RESILIENT VIDEO STREAM SWITCHING FOR MOBILE WIRELESS CHANNELS

MOHAMMAD ALTAF

COMSATS Institute of Information Technology, Attock, Pakistan
mohammadaltaf@gmail.com

MARTIN FLEURY

University of Essex, Colchester, United Kingdom
fleum@essex.ac.uk

MOHAMMAD GHANBARI

University of Essex, Colchester, United Kingdom
ghan@essex.ac.uk

In this paper, several error-resiliency techniques are combined with the H.264/AVC (Advanced Video Coding) codec's switching frames to adaptively switch between video streams, depending on wireless channel conditions. Switching frames are a feature of H.264/AVC that allows smooth transitions between streams without the overhead of periodic intra-coded I-frames. Because video streaming over bandwidth-limited mobile networks requires higher compression ratios, when 'lossy' channel conditions occur it is advisable to provide error resiliency at the application layer in order to avoid degradation of video quality at the receiver. In the scheme introduced in this paper, when a change in channel conditions occurs, protection to the video stream is provided by switching to an alternative stream with error resiliency protection. To accomplish stream switching to an error-resilient video stream, minimal feedback is necessary. In this way, robust streaming is confined to periods of poor channel quality, which results in up to 3-4 dB increase in video quality (PSNR) compared to using a single robust scheme over the entire session irrespective of channel conditions. In particular, this scheme appears suited to conditions of slow fading, caused by changes in the environment as a mobile device user moves from one location to another. In the paper, the response of different error resiliency techniques to error patterns (isolated or 'bursty') is determined in order to adapt in a suitable way to channel conditions. Constant Bit Rate switching frames are also implemented in this paper. Compared to the usual Variable Bit Rate switching frames, CBR frames allow low-latency and low-bandwidth video services to be supported by H.264/AVC switching frames. The robust switching stream scheme can be potentially combined with adaptive stream bitrate switching. However, the main gain at low bitrates comes not from the reduced overhead of embedded switching frames but the increase in error robustness.

Key words: Error concealment, error resiliency, switching frames, video streaming
Communicated by: (to be filled by the JMM editorial)

1 Introduction

Maintaining an acceptable video quality for a packet-switched streaming service over bandwidth limited, time varying, and error-prone wireless channels is a challenging problem. Potentially, mobile networks allow users to stream video over the Internet through a variety of wireless networks [2], either 3G cellular, broadband access, or wireless LAN. In such environments the packet loss rate may be high, particularly in cellular networks [12] and of a 'bursty' nature due to fading [7], shadowing, and bandwidth variations over time. Video streaming is a real-time service, typically requiring frames to be displayed at a rate of 25 or 30 frame/s. Consequently, the increased latency resulting from the retransmission of lost packets is not an attractive form of error protection. This is particularly so because heavy packet loss rates over mobile wireless channels would require frequent retransmissions of lost packets. In fact, because video coding exploits temporal redundancy through motion estimation between frames [6], the loss of a single packet can cause drift between the encoder's and decoder's view of the compressed stream until the decoder is next resynchronized. The result is a period of degraded video quality. Fast fading caused by multipath reflections in a built environment may be counteracted in part by forward error correction (FEC) at the physical (PHY) layer. However, PHY layer FEC may not be sufficient by itself and in this paper isolated errors are further counteracted by error resilience at the application layer, though the response is dependent on stream switching latency. To counteract slow fading caused by changes in the environment as a mobile device user moves from one location to another, the scheme proposed in this paper allows the video stream to be switched to a version with protection from burst errors. Thus because video streams are especially sensitive to errors, this paper's scheme provides additional protection over and above the normal PHY layer FEC.

Fortunately, in the H.264/Advanced Video Coding (AVC) codec a range of application-layer error resilience measures have been provided for wireless communication [18] and this paper's scheme exploits these facilities. Error resilience techniques are source coding solutions to the problem of transmission error that act independently of channel coding, as suggested by Shannon's source-channel separation theory. However, as increased error robustness and high compression are two opposing concepts, increasing one may decrease the other [9]. Therefore, in an adaptive solution, the application of error resiliency should be limited to periods of poor channel conditions in order to increase the average quality of the overall streaming session.

The principal contribution of this paper is adaptive switching of video streams according to channel conditions, as each stream can provide appropriate error resilience according to the channel conditions. In other words, a stream is coded with various degrees of error resilience (refer forward to Figure 1) and, as the channel conditions deteriorate, through a feedback request, a more error resilient stream is requested. An already received video stream at one level of error resilience will be switched to another of another level of error robustness. To provide smooth switching between the streams we employ switching frames [8], which are a built-in feature of an H.264/AVC codec. The switching frame configuration for each stream is set in the configuration file of the reference JM codec. To allow switching between Constant Bit Rate streams, the codec was altered in the way described in Section 4.4. Tests used Flexible Macroblock Ordering, as described in Section 5.1, to provide a more error resilient bitstream, after experiments in Section 5 to decide the most suitable means of error resilience. Feedback messages at each switching interval were delivered by a routine method developed for the Third Generation Partnership Project (3GPP), as described in Section 2. The switching interval is described in the next paragraph. Previous work such as [21] has concentrated on switching as a way of varying the bitrate to avoid congestion, while this paper proposes switching for an entirely different purpose, namely error protection. In a further departure from previous usage [19], switching frames are implemented by us for Constant Bit Rate (CBR) video. Unlike Variable Bit Rate (VBR) the Quantization Parameter (QP) of CBR video changes over time. Additionally, CBR switching frames employ two QPs rather than the usual one QP. Therefore, selection of the appropriate QP of the switching frames at a macroblock level for CBR switching is a more difficult procedure compared to VBR switching.

As previously mentioned, in the proposed scheme, a server stores different pre-coded versions of the same video sequence (clip or film) with varying degrees of error robustness (including no error robustness) and adaptively selects from these according to channel conditions. Accompanying pre-coded H.264/AVC secondary switching frames [8] are also generated for insertion in the sequence, should the need arise. Therefore, switching frames are the mechanism of adaptation to wireless channel conditions. Just as for Intra-coded (I)-frames, secondary switching frames also allow the decoder to resynchronize in the event of packet loss, thus helping to prevent error drift. Pre-coding switching frames does not result in much storage overhead, as these frames take up a relatively small fraction of the compressed video storage. Additionally, if switching frames were to be created dynamically this would cause delay which would restrict the applications they could be used for. The overhead is governed by the frequency of switching frames. This must be offset against the maximum tolerable latency before a switch to a more robust version of the stream occurs. For example, at 30 frame/s (Hz), the minimum latency is 33 ms, but if switching frames are placed realistically every eight frames, then the maximum latency is 264 ms. However, this is still half the usual latency of 500 ms when the intra-refresh period is 15 frames for a frame rate of 30 Hz. Further consideration of these implementation issues is beyond the scope of this paper.

The remainder of this paper is organized as follows. Section 2 considers previous research in this area. Section 3 is an overview of the scheme before further detailing the adaptive switching scheme in Section 4. Section 5 describes the video error resiliency techniques that form the basis of robust switching. Section 6 presents the results of evaluating the scheme. Finally, Section 7 summarizes and indicates further lines of research.

2 Related work

The adaptive element of the packet-switched streaming (PSS) system (audio and video) for the Third Generation Partnership Project (3GPP) is described in [5]. However, adaptation of 3GPP PSS is aimed at changing the throughput according to: changing the capacity across links formed by different wireless technologies, e.g. from WCDMA to GPRS (intersystem handover); changing cell user population; and lowered bandwidth as a result of channel conditions. Therefore, this generic system using a feedback channel responds indirectly to channel conditions by reducing the bitrate rather than directly by increasing the stream protection. The main aim of adaptation in that scheme [5] is to avoid receiving device buffer underflow or overflow, as well as to avoid exceeding the link bitrate. As such the system in our paper is complementary to 3GPP PSS as it adds robustness and additionally it could provide smooth bitrate transitions, without the need for extensive buffering. The main weakness of increasing buffer sizes in mobile devices is the impact on energy consumption, though this is small in comparison to power amplification.

In [17], refinements to the generic 3GPP PSS system are reported. A frame type prioritization scheme is introduced so that if the bandwidth drops or there is an outage the more important frames for decoding are available first. It is also possible to drop non-referenced predictively-coded (P)-frames in H.264/AVC if the available bandwidth requires this in a form of temporal scalability. The work in [19] adds H.264/AVC switching frames to the temporal scalability scheme for GPRS and EGPRS cellular wireless systems. In addition, H.264/AVC generalized bi-predictive (B)-frames are added as a further temporal scalability feature, though, in fact, B-frames are not supported in 3GPP's adoption of the H.264/AVC Baseline Profile. This is a VBR streaming system which relies on comprehensive feedback information, rather than the minimal feedback messages required by our scheme. It also differs from this paper's scheme as it relies on bitrate switching and does not include robustness switching. Consequently, this scheme [19] is also complementary to the one introduced in this paper.

Research in [2] was concerned with how to provide a seamless multicast streaming service in the event of a vertical handoff from one wireless technology to another. Multi-homed devices allow continuous wireless coverage when a currently available wireless network becomes unavailable. The paper [2] contains a detailed consideration of the signaling process that is required so that when a vertical handoff is anticipated streams of packets are duplicated at each network interface. As such this type of adaptive streaming would also complement our scheme.

3. Overview of the scheme

In the proposed scheme, a server will initially start streaming assuming the channel to be error-free. As no error-resilience is applied, the reconstructed video is of better quality than the robust stream versions at the same data-rate, because there is no coding overhead arising from resilience. At times when the channel is poor, with a high packet loss rate, the server will switch to a stream with higher robustness at the next switching point using a switching frame. At times when there are no packet losses, it is possible to switch back to the stream without error resiliency (or one with a reduced level of error resilience) with higher resulting video quality. In a demonstration of the scheme in which switching takes place between a 'no error resilience' stream and a stream with strong error resilience, simulations results showed (Section 6.3) that robust switching results in up to 3–4 dB improvement in Peak Signal-to-Noise Ratio (PSNR) for both isolated and burst packet losses compared to using error resiliency throughout without stream switching. Estimation of the channel conditions is commonly specified in standards documents. For example, in IEEE 802.16e (mobile WiMAX) [14] a mobile subscriber station can provide a received signal strength indicator to the base station based on the power level or it can estimate the carrier to interference noise level by taking measurements of the modulation preamble or of pilot tones.

The scheme will find immediate application to streaming over wireless networks, as existing simulcast schemes such as 'SureStream' of RealNetworks and 'Intelligent Streaming' from Windows Media employ inserted I-frames for switching and are mainly intended for the wired Internet in which traffic congestion rather than 'lossy' channels are the principal threat. In fact, simulcast is a bitrate switching scheme and does not involve changes in error robustness, whereas our scheme could potentially change both the bitrate and provide differing levels of error resilience. As previously remarked, an adaptive switching frame solution also brings greater compression efficiency compared to I-frame switching. However, the main gain at low bitrates potentially comes not from the reduced overhead of embedded switching frames but the increase in error robustness.

4. Adaptive video stream switching

4.1 Video background

H.264/AVC consists of a Video Coding Layer (VCL) and a Network Abstraction Layer (NAL). The VCL is responsible for generating the source encoded bit streams and has several techniques or tools for error resiliency [23], while the NAL adapts the bit stream for network transport. Video coding firstly takes place at the macroblock level but then macroblocks are consolidated into slices which contain entropy decoder resynchronization markers. Therefore, slices are the simplest form of error resilience. Each slice is placed in a NAL unit prior to encapsulation in a Real-Time Transport Protocol (RTP) header under IP/UDP transport. Feedback commands might be encapsulated in Real Time Control Protocol (RTCP) packets running on top of UDP [4].

In H.264/AVC, a new type of frame, namely Switching Predictive/Intra frame (SP/SI-frame), has been defined for switching purposes. There are two types of SP-frame, namely primary and secondary SP-frames. In this paper, Primary SP frames are denoted as 'PSP-frames' and Secondary or Switching SP frame are denoted as 'SSP-frames'. The intra-coded version of the SSP frame will be called an SI-frame, while 'switching frames' will signify the overall concept. An SI-frame does not reference a previous frame as it does not use predictive coding, while an SSP-frame does require a reference frame. Therefore, in the event of a NAK feedback message, a robust option is to use an SI-frame to switch streams to prevent any possibility of error drift. Using an SI-frame also avoids problems in feedback delay, were a message detailing which packets had been lost to be held up.

H.264/AVC switching frames are currently supported by the Extended Profile and not the low-complexity Baseline Profile. However, as hardware implementations of H.264/AVC were quickly implemented [13], the authors expect that low-power Application Specific Integrated Circuit (ASIC) decoders supporting switching frames will be implemented should the need arise. Other error resilience techniques such as Flexible Macroblock Ordering (FMO) [11] (see Section 5.1) are already part of the Baseline profile.

4.2. Robust switching-frame system

Figure 1 is a diagram of the proposed system. At the streaming server, a set of pre-encoded videos are stored with varying degrees of error protection through resiliency. After adaptively switched selection of the stream, generic PHY layer FEC is applied before packetization and transmission over the wireless packet loss channel. For simplicity of evaluation, this paper assumes a single wireless link employed for messaging. After transmission over a potentially 'lossy' channel, PHY layer channel decoding takes place. Error detection of remaining errors then occurs typically through the use of Cyclic Redundancy Checks (CRC). At this point in time, packets are declared in error. If a packet is in error then the process of error concealment (Error Con. in Figure 1) takes place at the decoder, making use of a previously decoded frame. Otherwise, the usual decoding processes of variable length decoding, dequantisation (Deq.) and inverse transformation applied to the residual predication data. If a frame is inter-coded then motion compensation prediction (MCP) from a previously decoded frame takes place using motion vectors to match the decoded residual data with a previously decoded frame. However, the main point as far as this paper is concerned is that upon detection of packet loss a feedback route exists to the server in the event of a stream switch being required. An application-layer Negative Acknowledgment (NAK) from the receiving device to the server is sent. In order to avoid delay, it is normal in an IP network to transport video streams through unreliable UDP, without packet retransmission.

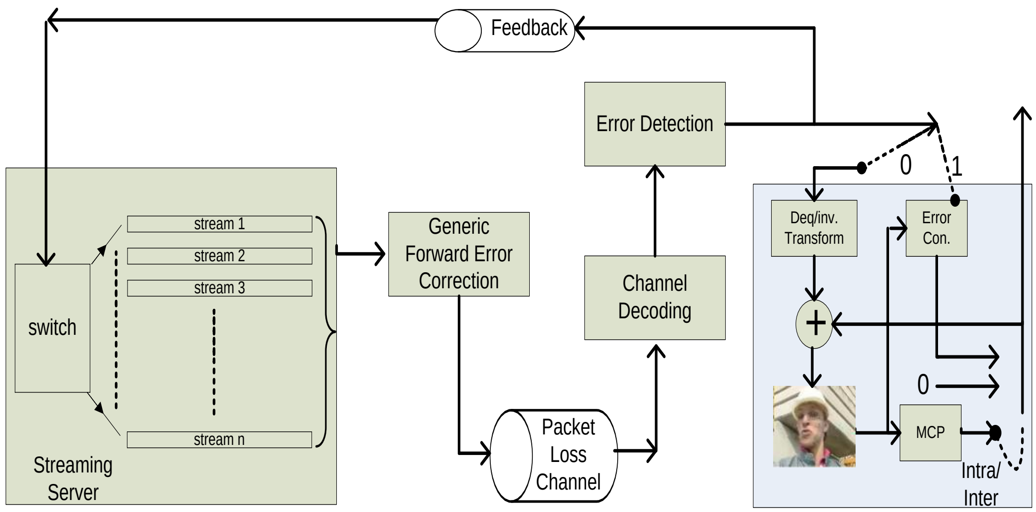


Figure 1 Switched video stream system for packet loss wireless channels

4.3. Switching frame procedure

In the tests in this paper, SI-frames have been used. However, if it is possible to specify which frame has been lost then a SSP-frame could be used instead, as it can be reconstructed from a different P-frame (not one that has been lost). Because we consider error bursts in which SSP-frames are unlikely to be of use (as multiple P-frames could be lost), this paper confines itself to testing with SI-frames. Though in some circumstances depending on choice of QP an SI-frame can be coded more efficiently than an I-frame, it is probably best to regard their coding penalty as essentially equivalent. The main reduction in overhead from using switching frames rather than I-frames is that PSP-frames can be coded significantly more efficiently, as they exploit temporal redundancy, which is not available to I-frames. Therefore, switching frame solutions increase coding efficiency in poor channel conditions over simulcast with I-frames (refer back to Section 4.1). The stream switching mechanism is now described in more detail.

PSP-frames can be inserted at various pre-determined periodic locations in the frame sequence. SSP- or SI-frames (or both), which are both a prediction mismatch-free version of PSP-frames, are created ready to be used should the need arise. If switching becomes necessary, an SSP/SI-frame is transmitted at the switching instance, replacing the PSP-frame at that position. In the event of one or more packet losses in a normal video stream without switching, the loss of synchronization normally results in drift error until the next synchronization point, which is either the next I-frame or could occur over a period of time if gradual decoder refresh is in place (refer to Section 5.2). However, if a feedback channel is available the decoder can signal the presence of error to the server, and an SI-frame can be transmitted without the need for I-frame synchronization.

In Figure 2, to enable drift-free switching, the streaming server stores the same sequences encoded either at different datarates or in this paper's scheme with different levels of error resilience or with a combination of both provisions. These bitstreams are populated with PSP-frames at the locations where switching is allowed, as shown in Figure 2. Notice though that Figure 2 is an illustrative example only and the periodicity of PSP frames can vary just as the I-frames they replace can. The central arrowed line in Figure 2 indicates that transmission starts with bitstream one and that all frames before the second PSP-frame of bitstream one are transmitted, followed by an SI-frame. From then onwards the rest of the transmitted frames are from bitstream two, omitting the PSP-frame in bitstream two, as the SI-frame has substituted for it. Therefore, in Figure 2 the bitstream two data from the start to the second PSP frame is never transmitted.

Of course, multiple switching sequences could take place in practice, switching back and forth between several streams. For simplicity Figure 2 shows only two streams. To achieve switching in both directions switching frames for both switching directions need to be generated beforehand, unless dynamic switching is implemented. The SI-frame shown in Figure 2 is the secondary representation of the corresponding PSP-frame in bitstream two. However, the SI-frame is generated for bitstream one rather than bitstream two. Thus it is generated as part of the same encoding process for the corresponding PSP-frame in bitstream one. As the SI-frame in Figure 2 is not predictively-coded from a prior P-frame, the result is that error propagation does not occur.

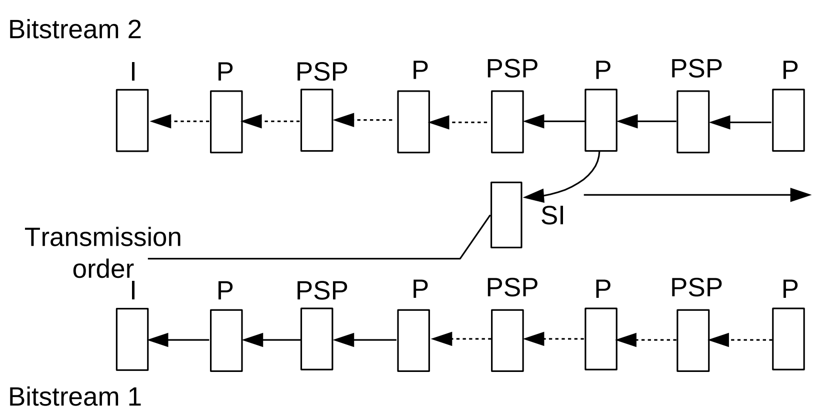


Figure 2 Switching between streams using SI-frames, showing transmission order and predictive dependencies between successive frames

Various implementations of switching back to a non-error resilient stream are possible depending on how frequently packet-loss NAKs are received. A server could either switch back if feedback is not received before the expiry of a pre-defined threshold time or switch back after the reception of an ACK command following the expiry of some pre-defined threshold time without any packet loss recorded at the receiving device. The former technique can be used to minimize the feedback traffic, as only one message is needed per switch and consequently in this paper the former method was used.

4.4. CBR switching

Switching frames are employed in this paper for error resiliency purposes. However, these frames are normally applied to VBR-encoded video, which limits their deployment, especially in delay-sensitive services, as outlined in Section 1. In this paper, the emphasis is on CBR-encoded video. Therefore, a CBR counterpart of VBR switching frames has been introduced in prior work [1] by two of the authors.

A PSP-frame is quantized with two quantization parameters: the first quantization parameter (QSP) is employed for coding prediction error and the second quantization parameter (QSSP) in coding predicted blocks before forwarding them to the reference frame buffer. This second quantization parameter enables drift-free reconstruction of a PSP-frame by an SSP-frame from different reference frames or an SI-frame without any reference frame at all. As employing two quantization parameters increases the quantization errors the quality of PSP reconstructed frames is usually lower than that of P-frames, but the coding overhead is less than that of I-frames.

QSP is the main quantization parameter of the PSP-frame, similar to the QP in other frames. Therefore, its behavior is straightforward and is not changed for CBR encoded video. The QSP is selected by the encoder according to the encoder rate distortion optimization (RDO) tools while the QSSP is taken as a function of the QSP for each macroblock as in (1)

$$\begin{aligned} QSP &= QP \\ QSSP &= QSP - j \end{aligned} \tag{1}$$

where QP is the quantization parameter of a predictive (P-) frame and *j* is any integer ranging from 0 to QSP. In this way, both the QSP and QSSP are selected without sacrificing the varying quantization choices for CBR encoding or RDO. In tests in Section 6 as SI frames are employed, j was set to one so that QSP and QSSP were close together.

5. Error resiliency and error concealment

Adaptive stream switching using various error resilience techniques and this Section outlines the methods that were applied in tests.

5.1. *Flexible Macroblock Ordering*

For FMO error resilience [11], compressed frame data is normally split into a number of slices each consisting of a set of macroblocks. In H.264/AVC, by varying the way in which the macroblocks are assigned to a slice (or rather group of slices), FMO gives a way of reconstructing a frame even if one or more slices are lost. The checkerboard FMO type does *not* employ adjacent macroblocks as coding references, which decreases its compression efficiency and the relative video quality after decode. However, if there are safely decoded macroblocks in the vicinity of a lost packet error concealment can be applied. For predictively coded P- or B-frames error concealment of lost macroblocks would normally occur through interpolation of the motion vectors of adjacent macroblocks. Because an IPPP… GoP structure is employed in tests then most error concealment occurs in this manner. However, if I-frames need to be reconstructed then spatial interpolation will be necessary. Further description of error concealment is given in Section 5.3.

5.2. *Other types of error resilience*

Prior to FMO slicing in H.264/AVC, slice structuring was also possible but this could not occur without breaching the raster-scan order of macroblock formation. This simple slicing scheme maintains the syntactic and semantic resynchronization information in slice headers but without the macroblock assignment mapping overhead from FMO. Simple slicing is still available in H.264/AVC [20].

Data partitioning is an error resilience scheme in H.264/AVC [12] that separates the compressed bitstream of P- and B-pictures into: A) configuration data and motion vectors; B) intra-coded transform coefficients; and C) inter-coded coefficients. This data forms A, B, and C partitions which are packetized as three separate NAL units. This arrangement allows a frame to be effectively reconstructed even if the inter-coded macroblocks in partition C. are lost, provided the motion vectors in partition A survive. Partition A is normally strongly FEC-protected and in our simulations this was assumed.

To reduce error propagation over time within predictively-coded P-frames it is possible to include some intra-coded MBs, as well as the usual inter-coded macroblocks and some SKIP macroblocks. When no I-frames are provided as anchor frames, this procedure can act as a form of gradual decoder refresh, though an appropriate refresh pattern should be chosen. In the JM implementation of H.264/AVC, intra-coded macroblocks can be placed randomly as a percentage of the whole. Alternatively in our simulations, each row of macroblocks within a slice was intra-coded in turn in a rotating order on a frame-by-frame basis.

5.3. *Error concealment*

As introduced in Section 5.1, error concealment, which is a non-normative feature of H.264/AVC [22], allows superior reconstruction of video in the event of packet loss, especially if checkerboard FMO error resiliency is used. For error concealment in H.264/AVC of predictively-coded frames, the motion vectors of correctly received slices (or prior concealed slices) are used in boundary matching motion vector recovery [10] if the average motion activity is sufficient (more than a quarter pixel). Research in [22] gives details of which motion vector to select to give the smoothest block transition. It is also possible to select the intra-coded frame method of spatial interpolation for predictively coded frames as well as their obligatory use for intra-coded frames. Spatial interpolation provides smooth and consistent edges at an increased computational cost through weighted pixel-value averaging [16]. In our tests, according to measured picture continuity, the error concealment method that best reduces 'blockiness' at macroblock boundaries was dynamically selected for predictively coded frames by the decoder.

6. Evaluation of robust switching

Tests were carried out based on the Foreman and Coastguard sequences. Foreman is a typical sequence taken from a handheld camera as might appear in YouTube, with a close-up sequence followed by a rapid pan. Coastguard is a view of a speeding coastguard boat as might be taken from a surveillance camera. Both the sequences are of Quarter Common Intermediate Format (QCIF) resolution ((176×144 pixel/frame) at a frame rate of 15 Hz for simplicity of testing. The CBR target bitrate was 64 kbps, which is a typical rate for 3G systems [18].

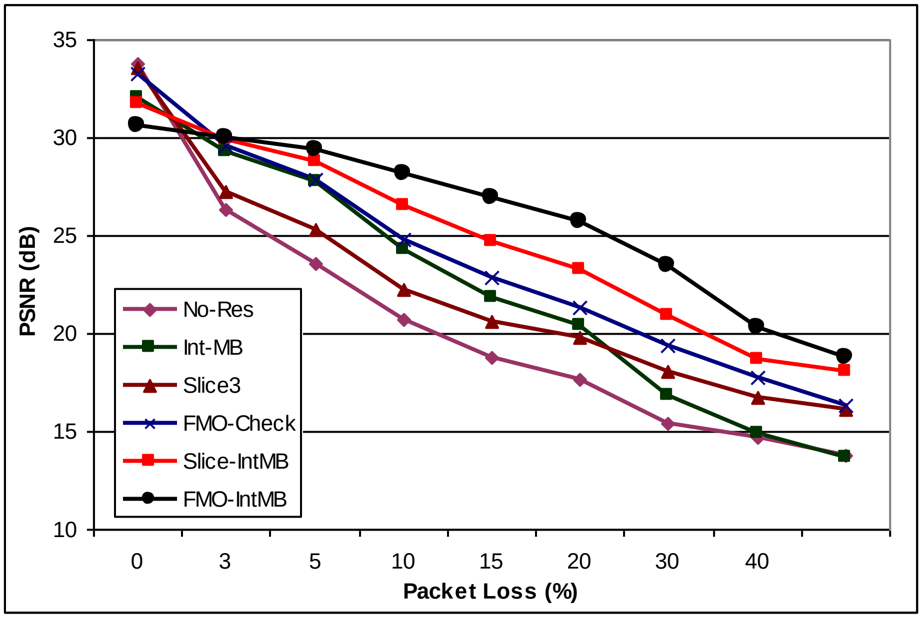
For simulation purposes, the feedback time threshold before switching back to a non-error resilient stream was taken to be five frames or 330 ms for a sequence with 15 fps, though in an implementation this value could be decided by the service provider. When there was no error resiliency or error resiliency without switching then the frame structure was IPPPPP…………, that is an initial I-frame followed by all P-frames. When switching was used a PSP-frame replaced a P-frame at every 10th location. It is normal not to include periodic I-frames over wireless links [18] but instead it is usual to employ conditional intra-coded rows of macroblocks within slices (refer to Section 5.2) to avoid sudden changes to the instantaneous bitrate and possible delay.

All data points in the following experiments are the arithmetic mean of 100 runs to ensure convergence. Preliminary tests assessed the relative merits of the error resilience techniques discussed in Section 5, before selection of the stream types in robust streaming.

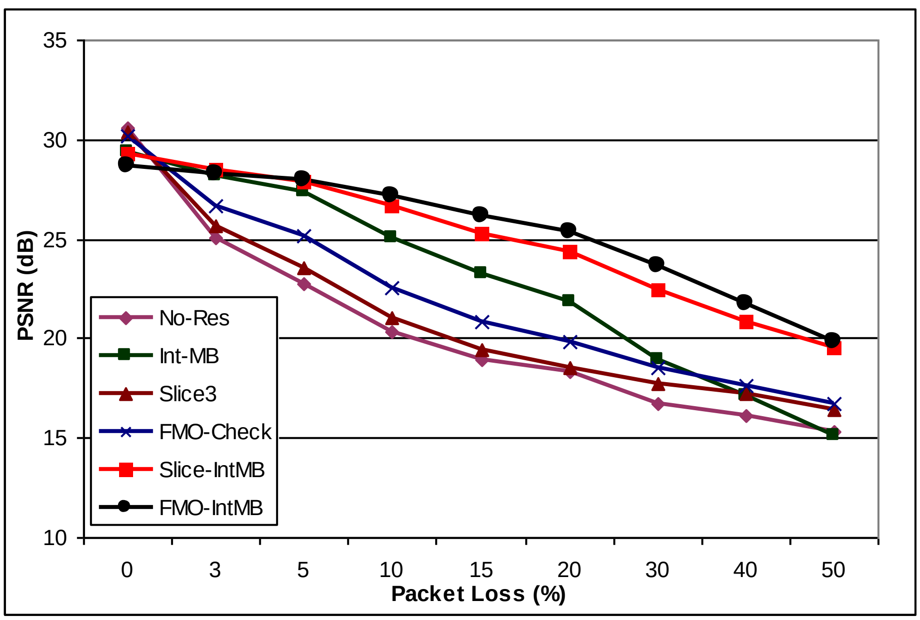
6.1. *Response to isolated errors*

In Figure 3, these techniques are compared against no error resilience (No-Res) for isolated losses (generated from a Uniform distribution up to a given percentage of packet losses). Clearly, no error resilience is inferior except when there is a zero packet loss rate, as then the lack of coding overhead for an equivalent datarate results in good quality. Notice that PSNR between 25 and 30 dB is weaker than would normally be accepted in wired Internet streaming but that this quality is generally acceptable [15] to users of mobile devices. Below 25 dB plots are included to show trends but would normally be unacceptable for viewing. Notice that in Figure 3 the packet loss percentage rather than the data lost is reported. Therefore, less data is lost in a scheme with smaller packets, though it is actually the affect on decoder synchronization of packet loss rather than quantity of data lost that is in general important in decoding of compressed data. The intention of Figure 3 is to show the effect of packet loss numbers on practical schemes as differing packet sizes undoubtedly occur in these schemes.

When error resilience is added one technique at a time then simple three slices per frame generally results in the worst quality. Adding intra-coded macroblocks (Int-MB) results in an improvement with two slice checkerboard FMO (FMO-Check) being better still. Two-slice checkerboard FMO combined with intra-coded macroblocks has the best performance followed by three slices combined with intra-coded macroblocks. Again ordering at zero packet loss reflects the relative coding overhead from each technique. In general, it is best to combine at least two error resilience techniques. For isolated errors, it is important to notice that in these circumstances utilizing checkerboard FMO with larger packet sizes is better than smaller packet sizes with more slices.



(a)



(b)

Figure 3 Different error resiliency schemes with isolated errors losses for (a) Foreman, (b) Coastguard video sequences

Notice that PSNR figures alone may exaggerate the gain from error resiliency, as visual inspection suggests that most of the distortion is in areas that it is difficult for error concealment to reconstruct accurately, i.e. movement of the head in the "Foreman" sequence shown in Figure 4. Here, the "Foreman" sequence is shown for the streams without any error resilience and with FMO combined with intra MB refresh for 20% isolated packet loss. From Figure 3, the version with FMO combined with intra MB refresh has a PSNR of just over 25 dB.

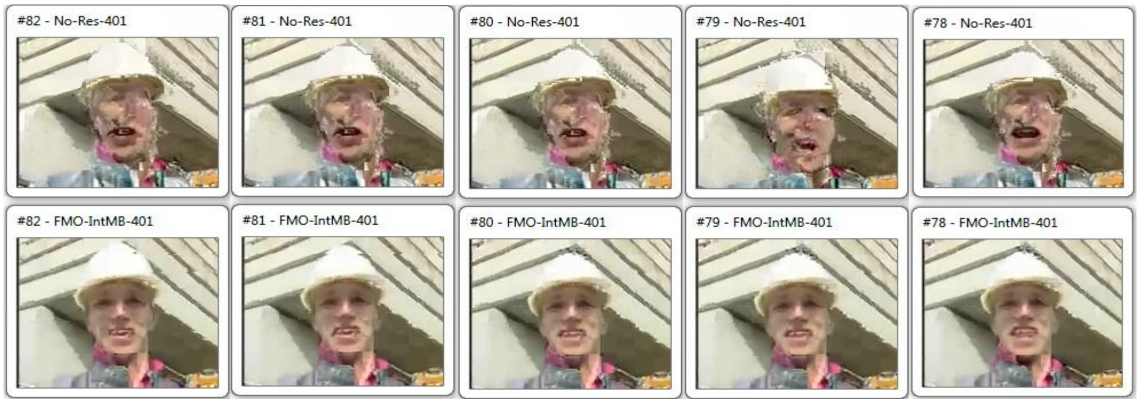


Figure 4 Visual comparison of No-res and FMO-IntMB for selected frames of the "Foreman" sequence.

A check was also made that the checkerboard FMO pattern of those supported in H.264/AVC is indeed superior to other patterns in these circumstances. As an example, Figure 5 shows the relative advantage of FMO-disp (dispersed — which is the name given to checkerboard FMO in the JM implementation of the codec). As it is not directly relevant, a description of the other patterns is not given here, with further information available in [11].

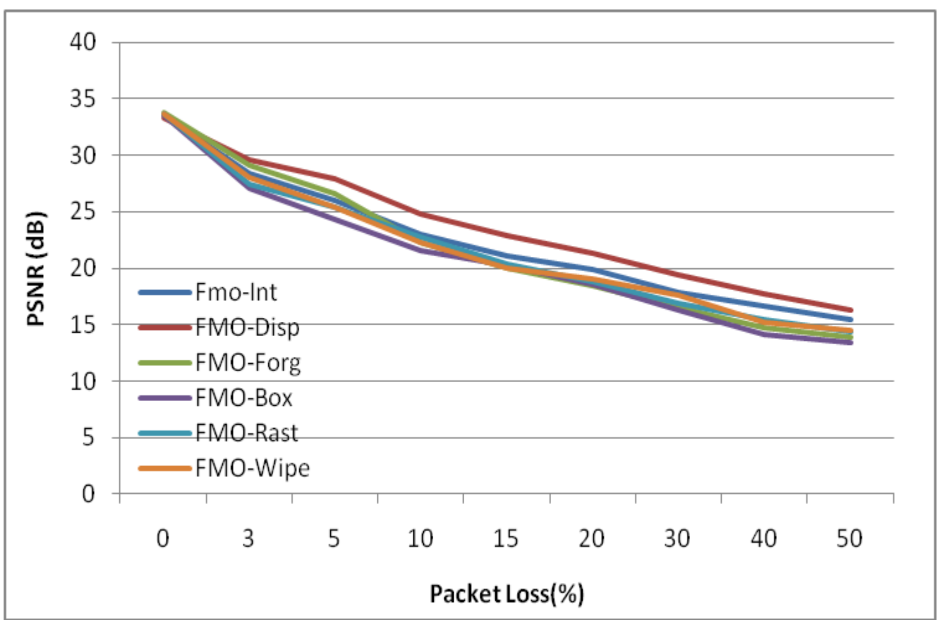


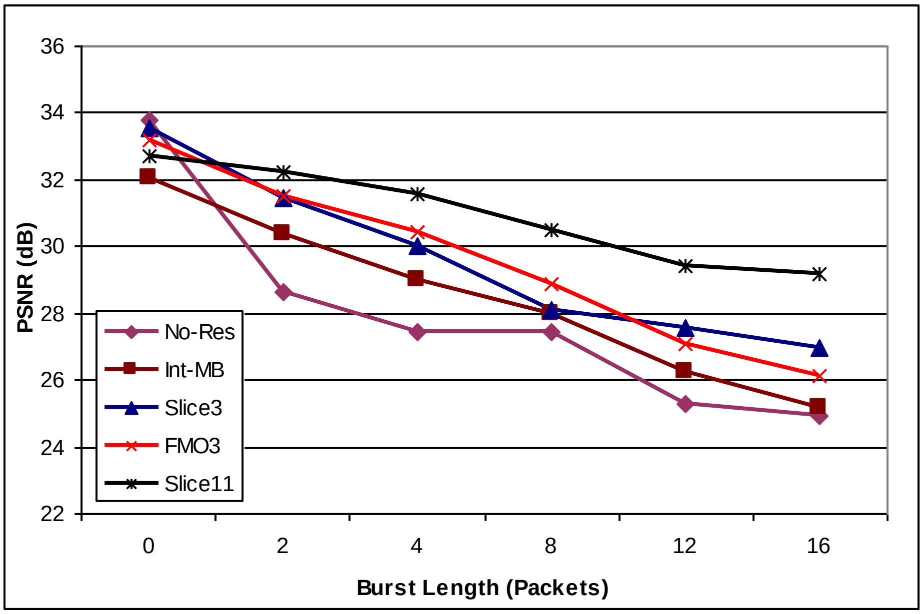
Figure 5 Different FMO schemes with isolated losses for the Foreman sequence

6.2. *Response to burst errors*

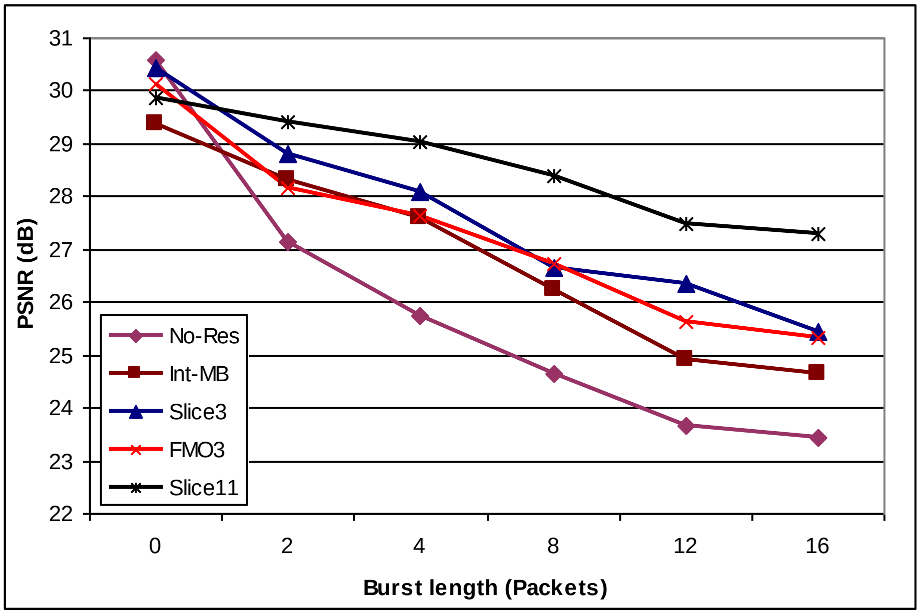
However, when burst errors were simulated then packet size has an important effect, as is summarized in Figure 6. Clearly in the multi-slice schemes the packet size is much reduced. However, unless an intelligent packetization scheme is in place such as [3] that uses aggregation and segmentation, small packet sizes are a risk for many coding schemes. Unfortunately, these packetization schemes are affected by the maximum transport unit size and, hence, are technology dependent. For example, in IEEE 802.15.4 (Zigbee) the maximum packet size is just 128 B, yet the UWB extension to Zigbee gives the capacity to support video streaming. It is unrealistic to expect that a general-purpose network will make special provision for one particular application such as video streaming. Therefore, in this paper a more general investigation occurs.

Increasing the number of slices to eleven slices results in good-quality video for burst lengths of up to six packets and results in superior quality to all other schemes except at lower burst lengths due to the overhead from slicing. H.264/AVC does not allow more than eight FMO groups and, therefore, increasing checkerboard FMO to eleven slices is not possible. FMO recovery is affected by the length of bursts, because error concealment through interpolation becomes less effective when two successive slices from the same frame are lost. Thereafter, the small extra overhead from FMO mapping compared to simple slicing should have an effect. At zero packet loss, the gain from less overhead using simple slicing is apparent in Figure 6a and 6b is apparent for the Slice3 and FMO3 example. More generally, Figure 7 shows the equivalent PSNR for different slicing and FMO schemes, when the overhead taken up by header information reduces the coding efficiency. However, between the two sequences there is no consistency between the ordering of Slice3 and FMO3 except for burst lengths of 12 and above. This may be an artifact of the number of tests conducted, which was actually large being an average of 100 tests in all. In a practical situation, a choice is made from just one of these instances and, therefore, one can conclude that except for large burst lengths, the likely performance is similar. It is preferable to use more slices rather than fewer in a ‘lossy’ channel and as the burst position cannot be predicted it is preferable to use simple slicing rather than FMO slicing, because at least less overhead is carried.

For both isolated and burst errors the relative ordering of the error resilience techniques is similar but the PSNR level is affected by the coding complexity, such as the affect of the complex texture of the background to the Coastguard sequence, which increases the spatial coding complexity.



(a)



(b)

Figure 6 Video quality with different error resiliency schemes with burst losses for (a) Foreman, (b) Coastguard video sequences

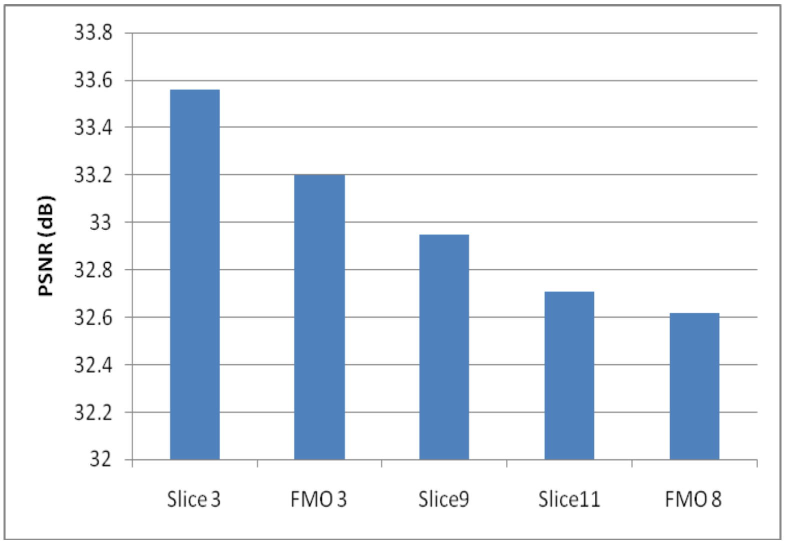
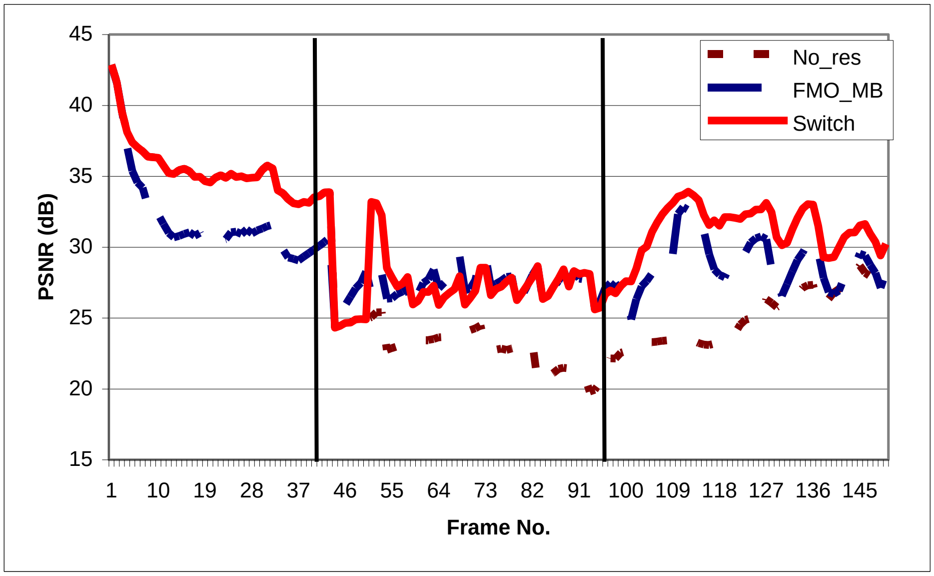


Figure 7 Zero packet loss PSNRs for equivalent schemes for the Foreman sequence, showing the effect of header information overhead on coding efficiency

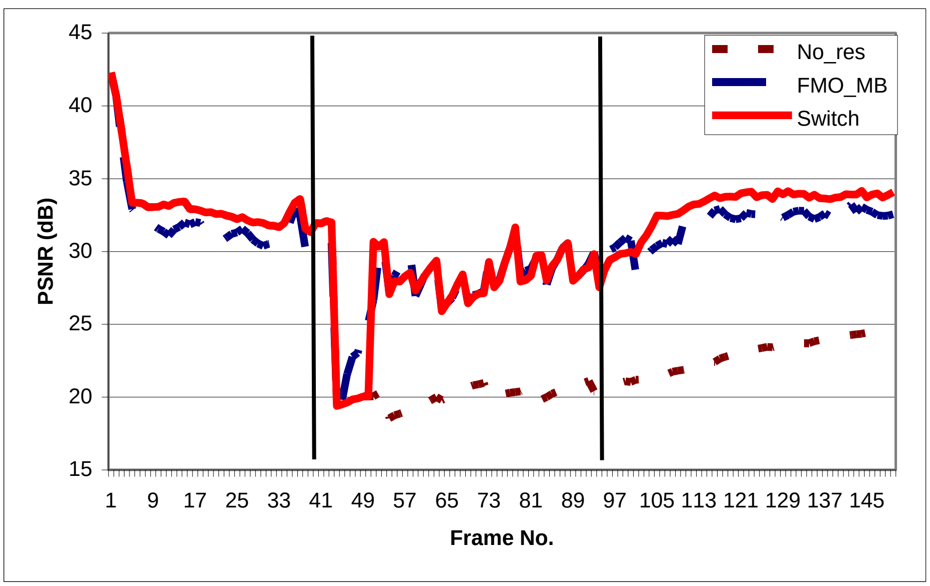
6.3. *Robust switching tests*

In comparing the relative impact of randomly occurring bit errors, a comparison was made between: 1) streaming with no error resilience; 2) streaming with the best of the schemes in Figure 3, FMO with intra-macroblock refresh; and 3) switching at the start of poor channel conditions to FMO with intra-macroblock with refresh. A period of poor channel conditions with a rate of 10% isolated packet losses was created. To compare results the same pattern of losses was replicated for all three tests. In Figure 8, the graph is divided into three regions: 1) before frame 45, when no errors occur; 2) after frame 45 when random packet loss occurs; and 3) after frame 100 when no errors occur. In the case of switching, an SI frame was used to transition between the streams. However, the bitrate for all three streams was the same. In fact, the first packet loss actually occurs at frame 45, but switching occurs only at frame 50 because of the positioning of the PSP-frames (every 10 frames).

Before frame 45, the video quality plots in Figure 8 for no-resilience and switching are identical with video quality declining once the effect of the initial I-frame has faded. Using error resilience during this period results in worst coding efficiency because of the overhead involved in providing resilience. When the first packet loss occurs, both the no resilience and switching curves drop in quality but as a result of the first SI-frame at frame 50, the video quality of the switched stream recovers. The decoder is resynchronized as a result of the switching frame and as a result the quality becomes equivalent to the protected stream without switching. On entering the good period after frame 100, an SI-frame resynchronizes the decoder. However, the no-resilience stream does not recover in quality for some time due to error drift caused by the loss of packets during the period of poor channel conditions.



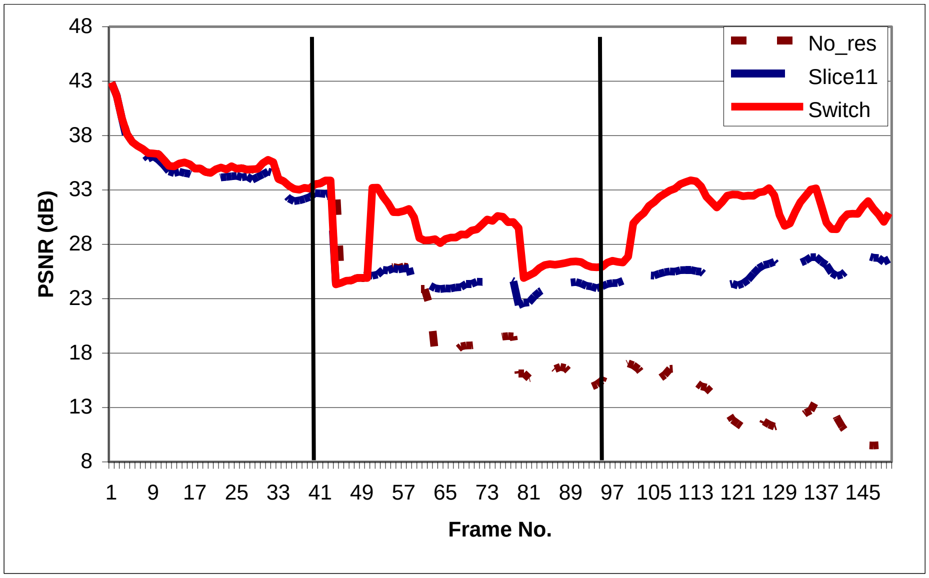
(a)



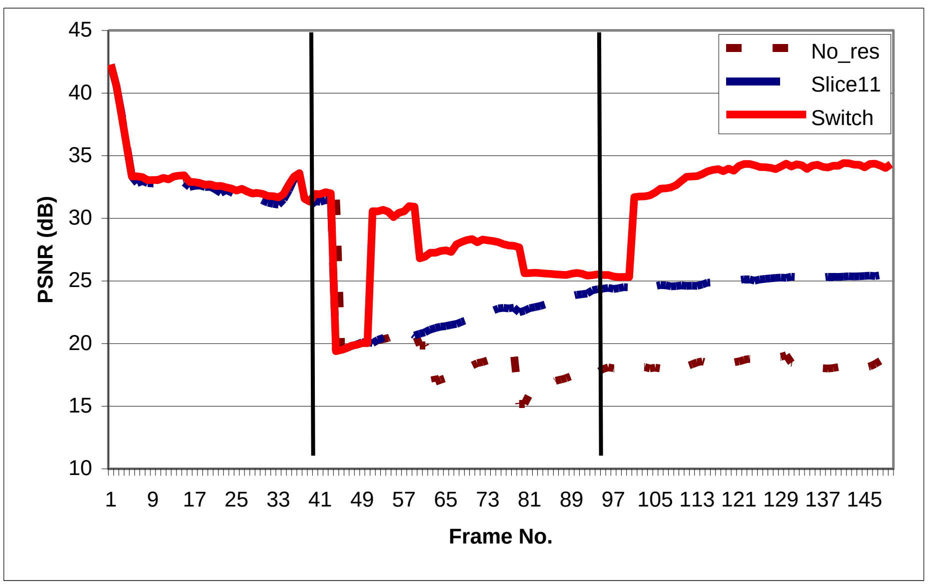
(b)

Figure 8 Video quality for robust switching in a channel with isolated losses for (a) Foreman, (b) Coastguard video sequences

Figure 9 is a comparison of switching in the presence of burst errors. During the poor channel period between frames 44 and 100 two packet burst errors of length four were introduced. Similarly to the random error tests, three schemes were compared. However, error resilience was through the best of the schemes in Figure 6, i.e. through simple slicing with 11 slices, as a result the behavior of the schemes using error resilience will be different from that of Figure 8. In Figure 9a, the robust switched scheme is superior throughout, while the no error resilience scheme fails to recover from lost packets during the period of poor channel conditions. Therefore, packet loss bursts can cause a breakdown in video quality if intra updates are not used. In Figure 9b, it is also clear how the transitions into and out of the period of packet losses affect the PSNR.



(a)



(b)

Figure 9 Video quality for robust switching in a channel with burst losses for (a) Foreman, (b) Coastguard video sequences

Table 1 summarizes a set of extensive tests in which poor channel conditions, isolated or burst errors, were simulated during the given range of frames (as annotated under the sequence name in the Table). The emboldened PSNRs highlight the advantage of the robust switching scheme, which is up to 3-4 dB in PSNR compared to using a robust scheme without switching, a considerable improvement given the logarithmic vertical scale. The mean video quality of all video sequences under robust switching is also good except for Mobile, which is a sequence with a high coding complexity that is normally is difficult to code efficiently.

Stream:	*Carphone (70-140)*	*Claire (80-160)*	*Mobile (50-100)*	*Salesman (60-120)*	*Foreman (50-100)*	*Coastguard (50-100)*
No-res (Random)	30.41	37.80	20.65	31.59	27.14	24.76
FMO_MB (Isolated)	30.77	38.44	24.62	32.09	29.14	30.48
Switch (Isolated)	**32.17**	**41.24**	**26.29**	**34.74**	**30.96**	**31.10**
No-res (Burst)	29.83	33.44	18.55	31.94	21.69	22.45
Slice11 (Burst)	29.88	40.86	23.90	32.05	27.73	26.13
Switch (Burst)	**32.83**	**43.45**	**26.83**	**36.88**	**31.33**	**30.76**

Table 1 Comparison of mean PSNR between robust switching and other schemes for various test video sequences, with range of frames for which poor channel conditions were simulated

7. Conclusion

In this paper different error resiliency schemes of the H.264 were presented for video streaming over wireless/mobile channels. The main contribution was to combine the switching frame concept with error resiliency schemes to create robustness scalability. It was observed that the Flexible Macroblock Ordering combined with intra-macroblock update works better in the case of isolated losses while increased levels of slicing is the best protection against burst losses.

These error resiliency techniques were than combined with the switching frames concept of H.264. It was shown that better video quality (PSNR) can be achieved if the stream without error resiliency is used for the time when channel is free of errors and robust streaming is limited to the periods when the channel is error prone. Feedback messages were used to indicate the channel state and SI-frames were used for switching. An advantage from using SI frames is that they also correct error drift.

We have reserved future research for a method of detecting which error resiliency scheme should be applied by finding a way of predicting the type of channel conditions. A developer would also wish to determine how many different robust streams are required for a particular channel type. However, these are problems outside the scope of the present paper, which concentrated on adaptive error resilience coding and which has established the concept and value of the robust streaming scheme.

References

1. Altaf, M., Khan, E. and Ghanbari, M., Switching frames for constant bit rate video streams. In 5th International ICST MobiMedia Conference, (2009), article no. 5.
2. Chou, L.-D., Chen, J.-M., Kao, H.-S., Wu, S.-F. and Lai, W., Seamless streaming media for heterogeneous mobile networks. Mobile Networks and Applications, 11, (2006), 873-887.
3. Ferré, P., Doufexi, A., Cang-How, J., Nix, A. R., and Bull, D. R., Robust video transmission over wireless LANs. IEEE Transactions on Vehicular Technology, 57(4), (2008), 2596-2602.
4. Fitzek, F., Seeling, P. and Reisslein, M.,Video streaming in wireless internet. In Wireless Internet: Technologies and Applications, A. Salkintzis and A. Poularikas (eds.) CRC Press, Boca Raton: FL, 2004, pp. 1-41.
5. Fröjdh, P., Horn, U., Kampmann, M., Nohlgren, A. and Westerlund, M., Adaptive streaming within the 3GPP packet-switched streaming service. IEEE Network, 20(2), (2006), 34-40.
6. Ghanbari, M. Standard Codecs: Image Compression to Advanced Video Coding. IET Press, London, UK, 2003.
7. Hantanong, W. and Aramvith, S., Analysis of macroblock-to-slice group mapping for H.264 video transmission over packet-based wireless fading channel. 45th Mid-West Symposium on Circuits and Systems, (2005), 1541-1544.
8. Karczewicz, R. K. M. and Kurceren, R., The SP- and SI-frames design for H.264/AVC. IEEE Transactions on Circuits and Systems for Video Technology, 13(7), (2003), 637-644.
9. Kumar, S., Xu, L., Mandal, M. K., and Panchanathan, S., Error resiliency schemes in H.264/AVC standard. Journal of Visual Representation & Image Representation, 17(2), (2006), 1-26.
10. Lam, W. M., Reibman, A. R. and Liu, B., Recovery of lost or erroneously received motion vectors. In IEEE Int. Conf on Acoustics, Speech, and Signal Processing, (1993), 417–420.
11. Lambert, P., Deneve, W., Dhondt, Y., and Vandewalle, R., Flexible macroblock ordering in H.264/AVC. Journal of Visual Communication and Image Representation, 17, (2006), 358-375.
12. Liu, L., Ye, X.-J., Zhang, S.-Y. and Zhang, Y., H.264/AVC error resilience tools suitable for 3G mobile video services. Journal of Zhejiang University SCIENCE, 6(4), (2005), 1-46.
13. Marpe, D., Wiegand, T. and Sullivan, G. V., The H.264/MPEG4 Advanced Video Coding standard and its applications. IEEE Communications Magazine, 44(8), (2006), 134-142.
14. Nuaymi, L. WiMAX: Technology for broadband wireless access. J. Wiley & Sons Ltd, Chichester, UK, 2007.
15. Sadka, A. Compressed Video Communications. J. Wiley & Sons, Chichester, UK, 2002.
16. Salama, F., Shroff, N. B. and Delp, E. J., Error concealment in encoded video. In Image Recovery Techniques for Image Compression Applications. Kluwer, Norwell, MA, 1998.
17. Schierl, T., Wiegand, T. and Kampman, M., 3GPP compliant adaptive wireless video streaming using H.264/AVC. In IEEE International Conference on Image Processing, (2005), 696-699.
18. Stockhammer, T., Hannuksela, M. M. and Wiegand, T., H.264/AVC in wireless environments. IEEE Transactions on Circuits and Systems for Video Technology, 13(7), (2003), 57-673.
19. Stockhammer, T., Liebl, G. and Walter, M., Optimized H.264/AVC-based bit stream switching for mobile video streaming. EURASIP Journal on Applied Signal Processing, [online journal], (2006), 1-19.

20. Stockhammer, T. and Zia, W., Error-resilient coding and decoding strategies for video communication, in Multimedia over IP and Wireless Networks, M. van der Schaar and P. A. Chou (eds.), Academic Press, Amsterdam, 2007, 59-80.
21. Sun, X., Wu, F., Li, S., Gao, W., Zhang, Y.-Q., Seamless switching of scalable video bitstreams for efficient streaming. IEEE Transactions on Multimedia, 6(2), (2004), 291-303.
22. Vars, V. and Hannuksela, M. N., Non-normative error concealment algorithms. (2001), ITU-T SGI6 Doc., VCEG-N62.
23. Wenger, S., H.264/AVC over IP. IEEE Transactions on Circuits and Systems for Video Technology, 13(7), (2003), 645-656.