# PROXY-BASED REFERENCE PICTURE SELECTION FOR REAL-TIME VIDEO TRANSMISSION OVER MOBILE NETWORKS

Wei Tu and Eckehard Steinbach
Institute of Communication Networks, Media Technology Group
Technische Universität München
{wei.tu, eckehard.steinbach}@tum.de

**ABSTRACT**

We propose a framework for error robust real-time video transmission over wireless networks. In our approach, we cope with packet loss on the downlink by retransmitting lost packets from the base station (BS) to the receiver for error recovery. Retransmissions are enabled by using fixed-distance reference picture selection during encoding with a prediction distance that corresponds to the round-trip-time between the BS and the receiver. We deal with transmission errors on the uplink by sending acknowledgements and predicting the next frame from the most recent frame that has been positively acknowledged by the BS. We show that these two separate approaches for uplink and downlink nicely fit together. We compare our approach to state-of-the art error resilience approaches that employ random Intra update of macroblocks and FEC across packets for error resilience. At the same bitrate and packet loss rate we observe improvements of up to 4.5 dB for our scheme.

## 1. INTRODUCTION

In 3G networks, video services are expected to be among the most popular ones and may be the key factor for success. Wireless video applications without real-time constraints (e.g. Multimedia Messaging Service) have been introduced to the market. However, real-time video communication over wireless networks remains challenging. Decoding of erroneous or incomplete video bit-streams leads to severe quality degradations. Because of motion compensated prediction, these impairments also propagate in space and time and therefore stay visible for a significant amount of time. Hence, an error resilient transmission scheme is essential to achieve good quality in a wireless multimedia communication system. A recent tutorial overview of approaches for error resilient video transmission can be found in [1],[2]. The specific error resiliency tools defined for the recently defined standard H.264/MPEG-4 AVC are described for instance in [3]. Most error resiliency tools require at least partial information about the current transmission situation in order to optimally trade-off error resilience versus coding efficiency.

For traditional data communication applications, retransmission of lost information triggered by feedback from the receiver is considered to be the most suitable approach for error resilience. The big advantage of feedback-based retransmission is its inherent adaptiveness to varying loss rates. Retransmissions are only triggered if information is actually lost. The overhead encountered is therefore a direct function of loss rate and the sender does not need to receive explicit information about the expected channel condition. For bi-directional conversational services like video telephony, however, the benefit of traditional packet retransmission is limited because of the stringent one-way latency requirement which is typically in the range of 150-250 ms.

Because of this limitation of retransmission-based approaches, alternative ways to exploit feedback information from the receiver have been proposed. Error tracking [4],[5], for instance, uses feedback about lost packets at the sender to reconstruct error propagation. Corrupted areas are encoded in Intra mode which leads to error recovery without introducing additional delay. Another proposal that uses feedback information to stop error propagation is NEWPRED [6],[7]. Here, feedback about lost packets or correctly received packets is used to restrict the prediction from those image areas that have been decoded successfully. The reference picture selection (RPS) concept introduced in H.263 Annex U which has been adopted in H.264/AVC supports a standard-compatible implementation of NEWPRED. The feedback information can also be incorporated into a RD-optimized mode decision as shown in [8],[9].

In [10] an elegant retransmission-based approach for end-to-end video error recovery called RESCU has been proposed. The main idea of RESCU is to change the frame dependencies in a video sequence such that a retransmission of lost information can be used for error recovery despite the low delay requirements of real-time video communication. In RESCU every p-th frame is a so called periodic frame that references another periodic frame p frame intervals away. The frames that are in-between two consecutive periodic frames are encoded using regular previous-frame prediction. If a non-periodic frame is lost, the error propagation stops at the next periodic frame. If a periodic frame is lost it is displayed using error concealment and a retransmission request is sent to the sender. If the retransmission arrives before the next periodic frame is to be displayed, fast decoding produces an error free reference frame for it.

In order to make RESCU work, the receiver has to be able to decode retransmitted information faster than real-time. Accelerated Retroactive Decoding (ARD) has been proposed in another context in [11]. ARD makes use of the ability of many streaming clients (especially those running on PCs) to decode video faster than real-time. Ordinarily, when packets used to decode a frame of video arrive late, an entire series of frames is distorted due to prediction dependencies among encoded frames. With ARD, when late-arriving packets finally do arrive, the decoder goes back to the frames corresponding to the late arriving packets and quickly re-decodes the dependency chain up to the current playout position without error. The remaining pictures in the GOP can then be decoded without error.

In this work, we propose a framework for mobile video telephony which deals with downlink packet loss by using fixed-distance reference picture selection (FDRPS) in combination with proxy-based retransmission of lost packets. The reference frame to be used for prediction is determined as a function of round-trip-time between base station and receiver on the downlink. To cope with losses on the uplink, we send feedback about lost packets from the base station to the sender. The encoder uses this information to predict the current frame from the most recent successfully uploaded reference frame which is older than the fixed-distance reference frame picked by the downlink strategy.

In comparison with [10], we consider proxy-based retransmission of lost information on the downlink and therefore avoid low coding efficiency that would be encountered when using RESCU for large end-to-end delays. Also, different to RESCU, in our scheme it does not matter which frame is lost. For every lost frame on the downlink we perform one retransmission and if that retransmission arrives in time, the error propagation is stopped. In case the retransmission also gets lost, in our scheme only a subsequence of the video will be affected by error propagation. This allows us to stop displaying this subsequence and to ask the sender for a resynchronisation frame. In RESCU, all frames will be affected. Similar to NEWPRED, we select reference pictures for prediction that have been successfully uploaded to the base station. However, different to NEWPRED, the choice of reference picture is additionally constraint by the fixed-distance encoding for downlink adaptation. Besides these major differences to previous work, the main contribution of our work is to extend these ideas towards an end-to-end scenario where we explicitly deal with losses on the uplink and/or downlink and take advantage of the comparatively small but potentially different round-trip times on the uplink and downlink.

The paper is organized as follows. In Section 2, our proposed framework for error robust real-time video telephony is described. Section 3 presents simulation results that show the improvements achieved by our proposal compared to standard error-resilient transmission approaches. Section 4 concludes the paper.

## 2. ERROR ROBUST VIDEO TRANSMISSION

Fig.1 illustrates our proposed framework for error robust real-time video transmission.

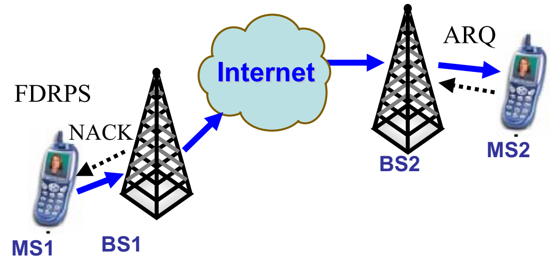


**Figure 1: Mobile video telephony scenario with FDRPS at the encoder, retransmission on the downlink, and adaptive RPS triggered by NACKs on the uplink.**

Let us assume that MS1 is the sender and MS2 is the receiver. Video is captured by the camera on MS1 and compressed using FDRPS. The video packets are sent uplink to base station BS1 and from there to BS2. BS2 sends the video stream downlink to the receiver MS2.

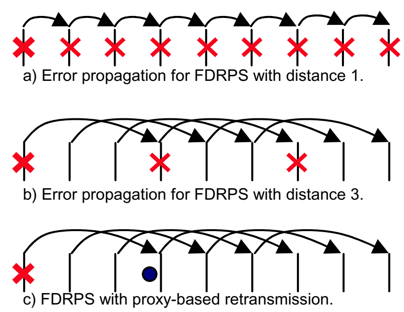


**Figure 2: Error propagation for FDRPS. The first frame marked with a thick line cross is lost. Frames corrupted by error propagation are marked with a thin line cross.**

The distance between the reference frame and the current frame for FDRPS is adjusted to the round-trip-time on the downlink. In case a packet is lost on the downlink, MS2 sends a NACK to BS2 via a feed-back channel and BS2 retransmits the lost packet once. The corrupted frame is displayed using error concealment. If the retransmitted packet arrives before the next frame has to be decoded that uses the lost frame as a reference, error propagation is stopped. Fig. 2 illustrates error propagation for FDRPS with and without retransmission of lost packets. In Fig. 2a) we show the effect of a lost packet if the current frame is predicted from the most recent frame (FDRPS distance 1). The loss of one frame affects all following frames. In Fig. 2b) FDRPS encoding with distance 3 is used. In case of frame loss only one third of the following frames are affected. FDRPS with distance 3 combined with proxy-based retransmission of lost packets is shown in Fig. 2c). Only one frame is corrupted and fast retransmission in combination with accelerated decoding stop error propagation. The arrival of a successful retransmission is shown as a solid circle.

While FDRPS with a distance adjusted to the round-trip-time of the downlink in combination with retransmission effectively stops error propagation as shown in Fig. 2c), the coding efficiency in the error-free case is decreased. We will show in Section 3 that even for low loss rates this effect is compensated by the greatly improved error-resilience.

We assume that lost packets on the downlink are retransmitted only once. The overhead caused by retransmission therefore corresponds to the packet loss rate of the downlink channel.

The error recovery strategy described so far is adapted to losses on the downlink. In case a packet is lost on the uplink, BS1 returns a NACK to MS1 and the sender reacts to the NACK by changing the reference frame for the next frame to be encoded as shown in Fig. 3. The lost frame is not used as a reference for any following frame. As a result, the frame loss on the uplink will only affect one single frame at the receiver. This frame is displayed at the receiver using error concealment. In Fig. 3 it is assumed that the round-trip-time on the downlink corresponds to three frame intervals. FDRPS therefore predicts frame *i* from frame *i-3*. After the NACK arrives for the lost frame, the reference picture used to predict the current frame is four frames back (thick arrow).

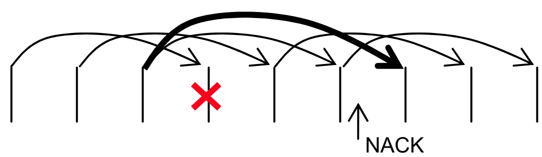


**Figure 3: Adaptive RPS triggered by feedback from the base station BS1 to the sender MS1.**

Packets that have been lost on the uplink from MS1 to BS1 do not arrive at BS2. This saves some transmission rate on the downlink which reduces the overhead encountered for retransmission of those packets that are lost on the downlink.

Please note that the description in this section assumes for illustration purposes that all slices of a video frame use the same reference frame for prediction. This restriction is not necessary. The adaptive RPS on the uplink and the retransmission and accelerated decoding on the downlink can also be done only for those slices that are lost. If one of the terminals is not mobile, the corresponding uplink or downlink error recovery strategy can be waived.

## 3. SIMULATION RESULTS

We simulate the burst packet loss behavior of a wireless channel using the well known two-state Gilbert-Elliott burst loss model. Foreman in QCIF resolution is used as the test sequence and 10 fps is selected as the frame rate. The H.264 test software version JM 8.4 [12] is employed as the video codec. We select a slice to correspond to one row of MBs. For transmission, one slice is put into one packet. The H.264 decoder performs error concealment when a slice is lost using both temporal and spatial information. In all simulations, the transmission rate for a video stream is fixed to 90kbps including all overhead. We compare our approach with two standard error resilience approaches. The first one is random Intra-update where we vary the number of Intra-MBs per frame from 0 to 60% and pick for every simulation run the random update rate which leads to the best reconstruction quality at the receiver. The second scheme used for comparison employs FEC across packets for protection against packet losses [13]. We assume that for every *k* video packets, *n-k* redundancy packets are transmitted. The receiver can decode the *k* video packets error free if any *k* out of the *n* transmitted packets are received. For the FEC-based scheme we vary the redundancy (*n/k*) from 0 to 250% and again pick the code rate that leads to the best result. The two schemes used here for comparison therefore work better than they would in practice where the current loss rate will not be known to the sender and picking the optimum Intra refresh rate or the optimum code rate would not be possible. For all schemes the QP of the encoded stream is adjusted to meet the bitrate constraint.

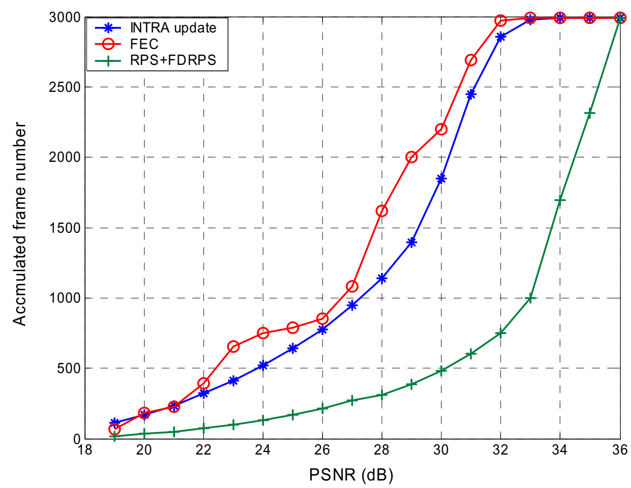


**Figure 4: Number of frames reconstructed with a quality lower than the PSNR values on the x-axis.**

In Fig. 4, we show experimental results for a test sequence of 3000 frames and a packet loss rate of 5% and mean burst length of 5 packets on both the uplink and downlink. The 3000 frame sequence is constructed by continuously repeating the Foreman test sequence. To avoid a scene cut, the sequence is played alternately in forward and backward direction. Fig. 4 shows how many frames out of the 3000 are reconstructed with a quality lower than the PSNR values shown on the x-axis.

It can be seen from Fig. 4 that random Intra-update and FEC-based scheme perform very similar. We set the prediction distance for our scheme (RPS+FDRPS) to be 3 frames, which corresponds to a round-trip-time on the downlink of 300ms. We set the QP to be 28, so that the encoded stream has a bitrate of 85 kbps and about 5kbps are left for the adaptive RPS on the uplink and the retransmissions on the downlink. Fig. 4 shows that the proposed scheme leads to significantly improved reconstruction quality in comparison to the other two schemes.

Fig. 5 shows the mean reconstruction quality in PSNR for the three schemes as a function of loss rate. We use the Gilbert-Elliott model to generate the test channels with 1%, 3%, 5%, 7%, 10% packet loss rate and keep the average burst length to be 5 packets. As we always pick the simulation runs with the best Intra-update rate and the optimum amount of FEC redundancy the two reference schemes outperform our scheme in the error free case. For error free transmission the optimum Intra-update rate and redundancy for the FEC-based scheme are 0%. However, even for very small loss rates the error recovery properties of our scheme lead to significantly improved performance. For packet loss rates between 5-10% we see up to 4.5 dB improvement in mean reconstruction quality.

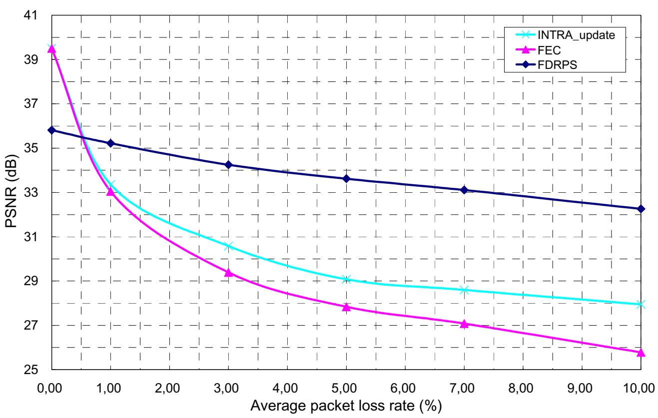


**Figure 5: Mean reconstruction quality as a function of loss rate for a mean packet burst loss length of 5.**

## 4. CONCLUSIONS

We have presented a framework for error resilient real-time video transmission in wireless environments. We combine fixed distance reference picture selection with retransmission of lost packets to deal with losses on the downlink. The distance of FDRPS is adjusted to the round-trip time of the downlink which gives us the opportunity to retransmit lost packets once and to use successfully retransmitted packet to stop error propagation by accelerated decoding. This strategy is combined with adaptive RPS on the uplink triggered by feedback from the base station to the sender. Please note that our proposal is fully standard-compatible when using H.264/AVC. The two major assumptions we make are the following. We assume that the base stations can send feedback about lost packets to the sender and can retransmit lost packets to the receiver. The second assumption is that the decoder has enough computational resources to decode retransmitted slices or frames fast enough to use them to stop error propagation. As a final point, we would like to mention that our approach can be combined with the two approaches used for comparison in this paper or other error resiliency approaches for which we expect even better performance.

## 5. REFERENCES


[1] Y. Wang, S. Wenger, J.T. Wen, A.K. Katsaggelos, “Review of Error Resilient Coding Techniques for Real-Time Video Communications,” IEEE Signal Proc. Magazine, vol. 17, no. 4, pp. 61-82, Jul. 2000.

[2] Y. Wang and Q. Zhu, “Error control and concealment for video communications: A review,” Proceedings of IEEE, Special Issue on Multimedia Signal Processing, pp. 974 - 997, May 1998.

[3] T. Stockhammer, M. Hannuksela, and T. Wiegand, "H.264/AVC in Wireless Environments", IEEE Transactions on Circuits and Systems for Video Technology, vol. 13, no. 7, pp. 657-673, July 2003.

[4] W. Wada, “Selective recovery of video packet losses using error concealment,” IEEE J. Select. Areas Commun., vol. 7, pp. 807–814, June 1989.

[5] E. Steinbach, N. Färber, and B. Girod, “Standard compatible extension of H.263 for robust video transmission in mobile environments,” IEEE Trans. Circuits Syst. Video Technol., vol. 7, pp. 872–881, Dec. 1997.

[6] ITU-T/SG15/LBC-96-033, “An error resilience method based on back channel signalling and FEC,” Telenor R&D, San Jose, Jan. 1996.

[7] S. Fukunaga, T. Nakai, and H. Inoue, “Error resilient video coding by dynamic replacing of reference pictures,” in Proc. IEEE GLOBECOM, vol. 3, Nov. 1996, pp. 1503–1508.

[8] T. Wiegand, N. Faerber, K. Stuhlmueller, “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, Jun. 2000.

[9] Y. J. Liang, M. Flierl, and B. Girod, “Low-Latency Video Transmission over Lossy Packet Networks Using Rate-Distortion Optimized Reference Picture Selection,” Proc. IEEE International Conference on Image Processing, vol. 2, pp. 181 -184, Rochester, NY, Sep. 2002.

[10] Injong Rhee and Srinath Joshi, “Error Recovery for Interactive Video Transmission over the Internet,” IEEE Journal on Selected Areas in Communications, Special Issue on Error Robust Transmission of Images and Video, vol.18 , no. 6 , pp. 1033-1049, June 2000

[11] M. Ghanbari, “Postprocessing of late cells for packet video,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 6, no. 6, pp. 669–678, Dec. 1996.

[12] http://iphome.hhi.de/suehring/tml/index.htm

[13] J. Rosenberg, H. Schulzrine, “An RTP payload format for generic forward error correction,” RFC 2733, Dec. 1999.