QUALITY ESTIMATION OF DAMAGED H.264/AVC SEQUENCES

Josep Colom Ikuno, Luca Superiori, Markus Rupp

Institute of Communications and Radio-Frequency Engineering
Vienna University of Technology, Austria
Gusshausstrasse 25/389, A-1040 Vienna, Austria
Email: {jcolom, lsuper, mrupp}@nt.tuwien.ac.at

ABSTRACT

In this paper we propose a subjective quality predictor for damaged low resolution H.264/AVC encoded sequences. The predictor utilizes the information relayed by a visual artifacts detection algorithm working at pixel level and is supported by a syntax analysis error detector working at bit level. The algorithm is capable of predicting the Mean Opinion Score (MOS) of the received damaged sequence. Preliminary results show a strong correlation between the predicted data and subjective measurement tests with volunteers. Such a method could be used over video transmissions in error-prone channels, such as in wireless systems to adjust the transmission to the estimated received subjective quality.

1. INTRODUCTION

H.264/AVC (Advanced Video Coding) [1] is nowadays the best performing video coding standard. Due to its enhanced coding gain and error resilience tools, it is widely employed for transmission of video over packet networks.

The units containing H.264/AVC an stream are the Network Abstraction Layer Units (NALUs). In this paper we will consider H.264/AVC encoded sequences streamed over a wireless packet network, with each NALU encapsulated into RTP/UDP/IP (Real Time Protocol - User Datagram Procol - Internet Protocol). The UDP packet contains a simple checksum capable of detecting errors. If an error is detected at the receiver, the entire packet is usually discarded and the missing part of picture is concealed. In [2] and [3], a combined method capable of more accurately detecting the erroneously received parts of a damaged H.264/AVC encoded frame was presented. By using the information output by those algorithms, we propose a quality estimator capable of estimating the subjective quality, measured as MOS of the decoded sequence.

H.264/AVC works by dividing the picture frame into 16×16 pixel macroblocks (MBs). Those MBs are in turn grouped into slices, which in our case are made to fit each into one RTP packet. Thus, every transmitted RTP packet contains a slice.

By analyzing the H.264/AVC bitstream and the decoded frame and feeding this information into the quality predictor, a cross-layer system capable of adapting the transmission to the subjective quality of the underlying video stream could be implemented.

In this work, the "foreman" video at QCIF resolution (176×144 pixels) and 20 fps frame rate has been used.

The paper is organized as follows: in Section 2 the error detection strategy will be explained, after which the quality estimation process will be explained in Section 3. In Section 4 the MOS tests on which the estimator is based are explained. With these findings, we are able to derive our analytical model in Section 5. Finally Section 6 concludes the paper.

2. ERROR DETECTION

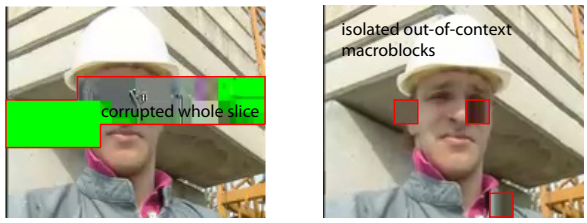
The proposed error detection method works by analyzing the received H.264/AVC stream at both bit and pixel level. We have called these methods Syntax Check (SC) [2] and Visual Impairments Detection and Concealment (VIDC) [3], respectively.

In SC, the error detection occurs during the reading and interpretation of the code associated to a MB. If an error is detected while decoding one of them, all the following ones up to the end of the slice are marked as flawed.

Syntax errors can be caused by different reasons, as defined in [2]. However, there are still errors that cannot be detected by means of SC. Moreover, an error is usually not detected immediately, but after some desynchronized decoding. In such cases, the errors can propagate over the slice causing the following MBs to be decoded erroneously. This behavior is typically seen in intra predicted frames (I frames). Inter predicted frames (P frames) do not generally suffer from this spatial error propagation. Figure 1 depicts the typical effects of desynchronized decoding on I and P frames.

VIDC works by enhancing the detection capabilities of SC by performing an additional visual analysis of the decoded

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.



(a) (b)

Fig. 1. Visual impairments caused by transmission errors. (a) depicts the typical result of transmission errors in I frames, a whole corrupted slice. In (b) it can be seen that for P frames, transmission errors produce isolated corrupted MBs

frame.

For I frames, the detection works on top of SC. To ensure more resistance against false positives, the decision is made depending on the result of a voting system. This voting system is designed to find the characteristic sequence of corrupted MBs that result from bit desynchronization.

As for P frames, the detection is performed separately for each 8×8 and 4×4 block, exploiting the isolated-block nature of the visual artifacts. The decision is taken based on the combined observation of edge and difference characteristics. In P frames analysis SC information is not used.

Optionally, a concealment of the detected erroneous MBs can be performed. A simple zero motion temporal error concealment has been used. Although more complex concealment strategies could have been used, with better results [4], since the focus of SC and VIDC is on detection rather than concealment, the mentioned simple copy-paste concealment method has been used. With it, each corrupted MB in the current frame is replaced with the spatially corresponding one in the previous frame.

The detection algorithm is embedded into a JM v.10.2 H.264/AVC decoder [5], while the quality estimator works outside of the decoding process, using the files output by the artifact detection algorithm as input. Figure 2 shows the structure of the modified JM decoder.

3. QUALITY ESTIMATION

The proposed quality estimator uses the artifact information output by SC and VIDC and uses it to estimate the subjective quality of the decoded H.264/AVC sequence.

The detection algorithm outputs the following information for each detected artifact: position in the frame, artifact size in MBs, frame type (I or P) and the $\mathrm{diff}_{\mathrm{frame}}$ parameter, which quantifies the amount of movement detected in the frame by averaging the measured difference between the current and previous frames.

The quality estimator has been realized by performing MOS tests of damaged concealed and unconcealed sequences

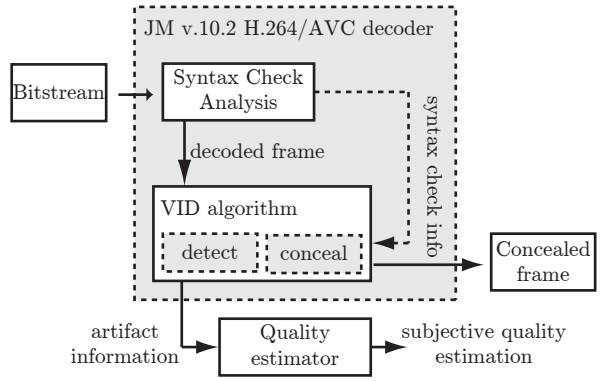


Fig. 2. Decoder and quality estimator structure

and then finding an analytical model that relates the characteristics the detected visual impairments to a MOS score.

MOS estimation was been split in the following steps:

1. Estimation of the MOS in the case no concealment is introduced (f_{MOS}).

2. Estimation of the MOS gain the concealment strategy will provide (f_{gain}). In this paper we used the mentioned copy-paste algorithm.

The resulting structure of the quality estimator can be seen in Figure 3.

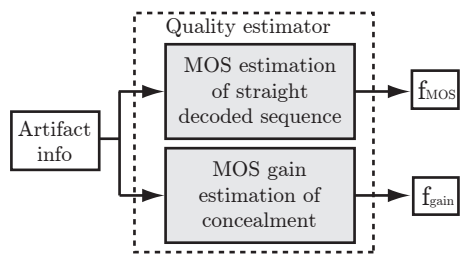


Fig. 3. Quality estimator structure

f_{MOS} predicts the MOS from the unconcealed sequences, what we have called straight decoded (SD) sequences, while f_{gain} predicts the MOS gain that the applied copy-paste error-concealment mechanism will provide. To be able to separate the quality estimation, the MOS tests contained both SD and concealed sequences, so both f_{MOS} and f_{gain} were obtained.

This partitioning makes it possible to estimate the perceptual quality of either concealed or unconcealed sequences. By using only f_{MOS}, our method can be potentially used on any H.264/AVC video sequence to estimate its degradation due to transmission errors. If error concealment is also used, our method can also output the predicted MOS score of the concealed sequence ($f_{\mathrm{gain}} + f_{\mathrm{gain}}$).

4. MOS TESTS

MOS has been utilized as the perceived subjective quality measure that our proposed method estimates. Test subjects rated each sequence from 1 (worst) to 5 (best). The MOS of each is then arithmetic mean of all the individual scores, and will thus also range from 1 to 5 [6].

The individual foreman sequences were encoded with a quantization parameter of 28, frame buffer size of 1, GOP=25 and slice size=700 bytes. With these settings, an I frame typically contains 4-5 slices (is contained in 4-5 RTP packets), while a P frame only one (one RTP packet).

The sequences used for the testing were two fragments of the foreman video. A slow-moving one (frames 0-74) and a fast-moving one (frames 260-334). Each sequence was comprised of three GOPs (75 frames), with the error being always located in the middle GOP. Thus, each sequence begins with an error-free GOP followed by one containing an artifact and finishes with an error-free GOP.

Three different GOP positions have been used: frame 25 (I frame), 34 and 43 (P frames). For I frames, different artifact sizes have also been used: 1/3, 2/3 and whole slice. Typical size of an I frame slice with our used encoding settings was 22 MBs.

The test consisted of 46 sequences which were shown in random order: 16 slow-moving SD sequences, 16 slow-moving concealed sequences, six fast-moving SD sequences, six fast-moving concealed sequences and two error-free control sequences (fast and slow sequences).

5. MOS ESTIMATION

MOS estimation is performed by an analytical model that can estimate the MOS of the resulting video as a function of the parameters output by the detection algorithm.

The scatter plots in Figs. 4 and 5, depict the relation between the different variables the estimator takes into account and (f_{MOS}) and f_{gain} respectively.

For f_{MOS} an exponential model has been chosen due to the exponential effect of artifact propagation [7]. For the MOS gain that the concealment introduces (f_{gain}), a simple quadratic model has been used.

When estimating the effect of an error or when concealing it the most important factor is its size. Hence, the detected artifact size determines the model that is used for f_{MOS}. The distribution of the dots in Fig. 4 (lower plot) conforms to the exponential decay model in [7]. In the case of f_{gain}, if too much concealment is performed, the user may perceive a decrease in quality, thus the concave distribution of the dots in Fig. 5 (uppermost plot).

Also as expected, the more movement there is in the sequence ($\mathrm{diff}_{\mathrm{frame}}$), the worse the effect of a transmission error (the movement spreads the error to a bigger area in following frames) and the poorer the performance of the applied concealment (it is a simple copy-paste algorithm, hence the more movement the more the concealed MBs appear out of context).

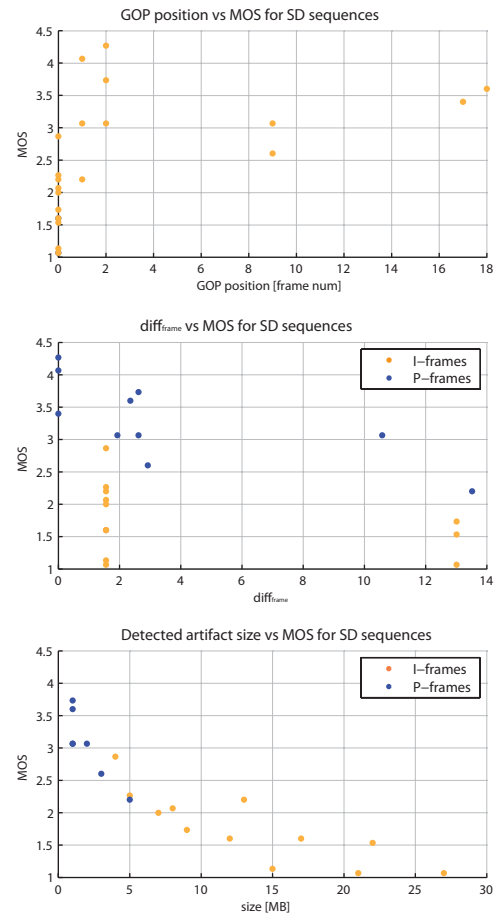


Fig. 4. f_{MOS} related parameters. An exponential model has been chosen to model it.

Equations (1) and (2) show the model that has been used to estimate the MOS. For f_{MOS}, n is the size of the impairment in MBs, d $\mathrm{diff}_{\mathrm{frame}}$ and g the position of the frame in the GOP. For f_{gain}, n is the number of concealed MBs and d $\mathrm{diff}_{\mathrm{frame}}$. The values of μ_f, δ_f, υ_f, ν_f and ε_f can be found in Table 1 (f_{MOS}) and Table 2 (f_{gain}) for $f = \mathrm{I\,frame}$ and $f = \mathrm{P\,frame}$

$$f_{\mathrm{MOS}} = \mu_f n + \upsilon_f g + \delta_f d + \varepsilon_f \quad (1)$$

$$f_{\mathrm{gain}} = \mu_f n^2 + \nu_f n + \delta_f d + \varepsilon_f \quad (2)$$

To test the accuracy of the proposed quality estimator, the estimations were correlated to the actual measured values from the MOS tests. Results show a good correlation between the output of the estimator and the measured MOS from the tests. The results of the correlation of the quality predictor with the data from the MOS tests can be found in Table 3.

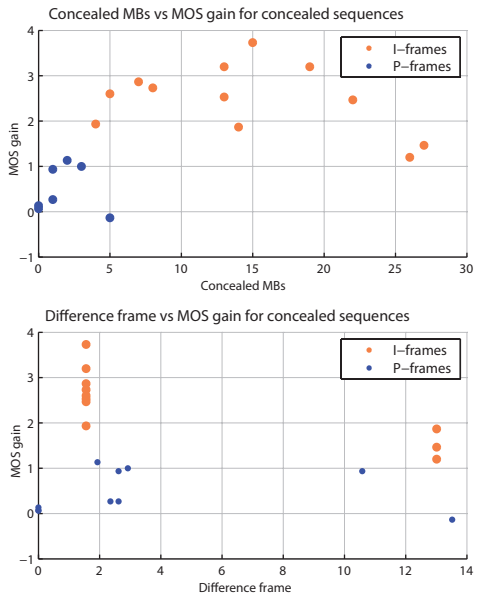


Fig. 5. f_{gain} related parameters. A quadratic model has been chosen to model it.

	I frame	P frame
μ_f	$-3.93 \cdot 10^{-2}$	$-9.76 \cdot 10^{-2}$
δ_f	$-1.66 \cdot 10^{-3}$	$8.93 \cdot 10^{-3}$
υ_f	0	$1.90 \cdot 10^{-3}$
ε_f	1.06	1.33

Table 1. Parameters for f_{MOS}

	I frame	P frame
μ_f	$-6.48 \cdot 10^{-3}$	-0.17
ν_f	0.20	-0.79
δ_f	$-9.73 \cdot 10^{-2}$	$1.54 \cdot 10^{-2}$
ε_f	1.68	$2.38 \cdot 10^{-2}$

Table 2. Parameters for f_{gain}

	I frames	P frames
f_{MOS}	0.850	0.905
f_{gain}	0.900	0.830

Table 3. Correlation of the estimations with the data

Further work will involve extending the estimator to make it content-aware, as the impact of some specific MBs can greatly vary (like players on a soccer field). For this, frame segmentation information, such as the proposed in [8] will be used.

6. CONCLUSIONS

In this paper, an H.264/AVC video subjective quality estimator has been proposed. Our proposed method is capable of estimating the subjective quality (measured by the MOS) of decoded erroneous H.264/AVC sequences via an analytical model obtained from MOS tests.

The estimator separately predicts the MOS of the unconcealed sequence (f_{MOS}) and then MOS gain that the error concealment will provide (f_{gain}). The employed two-step approach makes it possible to use the algorithm not only to estimate the quality of error-concealed sequences but also unconcealed sequences and keeps f_{MOS} valid even if the concealment strategy is changed (in this paper a simple zero motion temporal error concealment has been used).

The MOS estimation correlates well with the results from the subjective tests. A correlation of up to 0.9 was observed between the estimations and the measured MOS values.

The proposed method allows terminals in error-prone channels such as the wireless networks to feedback subjective quality information of the received video stream so the network can then use this cross-layer information to adapt the transmission to optimize the user-perceived quality.

7. REFERENCES

[1] "Advanced Video Coding, Final Committee Draft, ITU-T H.264, ISO/IEC 11496-10,document JVTE022," 2002.

[2] L. Superiori, O. Nemethova, and M. Rupp, "Performance of a H.264/AVC error detection algorithm based on syntax analysis," *Fourth International Conference on Advances in Mobile Computing and Multimedia*, Dec. 2006.

[3] L. Superiori, O. Nemethova, and M. Rupp, "Detection of visual impairments in the visual domain," *Picture Coding Symposium, Lisbon, Portugal*, July 2007.

[4] O. Nemethova, A. Al Moghrabi, and M. Rupp, "An adaptive error concealment mechanism for H.264/AVC encoded low-resolution video streming," *14th European Signal Processing Conference (EUSIPCO)*, Aug. 2006.

[5] H.264/AVC Software Coordination, "Joint model software, ver.10.2, http://iphome.hhi.de/suehring/tml/," .

[6] "ITU-T Rec. P.800, methods for subjective determination of transmission quality," Aug. 2006.

[7] O. Nemethova, W. Karner, C. Weidmann, and M. Rupp, "Distortion-minimizing network-aware scheduling for UMTS video streaming," *15th European Signal Processing Conference (EUSIPCO), Poznan, Poland*, Sept. 2007.

[8] L. Superiori and M. Rupp, "Encoding optimization of low resolution soccer video sequences," *International Conference on Multimedia and Expo 2008, Hannover, Germany*, June 2008.