1st International Workshop on Cross Layer Design (IWCLD 2007), Sept. 20-21, 2007 Jinan, China

Cross-Layer Detection of Visual Impairments in H.264/AVC Video Sequences streamed over UMTS Networks

Luca Superiori, Olivia Nemethova, Wolfgang Karner, Markus Rupp
Institute of Communications and Radio-Frequency Engineering
Vienna University of Technology, Austria
Gusshausstrasse 25/389, A-1040 Vienna, Austria
Email: {lsuper, onemeth, wkarner, mrupp}@nt.tuwien.ac.at

***Abstract*—Incorrectly received packets in low-rate video sequences result in the loss of considerably large picture areas that have to be concealed. The performance of error concealment decreases with the size of the interpolated picture area. The incorrectly received packets may still contain some correct information that can be exploited at the decoder. In this work we propose the utilization of information from the link layer of UMTS (Universal Mobile Telecommunications System) at the application layer for better pre-localization of errors in the bitstream domain. Syntax check together with the detection of impairments in the pixel domain decide which part of the incorrectly received packets will be concealed. We evaluate the results using error traces from the live UMTS network and H.264/AVC (Advanced Video Coding) encoded video stream. Apart from reduced complexity facilitated by the cross-layer approach, the proposed method gains 1.09 dBs of Y-PSNR compared to a slice rejection mechanism.**

***Index Terms*—Video streaming, UMTS, H.264/AVC, error resilience, error detection, cross-layer.**

I. Introduction

H.264/AVC (Advanced Video Coding) [1], [2] is currently the newest and best performing video codec. It has been developed by the Joint Video Team (JVT) of ITU (International Telecommunication Union) and MPEG (Moving Picture Expert Group). The codec defines several *profiles* covering a wide range of applications. In the following we will refer to its applications for video transmission over UMTS. The 3GPP (3rd Generation Partnership Projects) specification [3] requires the mobile terminal to support H.264/AVC in its baseline profile.

H.264/AVC is a hybrid block-based video codec. Each frame is subdivided into *macroblocks* of 16 × 16 pixel. A macroblock is encoded by means of spatial and temporal prediction and entropy coding. The encoded video data is segmented in NALUs (Network Abstraction Layer Units), each containing a video *slice*.

For streaming and real time transmission, each NALU is further encapsulated into one RTP (Real Time Protocol)/ UDP (Universal Datagram Protocol)/ IP (Internet Protocol) packet. This protocol stack is recommended in [4] for continuos media playback at the receiver side. The UDP is an unreliable protocol that does not allow retransmissions. It contains 16 bits checksum for error detection. In case a UDP packet fails the checksum test, it is usually discarded [4], [5]. In the considered low resolution (usually CIF (352×288 pixels) or even QCIF (176×144)), and low bit-rate (64-384 kbps) scenario, this could result in losing a considerable part of the frame.

The encoded video data preceding the error occurrence in the damaged packet is still valid and can be used to reconstruct correctly a part of the encoded frame. In [6] the authors proposed a smart decoder able to detect syntax errors at macroblock level. The packets that failed the UDP checksum, are decoded until a syntax error arises. Only the following macroblocks are concealed. The method showed significant improvement compared to the classical slice rejection mechanism. It still suffers from a limited detection capability and a distance between the error occurrence and the error detection.

In order to enhance the performance of the syntax analysis, in [8] a visual impairments detection mechanism was proposed. The characteristics of the visual artifacts remaining after syntax analysis were examined and, by means of local image statistics analysis, both the detection distance and the detection probability were improved. The method, however, requires the visual analysis of the whole decoded NALU. In the considered scenario a NALU can contain a whole frame.

The smallest unit in which an error can be detected without additional mechanisms (assuming UMTS as the underlying system), is the RLC Packet Data Unit (PDU). The information from the RLC (Radio Link Control) layer can be exploited by the decoder to reduce the visual artifacts search area. In this article we present a cross-layer mechanism capable of detecting visual impairments in the decoded frame. We limit the search region profiting of the information coming from the lower layers.

The paper is organized as follows. In Section II, the visual detection mechanism is briefly described. Section III presents the proposed cross-layer mechanism. The simulation scenario is described in Section IV. An evaluation of the results obtained is performed in Section V. Final remarks and conclusions are provided in Section VI.

1-4244-1284-6/07/$25.00 ©2007 IEEE

II. Detection of Visual Impairments

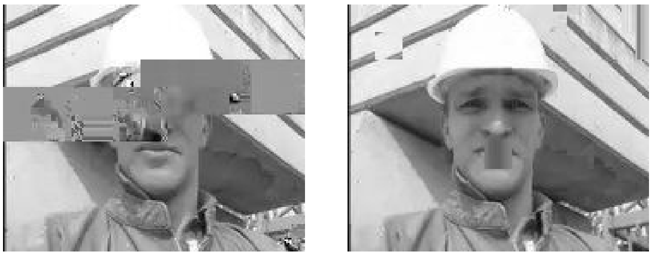
Focusing on the wireless link, the decoder of the mobile equipment has to cope with damaged packets. Since retransmission is not allowed, a damaged packet is considered as ”lost”, even if only few bits of its payload are faulty. In the following, we will refer to damaged packets as packets affected by *bit inversion*. The encoded data preceding the first error is still valid. A smart decoder can possibly decode damaged packets [5].

The decoding of a packet affected by bit inversion yields two drawbacks. On the one hand, the decoding of a faulty codeword leads to a wrong decoded value. The decoded parameter will deviate from the encoded one, possibly causing the misinterpretation of the following syntax elements. On the other hand, the entropy variable length encoding style used in H.264/AVC is sensitive to desynchronization. Even a single bit inversion can cause the deviation of the boundaries of a codeword, affecting the following ones up to the end of the slice.

In [6], a method for detecting errors in the bitstream syntax has been presented. During the decoding, the range and significance of the data associated to an encoded macroblock are analyzed. An error is detected in case invalid codewords or illegal decoding actions arise during the decoding of a macroblock. Therefore, false detection cannot occur. The macroblock considered as erroneous and the following ones up to the end of the slice are marked as faulty. The method was implemented as improvement of the reference H.264/AVC software [7] ver. 10.2.

The simulations showed that an error is usually detected later than its real occurrence. Between the error occurrence and the error detection the sequence is incorrectly decoded. This can result in strong artifacts. Such impairments, however, possess some characteristic visual features, allowing us to perform a refined search in the pixel domain.

A visual impairments detection algorithm is described in [8]. Simulations showed considerable difference for errors occurring in packets containing intra (I) or inter (P) predicted slices, as shown in Fig. 1. Inter predicted slices are encoded reducing the temporal redundancy, referencing to a macroblock belonging to an already decoded frame. Intra predicted slices are self contained, the encoded information references only other blocks belonging to the same slice, reducing the spatial redundancy. Therefore, two different detection mechanisms have been developed.


(a) I frames

(b) P frames

Fig. 1. Visual impairments caused by undetected errors.

The encoded data associated to I frames is highly susceptible to desynchronization. Usually, in case of bit inversions, the faulty macroblock and the following one result in annoying artifacts. In order to detect them, we proposed a visual impairments detection mechanism based on the information provided by the syntax analysis.

Our approach consists on the evaluation of the frame characteristics in the pixel domain. Artifacts can be noticed observing the elementwise difference between the examined frame and the previous ones, assumed to be error-free. The resulting difference map contains high magnitude components generated by movement within two frames as well. However, the considered artifacts lead to blockiness.

The detection of such impairments can be aided by means of an edge detector. In order to improve the robustness of the decision, for I frames we implemented a voting system considering sequences of decoded macroblocks. For each evaluated macroblock, the sequence vote is increased if its characteristics are similar to those of a visual artifact, decreased otherwise. The sequence vote then drives the decision, whether a series of artifacts has been found or not.

The mentioned mechanism works on top of the syntax analysis. Once a macroblock is recognized as faulty at bit level, this information is forwarded to the visual impairment detector. The method in [6] cannot lead to false detection, but the detection is performed later than the real error occurrence. Therefore, with the additional method of [8] possible impairments preceding the macroblock where a syntax error has been found are considered as well.

The information encoded in a P frame consists mainly of the position of the reference macroblock (motion vector) and of the transformed difference block (residual) between the predicted and the block to be encoded. Occurring errors lead to artifacts similar to those of temporal error concealment, i.e. spatial shifting of the affected macroblock. The entropic Variable Length Coding (VLC) desynchronizes rarely. The impairments remain usually isolated as shown in Fig. 1(b), the macroblocks following the error can still be valid. In the considered scenario, a whole P frame usually consists of a NALU. Concealing all the macroblocks following an erroneous one can cause some valid macroblocks to be concealed as well. Therefore, for P frames, the decision is taken independently for each macroblock.

III. Cross Layer Assisted Detection

The video slices generated by the H.264/AVC Network Abstraction Layer (NAL) have to be further encapsulated in appropriate transport protocols. For streaming or real time applications, the usage of RTP over UDP/IP is suggested in [4]. Each IP packet is then segmented in the UMTS Terrestrial Radio Access Network (UTRAN) Radio Link Control (RLC) [10] layer and mapped onto the transport channel by the Media Access Control (MAC) as shown in Fig. 2. The payload of the RLC packet consists of q bits. The value of q is set typically to

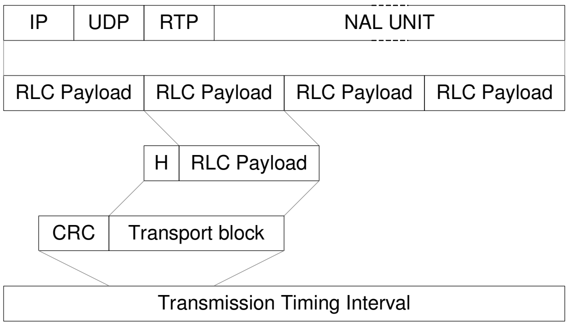


Fig. 2. Protocol stack

320 bits for data rate below 384 kbps and 640 bits otherwise. The RLC of UTRAN can both work in Acknowledged Mode (AM) or Unacknowledged Mode (UM). The latter allows error detection but no feedback and no retransmission. A *transport block* (TB) is obtained by adding an 8/16 bits header (for UM/AM mode, respectively) to the RLC packet. Cyclic Redundancy Check (CRC) info bits are finally attached to each transport block. The number of transport blocks in a TB set, interleaved over one Transmission Timing Interval (TTI) is defined by the transport format.

In [9] a cross-layer detection method for H.264/AVC encoded video over UMTS has been presented. The information about the correctness of the RLC packets at the radio link layer, are used at the application layer to reduce the impact of erroneous code segments. The decoder is intended to conceal the macroblocks since the first incorrectly received RLC packet. In this work we will exploit such information to enhance the detection mechanism described in Sec. II.

Instead of performing the visual detection over the whole damaged UDP payload (usually over 700 bytes) we limit the search area to the smaller area resulting by the decoding of the damaged RLC packets. The advantages are twofold. On the one hand we are reducing the detection mechanism complexity. On the other hand we reduce the number of possible false detection of visual artifacts.

IV. Simulation Scenario

In order to simulate the considered wireless transmission scenario, we based our simulations on the link error characteristics measured in live networks [11] for the UMTS Dedicated CHannel (DCH) in a mobile scenario. The error occurrences are correlated, indicating memory in the link. The network used as a reference allows three bearer data-rates (384, 128 and 64 kbps) with 12, 8 and 4 TBs per TTI respectively.

The measurements demonstrate that the error distribution has a bursty nature. Defining TTI_d as a TTI containing at least one damaged TB, the distribution of the number of erroneous transport blocks within a TTI_d can be approximated by a binomial distribution (see Fig. 3) with parameter $p_{eTB} = 0.98$ in the dynamic case.

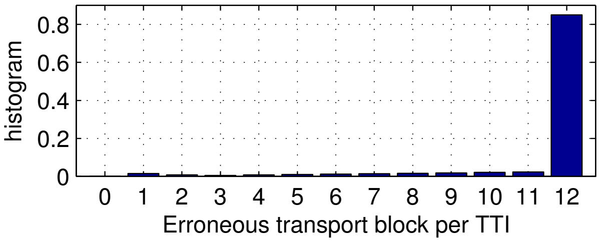


Fig. 3. Distribution of the number of incorrect TBs within a TTI_d

The distribution of the damaged TB is known from the measurement trace. Thus, within a TTI_d, each single TB has a probability of $p_{\text{eTB}} = 0.98$ to be damaged. Consequently, the probability that all the bits within a TB in a TTI_d are correctly received is $p_{\text{cTB}} = 1 - 0.98$, where

$$p_{\text{cTB}} = (p_{\text{cb}})^q, \tag{1}$$

and p_{cb} being the probability that a bit is correctly received.

This results into a bit error rate (BER) within a TB belonging to a TTI_d equal to

$$p_{\text{eb}} = 1 - \sqrt[q]{1 - p_{\text{eTB}}}, \tag{2}$$

considering $q = 320$, it follows $p_{\text{eb}} = 0.01215$.

The video used for the simulation is ”Foreman” in QCIF resolution. The sequence was encoded in baseline profile, using quantization parameter (QP) set to 26. In order to minimize the overhead caused by padding, the size of the NALU packets was limited to 680 bytes. Adding the IP/UDP/RTP 40 bytes long header, an IP packet of maximal 720 bytes (multiple of 320 bits) is obtained.

During the simulated transmissions, the resulting bit errors were inserted at RLC level. Errors occurring in the headers (IP,UDP,RTP, and NALU header) cause the loss of the entire considered packet. Errors affecting the NALU payload are handled as described in Sec. II.

V. Experimental Results

As first investigation we measured the complexity reduction of the proposed method compared to the implementation of visual impairment detection without the aid of the cross-layer information. The latter is based on the UDP checksum test, giving information about the correctness of the whole UDP payload (692 bytes). The proposed cross-layer detection mechanism relies on the RLC CRC, deciding the validity of the 320 bits contained in a RLC packet.

In the considered resolution and quality scenario, an I frame is subdivided into a number of slices depending on the frame content, usually four to six. The implementation in [8] requires the algorithm to perform the detection over about 20 macroblocks. The size of an I macroblock, using a QP of 26, is, in average, bigger than an RLC packet. Thus, the cross-layer information allows the exact localization of the damaged macroblock position. The detection of visual artifacts can therefore be performed beginning with the first damaged macroblock without investigating the previous ones that are assumed to be correctly received. The visual detection is still necessary, since the error could not have caused desynchronization and, therefore, resulted in visible artifacts. In case the visual artifact detector recognizes impairments, the position of the first damaged macroblock can be exactly determined using the RLC informations, minimizing the detection distance.

For P frames the detection of visual impairments is performed independently for each macroblock. A P frame in the considered scenario usually consists of a single RTP packet. Relying only on the information of the UDP checksum, the visual detection has to be performed over the whole frame. As discussed for I frames, we exploit the cross-layer information in order to start the search since the macroblocks resulting from the decoding of the damaged RLC packet. Considering the bursty nature of the occurring error, the average complexity decreases by a factor of two.

As final comparison we considered the time variant quality resulting from the decoding of damaged sequences. We compared the performance of the following three handling strategies:

H.1 The standard slice rejection mechanism. In case a UDP packet fails the checksum test, it is discarded and the whole contained video slice is concealed.

H.2 The cross-layer mechanism as proposed in [9]. The macroblocks within a frame are concealed starting from the first damaged RLC packet.

H.3 The visual impairment detection mechanism proposed in [8], relying on the additional cross-layer information as presented in this article.

We focused our investigation only on the detection of errors, as the most appropriate concealment method can be chosen depending on the sequence characteristics. For all the three handling strategies the concealment is a zero motion temporal error concealment. The results are plotted in Fig. 4 in the form of quality ECDFs (Empirical Cumulative Distribution Function).

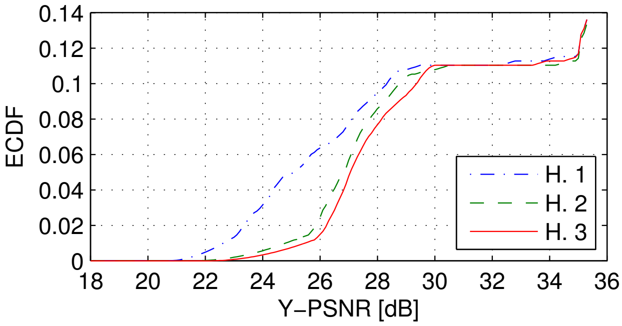


Fig. 4. Quality ECDF

The quality is expressed in terms of luminance Peak Signal to Noise Ratio (Y-PSNR). The plot is zoomed on the area representing the damaged frames. For the considered mobile scenario, they represent about 11% of the total number of frames. We observe enhancement using the cross-layer detection of visual impairments (H.3) with respect to the cross-layer mechanism in [9] (H.2). The method in [9] performs the concealment relying only on the RLC informations. The proposed algorithm evaluates additionally the impact of the error consequences. The occurring errors that do not cause distortion, do not result in unnecessary concealment. Considering only the damaged frames, the average improvement of Y-PSNR with respect H.2 and the slice rejection mechanism (H.1) are 0.19 dB and 1.07 dB, respectively.

VI. Conclusions

In this article we presented a cross-layer assisted error detection mechanism for H.264/AVC sequences transmitted over UMTS. Since the valid information preceding an error within a packet can still be exploited, we propose a cross-layer error localization mechanism. After a bitstream syntax analysis, the decoded frame is further investigated in the pixel domain, where visual artifacts are identified. The area, where such investigation is performed, is reduced exploiting the CRC information of the RLC layer. After a damaged macroblock is recognized, appropriate concealment methods can be used. The simulations, performed using the link error characteristics measured in a live scenario, confirm the benefit of the proposed method with respect to the standard mechanisms.

Aknowledgment

The authors thank mobilkom austria AG for technical and financial support of this work. The views expressed in this paper are those of the authors and do not necessarily reflect the views within mobilkom austria AG. Our thanks also go to Josep Colom Ikuno who helps during the implementation of the detector.

References

[1] ITU-T Recommendation H.264 and ISO/IEC 11496-10 (MPEG-4), “AVC: Advanced Video Coding for Generic Audiovisual Services,” version 3, 2005.

[2] I.Richardson: “H.264 and MPEG-4,” John Wiley & Sons Ltd, UK, 2005.

[3] “Packet switched conversational multimedia applications; Default codecs,” 3GPP TSG TS 26.235 ver. 6.4.0

[4] “Transparent end-to-end Packet-switched Streaming Service (PSS); Protocol and Codecs,” 3GPP TSG TS 26.234, ver. 5.7.0.

[5] S. Wenger, “H.264/AVC Over IP,” IEEE Trans. On Circuits And Systems for Video Technology, 13, no. 7, pp. 645-656, Jul. 2003.

[6] L. Superiori, O. Nemethova, M. Rupp, “Performance of an H.264/AVC Error Detection Algorithm Based on Syntax Analysis,” in Proc. of the Int. Conf. on Advances in Mobile Computing and Multimedia (MoMM 2006), Yogyakarta, Indonesia, Dec. 2006.

[7] H.264/AVC Software Coordination, “Joint Model Software,” ver.10.2, available in
http://iphome.hhi.de/suehring/tml/.

[8] L.Superiori, O. Nemethova, M. Rupp, “Detection of Visual Impairments in the Pixel Domain,” submitted to Picture Coding Symposium (PCS 2007), Lisboa, Portugal, Nov. 2007.

[9] O. Nemethova , W. Karner, A. Al Moghrabi , M. Rupp, “Cross-Layer Error Detection for H.264 Video over UMTS,” in Proc. of the International Wireless Summit (WPMC 2005), Aalborg, Denmark, Sep. 2005.

[10] H. Holma, A. Toskala: “WCDMA for UMTS: Radio Access for Third Generation Mobile Communications,” John Wiley & Sons Ltd, UK, 2007.

[11] W. Karner, P. Svoboda, M. Rupp: “A UMTS DL DCH Error Model Based on Measurements in Live Networks,” in Proc. of 12th International Conference on Telecommunications, Capetown, South Africa, May 2005.