264/AVC packetized video data is done via RTP along with RTCP control information. The lower protocol layers at the transceiver detect and discard corrupted data packets at reception.

The conversational application is subject to strict end-to-end delay requirements, e.g. 100 ms [9]. At the same time, real time video coding along with speech processing etc. on hand-held device leaves little room for employing error-resilience techniques with considerable complexity. On the other hand, stringent delay requirement also provides a window of op-

1-4244-1437-7/07/$20.00 ©2007 IEEE

portunity. It is known that feedback-based error resilience techniques perform better for smaller delays [10]. In addition to this, a bidirectional communication link with possibility of control traffic makes feedback based techniques an ideal choice. The following section describes the proposed techniques suitable for this system.

3. IEC STRATEGIES

The IEC techniques introduced below work on packet loss report from the receiver, which is translated to lost reference regions at the encoder by using IET described below.

3.1. Error Tracking

The system is depicted in Figure 1. In this example, a packet transmitted by the encoder at time t-$3T$ is lost, and the loss report is received at time t. The encoder keeps a record of the recent packets it has transmitted and the corresponding reference area in each such packet. Hence the lost packet number is translated into the corresponding lost reference region of frame t-$3T$.

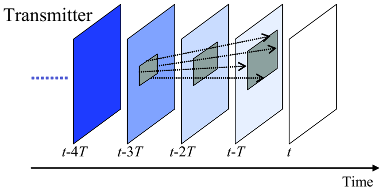


Fig. 1. IEC with error tracking. The video frame rate is $1/T$

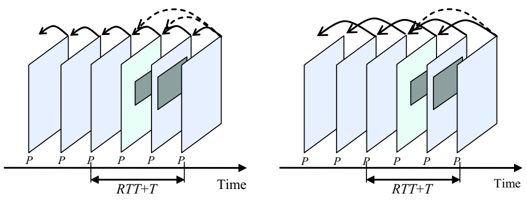
After this, error tracking is applied. We propose a technique in which lost area in a reference frame is assumed to grow at a rate equal to the motion vector (MV) search range plus 2 pixels in each temporally predicted frame. The additional 2 pixels are to compensate the effects of H.264/AVC fractional-pel interpolation. Simulation results will show the suitability of this technique. The shaded regions in Figure 1 are lost.

3.2. IEC with Error Tracking

For the proposed system, for all the blocks b in a given access unit (AU), the rate (r) distortion (d) minimization problem is stated as:

$$\forall_b \quad m_b^* = \arg\min_{m \in \mathcal{O}} (d_{b,m} + \lambda_{\mathcal{O}} r_{b,m}) \tag{1}$$

The minimization is done for the usable option set $\mathcal{O}$ with a lagrangian multiplier $\lambda_{\mathcal{O}}$. Since long term memory (LTM) motion compensated prediction (MCP) takes a considerable part of the total resources [11], the option set is restricted to only one frame for MCP process. Two complexity constrained IEC techniques will be employed for the study, as shown in Figure 2.



(a) Most-recent frame MCP (b) Sub-sequences

Fig. 2. Proposed IEC configurations.

In this system, the average round-trip-time ($\overline{RTT}$) of the feedback messages is measured dynamically. For both the configurations, reference frames up to $\overline{RTT} + T$ are kept in the reference frame buffer.

For the configuration in Figure 2(a), $\mathcal{O}$ consists of inter-coding modes from the most recent reference frame and intra coding options. The distortion d corresponding to inter-prediction from lost regions is infinite, and hence such modes are *invalid*. If none of the inter-coding modes within $\mathcal{O}$ is valid, only then a modified option set $\hat{\mathcal{O}}$ is used by recursively including temporally older reference frames one by one until at least one valid inter-coding mode is found. Intra-coding modes are unaffected by this process. Hence the modes in $\hat{\mathcal{O}}$ are less than or equal to the modes for error-free case $\mathcal{O}$. Since the implementation does not require actual distortion calculations for invalid modes, this technique, referred to as IEC1, ensures that the complexity does not increase in the case of error reports.

For the second configuration as depicted in Figure 2(b), the inter-coding option set is limited to the reference frame that temporally precedes the frame being encoded by $\overline{RTT}$. The modification of option set is done in the same way as for IEC1, except that there is only one additional, temporally preceding, reference frame available. This technique will be referred to as IEC2. The additional robustness by using reference frames $\overline{RTT}$ away from the current frame is that the loss report typically arrives before the reference is used for MCP and the probability of error propagation is reduced as such. However, temporally older reference frames reduce the compression efficiency, hence the cost-benefit analysis will be provided by this work.

It should be noted that the loss report handling and IET are packet-based processes. For the practical system configuration as investigated later, there are typically less than 10 packets per frame. Hence the computational overhead of such processing is negligible compared to the rest of the codec complexity.

The other reference techniques consist of random intra-MB refresh (RIR) [12]. Finally, instantaneous RIR tuning technique is used, which is expressed as

$$\rho = \alpha \cdot \beta^{s-s'} \tag{2}$$

where on receiving the feedback of a lost packet, the instantaneous RIR rate ρ is increased instantaneously to a peak value α and is then reduced with each frame according to β. Here, the latest loss report is received while encoding frame s' while the current frame being encoded is s. α and β are tuned experimentally, and the selected value of both is 0.5. This technique expedites error recovery compared to RIR, while avoiding the buffering overheads of transmitting a complete intra frame.

4. PERFORMANCE RESULTS

The simulation environment used for generating the results has been documented in [13]. The video sequences have been selected by the video adhoc group within 3GPP [14]. The results are reported for QCIF sized sequences "stunt" at 15 fps and "party" at 12 fps. MV search range of 8 pels is used. A 3GPP channel simulator [14] with realistic loss patterns is employed. The radio access bearer supports 128 kbps. In order to achieve better statistical significance, each test is repeated with 128 different channel realizations, and the readings averaged. In a realistic depiction, the feedback traffic is multiplexed along with normal video traffic, which is exposed to a RLC-PDU loss rate of 0.5%.

The assessment metrics are PSNR and percentage of degraded video quality ($PDVD$), defined as:

$$\begin{aligned} PDVD &= \frac{\sum_{i=1}^{N} f(\hat{d}_i, \tilde{d}_i)}{N}\%, \\ f(\hat{d}_i, \tilde{d}_i) &= 1 \text{ if } (\hat{d}_i - \tilde{d}_i) > 2 \text{ dB}, 0 \text{ otherwise} \qquad (3) \end{aligned}$$

whereby $\hat{d}_i$ is the reconstructed PSNR of i^{th} frame, and $\tilde{d}$ is the decoded PSNR of the corresponding frame for a sequence with N frames. This metric represents what fraction of decoded video has considerable distortion added because of losses, and hence tends to decouple this distortion from the compression losses. The threshold of this is selected as 2 dB.

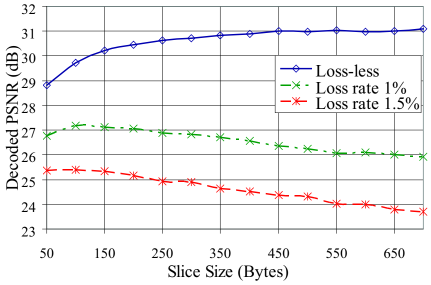
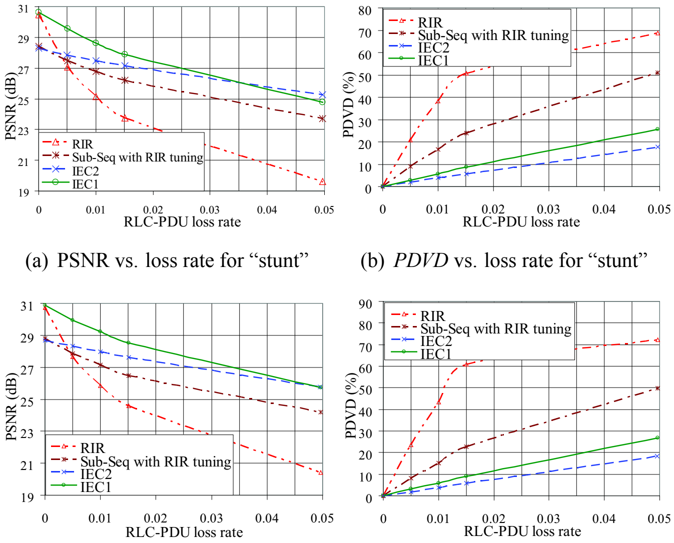


Fig. 3. Decoded quality vs. Slice size

To configure the system, appropriate slice size is selected. Typically it is close to half the radio frame size (RFS), but too small a slice size can result in considerable degradation of quality because of added restrictions on the prediction syntax. Figure 3 shows the decoded PSNR vs. slice size for various RLC-PDU loss rates for the sequence "stunt". The RFS for this experiment was set to 320 bytes. A suitable slice size observable from graph is 200 bytes.



(a) PSNR vs. loss rate for "stunt"

(b) *PDVD* vs. loss rate for "stunt"

(c) PSNR vs. loss rate for "party"

(d) *PDVD* vs. loss rate for "party"

Fig. 4. Performance results

Figure 4(a) show the decoded PSNR vs. RLC-PDU loss results for sequence "stunt." The measured $\overline{RTT}$ for the experiments was less than 130 ms and hence 3 reference frames were used. RIR rate was set to 5% of the total MBs in a frame. It can be seen that RIR technique performs worst. Using temporal sub-sequences along with instantaneous RIR tuning performs reasonably better for lossy channel conditions. IEC2 gives better performance, however the best configuration is IEC1. It is evident that the quality loss incurred by using sub-sequences, because of temporally older references, can not be compensated by the error robustness it adds for all practical loss scenarios. IEC1 shows an improvement of 4 dB at a loss rate of 1.5%.

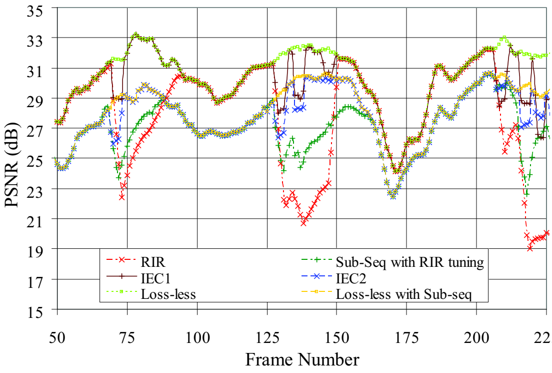
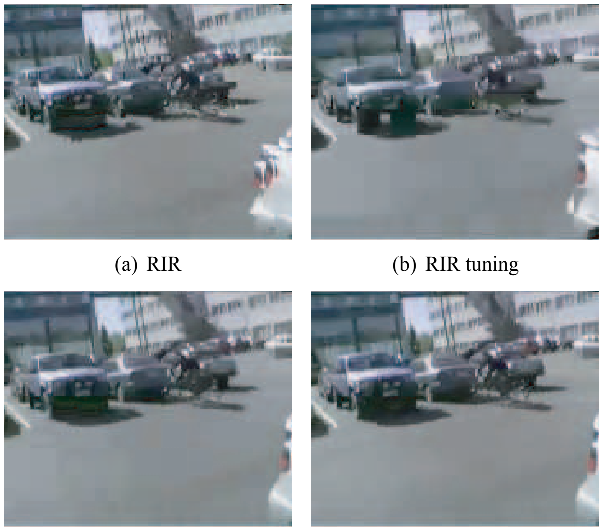


Fig. 5. PSNR variation within sequence "stunt"

Figure 4(b) shows that the proposed configurations give a significantly robust system with only about 20% of the video affected under the worst case losses of 5%. The feedback overhead only amounts to a maximum of 0.8% of the 128 kbps channel at a loss rate of 5%. Similar results were obtained for other test sequence, the results for sequence “party” presented in Figure 4(c) and 4(d).

Figure 5 shows the variation of PSNR for the sequence “stunt” for one channel realization. In addition to the results for all the techniques at an RLC-PDU loss rate of 1.5%, error-free results with *and* without using sub-sequences are plotted. Only a selected portion of sequence is plotted for better viewing. IEC techniques show very fast recovery from the losses. The results are in harmony with the average metrics.

Figure 6 shows the instantaneous visual comparison of frame number 220 from Figure 5. The subjective results follow closely to the objective analysis.


(a) RIR

(b) RIR tuning

(c) IEC2

(d) IEC1

Fig. 6. Subjective comparison for the investigated techniques.

5. CONCLUSIONS

A robust video codec configuration is essential for QoS provisioning for a mobile communication environment. The effects of packetization sizes is studied on system performance to select the appropriate configuration. Various feedback based interactive error control techniques were assessed. Results show that employing feedback in the target system gives significant advantage. However, the loss of compression caused by temporal sub-sequences cannot be offset by its added robustness. Hence the IEC technique using most recent reference frame is best performing solution for the target system and it improves the overall performance of the system by 4 dB for a moderate RLC-PDU loss rate of 1.5%, with more advantage for higher losses. An extension of this work can be to investigate techniques to reduce the memory overhead required for storing multiple reference frames for IEC.

6. REFERENCES

[1] T. Wiegand, G.J. Sullivan, G. Bjøntegaard, and A. Luthra, “Overview of the H.264/AVC video coding standard,” *IEEE Trans. on Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, July 2003.

[2] R. Zhang, S.L. Regunthan, 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, pp. 966–976, June 2000.

[3] T. Wiegand, N. Färber, K. Stuhlmller, and B. Girod, “Error–resilient video transmission using long-term memory motion–compensated prediction,” *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 6, pp. 1050–1062, June 2000.

[4] B. Girod and N. Färber, “Feedback-based error control for mobile video transmission,” *Proceeding of the IEEE*, vol. 97, pp. 1707–1723, Oct. 1999.

[5] I. Rhee and S. Joshi, “Error recovery for interactive video transmission over the internet,” *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 6, pp. 1033–1049, June 2000.

[6] M. Kalman, P. Ramanathan, and B. Girod, “Rate–distortion optimized streaming with multiple deadlines,” in *Proceedings IEEE International Conference on Image Processing*, Barcelona, Spain, Sept. 2003.

[7] W. Tu and E. Steinbach, “Proxy-based reference picture selection for real-time video transmission over mobile networks,” in *Proceedings IEEE ICME*, Amsterdam, Netherlands, July 2005, pp. 309–312.

[8] “Packet switched conversational multimedia applications; default codecs,” 3GPP Technical Specification TS 26.235, 3GPP, Mar. 2006.

[9] “Delay budget within the access stratum,” 3GPP Technical Specification TS 25.853, 3GPP, Mar. 2001.

[10] T. Stockhammer and S. Wenger, “Standard-compliant enhancement of jvt coded video for transmission over fixed and wireless ip,” in *Fourth International Workshop on Distributed Computing*, Capri, Italy, Sept. 2002.

[11] Z. He, Y. Liang, L. Chen, I. Ahmad, and D. Wu, “Power-rate-distortion analysis for wireless video communication under energy constraints,” *IEEE Trans. on Circuits Syst. Video Technol.*, vol. 15, no. 5, pp. 645–658, May 2005.

[12] G. Cote and F. Kossentini, “Optimal intra coding of blocks for robust video communication over the internet,” *Signal Processing: Image Commun., Special Issue on Real-time Video over Internet*, vol. 15, pp. 25–34, Sept. 1999.

[13] W. Zia, T. Stockhammer, T. Afzal, and W. Xu, “Time-sliced simulation and testing framework for mobile video applications,” in *9-th ACM/IEEE International Symposium on Modeling, Analysis and Simulation of Wireless and Mobile Systems, Tools/Demos*, Torremolinos, Spain, Oct. 2006.

[14] “Permanent document on test components,” 3GPP Permanent Document S4-060515, 3GPP, Aug. 2006.