Telecommun Syst (2009) 41: 65–76
DOI 10.1007/s11235-009-9151-3

Efficient error resilient algorithm for H.264/AVC: mobility management in wireless video streaming

Kostas E. Psannis · Yutaka Ishibashi

Published online: 15 April 2009
© Springer Science+Business Media, LLC 2009

Abstract The H.264/AVC standard introduces enhanced error robustness capabilities enabling resilient and reliable transmission of compressed video signals over wireless lossy packet networks. Those robustness capabilities are achieved by integrating some new error resilience tools that are essential for a proper delivery of real-time video services. Those tools include the Intra Refreshing (IR), Arbitrary Slice Ordering (ASO), Sequence Picture Parameter Sets (PPS), Redundant Slices (RS) tools and Flexible Macroblock Ordering (FMO). This paper presents an error resilient algorithm in wireless H.264/AVC streaming. The proposed method merges Reference Frame Selection (RFS), Intra Redundancy Slice and Adaptive Intra Refreshment techniques in order to prevent temporal error propagation in error-phone wireless video streaming. The coding standards only specify the decoding process and the bitstream syntax to allow considerable flexibility for the designers to optimize the encoder for coding performance improvement and complexity reduction. Performance evaluations demonstrate that the proposed encoding algorithm outperforms the conventional H.264/AVC standard. Both subjective and objective visual quality comparative study has been also carried out in order to validate the proposed approach. The proposed method can be used and integrated into H264/AVC without violating the standard.

K.E. Psannis (✉)
Department of Technology Management, University of Macedonia, Thessaloniki, Greece
e-mail: mobility2net@gmail.com

Y. Ishibashi
Department of Scientific and Engineering Simulation, Nagoya Institute of Technology, Nagoya 466-8555, Japan
e-mail: ishibasi@nitech.ac.jp

Keywords Error resilient · Encoding and transmission algorithm · Wireless video · H.264/AVC · Smoothness of the video

1 Introduction

Today's multimedia technology heralds an exciting era that will enormously impact daily life. On the other hand, the rapid growth of wireless communications and networking protocols will ultimately bring video to our lives anytime, anywhere, and on any device. Until this goal is achieved, wireless video delivery faces numerous challenges, among them highly dynamic network topology, high error rates, limited and unpredictably varying bit rates, and scarcity of battery power. Most emerging and future mobile client devices will significantly differ from those used for speech communications only; handheld devices will be equipped with color display and a camera, and have sufficient processing power to allows presentation, recording, and encoding/decoding of video sequences. In addition, emerging and future wireless systems will provide sufficient bit rates to support video communication applications. Nevertheless, bit rates will always be scarce in wireless transmission environments due to physical bandwidth and power limitations; thus, efficient video compression is required [1–4].

In the last decade video compression technologies have evolved in the series of MPEG-1, MPEG-2, MPEG-4 and H.264 [5]. Given a bandwidth of several hundred of kilobits per second, the recent codecs, such as H.264, can efficiently transmit quality video.

A video stream comprises Intra (I)-frames, Predicted (P)-frames, and interpolated-Bidirectional (B)-frames. According to H.264/AVC, I- , P- and B-frames have been extended with new coding features, which lead to a significant increase in coding efficiency. For example, H.264/AVC allows using more than one prior coded frame as a reference for P- and B-frames. Furthermore, in H.264/AVC, P- frames and B-frames can use prediction for subsequent frames. The H.264/AVC syntax permits the use of B-frames or P-frames as reference frames with the feature called stored B- or P-frames. These new features are described in detail in [1]. Moreover an H.264/AVC coded video sequence is typically partitioned into small intervals called GOP (Group Of Pictures).

In recent years several techniques for streaming wireless video have been devised [6–25]. A TCP friendly rate control which is equation based over wired networks have been proposed in [6, 7]. Several research techniques have been reported to improve the performance of TCP friendly rate control over wireless networks. These methods either hide end-hosts from packet loss caused by wireless channel error, or provide end hosts the ability to distinguish between packet caused by congestion, and that caused by wireless channel error [8]. One other approach that has been shown to improve wireless video quality by efficiently utilizing transmission energy is joint source coding and transmission power to different regions of a sequence based on their relative importance. For example, less power may be used to transmit packets in a static background region in order to allocate more power to regions of the sequence that are more difficult to conceal. Several research works on the region based rate control have been reported [9–13]. The Lagrange multiplier rate approach is employed for rate control in region-based coding [9]. Although the complexity of Lagrange multiplier was reduced it is still major concern in real time wireless video. Other methods [10–12] adopted a heuristic scheme to decide the quantization parameters for different region in a frame. These approaches lack a quantitative method to perform bit allocation among different regions. This may cause improper quantization parameters which imposes unreasonable bits used for different regions. Another region-based approach [13] adopts the most effective criteria as quantitative factors to directly control bit allocations among different regions and employs the priority concept to further adjust bit allocation under different channel conditions. Moreover recent research shows that video traffic under time varying channels may be modeled as multifractal cascade accounting for the multiscaling property of the traffic [14, 15]. A rate-distortion optimized packet scheduling algorithm has been proposed in [16], for streaming media by generating a number of nested sub streams, with more important streams embedding less important ones in a progressive manner. On the other hand, several works on H.264 Flexible Macroblock Ordering (FMO) techniques have been reported [17–22]. FMO is a very powerful tool for error resilience. For instance, slice groups can be constructed in such a way that, if one slice group is not available at the decoder, each lost Macroblock (MB) may be surrounded by macroblocks of other slice groups (above, below, right, and left). In that case, the missing Macroblock (MB) can be reconstructed in a very effective way using interpolation based on surrounding (available) sample values [17, 18]. In [19] the authors evaluated the H.264 codec specifically concerning the error resilience features available. Their analysis was centered on the types of error in wireless ad-hoc networks, which were modeled as random and burst packet losses. They suggest tuning the encoder according to the expected packet loss rates inside the network to increase the overall PSNR of the sequence. The authors also demonstrate that the chessboard like pattern FMO type achieves good PSNR recovery for different error bursts sizes. Another approach [20] tackles the problem of burst error in wireless video transmission using FMO. They propose an interleaving method where texture coefficients are sampled and grouped into several interleaved groups so that burst error cannot damage all groups from one block. Other method proposed an Explicit Spiral-Interleaved FMO technique, which improves video transmission qualities, under lossy transmission channels [21]. In [22] the authors present a three dimensional (3D) Flexible Macroblock Ordering (FMO) technique. The proposed 3D macroblocks allocation method results with PSNR gain of about 1–2 dB for broadcast video application for 10% packet loss.

H.264/AVC Redundant slices feature, is an error robustness feature allowing the encoder to send an extra representation of a picture region (typically at lower fidelity) that can be used if the primary representation is corrupted or lost. Redundant slices tool allow the insertion of primary slices and one or more additional secondary slices belongs to the original picture in the same bit stream. If a primary slice is affected by errors, it can be replaced by an error-free redundant one; otherwise the redundant slices are discarded. Moreover, messages to the decoder interleaved in the code stream containing supplemental enhancement information may contain further information about the bit stream, which can be utilized by an error concealment scheme. Coding of redundant slices may use different quantization parameters, different reference pictures, and different motion vectors than those used in the encoding of the primary slice. However, the parameters for encoding the redundant slices should be selected in such a way that there is no visual discrepancy between the primary and redundant slice representations [5]. Although redundant slices technique is part of the H.264/AVC standard, only few publications present actual results evaluating this powerful error resilience feature. In [23] the authors demonstrate how H.264/AVC redundant slices can be used to generate the Wyner-Ziv coding, and present simulation results to demonstrate the advantages of this method over traditional methods such as FEC. The Wyner-Ziv bit stream is decoded in order to recover the redundant video descriptions, which are used in lieu of portions lost from the original video signal due to channel errors. Another approach [24] employs redundant slices to protect important video frames, such as Intra-coded frames, by restricting the redundant slices to be identical copies of the primary slices. Alternatively [25] proposes H264/AVC standard features such as redundant slices and FMO to protect video packets, and demonstrate the advantage over FEC.

Another approach [26] represents an error resilient Packet Scheduling (PS) scheme based on the Reference Frame Selection (RFS) without feedback channel. The authors used the Gilbert channel model and as a result the case of burst packet loss is considered. The authors claim that by applying the delayed packet scheduling scheme in a GOP according to different delay requirements, the probability of simultaneous losses between different prediction chains can be reduced.

The paper is organized as follows. In Sect. 2 we analyze the effects of temporal error propagation imposed by the variable wireless channel conditions. Specifically extensive simulations have been carried out for different encoding schemes in order to verify the impact of the motion compensation reference frame in temporal error propagation under different packet loss rates. In Sect. 3 the preprocessing steps of the proposed approach is detailed. The proposed method merges Reference Frame Selection (RFS), Intra Redundancy Slice and Adaptive Intra Refreshment techniques in order to prevent temporal error propagation in error-phone wireless video streaming. Section 4 includes the performance evaluation of the proposed very efficient method. Both subjective and objective visual quality comparative study has been also carried out in order to validate the proposed approach. Moreover the coding efficiency of the different encoding schemes is analyzed. Section 5 concludes the paper with final observations.

2 Problem definition

Encoded video data consists of frames with different level of importance in terms of frame types Intra (I), Predicted (P) and Bidirectional (B). In addition there is a complex dependency relationship across the different video frames, and lost packets have impact on the video quality. Since mobility is expected of the wireless client, there is typically significant packet loss. If the packet loss occurs on an I-frame, it would affect all the P- and B-frames in a GOP that are predicted from the I-frame. In the case of P-frames packets loss the decoding distortion would continue to propagate until an I-frame is found. On the hand packet loss of B-frame limits the loss to that particular frame and does not result in error propagation.

The original H.264/AVC standard includes 3 profiles (baseline, main and extended), each having a different set of functionalities. The baseline profile is our main focus for this research, since it was designed primarily for low-cost applications which do not have a great amount of computational power. This profile is mainly used in mobile applications. For the baseline profile only I- and P-slices can be used and therefore only I- and P-macroblocks are supported. We consider only I- and P-frames in a GOP. Assume that N is the distance between two successive I-frames, defining a Group Of Pictures (GOP) and N_{GOP} is the total number of GOPs in the video sequence.

Figure 1 depicts three different encoding schemes for real time, low latency, applications with $N = 8$. Figure 1(a) depicts the conventional coding where the very first frame is Intra (I-) coded and each subsequent frame is inter-coded with the previous coded frame as its motion compensation reference. The number of I- and P-frames in a GOP can be defined as follows

$$I_{number} = 1,$$
$$P_{number} = N - 1$$

The Total Number (TN) of I- and P-frames in the video sequence can be computed as follows

$$\text{TN} = N_{\text{GOP}} + N_{\text{GOP}} * (N - 1)$$

Figure 1(b) and (c) depicts encoding patterns using different motion compensation reference frame through reference frame selection. Specifically, in Fig. 1(b), P_1 has a dependency to the preceding I_0, P_2 has a dependency to the

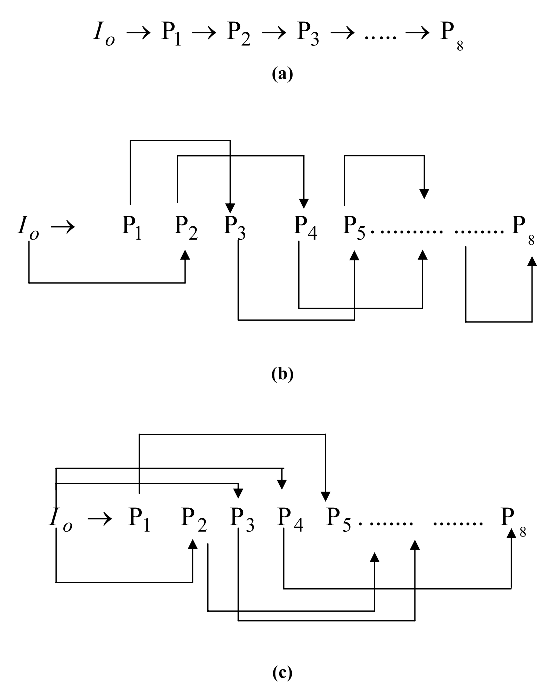


Fig. 1 Three different encoding schemes for H.264/AVC with $N = 8$. Arrows indicate motion compensation predictions relationships

preceding I_0, P_3 has a dependency to the preceding P_1, P_4 has a dependency to the preceding P_2, and for the final frame in a GOP, P_8 has a dependency to the preceding P_6. Furthermore, in Fig. 1(c), each of P_1, P_2, P_3 and P_4 frames have a dependency to the preceding I_0. P_5 has a dependency to the preceding P_1 and for the final frame in a GOP, P_8 has a dependency to the preceding P_4.

We adopt a modified version of the Decoded Frame Rate Metric [27, 28] in order to assess the effect of the temporal error propagation imposed by streaming H.264/AVC over error prone channels. The main modification is the use of the packet loss rate instead of frame loss for the different encoding schemes (Fig. 1). The extended successfully Decoded Frame Rate Metric (*DFRM*) can be defined as follows

$$DFRM = \frac{F_{dec}}{N_{\mathrm{GOP}} + N_{\mathrm{GOP}} * (N-1)}, \quad 0 \le DFRM \le 1$$

where F_{dec} is the summation of the successfully decoded I- and P-frames in the video sequence. $DFRM = 1$ implies no quality degradation and $DFRM = 0$ implies the most unpleasant case. It should be noted that I-frame in a GOP is successfully decoded if all the packets that belong to an I-frame received intact. On the other hand, P-frame is successfully decoded if the preceding motion compensation reference frame (I- or P-) has been successfully received.

Based on the probability theory, we compute for each frame (i.e., I- and P-) the probability to be successfully decoded, considering a p packet loss rate. We denote the C_I and C_p as the average number of packets in an I- and P-frame respectively [27].

Hence the number of the correctly decoded I- and P-frames in a GOP can be computed as follows for the three coding schemes (Fig. 1).

For the encoding scheme depicted in Fig. 1(a)

$$\begin{aligned}
S(I_0) &= (1-p)^{C_I}, \\
S(\mathrm{P}_1) &= (1-p)^{C_I} * (1-p)^{C_p} = S(I_0) * (1-p)^{C_p}, \\
S(\mathrm{P}_2) &= (1-p)^{C_I} * (1-p)^{C_p} * (1-p)^{C_p} \\
&= S(\mathrm{P}_1) * (1-p)^{C_p}, \\
S(\mathrm{P}_3) &= S(\mathrm{P}_2) * (1-p)^{C_p}, \\
&\vdots \\
S(\mathrm{P}_8) &= S(\mathrm{P}_7) * (1-p)^{C_p}
\end{aligned}$$

For the encoding scheme depicted in Fig. 1(b)

$$\begin{aligned}
S(I_0) &= (1-p)^{C_I}, \\
S(\mathrm{P}_1) &= (1-p)^{C_I} * (1-p)^{C_p} = S(I_0) * (1-p)^{C_p}, \\
S(\mathrm{P}_2) &= (1-p)^{C_I} * (1-p)^{C_p}, \\
S(\mathrm{P}_3) &= (1-p)^{C_I} * (1-p)^{C_p} * (1-p)^{C_p} \\
&= S(\mathrm{P}_1) * (1-p)^{C_p}, \\
S(\mathrm{P}_4) &= S(\mathrm{P}_2) * (1-p)^{C_p}, \\
&\vdots \\
S(\mathrm{P}_8) &= S(\mathrm{P}_6) * (1-p)^{C_p}
\end{aligned}$$

For the encoding scheme depicted in Fig. 1(c)

$$\begin{aligned}
S(I_0) &= (1-p)^{C_I}, \\
S(\mathrm{P}_1) &= (1-p)^{C_I} * (1-p)^{C_p} = S(I_0) * (1-p)^{C_p}, \\
S(\mathrm{P}_2) &= S(I_0) * (1-p)^{C_p}, \\
S(\mathrm{P}_3) &= S(I_0) * (1-p)^{C_p}, \\
S(\mathrm{P}_4) &= S(\mathrm{P}_0) * (1-p)^{C_p}, \\
&\vdots \\
S(\mathrm{P}_8) &= S(\mathrm{P}_4) * (1-p)^{C_p}
\end{aligned}$$

Consequently, the summation of the successfully decoded I- and P-frames for the entire video sequence is computed as follows

$$F_{dec} = \{S(I_0) + S(\mathrm{P}_1) + S(\mathrm{P}_2) + \cdots S(\mathrm{P}_{N-1})\} * N_{\mathrm{GOP}}$$

We simulate the scenario of the H.264-based video transmission for different encoding schemes and different packet loss rates. Simulations were done using the H.264 Test Model [36]. The test video sequence are in QCIF format (176×144 pixels/frame) and encoded at target frame rate 15 frames/s, the number of frames to be encoded/decoded is 150 frames. Frames were partitioned into slices and the slices are organized in packets for transmission where each slice is packed in one packet. We consider that a QCIF frame is constructed with nine slice groups and each of the nine slice groups consist of eleven Macroblocks (MBs). Assume that the maximum transmitting packet size is 1000 bytes. The packet loss situation is simulated according to the channel conditions specified in [26, 27]. Specifically we examined the packet loss rates from 5% up to 30% with step 5%. The simulations were carried out on the QCIF Foreman video trace at 144 kbps channel throughput under different packet loss rates obtained for error burst ranging from 5% to 30% frame loss (i.e. loss of 5 MBs to 30 MBs for a QCIF frame). We examine the effect of the error location by moving the starting point of the error by one Macroblock (MB) at time. This way we cover all missing slices possibilities due to an error burst with given length. Figure 2 depicts the percentage of the successfully Decoded Frame Rate Metric (DFRM) for the three encoding schemes (Fig. 1) under different packet loss rates.

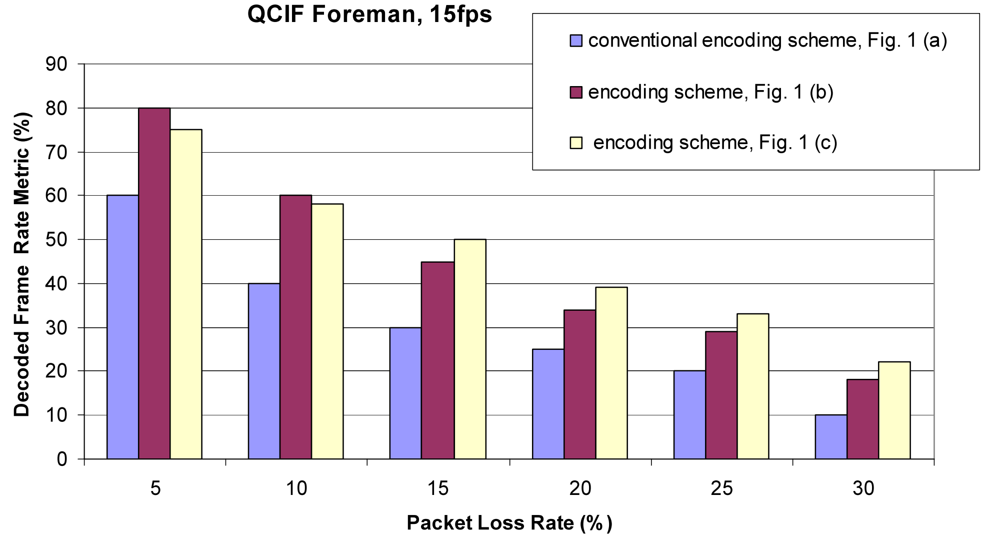


Fig. 2 Decoded Packet Rate Metric as a function of Packet Loss Rate

It can be observed that there is a significant increase in the successfully Decoded Frame Rate Metric (DFRM) for the encoding schemes (Fig. 1(b)–(c)) compared to the conventional encoding scheme (Fig. 1(a)) for different packet loss rates. Hence, it should be emphasized that the most important property to prevent temporal error propagation is the motion compensation reference frame.

3 Proposed method

The H.264/AVC standard introduces enhanced error robustness capabilities enabling resilient and reliable transmission of compressed video signals over wireless lossy packet networks. Those robustness capabilities are achieved by integrating some new error resilience tools that are essential for a proper delivery of real-time video services. Those tools include the Intra Refreshing (IR), Arbitrary Slice Ordering (ASO), Sequence Picture Parameter Sets (PPS), Redundant Slices (RS) tools and Flexible Macroblock Ordering (FMO). The transmission of compressed video streams over error-prone networks requires an encoder equipped with error resilient tools.

In order to prevent temporal error propagation imposed by wireless video streaming, we propose the use of differently encoded H.264/AVC sequence. The proposed method merges the proposed encoding pattern with frames protection. Frames protection employs Redundant Slices (RS) to protect important video frames, such as Intra-coded frames, by restricting the redundant slices to be identical copies of the primary slices. Moreover in order to further protect Predicted (P-) frames from temporal error propagation we use adaptive Intra Macroblocks Refreshment (IMR).

The coding standards only specify the decoding process and the bitstream syntax to allow considerable flexibility for the designers to optimize the encoder for coding performance improvement and complexity reduction. The corresponding encoding pattern is obtained by encoding the original uncompressed video data as follows

$$I_0 \mathrm{P}_1(1)\mathrm{P}_i(k)\cdots \mathrm{P}_{N-2}(k)\mathrm{P}_{N-1}(k)$$

subject to

$$2 \le i \le N-1,$$
$$1 \le k \le N-1$$

k denotes the Reference Frame Selection (RFS). $\mathrm{P}_i(k)$-frames are coded using motion estimation and each one has a dependency only to the preceding kth-frame. Effectively this results that loss packet of P-frame does not affect the next frame to be decoded. A $\mathrm{P}_i(k)$-frame is not a full frame. Instead, it is a Predictive video frame. This means that a $\mathrm{P}_i(k)$-frame only stores the data that has changed from the preceding kth-frame. Figure 3 depicts the proposed encoding pattern using motion compensation reference frame through reference frame selection (k). Arrows indicate motion compensation prediction relationships. For instance, for reference frame selection (RFS) $k=2$ (short term motion compensation reference), P_1 has a dependency to the preceding I_0, P_2 has a dependency to the preceding I_0, P_3 has a dependency to the preceding P_1, P_4 has a dependency to the preceding P_2, and for the final frame in a GOP, P_{14} has a dependency to the preceding P_{12}. Similarly for reference frame selection (RFS) $k=4$ (long term motion compensation reference), each of P_1, P_2, P_3 and P_4 frames have a dependency to the preceding I_0. P_5 has a dependency to the preceding P_1, P_6 has a dependency to the preceding P_2 and for the final frame in a GOP, P_{14} has a dependency to the preceding P_{10}. Note that for $k=1$ motion compensation reference frame implies the conventional H.264/AVC I-P structure.

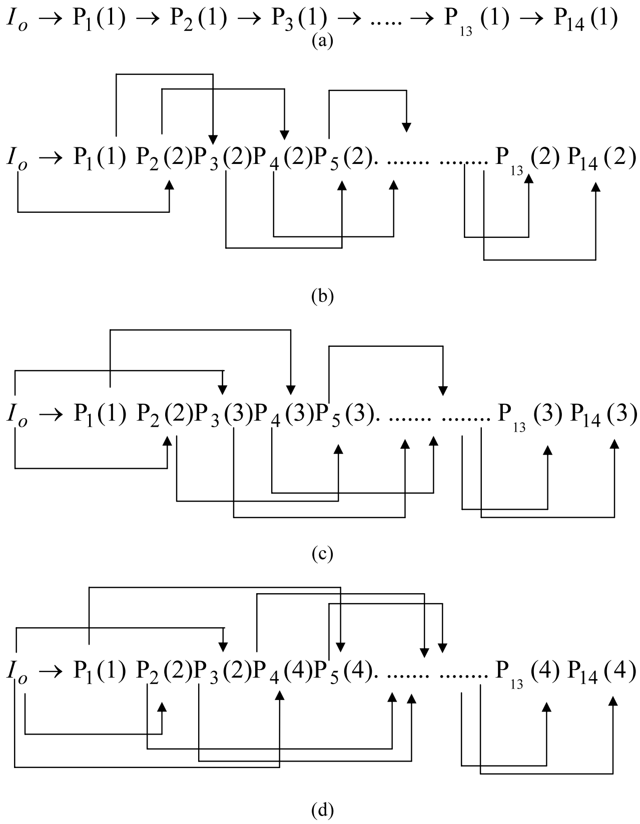


Fig. 3 Encoding pattern for GOP Length ($N = 15$) for different reference frame selection (k). Conventional H.264/AVC I-P structure, $k = 1$ (**a**). Short term motion compensation reference, $k = 2$ (**b**). Middle term motion compensation reference, $k = 3$ (**c**). Long term motion compensation reference, $k = 4$ (**d**)

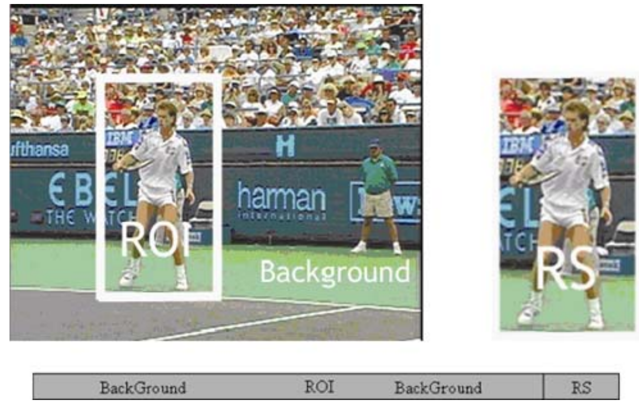


Fig. 4 Intra Frame with Redundant Slice Feature Transmission

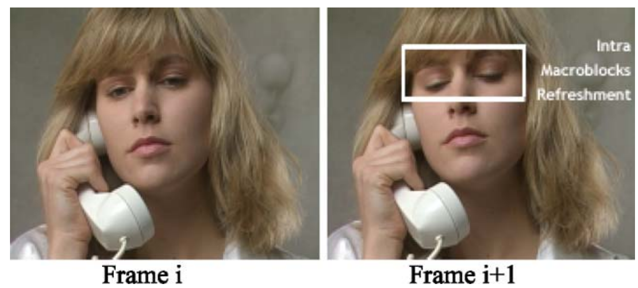


Fig. 5 Intra Macroblocks Refreshment (IMR) technique

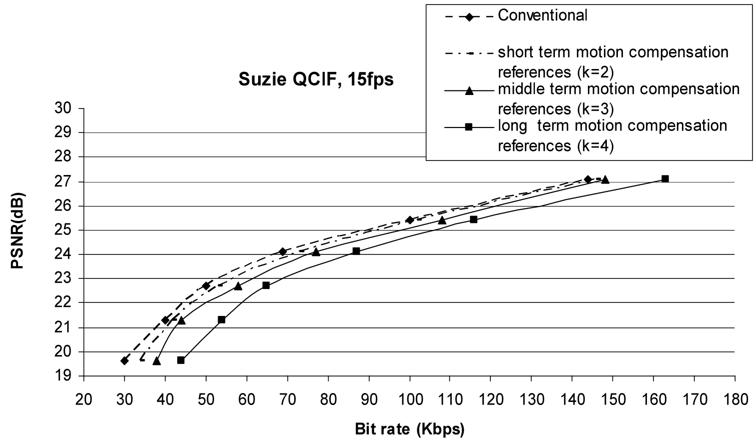
In order to minimize the effect of I-frame packets loss we proposed the use of Redundancy Slice (RS) for Intra (I-) frames. If a packet loss occurs on I-frame, it would affect all successive P-frames in a GOP that are predicted from the I-frame. Redundancy slices allow the encoder to include an additional representation of a slice with the original primary coded slice in the same bit stream. The parameters used for encoding the redundant slices may differ from those used for the primary slices, but they should be selected in such a way that there is no visual discrepancy between the two representations [5]. In order to highly protect the most important information in the Intra frame, some specific slices are dedicated for Region of Interest (ROI). Whenever any of the primary intra frames coded slices cannot correctly decoded the decoder can replace the primary slice with its corresponding redundant representation. The H.264/AVC standard states that redundant representation (RS in the figure) should follow the corresponding primary frame as depicted in Fig. 4.

Moreover in order to further prevent temporal error propagation we use adaptive Intra Macroblocks Refreshing (IMR) technique for the predicted frames $P_i(k)$. Specifically, only macroblocks with high activity (i.e., Macroblocks that are significantly different from those in the same position of the previous frame) are refreshed and coded as Intra Macroblocks (Fig. 5). In the H.264/AVC standard Intra Refreshing (IR) technique rate may be applied at the frame level or Macroblock (MB) level. At the frame level, the IR rate specifies the percentage of MBs to be Intra-coded within the frame. At the MB level, the IR rate defines a statistical probability that a particular MB is to be Intra-coded. Figure 5 depicts P- frame Macroblocks, which are significantly different from those in the same position of the previous frame, are Intra Coded.

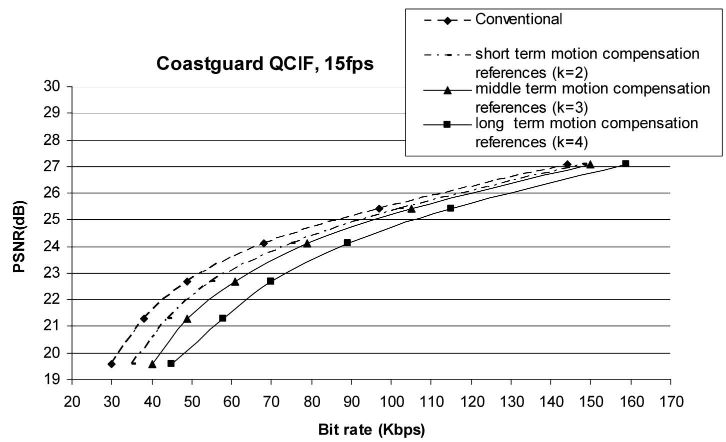
4 Performance evaluation

The proposed approach is evaluated and compared to H.264/AVC conventional scheme in terms of video quality and the ability to prevent temporal error propagation. We simulate the scenario of the H.264-based video transmission for different erroneous environments. Simulations were done using the H.264 Test Model (H.264/AVC Software Coordination, software version: JM 13.0) [36]. All test sequences are in QCIF format (176 × 144 pixels/frame) and encoded at target frame rate 15 frames/s, the number of frames to be encoded/decoded is 150 frames. The first frame in a GOP (Group of Pictures) is Intra (I-) frame

Fig. 6 The PSNR as a function of bit rate for various encoding schemes for the QCIF video traces (**a**) Suzie, (**b**) Coastguard, (**c**) Foreman and (**d**) Stefan



(a)

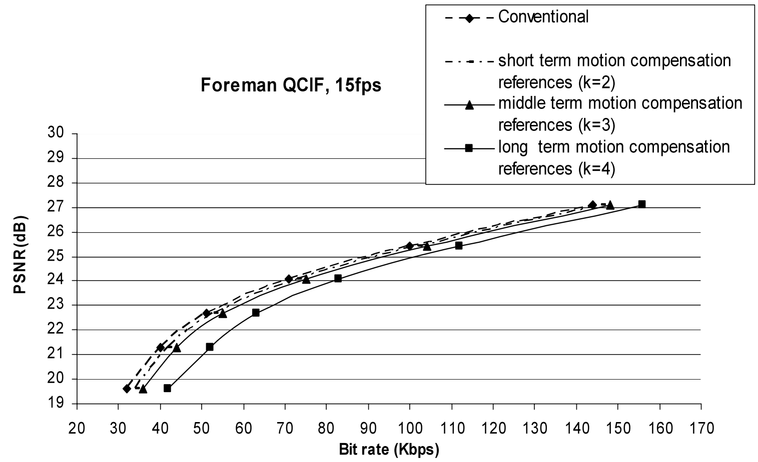


(b)

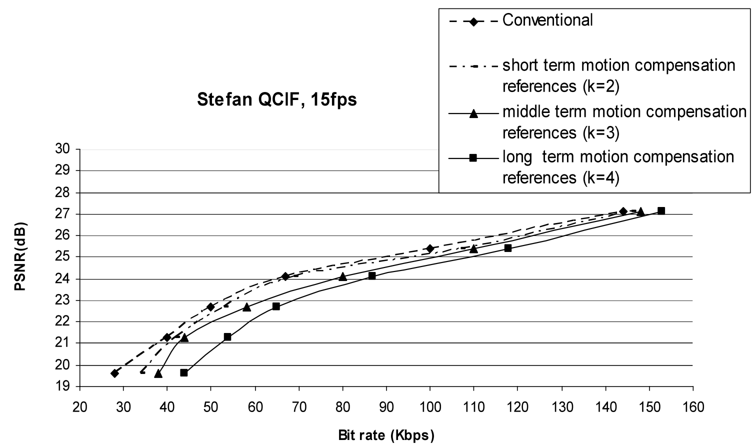
and the remaining consecutive frames are Predicted (P-) coded. Frames were partitioned into slices and the slices are organized in packets for transmission where each slice is packed in one packet. We consider that a QCIF frame is constructed with nine slice groups and each of the nine slice groups consist of eleven Macroblocks (MBs). The simulations were carried out on the following QCIF video traces: Suzie, Coastguard, Foreman and Stefan at 144 kbps channel throughput under different packet loss rates. The compression efficiency of the proposed encoding schemes is compared with the conversional encoding scheme for all the QCIF video traces. The encoding is performed using the H.264 reference encoder [36], at Variable Bit Rate (VBR) mode with constant frame rate 15 frames per second (fps). In the VBR mode the quantization parameters are maintained constant for the encoding process. Hence the video quality is almost sustained steady but the derived encoding bit rate fluctuates around a mean value. The impact of the encoding efficiency of the proposed encoding patterns is depicted in Fig. 6 for all the video traces (a) Suzie, (b) Coastguard (c) Foreman and (d) Stefan. Specifically Fig. 6 shows the PSNR values as a function of the bit rate. From Fig. 6, it is clearly observed that there is only a minimal bit rate overhead for the short term motion compensation reference ($k = 2$) and higher bit rate overhead for long term motion compensation reference ($k = 4$).

Video Communications are often afflicted by various forms of losses, such as burst packet loss or uniform packet loss. An understanding of the effect of packet loss on the reconstructed video quality, and developing accurate models for predicting the distortion for different loss events, is clearly very important for designing, analysing, and operating video communications systems over lossy networks. An important question is whether the expected distortion depends only on the average packet loss rate, or whether it also depends on the specific pattern of the loss. For example, does packet loss burst length matter, or is the resulting distortion equivalent to an equal number of isolated losses? Specifically in [29] the length of a burst loss was shown to have an important effect on the resulting distortions where longer burst length generally led to larger distortions. [30] provides a model that explains why a loss pattern, such as

Fig. 6 (*Continued*)



(c)



(d)

a burst loss, generally produces a larger distortion that an equal number of isolated losses. The packet loss situation is simulated according to the channel conditions specified in [26] and [27]. The average Y-PSNR (dB) values for the different packet loss rates obtained for error burst ranging from 0% to 30% frame loss (i.e. loss of 0 MBs to 30 MBs for a QCIF frame) are presented in Table 1.

It should be noted that as showed in Fig. 2, when packet loss rate is no higher than 10%, the encoding scheme with short term motion compensation reference {Fig. 1(b)} performs better than the encoding scheme with long term motion compensation reference {Fig. 1(c)}. On the other hand, if packet loss rate is higher than 10%, the encoding scheme depicted at Fig. 1(c) performs better than the encoding scheme depicted at Fig. 1(b). This is due to the mismatch between the short term and long term motion compensation reference which implies dissimilar performance for different packet loss rates.

Table 1 depicts the average Y-PSNR (dB) for the Suzie, Coastguard, Foreman and Stefan at 144 kbps channel throughput under the following packet loss rates, 0%, 5%, 15%, 20% and 30%. Packet Scheduling (PS) Approach represents an error resilient packet scheduling scheme based on the reference frame selection without back-channel [26], Reference Frame Selection (RFS) represents short ($k = 2$), middle ($k = 3$), long ($k = 4$) term motion compensation references. Frames Protection (FP) represents the Redundancy Slice (RS) for Intra (I-) frames and adaptive Intra Macroblocks Refreshing (IMR) technique for the Predicted $\{P_i(k)\}$ frames.

The largest Y-PSNR value in each column is shown in bold font. From the results shown in Table 1 the follow observations can be derived. Applying Frames Protection (FP) can improve error resilient for all the video traces and packet loss rates. All the cases with Frames Protection (FP) outperform the conventional H.264 coding standard and the Packet Scheduling (PS) approach. For instance, for packet loss rates 0% to 5% the short term motion compensation references ($k = 2$) with Frames Protection (FP) outperforms all the other FP representations, the short, middle, long terms motion compensation references without FP, the Packet Scheduling (PS) approach and the conventional H.264/AVC

Table 1 Average Y-PSNR (dB) of Coastguard (**a**), Suzie (**b**), Stefan (**c**), and Foreman (**d**) at 144 Kbps under different packet loss rates

Coding pattern	0%	5%	15%	20%	30%
(a)					
Conventional	27.12	26.24	25.08	24.11	22.88
PS Approach	26.81	26.41	25.42	25.16	23.72
RFS ($k=2$)	26.93	26.43	25.48	25.19	23.79
RFS ($k=3$)	26.88	26.44	25.51	25.22	23.81
RFS ($k=4$)	26.85	26.46	25.53	25.24	23.83
RFS ($k=2$) FP	**27.32**	**27.28**	26.35	26.06	24.65
RFS ($k=3$) FP	27.21	27.19	**26.76**	26.26	24.85
RFS ($k=4$) FP	27.18	27.16	26.43	**26.41**	**25.25**
(b)					
Conventional	27.11	26.26	25.17	24.16	23.01
PS Approach	26.71	26.37	25.41	25.17	23.78
RFS ($k=2$)	26.98	26.39	25.45	25.21	23.84
RFS ($k=3$)	26.88	26.43	25.49	25.25	23.88
RFS ($k=4$)	26.77	26.45	25.51	25.27	23.9
RFS ($k=2$) FP	**27.47**	**27.39**	26.88	26.28	24.91
RFS ($k=3$) FP	27.39	27.22	**26.96**	26.48	25.11
RFS ($k=4$) FP	27.28	27.19	26.92	**26.78**	**25.41**
(c)					
Conventional	27.13	26.25	25.09	24.12	22.89
PS Approach	26.68	26.71	26.27	25.85	24.51
RFS ($k=2$)	26.83	26.76	26.31	25.93	24.61
RFS ($k=3$)	26.77	26.79	26.61	26.23	24.91
RFS ($k=4$)	26.71	26.84	26.91	26.53	25.21
RFS ($k=2$) FP	**27.29**	**27.24**	27.01	26.63	25.31
RFS ($k=3$) FP	27.18	27.14	**27.09**	26.83	25.51
RFS ($k=4$) FP	27.14	27.08	27.03	**26.98**	**25.64**
(d)					
Conventional	27.14	26.29	25.2	24.19	23.04
PS Approach	26.62	26.61	26.12	25.74	24.31
RFS ($k=2$)	26.78	26.65	26.15	25.86	24.45
RFS ($k=3$)	26.71	26.69	26.45	26.16	24.75
RFS ($k=4$)	26.68	26.74	26.75	26.46	25.05
RFS ($k=2$) FP	**27.28**	**27.22**	26.95	26.66	25.25
RFS ($k=3$) FP	27.21	27.15	**27.08**	26.76	25.35
RFS ($k=4$) FP	27.17	27.01	27.02	**26.88**	**25.75**

coding. On the other hand, for packet loss rates 20% to 30% the long term motion compensation references ($k=4$) with Frames Protection (FP) outperforms all the other FP representations, the short, middle, long terms motion compensation references without FP, the Packet Scheduling (PS) approach and the conventional H.264/AVC coding. These results verify the effectiveness of the proposed approach compared to conventional H.264/AVC standard under different packet loss rates.

It should be noted that the Reference Frame Selection (RFS) mode can be applied both with and without feedback channel (channel which gives loss-packet report). The reference frame selection with feedback channel uses reference frame based on the feedback acknowledgement.

There are four possible modes defined in the reference frame mode [5]

1. Neither: no back-channel data is returned from the decoder to the encoder.
2. ACK: the decoder returns only acknowledgment messages. In this mode, the encoder uses only an acknowledged segment as a reference for inter- frame encoding.
3. NACK: the decoder returns only non-acknowledgment messages. Whenever NACK is received for a frame, the new reference frame is selected from amongst the frames coded before the NACK frame.
4. ACK + NACK: the decoder returns both acknowledgement and non acknowledgment messages. In the ACK + NACK mode, the encoder switches between two kinds of modes according to the upstream messages from the decoder.

Moreover RFS usually is adapted as an interactive error-resilience tool that depends on feedback channel. Generally the content of the feedback information may be, e.g. an ACK/NACK or a Channel Quality Indicator (CQI). On the other hand, the ACK/NACK feedback has to be avoided in many applications such as multicast or broadcast [26].

Channel Quality Indicator (CQI) is a measurement of the communication quality of wireless channels. CQI can be a value (or values) representing a measure of channel quality for a given channel. Typically, a high value of CQI is indicative of a channel with high quality and vice versa [31]. A CQI for a channel can be computed by making use of performance metric, such as a signal-to-noise ratio (SNR), signal-to-interference plus noise ratio (SINR), signal-to-noise plus distortion ratio (SNDR), and so forth of the channel. SNR determines several important attributes like packet loss that affects the network performance [32].

Specifically, based on a network monitoring system (i.e., packet loss range) we can further study the proposed encoding schemes. Assume that a network monitoring systems provides that the average packet loss range at the transmission network is for instance 0–5% then it can be predicted that the end user will experience acceptable video quality with short term motion compensation references ($k=2$) with Frames Protection (FP). On the other hand, assume that a network monitoring systems provides that the average packet loss range at the transmission network is for instance up 20% then it can be predicted that the end user will experience acceptable video quality with long term motion compensation references ($k=4$) with Frames Protection (FP).

It should be emphasized that in H.264/AVC standard the relation between the frames is content dependent and defined by the motion estimation algorithm [5]. In order to validate the efficiency of the proposed technique with the three different encoding patterns (short, middle, and long motion completion reference) extensive simulation were derived. Specifically four different QCIF video sequences, Suzie, Coastguard, Foreman and Stefan are used to evaluate the performance of the proposed scheme. These four sequences are chosen for their different texture complexity and motion activity. Suzie and Foreman are active sequences which include a moving background and a fair amount of motion of the foreground object. On contrary, Stefan and Coastguard sequences have rapid foreground motion with a fair amount of motion of the background object. From Table 1 and Fig. 6 it can be observed that for all the video sequences with different amount of spatiotemporal dynamics of the content the proposed encoding patterns outperform the conventional structure under different packet loss rates with minimal bit rate overhead.

The Structural SIMilarity (SSIM) index is a method for measuring the similarity between two images. The SSIM index is a full reference metric, in other words, the measuring of image quality based on an initial uncompressed or

Fig. 7 Error burst with 5% frame loss for Coastguard (*1*) and Suzie (*2*) video traces for the conventional H.264/AVC coding (1, 2)**a**, long term motion compensation references ($k = 4$) with FP (1, 2)**b** and short term motion compensation references ($k = 2$) with FP (1, 2)**c**

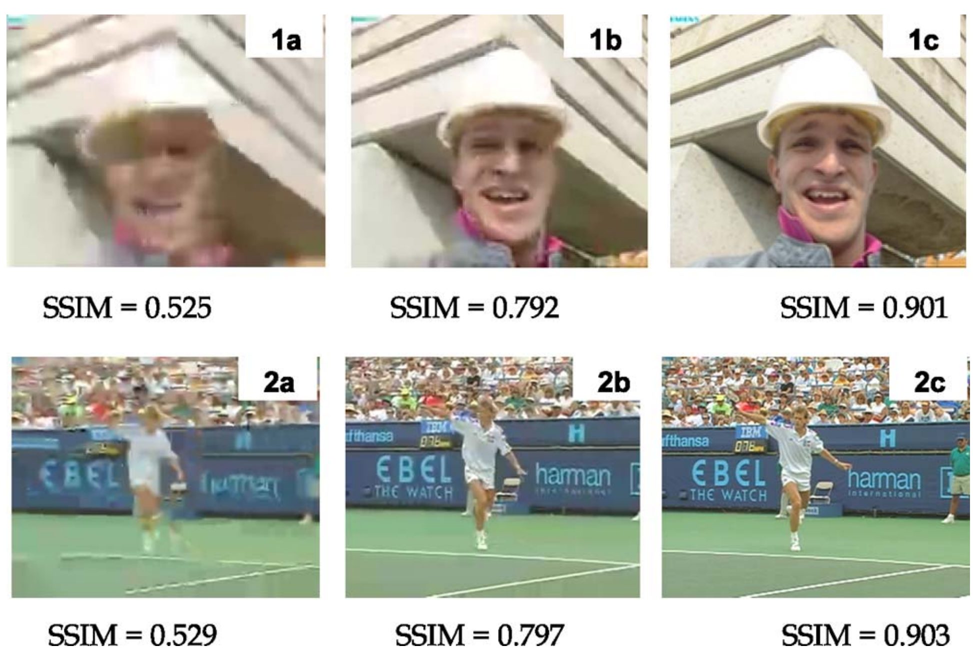


Fig. 8 Error burst with 30% frame loss for Foreman (*1*) and Stefan (*2*) video traces for the conventional H.264/AVC coding (1, 2)**a**, short term motion compensation references ($k = 2$) with FP (1, 2)**b** and long term motion compensation references ($k = 4$) with FP (1, 2)**c**

distortion-free image as reference. SSIM is designed to improve on traditional methods like PSNR and MSE, which have proved to be inconsistent with human eye perception. SSIM is also commonly used as a method of testing the quality of various lossy video compression methods. The SSIM emphasizes that the Human Visual System (HVS) is highly adapted to extract structural information from visual scenes. Therefore, a measurement of structural similarity (or difference) should provide a good approximation to perceptual image quality. The SSIM index is defined as a product of luminance, contrast and structural comparison functions. The SSIM index is a decimal value between 0 and 1. A value of 0 would mean zero correlation with the original image, and 1 means the exact same image. Through this index, image and video compression methods can be effectively compared [33–35].

The four QCIF video traces affected by burst error, 5% frame loss for the Coastguard and Suzie video traces and 30% frame loss for the Foreman and Stefan video traces. Figures 7 and 8 demonstrate the visual effect and the SSIM value. It can be seen that for different burst error the proposed approach outperforms the conventional H.264/AVC standard since it achieves better visual quality.

5 Conclusions

Error resilient is a key technique that enables robust streaming of stored video content over wireless lossy networks. It is particularly useful when content has been produced independently of the transmission network conditions. In this paper, we analyzed the constraints of temporal error propagation in error-prone video communications. In order to overcome the constraints we proposed the use of error resilient algorithms in H.264/AVC wireless streaming. The proposed method merges Reference Frame Selection (RFS), Intra Redundancy Slice and Adaptive Intra Refreshment techniques. Both subjective and objective visual quality comparative study demonstrates that the proposed encoding algorithm outperforms the conventional H.264/AVC standard. Moreover the coding efficiency of the proposed technique is evaluated. These results verify the effectiveness of the proposed approach. Future work will include the impact of the proposed encoding algorithm in the decoding efficiency, delay and random bit errors. It should be emphasized that if adaptive switching and VCR-like interactive functions were carried out the error resiliency performance of the proposed encoding pattern could be better enhanced. This is also a prospective research direction.

Acknowledgement This paper was supported in part by International Information Science Foundation (IISF), Japan (Grant. no. 2006.1.3.916).

References

1. Psannis, K., & Ishibashi, Y. (2008). Efficient flexible macroblock ordering technique. *IEICE Transactions on Communications*, *E91-B*(08), 2692–2701.
2. Psannis, K., Hadjinicolaou, M. G., & Krikelis, A. (2006). MPEG-2 streaming of full interactive content. *IEEE Transactions on Circuits and Systems for Video Technology*, *16*(2), 280–285.
3. Psannis, K., Ishibashi, Y., & Hadjinicolaou, M. G. (2007). QoS for wireless interactive multimedia streaming. In *Proceedings of the 3rd ACM workshop on QoS and security for wireless and mobile networks 2007*, Chania, Crete Island, Greece (pp. 168–171) October 22–24, 2007.
4. Psannis, K., Hadjinicolaou, M., & Ishibashi, Y. (2006). Efficient support of wireless video multicast services in 3G and beyond, advances in computer, information, and systems sciences, and engineering. In K. Elleithy, T. Sobh, A. Mahmood, M. Iskander & M. Karim (Eds.), *Springer signals and communications*, ISBN-13: 978-1-4020-5260-6 (pp. 121–127). Berlin: Springer.
5. Advanced Video Coding for Generic Audiovisual Services (2007). ITU-T Rec. H.264 and ISO/IEC 14496-10 (MPEG-4 AVC), ITU-T and ISO/IEC JTC 1, Version 1: May 2003, Version 2: May 2004, Version 3: Mar. 2005, Version 4: Sept. 2005, Version 5 and Version 6: June 2006, Version 7: Apr. 2007, Version 8 (including SVC extension): Consented in July 2007.
6. Floyd, S., Handley, M., Padhye, J., & Widmer, J. (2000). Equation-based congestion control for unicast applications. In *Proc. ACM SIGCOMM*, Stockholm, Sweden (pp. 43–56) Aug. 2000.
7. Floyd, S., & Fall, K. (1999). Promoting the use of end-to-end congestion control in the Internet. *IEEE/ACM Transactions on Networking*, *7*(4), 458–472.
8. Chen, M., & Zakhor, A. (2006). Multiple TFRC connections based rate control for wireless networks. *IEEE Transactions on Multimedia*, *8*(5), 1045–1062.
9. Song, H., & Jay Kuo, C.-C. (2004). A region-based H. 263+ codec and its rate control for low VBR video. *IEEE Transactions Multimedia*, *6*(3), 489–500.
10. Lin, C.-H., & Wu, J.-L. (1997). Content-based rate control scheme for very low bit-rate video coding. *IEEE Transactions Consumer Electronics*, *43*(2), 123–133.
11. Aramvith, S., Kortrakulkij, H., Tancharoen, D., & Jitapankul, S. (2002). Joing source-channel coding using simplified block-based segmentation and content-based rate-control for wireless video transport. In *Proc. int. conf. information technology: coding and computing (ITCC) 2002*, Las Vegas (pp. 71–76) Apr. 2002.
12. Lin, C.-W., Chang, Y.-J., & Chen, Y.-C. (2003). A low-complexity face-assisted coding scheme for low-bit-rate video telephony. *IEICE Transactions on Information and Systems*, *E86-D*(1), 101–108.
13. Sun, Y., Ahmad, I., Li, D., & Zhang, Y.-Q. (2006). Region-based rate control and bit allocation for wireless video transmission. *IEEE Transactions on Multimedia*, *8*(1), 1–10.
14. Melo, C. A. V., & Fonseca, N. L. S. (2004). An envelope process s for multi-fractal traffic modeling. In *Proc. IEEE int. conf. communication (ICC) 2004* (Vol. 4, pp. 2168–2173) June 20–24, 2004.
15. Erramilli, A., Narayan, O., Neidhardt, A., & Saniee, I. (2001). Multi-scaling models of TCP/IP and sub-frame VBR video traffic. *Journal of Communications and Networks*, *3*(4), 383–395.
16. Kang, S. H., & Zakhor, A. (2005). Effective bandwidth based scheduling for streaming multimedia. *IEEE Transactions on Multimedia*, *7*(6), 1139–1148.
17. Kwon, S.-K., Tamhankar, A., & Rao, K. R. (2006). Overview of H.264/MPEG-4 part 10. *Elsevier Journal of Visual Communication and Image Representation*, *17*(2), 186–216.
18. Lambert, P., De Neve, W., Dhondt, Y., & Van de Walle, R. (2006). Flexible macroblock ordering in H.264/AVC. *Elsevier Journal of*

Visual Communication and Image Representation, *17*(2), 358–375.
19. Calafate, C. M., & Malumbres, M. P. (2003). *Testing the H.264 error-resilience on wireless ad-hoc networks*. Paper presented at the EURASIP video/image processing and multimedia communications conference, 2003.
20. Kim, M., Lee, S.-W., & Kim, S.-D. (2005). Error resilient texture coding scheme for wireless video transmission based on coefficient sampling and interleaving. *SPIE Optical Engineering, 44*.
21. Katz, B., Greenberg, S., Yarkoni, N., Blaunstien, N., & Giladi, R. (2007). New error-resilient scheme based on FMO and dynamic redundant slices allocation for wireless video transmission. *IEEE Transactions on Broadcasting*, *53*(1), 308–329.
22. Ogunfunmi, T., & Huang, W. C. (2005). A new flexible macro block ordering with 3D MBAmap for H.264/AVC. In *IEEE ISCAS* 2005.
23. Rane, S., & Girod, B. (2005). Systematic lossy error protection of video based on H.264/AVC redundant slices. In *Proc. visual communication and image processing VCIP-2006*, San Jose, CA, Jan. 2005.
24. Wang, Y.-K., Hannuksela, M. M., & Gabbouj, M. (2003). Error resilient video coding using unequally protected key pictures. In *Proc. international workshop VLBV03*, Sep. 2003.
25. Baccichet, P., Rane, S., & Girod, B. (2006). Systematic lossy error protection based on H.264/AVC redundant slices and flexible macroblock ordering. In *Proc. 15th international packet videoworkshop*, Hangzhou, P.R. China, Apr. 2006.
26. Liu, H., Zhang, W., & Yang, X. (2006). Error-resilience packet scheduling for low bit-rate video streaming over wireless channels. In *IEEE international symposium on circuits and systems*, 2006.
27. Koumaras, H., Kourtis, A., Lin, C.-H., & Shieh, C.-K. (2007). A theoretical framework for end-to-end video quality prediction of MPEG-based sequences. In *The third inter. conf. on networking and services—ICNS07*, Athens, Greece, June 19–25, 2007.
28. Ziviani, A., Wolfinger, B. E., de Rezende, J. F., Carlos, O., Duarte, M. B., & Fdida, S. (2005). Joint adoption of QoS schemes for MPEG streams. In *Multimedia tools and applications*, April, 2005.
29. Liang, Y. J., Apostolopoulos, J. G., & Girod, B. (2003). Analysis of packet loss for compressed video: does burst-length matter? In *IEEE international conference on acoustics, speech, and signal processing* (pp. 684–687) 2003.
30. Apostolopoulos, J. G. (2001). Reliable video communication over lossy packet networks using multiple state encoding and path diversity. In *VCIP* (pp. 392–409). Bellingham: SPIE Press.
31. Hu, J., Choudhury, S., & Gibson, J. D. (2007). H.264 Video over 802.11a WLANs with multipath fading: parameter interplay and delivered quality. In *IEEE international conference on multimedia and expo* (pp. 1790–1793) 2007.
32. Guha, R. K., & Sarkar, S. (2006). Characterizing temporal SNR variation in 802.11 networks. In *IEEE wireless communications and networking conference* (Vol. 3, pp. 1408–1413) 2006.
33. Koumaras, H., Kourtis, A., & Martakos, D. (2005). Evaluation of video quality based on objectively estimated metric. *Journal of Communications and Networking*, *7*(3), 235–242.
34. Koumaras, H., Kourtis, A., Martakos, D., & Lauterjung, J. (2007). Quantified PQoS assessment based on fast estimation of the spatial and temporal activity level. *Multimedia Tools and Applications*, *34*(3), 355–374.
35. Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, *13*(4), 600–612.
36. H.264/AVC Software Coordination (2009). Software version: JM 13.0. http://iphome.hhi.de/suehring/tml/.

Kostas E. Psannis was born in Thessaloniki, Greece. He was awarded, in the year 2006, a research grant by International Information Science Foundation (Grant No. 2006.1.3.916), Japan. Since 2004 he has been a (Visiting) Assistant Professor in the Department of Technology Management, University of Macedonia, Greece. Since 2005 he has been a (Visiting) Assistant Professor in the Department of Computer Engineering and Telecommunications, University of Western Macedonia, Greece. From 2002 to 2004, he was a Visiting Postdoctoral Research Fellow in the Department of Computer Science and Engineering, Graduate School of Engineering, Nagoya Institute of Technology, Japan. He has extensive research, development, and consulting experience in the area of telecommunications technologies. Since 1999, he has participated in several R&D funded projects as a Research Assistant in the Department of Electronic and Computer Engineering, School of Engineering and Design, Brunel University, UK. From 2001 to 2002 he was awarded the British Chevening scholarship sponsored by the British Government. He has several publications in Conferences and peer-reviewed journals. He received a degree in physics from Aristotle University of Thessaloniki (Greece), and the Ph.D. degree from the Department of Electronic and Computer Engineering of Brunel University (UK). He is a Member of the IEEE, ACM and IET.

Yutaka Ishibashi received the B.S., M.S., and Ph.D. degrees from Nagoya Institute of Technology, Nagoya, Japan, in 1981, 1983, and 1990, respectively. From 1983 to 1993, he was with NTT Laboratories. In 1993, as an Associate Professor, he joined Nagoya Institute of Technology, in which he is now a Professor in the Department of Scientific and Engineering Simulation, Graduate School of Engineering. From June 2000 to March 2001, he was a Visiting Professor in the Department of Computer Science and Engineering at the University of South Florida. His research interests include networked multimedia applications, media synchronization algorithms, and QoS control. He is a Member of the IEEE, ACM, Information Processing Society of Japan, the Institute of Image Information and Television Engineers, and the Virtual Reality Society of Japan.