Analysing GOP Structure and Packet Loss Effects on Error Propagation in MPEG–4 Video Streams

Árpád Huszák, Sándor Imre

Abstract— **Video streaming applications are commonly used in both wired and wireless environment; however, wireless links are burden by higher packet loss ratio and delay variation. In order to make video transmission possible in wireless networks MPEG video coding is usually used to reach the bandwidth constrains of the links. The video quality and compression ratio depends on Group of Pictures (GOP) structure, but it also affects the distortion sensitivity of the video stream due to packet losses. In this paper the correlation between GOP size, packet loss ratio and video quality is investigated. By increasing the distance between the reference frames the effectiveness of coding can be improved, but on the other hand the effect of error propagation due to packet losses also increases. Our aim was to find the optimal GOP structure to maximize the coding efficiency and minimize the quality distortion due to error propagation. We have implemented a simulation tool to make the analysis of differently structured video streams transmitted over lossy channels.**

Keywords – video streaming; MPEG; Group of Pictures; QoS, error propagation

I. INTRODUCTION

The demand for multimedia services is significantly increased in IP–based networks. Multimedia networking products like Internet telephony, Internet TV, video conferencing have appeared on the market. These applications are not only used in reliable wired networks, but also in wireless environment, where the obstacles of the expansion are the higher bit error ratio of the radio link and the limited bandwidth of the mobile links. Despite the development of new access technologies that provide higher bandwidth to the users, multimedia streaming applications still suffer from limited and highly varying bandwidth. Third–generation wireless networks are rapidly approaching reality, also providing higher bandwidth levels with the ability to transmit video streams in acceptable quality. The transmission of delay sensitive video streams over lossy wireless links needs special attention.

The real–time applications usually encode audio/video in a format that handles loss of packets, e.g. MPEG coding standard. MPEG uses intra–frame and inter–frame compression with different types of frames (I, P and B frames). The repeated pattern of I, P and B frames in an MPEG video stream is known as the Group of Pictures (GOP).

The choice of GOP structure affects static MPEG properties such as frame size and file size. This structure also impacts the streaming MPEG in terms of network bitrate and video quality.

The successful decoding ability of the compressed video stream with inter–frame dependencies depends heavily on the receipt of reference frames (I and P frames). While the loss of packets in a frame can degrade the video quality, the more problematic situation is the propagation of errors to dependent frames. By the increase of the packet loss rate, the quality of the decoded frames becomes too poor for viewing. The GOP structure defines the frame type layout of the video; therefore the error propagation extent depends heavily on this structure. The error will propagate till the next reference frame.

If the reference frames are close, the error will not spread widely, but in this case the video coding (compressing) efficiency will be also lower. If a fixed constant bit rate (CBR) video stream is considered, the lower coding efficiency means lower video quality. As the opposite of the previous case, if the reference fames (I and P frames) are far from each other, the error due to link failures will propagate to more frames. However, the coding efficiency will be higher, so the transmitted video quality will be better.

In this paper we have examined several GOP structures under different transmission conditions, and analyzed the optimal GOP structure, which provide the best video quality. Our aim was to find the optimal distance between the reference frame in order to maximize the coding efficiency and minimize the quality distortion due to error propagation. To make these analyses, we have implemented a simulation tool. As the result of our measurements, we were able to recommend the adequate GOP structure for the given link loss conditions, in order to achieve the best received video quality.

The rest of the paper is organized as follows. Section 2 presents an overview of the MPEG–4 video compression standard, and introduces some of the related works. In Section 3 we present a model for propagation of error due to packet losses and propose GOP structuring instructions, while in Section 4 we introduce our measurement results. Finally, conclusions and future work are drawn in Section 5.

Árpád Huszák is with the Budapest University of Technology and Economics, Department of Telecommunications, Budapest, Hungary (e-mail: huszak@hit.bme.hu).

Sándor Imre is with the Budapest University of Technology and Economics, Department of Telecommunications Budapest, Hungary (e-mail: imre@hit.bme.hu).

II. Background and Related Works

MPEG–4 is an encoding and compression system for digital multimedia content defined by the Motion Pictures Expert Group (MPEG) 0[2]. Inter–frame video compression algorithms such as MPEG–4 exploit temporal correlation between frames to achieve high levels of compression by independently coding reference frames. In this coding standard the majority of the frames are represented as the difference from each frame and one or more reference frames. However, these algorithms suffer from the well–known propagation of errors effect, because errors due to packet loss in a reference frame propagate to all of the dependent difference frames.

Intra–coded images (I frames) are coded independently of other frames in a manner similar to a JPEG coding. These frames are also called reference frames because they do not exploit temporal redundancy, but they are used as a reference in the prediction process. MPEG uses two types of dependent frames: predictively coded frames (P frames), and bi–directionally coded frames (B frames). The P frames are coded predictively from the closest previous reference frame (either an I frame or a preceding P frame), while B frames are coded bi–directionally from the preceding and succeeding reference frames.

The frame structure is specified by parameters N and M. These parameters are the intra–frame and inter–frame coding ratios, which define the sequence of I, P and B frames. Parameter N specifies the I frame interval whereas M determines the I or P frame interval.

Fig. 1. Frame sequence in MPEG video stream (N=9, M=3)

To limit the cascading effect that link errors create due to frame dependencies, the I frames frequency must be increased. However, the increased frequency of I frames must be traded off against the higher compression rates afforded by the P and B frames.

Errors in reference frames are more harmful than those in predicted frames due to error propagation. Several studies were made how to give higher level of protection to the key frames [3][4]. In these works redundancy was added to the important portions of the bitstream; however, this approach reduces compression gains. In [5] a retransmission based frame recovery was studied to minimize the error propagation effect.

In previous works the error propagation due to packet losses was analyzed in the GOP pattern. Maugey et al. [6] proposed a theoretical model for the error propagation phenomenon generated by a frame loss in a distributed video coding framework. Using rate–distortion functions, they analyzed the impact of a frame loss on the average distortion of a group of pictures depending on the position of the lost frame within the GOP.

Authors of [7] used an experimental approach to model the error propagation with commonly used error detection conditions. They have shown that errors detected in forward section of texture data may be propagated from motion data. Based on their experiments, they have proposed motion marker assumption and backtracking–based concealment strategies.

Lin et al. [8] analyzed the effect of wireless link characteristics on the video quality and found that burst packet losses on the video delivered quality is less than distributed packet losses in the same packet loss rate.

Some of the researchers have proposed guidelines for the GOP pattern determinations.

In [9] the authors analyzed a large range of GOPs to found the optimal GOP for MPEG streaming. They proposed a large number of P frames in one GOP to ameliorate the effects of the increased loss.

In paper [10], the number of B frames between two reference frames was investigated. According to the results the number of following B frames should be from 1 to 4, while in [11] the conclusion was that the number should be varied from 0 to 2. Paper [12] studies the impact of the choice of GOP by evaluating the effects of GOP on both static MPEG videos and on MPEG videos streaming over a lossy network. Their results consistently suggest two guidelines. First, the number of B frames between two reference frames should be close to 2. Second, the number of P frames should be 5 or fewer.

The presented related works tried to found a general GOP structure for different conditions. These solutions can be acceptable, if there is no information on the link loss ratio and an overall optimum is needed. However, with the new cross layer solutions the measured link characteristics can be handed over to the video coding application.

In our work we tried to find the most adequate GOP pattern if the loss ratio is known. For example, if there is no packet loss event during the transmission, the best quality can be achieved if the number of independently coded reference frames (I frame) is low, while the number of bi–directionally predicted frames (B frames) is high. With high compression rate (lot of B frames), the video quality is better if same coding bitrate is used. Using B frames the available bandwidth can be more efficiently utilized. The saved bandwidth can be used to transmit video frames with more details, so increasing the quality. This assumption is right, only if the coding rate is considered fixed.

III. Analyzing Relation of Error Propagation and GOP Structure

In mobile networks the strict bandwidth limitation determinates the video coding rate as well the video quality. The different MPEG frame types have dissimilar compression ratios and error propagation features. To

determinate the GOP structure, the frame features must be analyzed. In Table 1 the most important characteristics of I, P and B frames are represented.

The used GOP frame pattern significantly determinates the quality of the video, coded on the same bitrate. If only I frames are used the available link bitrate will reduce the efficiency of the coding and the streamed video quality will be lower compared to the case when B frames are also used. If fixed streaming rate is considered, the lowest quality is achieved if only I frames are used. By increasing the number of P and B frames the quality can be improved using the same streaming rate.

TABLE I
FRAME TYPE FEATURES

	I frame	P frame	B frame
Coding	intra–coded	forward prediction	bi–directional prediction
Reference frames	none	previous I or P frame	I and P frames
Compression ratio	low	medium	high
Frame size	high	medium	low
Error Propagation	high	medium	low

Unfortunately, an error in the I frame will be spread in the whole GOP causing significant distortion in the video stream. In order to reduce the effects of the error propagation, the distance between the reference frames must be reduced.

We have modeled error propagation when different frame types are damaged. The error in an I frame will cause errors in all frames in the GOP and in some frames in the previous GOP (till the last P frame). The first frame in a GOP is always a reference frame and there is only one I frame in the GOP, as Fig. 1 represents. We can consider that an error in the I frame will affect N (distance of I frames) frames in the actual GOP and M-1 frames in the previous one; therefore the distortion level is the highest when I frame is damaged.

The errors in a P frames also propagates to other frames. P frames are used for bi–directional prediction of B frames, and for the prediction of next P frames in the GOP. Hence the previous I or P frames stops the propagation backwards, but the forward propagation is stopped only by an I frame, as the next picture illustrates.

Fig. 2. Error propagation due to damaged P frame

The B frame is the only one, which is not used as reference; therefore it does not spread errors. Only the damaged frame will be distorted.

Using the distortion spreading characteristics of the different frame types, the expected number of distorted frames (Π) due to error propagation can be calculated as follows.

$$\mathbb{E}(\Pi) = p_I \cdot (N + M - 1) + p_P \cdot \pi + p_B \cdot 1 \qquad (1)$$

Parameter N specifies the I frame interval and M determines the I or P frame interval. In the equation p_I, p_P and p_B stands for the probability that error occurred in an I, P and B frames, respectively. These probabilities are correlated to the total size of different frames. The number of infected frames due to an error in a P frame depends on the position of the damaged P frame in the GOP; therefore in (1) we used variable π as the expected number of distorted frames. Considering that the number of P frames in a GOP is $N/M - 1$ and they are damaged uniformly, the expected number of distorted frames (π) can be given as the following equation shows.

$$\pi = \frac{1}{\frac{N}{M} - 1} \cdot (2M - 1) +$$
$$+ \frac{1}{\frac{N}{M} - 1} \cdot (3M - 1) + \dots + \frac{1}{\frac{N}{M} - 1} \cdot (\frac{N}{M} M - 1) \qquad (2)$$

If the last P frame in the GOP is damaged, beside the current P frame, 2(M-1) other B frame will be distorted. If the second last P frame is damaged, 3(M-1) other B frames, the last P frame and the current P frame will be distorted, so resulting 3M-1 frame distortions due to error propagation. By simplifying (2), the expected number of distorted frames (π) can be calculated as

$$\pi = \frac{2M + N - 2}{2}. \qquad (3)$$

The number of B frames also impacts the minimal playout delay because the next I or P frames in the sequence must be arrived to the receiver, in order to decode the bi–directionally coded B frames.

As a consequence we can say that the determination of the optimal GOP pattern must be based on the study of error propagation and coding efficiency assumptions. While the I frame thickening will reduce the error propagation, on the other hand it will increase the coding distortion. In the following section we will try to find compromise in the GOP structure determination to achieve the best quality at the video playout.

IV. MEASUREMENT RESULTS

In this section numerous GOP structure was analyzed in different transmission conditions in terms of loss rate. In our

measurements the *mother_and_daughter* QCIF sequence was used as the reference video. The video sequence was compressed with an MPEG–4 encoder – *ffmpeg* [13]. Each video stream was coded at 25fps and 100kbps.

First the coding efficiency was analyzed on fixed 100kbps coding rates. It can be seen in Fig. 3 that using only B frames and one I frame in the GOP results better video quality then using I frame and P frames. More significant improvement can be achieved if the length of the GOP is increased. Using the same coding rate (100kbps) even 8dB video quality improvement can be experienced. In the measurement we used the widely accepted peak signal–to–noise ratio (PSNR) video quality measurement algorithm.
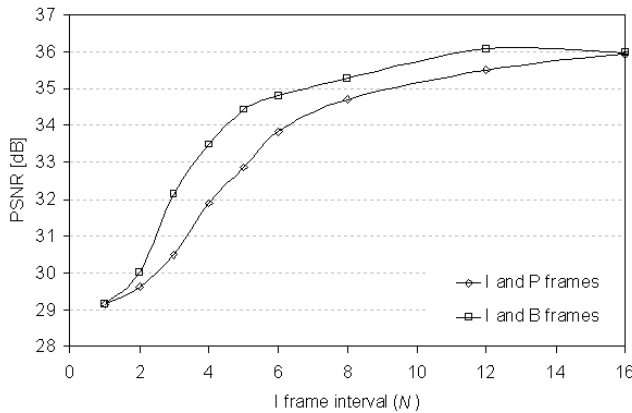

Fig. 3.   Coding efficiency in case of different GOP lenghts (N) using 100kbps coding rate

In order to simulate the packet losses, the coded video stream was split to 1kB sized packets and random packet loss ratio was set up. We used differently coded video streams to analyze the effect of error propagation and the level of quality distortions. In the examined streams the GOP length was changed from N=3 to N=40, while M=1 was kept unchanged. M=1 means that there are only P frames between the I frames. The GOP pattern without B frames made it possible to analyze the effect of GOP length variance without P frame interval influence on the measured results. As previously discussed the coding efficiency is the best when N=40 and the worst when N=3, while error propagation distortion is the highest when long GOPs are used and the lowest when short ones are used. The obtained results are presented in Fig. 4.

As it can be seen in the figure, the video quality of N=40 is the highest, while setting the I frame interval low (N=3), the PSNR value is the lowest when low packet loss ratios are examined. However, by increasing the packet loss ratio the order of the measured PSNR values changes and becomes the opposite. Among the four examined video coding settings, the best quality can be achieved by setting the GOP size to N=3 in case of high packet loss ratios. This behavior is due to the different error propagation characteristics of the examined video streams.
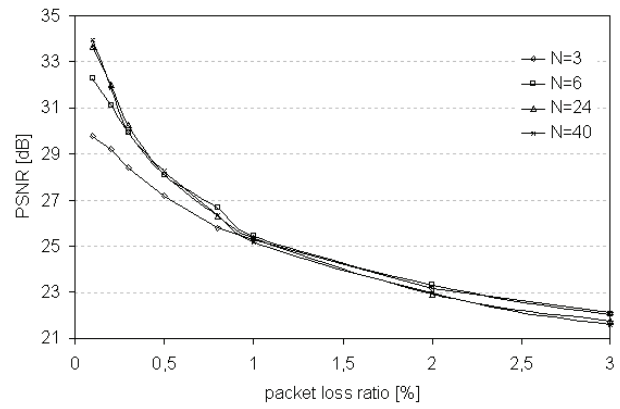

Fig. 4.   Measured video quality analyzing the quality distortion caused by packet losses

The differences between the PSNR values of the transmitted videos are bigger when the packet loss ratio is less then 1%, while above it the differences are less then 0.5dB. According to the obtained results, the coding efficiency is more important then the distortion effects of the error propagation. If the coding is not adaptive to the current link conditions, we propose to use bigger I frame intervals (N), because higher overall video quality will be obtained.

In some cases the expected packet loss rate is known, so the video coding parameters can be set according to the wireless channel conditions. We have also analyzed the quality of the video stream transmitted over a link with fixed packet loss rate. The obtained results are presented in Fig. 5.
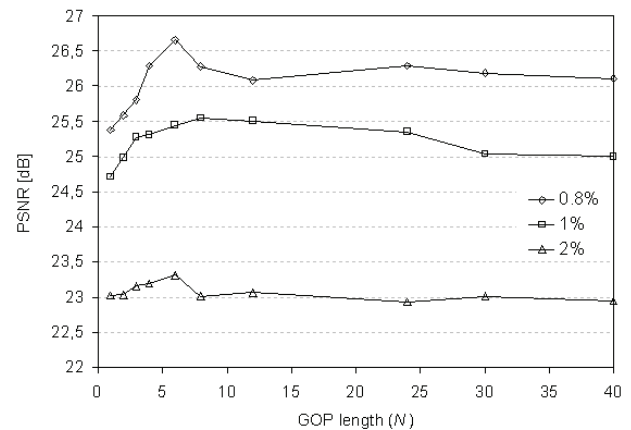

Fig. 5.   Measured video quality in function of GOP length

The figure shows that by increasing the GOP length the video quality first increases, but after reaching the highest PSNR value, it starts to decrease. In the first increasing period the efficient coding plays major role, so higher GOP lengths leads to better quality. After the optimal GOP settings, when the video quality is the highest, the error propagation effect becomes more significant. Hence, using higher I frame intervals, the error spreading will cause significant distortion.

In Fig. 5 our results show that the length of the optimal

GOP structure is never higher then 10. In most of the cases the optimal I frame interval setting was around $N=5$. However, as Fig. 5 shows the quality distortion due to error losses was less then 1dB when the examined video streams were transmitted over a lossy channel with fixed packet loss ratio.

Our joint analysis of error propagation and MPEG–4 coding proved that generally $N \geq 5$ I frame intervals should be used. Using this guidance, the coding efficiency is more beneficial then the distortion caused by error propagation, so the received video quality will be higher.

V. CONCLUSIONS

The increase of multimedia applications in mobile environment has placed new requirements on current video streaming solutions. The generally used MPEG video coding methods significantly reduce the bitrate of the transmitted video stream, but due to utilized dependencies between the frames the error propagation may become also notable.

In this paper we analyzed the distortion due to error propagation and the MPEG–4 coding efficiency together. We have modeled the error propagation and derived the expected number of infected frames due to the error spreading in the GOP. Besides analytical assumptions, we have performed numerous measurements. Based on our results, we proposed video coding guidelines for videos transmitted over lossy wireless links. Our measurement results showed that the coding efficiency is more beneficial then the distortion caused by error propagation. Hence, the GOP length should be increased to achieve higher streamed video quality improvement at the receiver.

ACKNOWLEDGMENT

REFERENCES

[1] L. Chiariglione. MPEG–4 FAQs. Technical report, ISO/IEQ JTC1/SC29/WG11, July 1997.

[2] International Organization for Standardization. Overview of the MPEG–4 Standard, December 1999.

[3] A. Albanese, J. Blomer, J. Edmonds, M. Luby, and M. Sudan. Priority encoding transmission. In Proc. 35th Ann. IEEE Symp. on Foundations of Computer Science, pages 604–612, November 1994.

[4] W. Heinzelman. Application–Specific Protocol Architectures for Wireless Networks. PhD thesis, Massachusetts Institute of Technology, June 2000.

[5] N. Feamster and H. Balakrishnan, "Packet Loss Recovery for Streaming Video," 12th International Packet VideoWorkshop, Apr. 2002

[6] T. Maugey, T. André, B. Pesquet–Popescu, J. Farah, "Analysis of Error Propagation due to frame losses in a distributed video coding system", EUSIPCO 2008, Lausanne, Switzerland, Aug. 2008

[7] Yung-Chi Chang, Chao-Chih Huang, Hao-Chieh Chang, Hung-Chi Fang, Liang-Gee Chen, "Error-Propagation Analysis and Concealment Strategy for MPEG-4 Video Bitstream with Data Partitioning," icme, pp.90, 2001 IEEE International Conference on Multimedia and Expo (ICME'01), 2001

[8] Cheng-Han Lin, Chih-Heng Ke, Ce-Kuen Shieh, Naveen K. Chilamkurti, "The Packet Loss Effect on MPEG Video Transmission in Wireless Networks", AINA, Vienna, Austria, 18-20 April 2006

[9] K. Mayer-Patel, L. Le, and G. Carle. An MPEG Performance Model and Its Application To Adaptive Forward Error Correction. In Proceedings of ACM Multimedia, December 2002.

[10] A. Dumitras and B. G. Haskell. I/P/B frame type decision by collinearity of displacements. In Proceedings of ICIP 2004, Singapore, Oct. 2004.

[11] Y. Yokoyama. Adaptive GOP structure selection for real-time MPEG-2 video encoding. In Proceedings of ICIP 2000, Vancouver, Canada, Sept. 2000.

[12] Huahui Wu, Mark Claypool, Robert E. Kinicki, "Guidelines for Selecting Practical MPEG Group of Pictures", EuroIMSA 2006, Innsbruck, Austria, 13-15 February 2006

[13] FFmpeg Project, http://ffmpeg.mplayerhq.hu/