Error Resiliency Transcoding and Decoding Solutions Using Distributed Video Coding Techniques

Tamer Shanableh[1], Tony May[2] and Faisal Ishtiaq[3]

[1]Department of Computer Science and Engineering
College of Engineering
American university of Sharjah
Fax: +971 6 515-2979
tshanableh@aus.edu

[2]Multimedia Research Lab
Motorola Ltd
Basingstoke, Hampshire
Fax: +44 1256 484438
tony.may@motorola.com

[3]Multimedia Research Lab
Motorola Inc
Schaumburg, IL 60196
Fax: +1 847 576-6030
faisal@motorola.com

Abstract

This paper proposes a number of video transcoding techniques for the purpose of adding error resiliency. The proposed solutions make use of distributed video coding technologies which were originally reported in the literature for distributing coding complexity between the encoder and the decoder. Three transcoding solutions are proposed: frequency-domain, time-domain and compressed-domain transcoding, and various decoding architectures are investigated. Error resiliency at the decoder is addressed as a post-process, an integrated process and a pre-process. It is shown that implementing error resiliency as a decoder pre-process and combining this with compressed-domain transcoding removes the need to interfere with the functionality of existing and compliant decoders whilst minimising complexity. The proposed solutions serve as a framework for boosting the error resiliency of pre-encoded video and can be applied to MPEG-2, and H.264/AVC coded streams.

Keywords: Distributed video coding, video transcoding, channel coding and video networking.

http://doi.org/10.1016/j.image.2008.05.005

1. Introduction

Distributed video coding is a technique in which the computational complexity is distributed between the encoder and the decoder. In some implementations the motion estimation is moved completely to the decoder side. This is a reversed coding technique which might be suitable for compressing content with computationally-limited encoding devices. In one of its forms, the encoder identifies and sends key frames using traditional intra-frame coding. The encoder then generates parity bits for the rest of the frames and sends them to the decoder as an additional bit stream. The decoder processing includes motion compensation interpolation guided by this additional bit stream. A full review of distributed coding theory, history, and practical implementation is available in [1].

Distributed video coding has also been applied to the problem of error resiliency. The work carried out by Stanford University [2] is based on coding raw video using H.264/AVC whilst employing distributed coding as an error resilience mechanism. Regions of interest are selected from the partially reconstructed video using Flexible Macroblock Ordering (FMO). Redundant slices are then used to re-encode the regions of interest using the coding modes of the original video albeit at a lower quality through coarser quantization. The coarse representation of the video is then channel coded and the parity bits, along with the new quantization parameters and region boundaries, comprise the Error Resiliency (ER) supplementary bit stream.

The decoder reconstructs the coarse video representation and applies channel decoding and correction to it. This coarse representation can be used to replace a part of the bit stream that is received in error. Clearly, a motion compensation loop-mismatch will be caused by the replacement, however such a mismatch has a less severe visual effect than that caused by the bit stream errors.

This paper proposes to apply the above distributed-coding error resiliency mechanism to pre-coded video using video transcoding techniques. The paper also proposes a pre-decoding approach to ensure the compliancy of the bit stream delivered to the decoder. The approach identifies regions of interest from pre-encoded video and reproduces them with a coarser representation. This is achieved through a number of transcoding techniques such as requantization, requantization with drift compensation, and dropping DCT coefficients. It is also proposed to retain regions of interest as is without reproduction. Following the work in [2], the ER supplementary bit stream is then generated. We complement this work with a pre-decoding solution that mirrors the transcoder by regenerating the regions of interest and correcting them using the ER supplementary bit stream. Non-regions of interest are concealed if needed. The corrected bits are then repacked and sent to the decoder. Unlike the work proposed in [2], the proposed transcoding and error correction solutions do not require modifications to the destination decoder.

The proposed solution can be used to boost the error resiliency of both H.264/AVC and MPEG-2 pre-encoded broadcast video. It can also be applied to a Digital Living Network Alliance (DLNA) network in a home [3]. Digital media servers and digital media players or renderers could be located anywhere in the home. Creating wired connections between all of these devices may not be convenient or economical and so it is likely that wireless Internet Protocol (IP) connections will exist as part of the network. Since there is no guarantee that content stored on a server has a particular set of error resiliency features, it is necessary to add error resiliency to the stored bit streams if they are to be streamed using User Datagram Protocol (UDP) over IP. The server may also provide live content, which is particularly suited to transmission using UDP, and this content may also not contain any error resiliency features.

Additionally, DLNA specifies network devices that allow mobile devices to access content stored on a home media server. The Media Interoperability Unit (MIU) is a DLNA device that transcodes content between different formats so that a player that only contains an H.264/AVC decoder, for example, is able to view content stored in MPEG-2 format. When streaming content to a mobile device, the MIU could also make use of the proposed work to add error resiliency to the bit stream to protect it against lost packets in the wireless part of the home network.

2. Related work

The idea of transcoding video to generate a supplementary error resilient bit stream using distributed-coding is rather novel. However, the general concept of transcoding video to boost error resiliency is reported in the literature. For instance, [4,5] proposed the use of Adaptive Intra Refresh (AIR) and feedback control signalling for video transmission over GPRS. A similar approach appeared in [6] which estimates the effect of error propagation and selectively applies Adaptive Intra Refresh. Recent work on adaptive error resiliency transcoding based on intra refresh algorithms is reported in [7, 8, 9]. A more complex transcoder was introduced in [10] which recalculates Motion Vectors (MVs) and coding modes based on a given error model and estimation of distortion at the decoder side. In [11] Xin et al. discussed the use of a number of error resiliency tools in terms of the benefits they provide and their impact on coding efficiency from a transcoding perspective. Some of the discussed tools provide error localization, such as the support for multiple reference pictures and the insertion of intra-coded MBs and the like. Other tools provide unequal error protection through data partitioning, reversible Variable Length Coding (VLCs), redundant slices, concealment motion vectors and flexible MB ordering. Forward Error Correction (FEC) is also used, for instance [12] proposed FEC based multiple description transcoding with unequal loss protection for embedded bit streams such as 3D-SPHIT.

From a different perspective, the work in [13] proposed a solution for transmitting pre-encoded video over wireless Local Area Networks (LANs). In their solutions, a media gateway gathers channel statistics from the base-stations. The media gateway then makes a decision on whether or not to transmit the coded video in a single description mode and use content-based prioritized packet retransmission. In the case of detecting mobile channel switching, the coded video is transcoded into multiple descriptors and the gateway will search for another channel to be used for simultaneous transmission of the multiple description video.

It is also worth noting the conceptual similarity between error resiliency based on distributed video coding and the work carried out by [14] on image authentication. The mentioned work generates an encrypted image signature based on the invariance of the relationships between Discrete Cosine Transform (DCT) coefficients at the same position in separate blocks of an image. This relationship is invariant to JPEG quantization. Thus malicious manipulations can be detected if the original image signature differs from that computed from a given image. In comparison to work proposed by [2] this is analogous to error detection at the receiver side, however the main difference is that the generated image signature proposed by [14] cannot be used for error recovery.

3. System Overview of Error Resiliency Scheme

Prior to proceeding with the proposed transcoding and decoding architectures, it is worthwhile illustrating the proposed error resilient system. In Figure 1, a transcoder is deployed to generate a supplementary error resilient bit stream from the original video. The coded video and the supplementary bit stream are then sent through the error prone channel. The details of the transcoder and its output stream shall be covered in the subsections to follow. The proposed pre-decoder then receives the encoded video alongside the supplementary bit stream. In case of channel errors, the decoder uses the

http://doi.org/10.1016/j.image.2008.05.005

supplementary bit stream to recover from errors and erasures. If errors occur in non-regions of interest of the video stream then the pre-decoder conceals the error. The corrected video stream is then repacked and sent to the destination decoder.

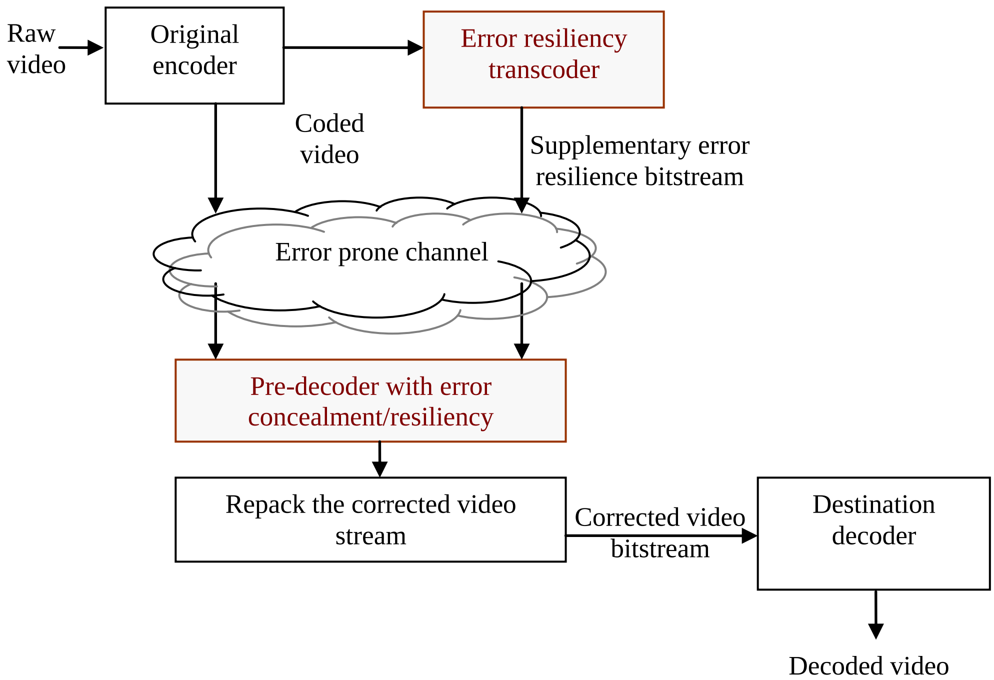


Figure 1. System overview of proposed solution.

The rest of this section proposes three approaches to error resilient transcoding based on distributed-coding namely; frequency-domain, pixel-domain and compressed-domain transcoding. Different decoding architectures are then discussed based upon the aforementioned processing domains.

In the three approaches the first step is to generate a coarse description of regions of interest of the pre-encoded video. Channel coding is then applied to this low fidelity video. The resultant parity bits and some additional information such as the boundaries of the areas of interest comprise the supplementary error resilient bit stream.

4. Identifying regions of interest

This section introduces approaches to identify regions of interest in pre-encoded video. In previous work, Baccichet *et al.* proposed to classify video slices as regions of interest if they cannot be reasonably concealed at the decoder [15]. An estimated version of a pixel-domain decoded and concealed image is produced at the encoder. This is then subtracted from the original image. Pixel blocks of high mean absolute errors are consequently labelled as regions of interest. Note that this approach is inapplicable to video transcoding because there is no access to pixel-domain reconstructions or original images. Rather, in video transcoding, regions of interest are selected based on information available from the coded bit stream without reconstructing the video. This includes motion information, MB coding modes, presence of high frequency DCT coefficients and quantization indices. Note that a more sophisticated technique for determining regions of interest from pre-encoded video based on perceptual importance is reported in [16]. Likewise, [17] proposed a solution in which an encoder identifies regions of interest using a so called attention model and embeds attention information in the coded bit streams to guide the transcoder in determining the regions of interest.

The work in this paper uses a simpler approach that utilizes the information available in the pre-encoded bit stream. For instance in a predicted frame, a slice can be classified as non-concealable, and hence labelled as a region of interest, if any of the following conditions are met:

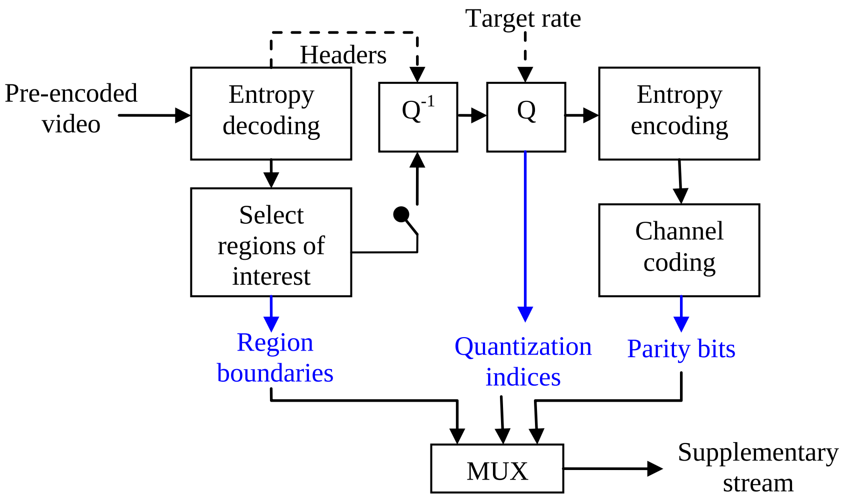
1. The motion vectors of the current slice are different from those of the above slice and so cannot be reliably estimated at the decoder. One way to detect this is to compare the average Euclidean distance of motion vectors between the current slice and its above slice. If this distance is greater than the overall average of all predicted slices in the same frame then it can be assumed that the motion vectors of the current slice cannot be reliably estimated at the decoder.
2. A large number of intra coded MBs are present in the slice. This can be detected if the number of intra coded MBs in a given slice is greater than 50%.
3. The slice contains significant high frequency coefficients in conjunction with relatively high quantization indices. This can be detected if the quantization indices and the sum of absolute DCT coefficients in a given slice are greater than the average values for the frame.
4. A low number of skipped MBs indicating significant changes in the slice. This can be detected if the percentage of skipped MBs in a given slice is less than 50%.

5. Proposed transcoding and decoding architectures

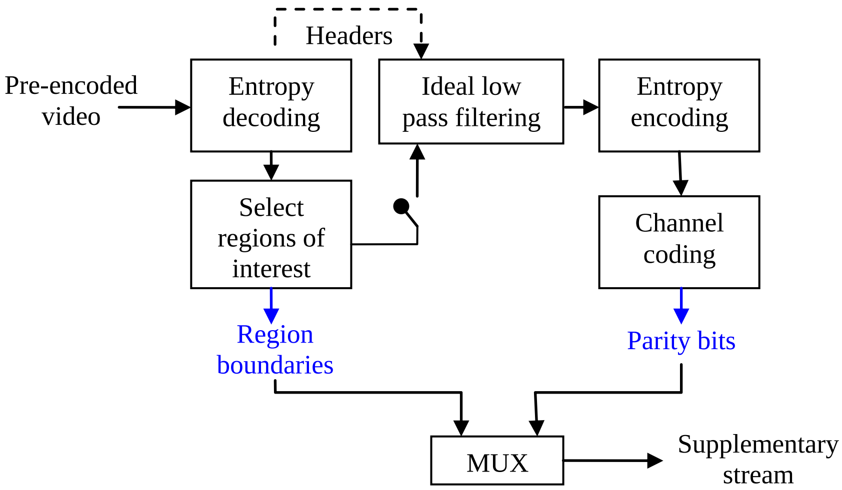
This section introduces a number of video transcoding techniques to identify regions of interest and produce the side information which is the coarse representation of the regions of interest. The side information is then channel coded to generate the supplementary ER bit stream. The section also proposes a number of decoding and pre-decoding arrangements that complement the proposed transcoders.

a. Frequency-domain ER transcoding and decoding architectures

In this approach the pre-encoded video is parsed and a coarser representation of the DCT coefficients is computed. To further reduce the volume of bits entering the channel coder, coarse reproduction of the coded coefficients is applied to selected regions of interest only. Block diagrams of the proposed transcoding architectures are shown in Figure 2.



(a) Requantization-based transcoder



(b) Ideal low pass filtering based transcoding.

Figure 2. Architectures for frequency-domain distributed-coding error resilient transcoding

The figure shows two different approaches to transcoding. In Figure 2(a), the coarse representation is achieved through requantization whilst in Figure 2(b), it is achieved by retaining a predefined percentage of low frequency DCT coefficients. To support the generation of a low fidelity version of the video, non-referenced frames can be excluded from the supplementary stream. For instance B-frames in MPEG-2 and non-reference B-slices in H.264/AVC can be skipped.

Note that in the presence of channel errors the destination decoder will try to recover the coarse video representation generated by the transcoder. This coarse representation will then replace the MC-loop buffer content. Since this is different to the original contents of the encoder buffer, a mismatch will occur. However this mismatch is insignificant when compared to the drastic effect of channels errors. On the other hand, if the channel error rate is rather high and many coarse representations of the video are being subsequently stored in the MC buffers of the destination decoder then a transcoder drift correction loop may be of some benefit.

In Figure 3, the transcoder computes the requantization error, motion compensates it and adds it to incoming frames that use the requantized frame as a source of prediction. More information on drift correction can be found in [18, 19, 20].

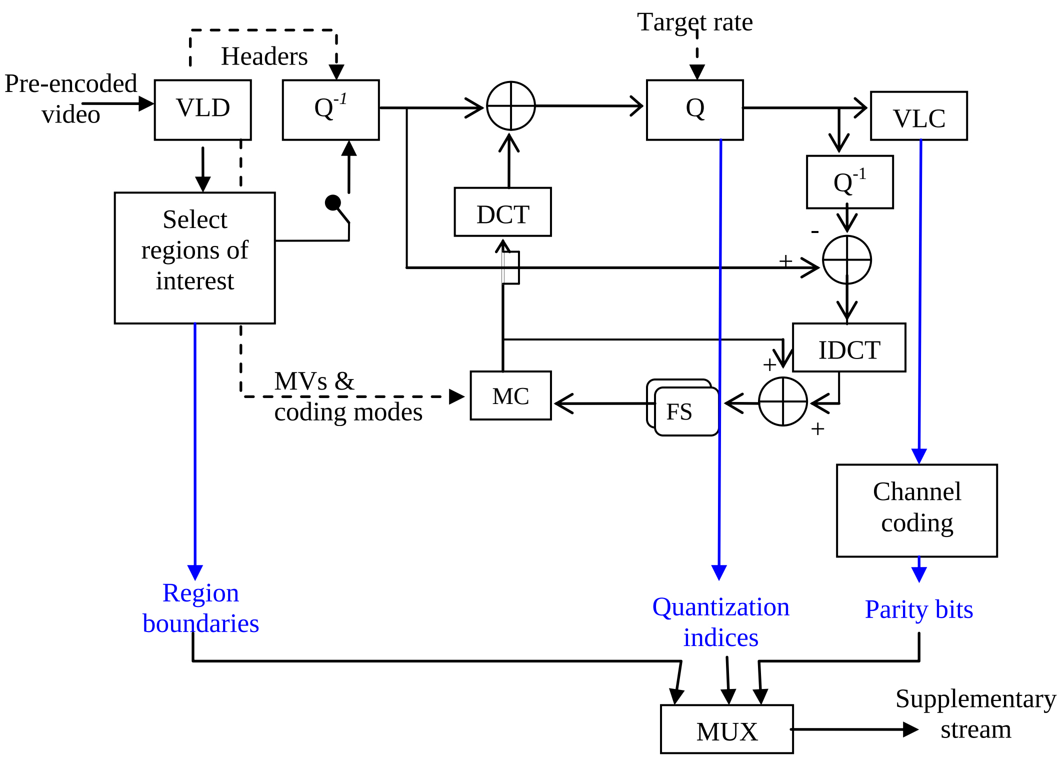


Figure 3. Architecture for frequency-domain distributed-coding error resilient transcoding with a drift correction loop

The downside of drift correction is that the destination decoder has to replicate the MC loop of the intermediate transcoder thus adding some complexity to the system. The extra complexity is elaborated upon in Section 7 .These details will be further discussed in the experimental results section.

Once the supplementary bit stream is generated, the decoding arrangement of Figure 4 is applicable to all of the aforementioned transcoding solutions. Note that the only difference between this decoder and the one proposed in [2] is that the 'Reproduced side information' is now generated by a transcoder function at the decoder. This transcoder should match the intermediate transcoder that generated the supplementary bit stream as elaborated upon in Figures 2 and 3 above.

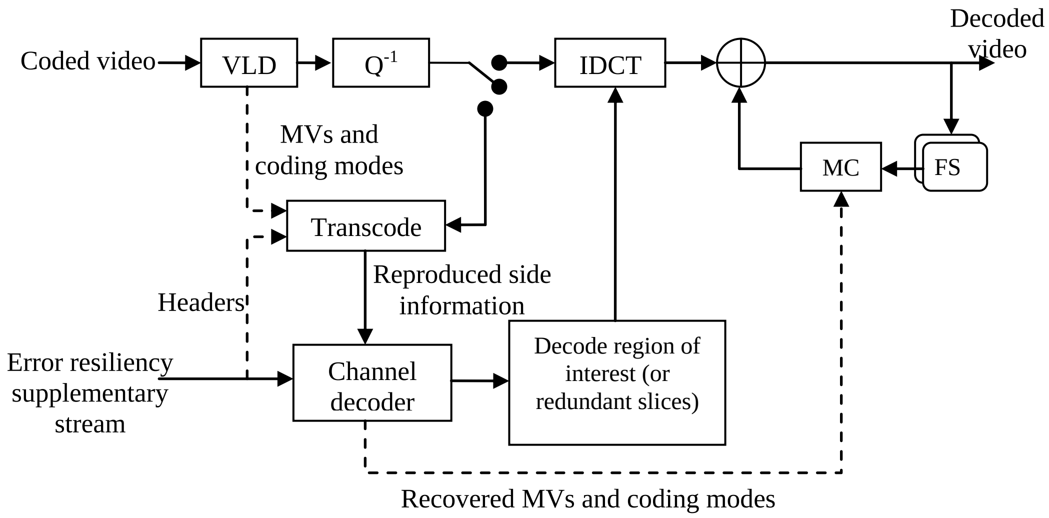


Figure 4. ER-decoding architecture for the frequency-domain transcoding solution.

The decoding solution is applicable to both MPEG-2 and H.264/AVC. The only difference is that in MPEG-2, the reproduced area of interest is a subset of the coded slices. Whereas in H.264/AVC the Flexible MB Ordering (FMO) feature can be employed for an arbitrary selection of macroblocks if needed.

The 'Transcode' function will reproduce the coarse representation of the video. In distributed coding jargon this is referred to as 'decoder side information'. The side information is then fed into the channel decoding to correct errors and recover erasures. If needed, the recovered coarse representation of the video is then fed into the destination decoder and decoding resumes. Clearly there might be cases where the error is more than what the channel decoder can handle.

A serious drawback of this decoding arrangement is that it mandates modifications to the decoding process as illustrated in Figure 4. Although no modifications are required to the bit stream syntax, the output of the channel decoding process might be required to replace the dequantized DCT coefficients of the destination decoder.

On the other hand, the Stanford distributed coding group proposed another decoding solution where error resiliency and concealment is carried out as a post-process [21]. The decoding architecture is illustrated in Figure 5 below.

The decoded pixel-domain images are transcoded to reproduce the coarser representation of the regions on interest. With the help of the ER supplementary bitstream, channel decoding will then correct errors in the pixel-domain regions of interest. However, this arrangement cannot be classified as a pure post-process. This is because the corrected regions of interest will then replace the contents of the frame buffers in the motion compensation loop as shown in the buffer switch of Figure 5.

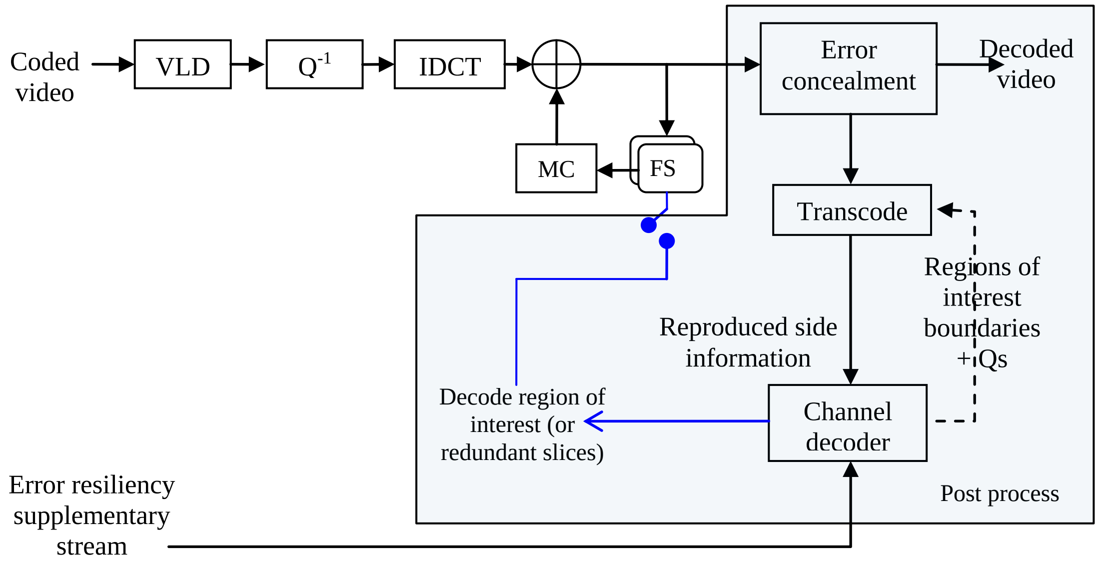


Figure 5. Post processing ER-decoding architecture for the pixel-domain transcoding solution.

Clearly, the matching transcoder of the above decoder should operate in the pixel-domain, including pixel-domain channel coding. This means that the pre-encoded video is decoded to the pixel-domain and requantized with intraframe coding to a coarser representation. This is followed by selecting regions of interest and channel coding. The complexity in this case is identical to the decoding process plus the additional computations required by the coarser intraframe codec and channel encoder.

The transcoding complexity can be reduced by restricting the above mentioned transcoding process to intra-coded macroblocks, removing the need for motion compensation at the transcoder. Reducing the complexity in this way comes at the expense of lower quality error resiliency and concealment at the destination decoder.

An additional drawback of the pixel-domain transcoding approach is due to applying the channel coding in the pixel-domain of the selected regions of interest. Protecting such a vast amount of information requires much more parity bits and therefore an increase in the overall bit rate.

b. Compressed-domain ER transcoding and decoding architectures

This section proposes a compressed-domain processing approach to overcome the limitations of modifying the destination decoders. The proposed solution also overcomes the limitations of the post-processing decoding approach of Figure 5 which requires either a complex transcoder or demands a noticeable increase in parity bits for the channel encoding to be efficient.

The aim here is to make the error resiliency approach transparent to the destination decoder and therefore be applicable to existing and compliant decoders. Since the post processing approach of Figure 5 has some limitations as previously discussed, this section proposes a pre-decoding approach to error resiliency. To achieve this we propose a compressed-domain transcoder is illustrated in Figure 6. Briefly, the regions of interest are identified from the parsed bit stream. The regions of interest are then protected using channel coding, and the resultant parity bits and regions of interest boundaries are multiplexed into one ER supplementary bit stream. Since DCT coefficients are not requantized or dropped, this architecture is given the name 'compressed-domain' processing.

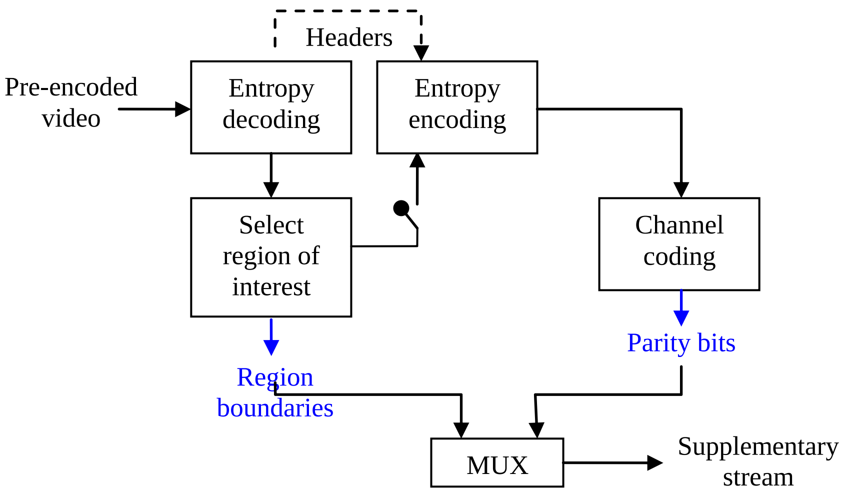


Figure 6. Proposed architecture for compressed-domain distributed-coding error resilient transcoding

In contrast to the proposed transcoders of Figures 2 and 3, in the compressed-domain transcoding approach, a coarser representation of the coded video is not needed. The volume of coded material is reduced by selecting specific regions of interest only. Non-referenced pictures or slices, e.g. B frames or some B-slices, can be excluded from channel coding as well.

The matching pre-decoding solution is illustrated in Figure 7.

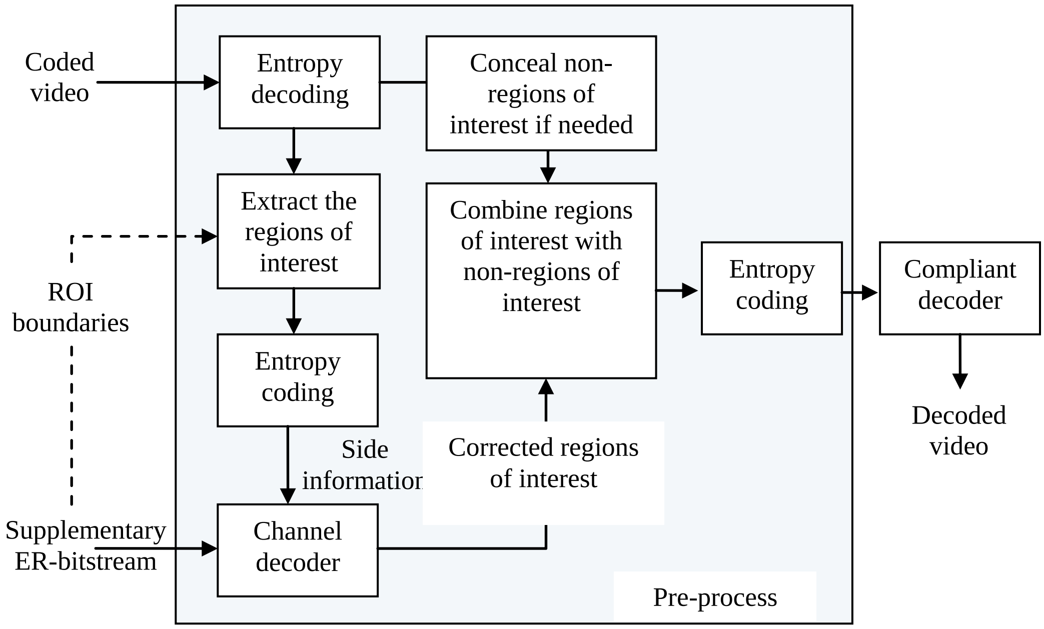


Figure 7. Pre-decoding ER architecture for the proposed compressed-domain transcoding solutions.

In this arrangement, the side information of the distributed coding approach is reproduced as part of the decoding pre-process. The regions of interest are extracted from the coded bit stream and channel decoded to correct any bit errors and/or erasures. The bit stream is then regenerated by combining the corrected regions of interest with the potentially concealed non-regions of interest. The repacked bit stream is then sent to a compliant decoder.

In addition to not interfering with the decoder's operations, the advantage of this solution over traditional Forward Error Correction is that channel coding is applied to a subset of the bit stream or the regions of interest, thus requiring fewer parity slices. Non-regions of interest can be concealed if needed through classical means.

However, if the coarser representation of the video is reproduced at the decoder's pre-processor then both transcoders of Figures 2 and 3 can communicate with the proposed pre-decoder of Figure 7. Clearly, in this case, the operations will be carried out in the frequency rather than the compressed-domain. For completeness such a decoding pre-processor is illustrated in Figure 8.

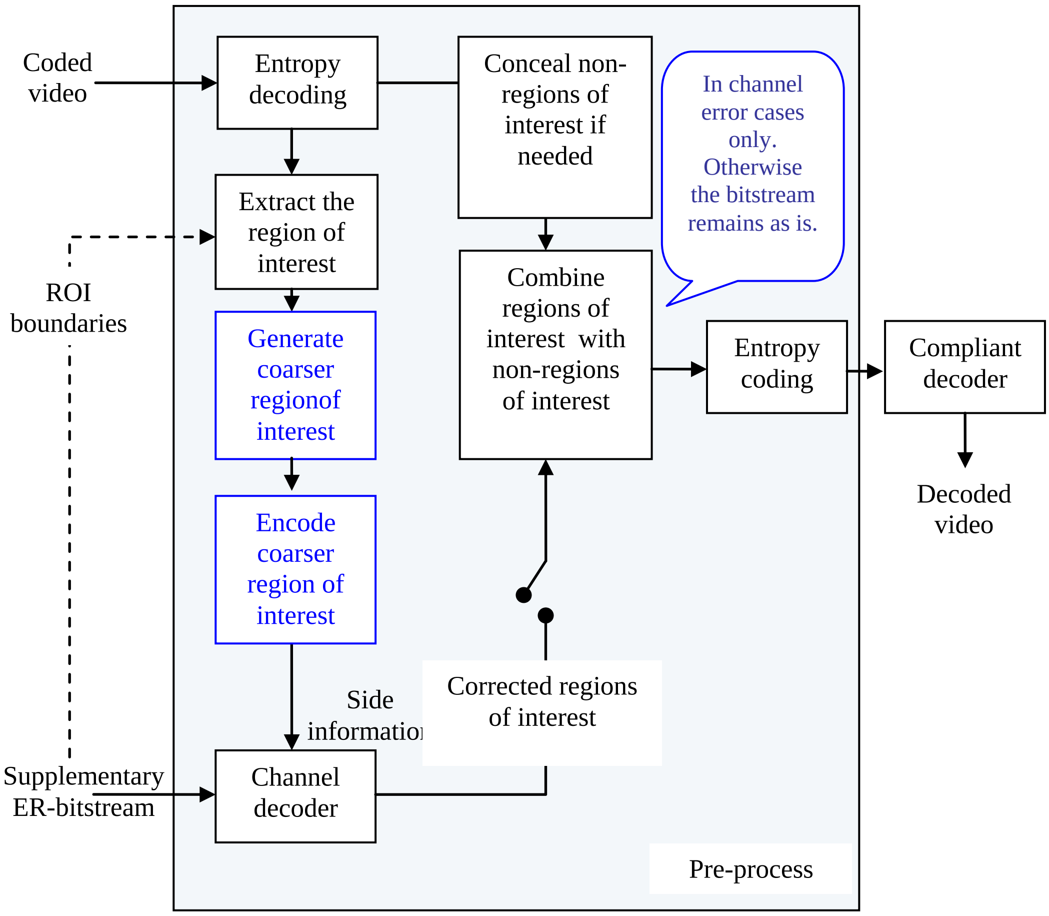


Figure 8. Pre-processing ER-decoding architecture for the frequency-domain transcoding solution.

Figure 8 adds the computational requirements of reproducing and re-encoding the regions of interest. In the presence of channel errors the corrected regions of interest will be injected in the original bit stream. Otherwise the original video slices will remain unchanged.

6. Channel coding

In the proposed ER transcoders, the regions of interest are identified and potentially converted to a coarser representation. The slices of the regions of interest are then aligned on top of each other to facilitate interleaved channel coding as shown in Figure 9. If slices and thus regions of interest are determined by the number of MBs rather than the number of coded bytes then zero padding may be required. In MPEG-2 video, slices are determined by the number of MBs whereas in H.264/AVC, both approaches are compliant with the standard.

Reed-Solomon (R-S) codes are then computed by vertically cutting through the aligned regions of interest as illustrated in Figure 9.

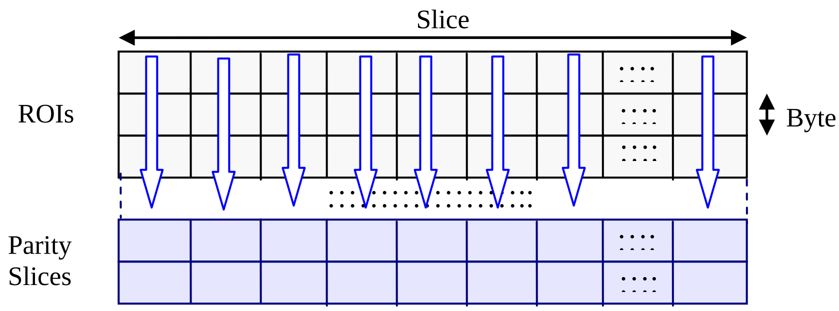


Figure 9. Arrangement for computing the Reed-Solomon parity slices.

The rate of the resultant parity slices is typically set to around 10%-15% of the total video coding rate. n parity bytes are capable of recovering either n erasure bytes or correcting $n/2$ erroneous bytes and are capable of dealing with both cases simultaneously as long as $n \geq$ (2 * number of erroneous bytes + number of erasure bytes). This arrangement of interleaved channel coding has been deployed in [2].

At the pre-decoder side, the regions of interest are reproduced again and the parity slices are loaded from the ER supplementary bit stream. Both are arranged in a similar manner to that of the ER transcoder. This is followed by channel decoding and possibly correction.

If the bitstream is transmitted over a packet switched network then regardless of the transcoding approach used, the channel decoding and correction scenario is as illustrated in Figure 10. The 'X' locations indicate that the corresponding byte within a region of interest have been lost.

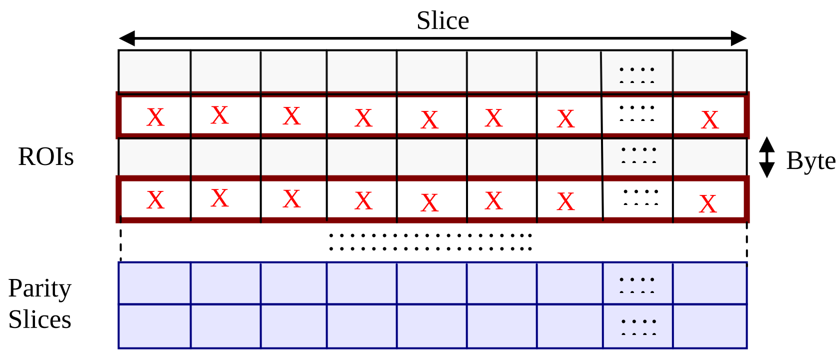
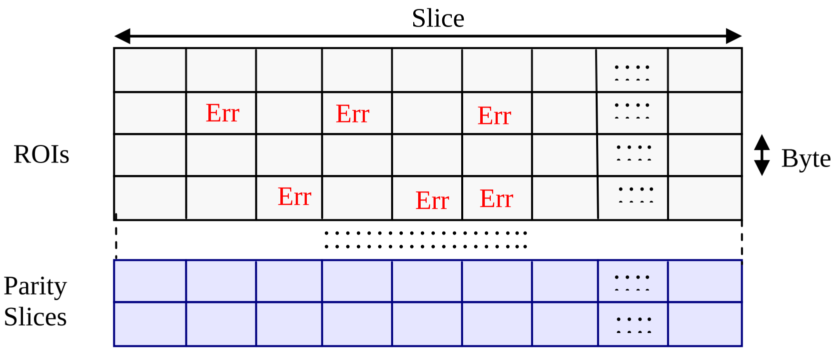
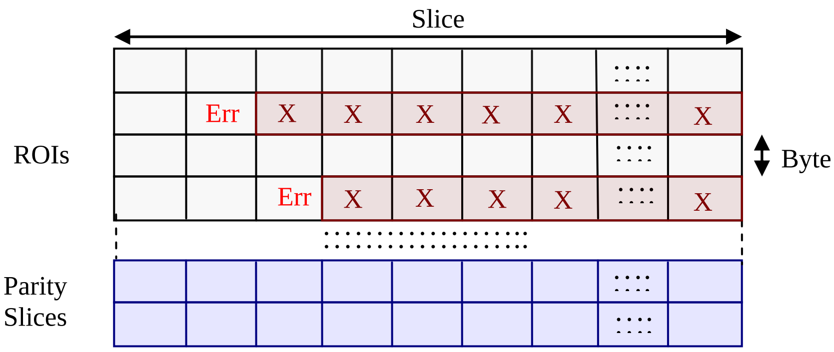


Figure 10. Channel correction for recovering from packet loss.

The situation is different in the case of bit errors. If at the ER transcoder the regions of interest are not reproduced by requantization or ideal low pass filtering then they can be extracted from the incoming video stream and sent to the channel decoder as is. The compressed-domain processing means that there is no error propagation due to the VLC because the regions of interest are extracted and not reproduced. If the regions of interest have been converted to a coarse representation then a bit error may cause VLC desynchronisation and the rest of the region of interest bits will be undecodable. These two scenarios are illustrated in Figures 11-a and 11-b. The 'Err' locations indicate that the corresponding bytes within the region of interest contain bit errors.



(a) Bit errors: Retain regions of interest and FEC approaches



(b) Bit error: Requantization and ideal low pass filtering approaches

Figure 11. Channel coding for recovering from bit errors

In part (a) of the figure, the extracted regions of interest might contain bit errors. These will be isolated and possibly scattered and hence easier for the channel decoder to correct. In part (b) of the figure, the incoming video bit stream is entropy decoded and processed in the frequency-domain to reproduce the regions of interest. The variable length decoding will cause bit errors to propagate until the end of the slice. Therefore, a region of interest will contain erroneous bits and erasure bytes as illustrated in part (b) of

http://doi.org/10.1016/j.image.2008.05.005

the figure. This can put more strain on the channel decoder and reduce its effectiveness if not enough parity slices are available. Packet loss and bit error simulations are elaborated upon further in the experimental results section.

7. Complexity considerations

Table 1 summarizes the complexity involved in each of the proposed transcoding architectures by identifying the functional blocks that each approach requires.

Approach	VLD	Identify region of interest	Drop coeff.	Q^{-1}	Q	DCT	DCT^{-1}	MC	VLC	RS coding
Requantization	✓	✓		✓	✓				✓	✓
Requantization and Drift MC	✓	✓		✓	✓	✓	✓	✓	✓	✓
Ideal low pass	✓	✓	✓						✓	✓
Retain regions of interest	✓	✓							✓	✓
FEC	Partial									✓

Table 1. Complexity considerations for the ER transcoding modes

Note that at the pre-decoder, the complexity is very much similar to that of the ER transcoder. Minor differences include loading regions of interest indices from the ER supplementary bit stream rather than identifying them, and replacing RS encoding by RS decoding and correction.

As the table indicates, the FEC approach is the least complex, followed by retaining the regions of interest (transcoder of Figure 6), ideal low pass filtering and requantization (transcoders of Figures 2.b, and 2.a respectively). The most complex approach is requantization with drift compensation (the transcoder of Figure 3).

8. Experimental results

This section presents the results of experiments on the proposed transcoding architectures and pre-decoding solution using a test system based on MPEG-2. An open source implementation of Reed-Solomon coding is employed for the channel coding [22]. Two 25Hz test sequences are used; Foreman (250 frames) and Football (118 frames). For all of the experiments, no intermediate I frames are used, thus the GOP structure is N=∞, M=3 i.e. "IBBPBBPBBP…". This emphasises the effect of error propagation and examines the effectiveness of the underlying ER approach. Test sequences with Common Image Format (CIF) resolution are used with a coding rate of

1.35Mbit/s. The bit rate of the ER supplementary bit stream is around 150Kbit/s. The total coding rate is 1.5Mbit/s.Two sets of experiments are carried out; packet loss and bit error simulations. A Gaussian random error generator is used to simulate the error conditions.

The proposed transcoding architectures are compared against the FEC approach where all slices are considered to be regions of interest regardless of scene activity or picture type. Regions of interest are not reproduced in this case. Thus only one parity slice is generated resulting in an ER supplementary bit stream of 150kbit/s. In the proposed transcoding approaches on the other hand, the slices of non-referenced B-frames are considered non-regions of interest. I and P frames are processed to determine regions of interest as described in Section 5.a above.

In the following experiments, the percentage of I and P slices that are identified as regions of interest is 54% for the Football sequence (18% of all I, P and B slices) and 45% for the Foreman sequence (15% of all I, P and B slices). The difference is caused by the fact that the former sequence is more temporally active and so fewer slices are concealable without channel protection.

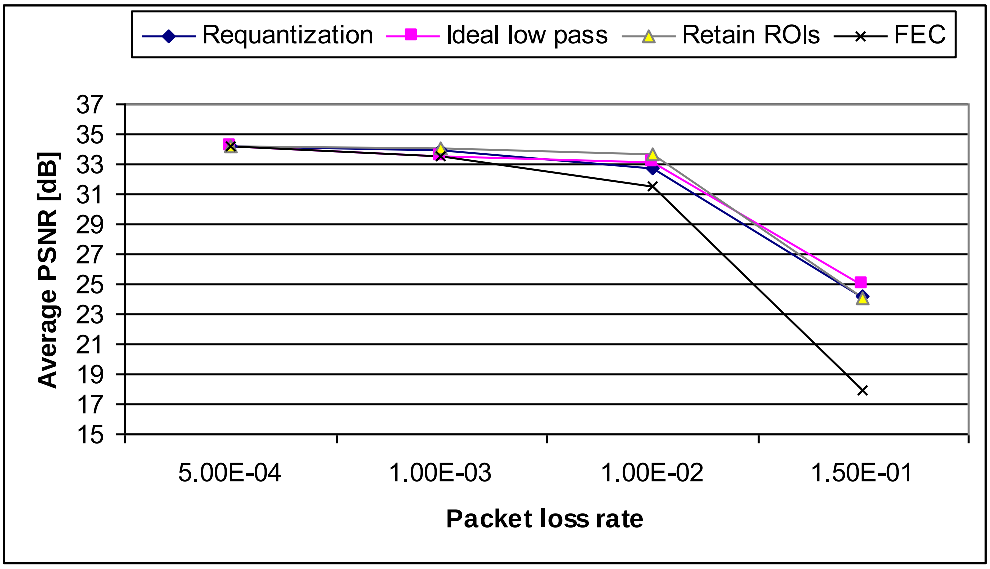
Common to this section, three ER transcoding approaches are employed; reproducing regions of interest through requantization, reproduction through ideal low pass filtering and retaining regions of interest without reproduction. In the requantization approach, once a region of interest is identified, it is scanned for the maximum quantization step size. A bias is added to this maximum and the new step size is then used to reproduce the coarser region of interest. In reproduction through ideal low pass filtering, the number of DCT coefficients to retain is set to 5.

For the requantization and ideal low pass filtering approaches, the 150kbit/s supplementary bit stream accommodates 6 Reed-Solomon parity slices, i.e. each vertical cut through the slice stack is protected by 6 parity bytes. For the case where the regions of interest of I and P slices are retained without reproduction, the 150kbit/s stream accommodated 3 and 2 parity slices for the Football the Foreman sequences accordingly.

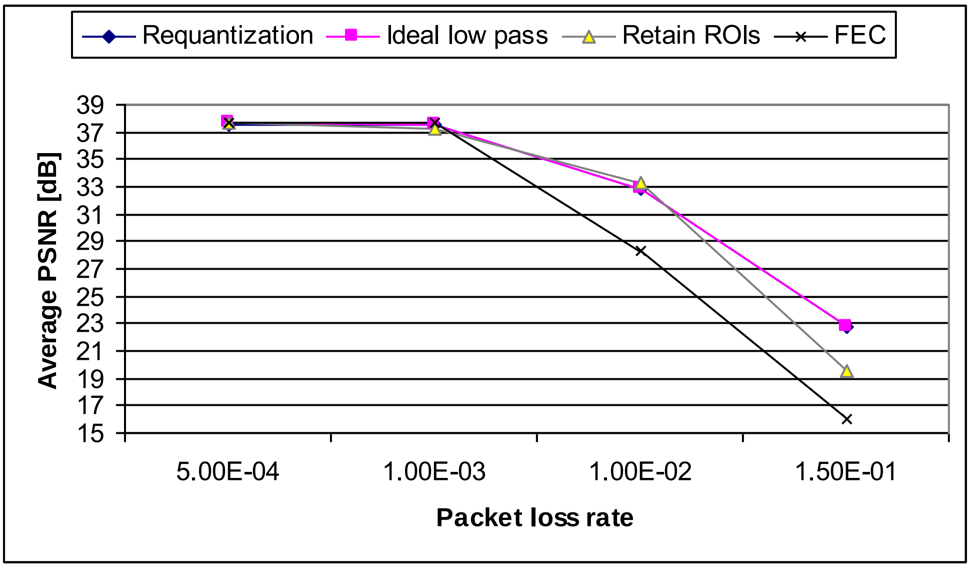
Following the work proposed in [2], we assume that the ER supplementary bit stream is transmitted over a reliable channel, therefore packet losses and bit errors are only applicable for the transmitted video stream.

Figure 12 shows the effect of various packet loss ratios on the average PSNR. Since this is MPEG-2 video then the packet size can be set to one slice. The figure shows that at low loss rates the performance of error recovery is very similar. In these cases packet losses are scattered and therefore even one parity slice is capable of fully recovering a lost region of interest. However as the packet loss rate increases, more than one region of interest per frame can be lost and so a small number of parity slices is insufficient for error recovery. In the Foreman sequence, since only 2 Reed-Solomon parity slices are possible in the 'Retain regions of interest' approach, the error recovery performance dropped at high packet loss rates.

The figure also shows that the recovery performance of the requantization and ideal low pass filtering approaches are very similar. Keep in mind that the latter approach is less complex as discussed in Section 7 and is thus preferred over the requantization approach.



(a) Football test sequence.



(b) Foreman test sequence

Figure 12. Packet loss simulations.

Figure 13 presents the results obtained by using the requantization ER transcoding approach with drift motion compensation. The poor performance of this approach is counter intuitive, but can be justified as follows. When packet loss occurs, non-regions of interest are recovered through simple concealment techniques as described previously. The transcoding drift of the frame that contains the concealed parts is then motion compensated for drift compensation. But in this case a mismatch between the MC buffers of the ER transcoder and the pre-decoder shall appear. This mismatch is unavoidable and causes a drop in picture quality as shown in the figure. The mismatch also means that the reproduced regions of interest at the pre-decoder side might be different than those of the transcoder side. Thus the Reed-Solomon decoder will attempt to correct non-erroneous bits. Add to this deficiency the complexity of the drift compensation approach as illustrated in Table 1 and it can be concluded that this approach is not worth pursuing.

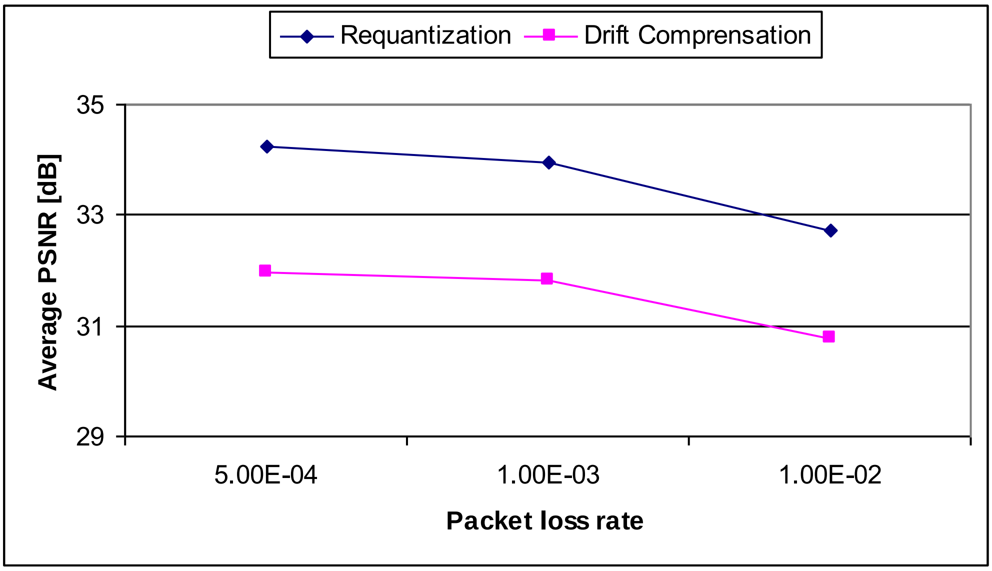
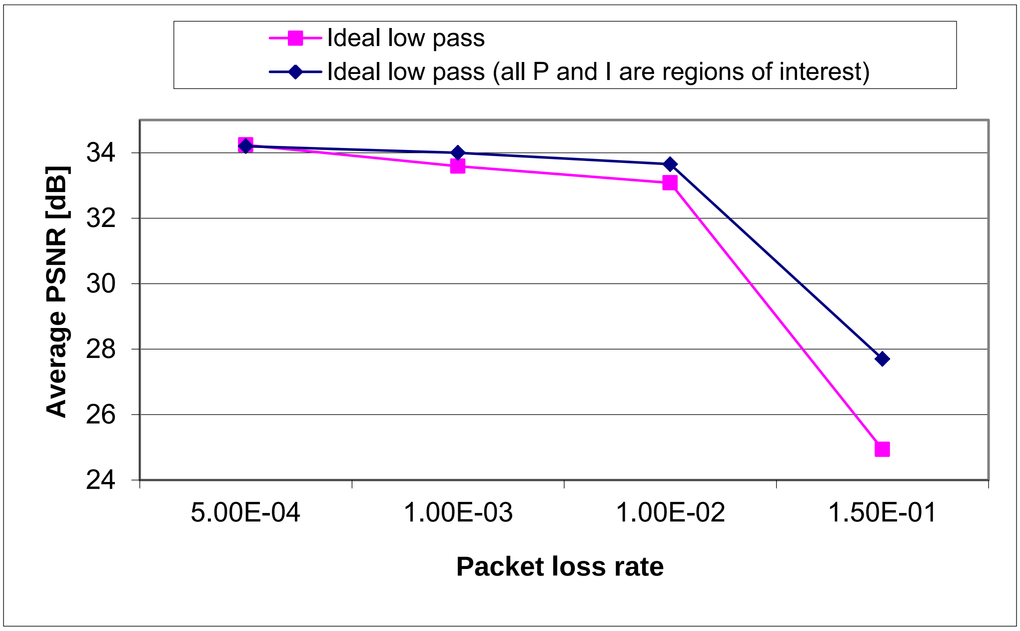
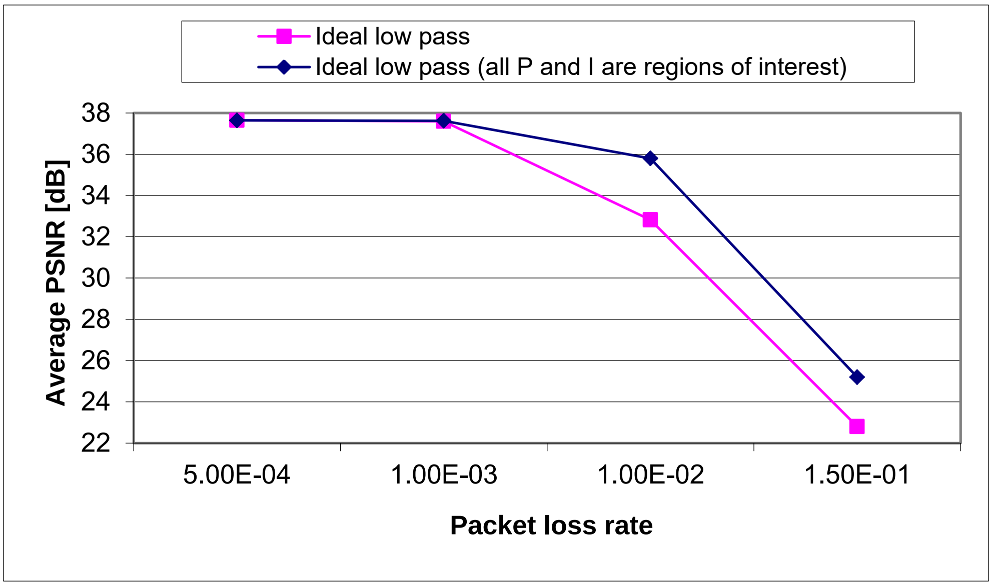


Figure 13. ER transcoding with drift compensation

Figure 14 presents an enhancement to the proposed ER transcoding solutions. Since it was shown that reproducing regions of interest from I and P frames allows for larger Reed-Solomon parity slices then one can design a solution where all I and P slices are reproduced and treated as regions of interest. This solution has an advantage of avoiding the concealment of I and P non-regions of interest at the pre-decoder thus producing higher quality images. To verify this, Figure 14 shows the results of simplifying the ideal low pass filtering approach so that all I and P slices are treated as regions of interest. The results in the figure show the superiority of this solution compared to the case of mixing regions of interest with non-regions of interest in both I and P frames. Another advantage of this approach is the reduction in ER transcoding complexity where regions of interest need not be identified.



(a) Football sequence

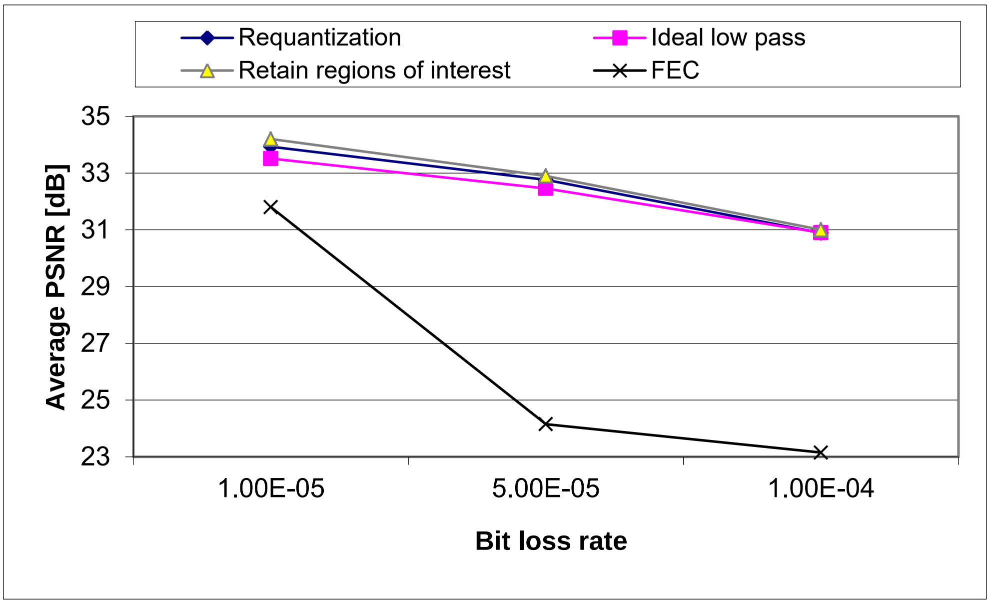


(b) Foreman sequence

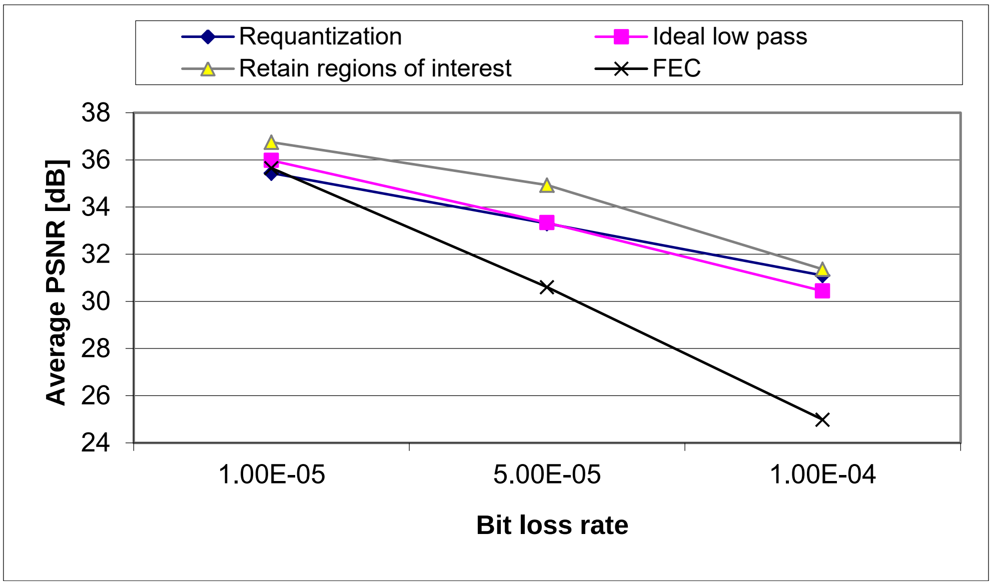
Figure 14. Packet loss simulations for simplified ER transcoding approach.

http://doi.org/10.1016/j.image.2008.05.005

Lastly, Figure 15 presents the effect of bit errors rather than packet losses on the proposed ER transcoding approaches. The figure shows that the relative performance of the approaches in this case is the same as that for packet losses, except for the approach of retaining all of the regions of interest. In Figure 11 it is shown that the combination of bit errors and the nature of VLC coding causes regions of interest to be partially decoded and thus partially reproduced. One bit error can cause desynchronization and cause the loss of a large part of a region of interest. On the other hand if regions of interest are retained without reproduction then bit errors are localized since no reproduction is needed as discussed in Section 6. As such the channel codec is capable of correcting more bits, resulting in a higher decoding quality as evident in Figure 15.



(a) Football sequence



(b) Foreman sequence

Figure 15. Affect of bit errors on proposed ER transcoding approaches.

9. Conclusion:

This paper proposed a novel ER transcoding and pre-decoding solution based on distributed video coding concepts. The main contributions of this work are:

- Applying a distributed video coding error resiliency approach to pre-encoded video.
- Correcting video streams using a pre-decoder rather than an in-decoder or post-decoder approach.
- Extending the concept of distributed coding using ideal low pass filtering of regions of interest and retaining regions of interest without reproduction.

An intermediate solution between reproducing regions of interest and FEC was proposed in which regions of interest are selected from I and P slices without reproduction. This reduces the computational complexity at both the transcoder and the pre-decoder. It was also shown that this approach does not suffer from the effect of propagating bit errors due to the nature of VLC coding. Thus bit errors are localized and error recovery is boosted.

Requantization as a means of reproducing regions of interest was also extended by adding drift compensation loops. However this solution proved to be counter effective in the presence of error concealment. The drift compensation causes a mismatch between the MC buffer content at the transcoder and at the pre-decoder. The solution was also disfavoured due to its high computational complexity.

Finally, it was shown that by treating all I and P slices as regions of interest and reproducing them through ideal low pass filtering, a higher decoding quality can be achieved. This result is caused by the elimination of the need for concealing non-regions of interest. These non-regions of interest cause error propagation when they are used for the reconstruction of future images. The proposed solution further reduces the transcoding complexity by eliminating the need for determining regions of interest.

In future work, it is worthwhile examining bit rate control techniques to adaptively reproduce the coarser representation of the regions of interest. Both requantization and ideal low pass filtering can be carried out adaptively according to the size of the regions of interest such that the generated bitrate-limited channel code would be adequate for error recovery at a given packet loss ratio or bit error rate.

10. References:

[1] B. Girod, A. M. Aaron, S. Rane and D. Rebollo-Monedero, "Distributed video coding," Proc. of the IEEE, 93(1), pp. 71-83, January, 2005.

[2] P. Baccichet, S. Rane, and B. Girod, "Systematic Lossy Error Protection based on H.264/AVC Redundant Slices and Flexible Macroblock Ordering," Journal of Zheijang University, Science A, 7(5), pp. 727-736, May, 2006.

[3] "DLNA overview and vision white paper," Digital living network alliance organization, white paper, available online http://www.dlna.org/en/industry/pressroom, January, 2007.

[4] S. Dogan, A. Cellatoglu, M. Uyguroglu and A.H. Sadka and A. M. Kondoz, "Error-resilient video transcoding for robust internetwork communications using GPRS," IEEE Transactions on Circuits and Systems for Video Technology, 12(6), pp. 453-464, June, 2002

[5] H.-J. Chiou, Y.R. Lee and C.W. Lin "Error-resilient transcoding using adaptive intra refresh for video streaming," Proc. IEEE ISCAS, v 3, pp. 777-780, 2004.

[6] H.-J. Chiou, Y.-R. Lee and C.-W. Lin, "Content-aware error-resilient transcoding using prioritized intra-refresh for video streaming," Journal of Visual Communication and Image Representation, 16(4-5), pp. 563-588, August/October, 2005

[7] Chen, C., Lin, C., and Chen, Y. "Adaptive error-resilience transcoding using prioritized intra-refresh for video multicast over wireless networks," Image Commun. 22(3), 277-297. DOI= http://dx.doi.org/10.1016/j.image.2006.12.010, March, 2007

[8] C.-M. Chen, Y.-C. Chen, and C.-W. Lin, "Error-resilience transcoding using content-aware intra-refresh based on profit tracing," Proc. of IEEE International Symposium on Circuits and Systems (ISCAS '06), pp. 5283–5286, Kos, Greece, May, 2006

[9] S. Eminsoy, S. Dogan, and A. M. Kondoz, "Transcoding-Based Error-Resilient Video Adaptation for 3G Wireless Networks ," EURASIP Journal on Advances in Signal Processing Volume 2007, Article ID 39586, 13 pages doi:10.1155/2007/39586

[10] J. Seong and N. Hwan, "An efficient error resilient technique for applications of one-way video using transcoding and analysis by synthesis," Proc. GLOBECOM*,* pp. 428-432, 2004

[11] A. Vetro, J. Xin, and H. Sun, "Error-resilience video transcoding for wireless communications," IEEE Wireless Communications, 12(4), pp. 14-21, Aug. 2005.

[12] T. Gan and K.-K. Ma, "Sliding-window packetization for forward error correction based multiple description transcoding," Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), vol. 5, pp. 756-759, 2003.

[13] C. M. Chen, C. W. Lin, H. C. Wei, and Y. C. Chen, "Robust Video Streaming over Wireless LANs Using Multiple Description Transcoding and Prioritized Retransmission," Journal of Visual Communication and Image Representation, 18(3), pp.191-206, 2007.

[14] C.-Y. Lin and S.-F. Chang, "A robust image authentication method distinguishing JPEG compression from malicious manipulation," IEEE Trans. on Circuits and Systems for Video Technology, 11(2), February, 2001.

[15] P. Baccichet and A. Chimienti, "Forward Selective Protection Exploiting Redundant Slices and FMO in H.264/AVC," Proc. IEEE International Conference on Image Processing (ICIP), Atlanta, USA, 2006

[16] S. Aniruddha, A. Gaurav and A. Alwin, "Region-of-interest based compressed domain video transcoding scheme," Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), v 3, pp. 161-164, 2004.

[17] H. Li, Y. Wang, and C. Chen, "An Attention-Information-Based Spatial Adaptation Framework for Browsing Videos via Mobile Devices," EURASIP Journal on Advances in Signal Processing, vol. 2007, Article ID 25415, 12 pages, 2007. doi:10.1155/2007/25415

[18] P. A. A. Assuncao and M. Ghanabari, "A frequency-domain video transcoder for dynamic bit-rate reduction of MPEG-2 bit streams," IEEE Trans. Circuits and Systems for Video Technology, 8(8), pp. 953-967, December, 1998

[19] T. Shanableh and M. Ghanbari "Multilayer Transcoding with format portability for multicasting of single-layered video," IEEE Trans. on multimedia, 7(1), pp. 1-15, February, 2005.

[20] J. Xin, C.-W. Lin, and M.-T. Sun, "Digital video transcoding," Proceedings of the IEEE, 93(1), pp. 84-97, January, 2005.

[21] S. Rane, A. Aaron and B. Girod, "Systematic lossy forward error protection for error-resilient digital video broadcasting - A Wyner-Ziv coding approach," Proc. IEEE ICIP, Singapore, vol. 5, pp. 3101- 3104, October, 2004.

[22] Software implemented by Henry Minsky. Available on http://sourceforge.net/projects/rscode.