



**HAL**  
open science

# A New H.264/AVC Error Resilience Model Based on Regions of Interest

Fadi Boulos, Wei Chen, Benoît Parrein, Patrick Le Callet

► **To cite this version:**

Fadi Boulos, Wei Chen, Benoît Parrein, Patrick Le Callet. A New H.264/AVC Error Resilience Model Based on Regions of Interest. Packet Video, May 2009, Seattle, Washington, United States. pp.1-9, 10.1109/PACKET.2009.5152159 . hal-00404333

**HAL Id: hal-00404333**

**<https://hal.science/hal-00404333>**

Submitted on 16 Jul 2009

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A NEW H.264/AVC ERROR RESILIENCE MODEL BASED ON REGIONS OF INTEREST

*Fadi Boulos, Wei Chen, Benoît Parrein and Patrick Le Callet*

Nantes Atlantique Universités  
IRCCyN, Polytech’Nantes  
Rue Christian Pauc, 44306 Nantes, France  
*firstname.lastname@univ-nantes.fr*

## ABSTRACT

Video transmission over the Internet can sometimes be subject to packet loss which reduces the end-user’s Quality of Experience (QoE). Solutions aiming at improving the robustness of a video bitstream can be used to subdue this problem. In this paper, we propose a new Region of Interest-based error resilience model to protect the most important part of the picture from distortions. We conduct eye tracking tests in order to collect the Region of Interest (RoI) data. Then, we apply in the encoder an intra-prediction restriction algorithm to the macroblocks belonging to the RoI. Results show that while no significant overhead is noted, the perceived quality of the video’s RoI, measured by means of a perceptual video quality metric, increases in the presence of packet loss compared to the traditional encoding approach.

**Index Terms**— Eye tracking, region of interest, packet loss, error resilience, perceptual quality.

## 1. INTRODUCTION

With Internet becoming the cheapest and preferred medium of communication, video traffic over IP is in constant and sharp increase. On one hand, this is made possible by the increasing broadband speeds and the variety of multimedia services offered by Internet Service Providers, *e.g.*, triple-play offers. On the other hand, recent video coding standards such as H.264/AVC [1] allow compression rates for up to twice those of their predecessors, thus making it possible to stream Standard Definition (SD) or High Definition (HD) video contents over the Internet.

However, packet loss still characterizes the best-effort Internet. To overcome this problem, several solutions have been proposed at both the channel and source levels. Mechanisms like Forward Error Correction (FEC) or Automatic Repeat reQuest (ARQ) can be used to compensate for the packets lost during transmission. At the source level, some error resilience features used during the encoding process can help in attenuating the impact of packet losses. In the H.264/AVC standard, Flexible Macroblock Ordering (FMO), Data Partitioning (DP) and Redundant Slice (RS) are examples of error

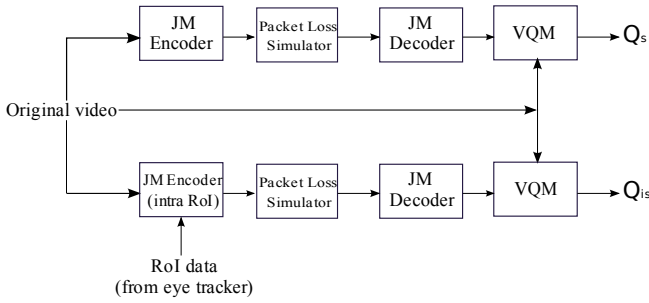
resilience features.

Due to the very nature of video compression techniques, a packet lost from the encoded bitstream has generally a spatio-temporal propagating effect. This is largely due to spatio-temporal dependency between parts of the bitstream. In an earlier study [2], we showed that the following two parameters have a great impact on the perceived quality: (1) the spatial position of the loss in the picture, *i.e.*, if it belongs or not to the Region of Interest (RoI) of the picture; and (2) the temporal position of the loss in the sequence. In this work, we propose to link these two parameters to prevent the error from propagating to the RoIs of the video sequence. To this end, we force the macroblocks that belong to an RoI to be coded in intra-prediction mode, thus removing their temporal dependency on macroblocks in other pictures. We also propose an extension to this algorithm to remove the spatial dependency.

The outline of the paper is as follows: in Section 2, we give an overview of some state-of-the-art RoI-based error resilience models. Then we describe the eye tracking tests we performed in order to determine the saliency maps of SD and HD video sequences in Section 3. We also present the methodology of transforming these maps into RoIs. In Section 4, we propose several variants of our method to take advantage of the RoIs by forcing RoI macroblocks to be coded in intra-prediction mode. We validate our method by simulating packet losses in the RoI and assessing the perceived video quality by means of a perceptual video quality metric, namely VQM [3]. The block diagram of the whole processing chain, from encoding to performance evaluation is depicted in Figure 1.

## 2. RELATED WORK

While RoI-based video coding is probably the most used application for deriving RoIs, *e.g.*, in [4, 5], using the RoI in error resilience models has also been investigated. In all of the following research works, RoI-based models have been coupled with an H.264/AVC error resilience feature, namely



**Fig. 1.** Block diagram of our work.

FMO. FMO allows the ordering of macroblocks in slices according to a predefined map rather than using the raster scan order, *e.g.*, to improve the robustness of the video against transmission errors or to apply an Unequal Error Protection (UEP). When coupled with RoI-based coding, FMO is generally used to assemble the RoI macroblocks into a single slice.

In [6], multiple RoI models are proposed to enhance the quality of video surveillance pictures. The RoIs are defined by the user in an interactive way: the mouse pointer is put over the RoI and the coordinates of the pixel position pointed to are transmitted to the encoder. Then, every model will build its own RoI (*e.g.*, square-shaped, diamond-shaped) coupled with an FMO type. For more information about FMO types, the reader is referred to [7]. Results show that a convenient selection of the RoI shape, the quantization parameter and the FMO type can reduce bandwidth usage while maintaining the same video quality.

In [8], an error resilience model where RoIs are derived on a per picture basis is proposed. Each picture RoI is determined by simulating slice losses and corresponding error concealment at the encoder to build a distortion map. Macroblocks with the highest distortion values are coded into Redundant Slices (RS), which is another H.264/AVC resilience feature and background macroblocks are signalled using FMO type 2. Simulation results demonstrate the efficiency of this method compared to the traditional FEC approach in the presence of packet loss.

The work reported in [9] aims at improving the robustness of the video by applying UEP wherein the RoI benefits from an increased protection rate along with a checkerboard FMO slicing. The authors conclude that their approach outperforms UEP and Equal Error Protection (EEP) for lower Signal-to-Noise-Ratio (SNR) values.

A robustness model for RoI-based scalable video coding is proposed in [10]. The model divides the video into two layers: the RoI layer and the background layer. Dependencies between the two layers are removed to stop the error from propagating from the background layer, which is less protected than the RoI layer, to the latter in case of packet loss. This process decreases coding efficiency in error-free environments but enhances the video robustness in the presence

of packet loss.

### 3. EYE TRACKING TESTS

The goal of performing eye tracking tests is to record the eye movement of the viewers while they are watching video sequences. These data can then be used to achieve a RoI-based error resilient video coding. In this section, we first describe the test setup and the set of videos used then we explain how the RoIs are generated for each video sequence. We also present the results of the tests and discuss them.

#### 3.1. Setup

We use a dual-Purkinje eye tracker (Figure 2) from Cambridge Research Systems. The eye tracker is mounted on a rigid EyeLock headrest that incorporates an infrared camera, an infrared mirror and two infrared illumination sources. The camera records a close-up image of the eye. To obtain an accurate measure of the subject’s pupil diameter, a calibration procedure is needed. The calibration requires the subject to stare at a number of screen targets from a known distance. Once the calibration procedure is completed and a stimulus has been displayed, the system is able to track a subject’s eye movement. Video is processed in real-time to extract the spatial location of the eye’s position. Both Purkinje reflections (from the two infrared sources) are used to calculate the location. The sampling frequency is 50 Hz and the tracking accuracy is in the range 0.25 – 0.5 degree.

The video testbed contained 30 SD and 38 HD source sequences, 23 of which were common to both resolutions. These 23 720 × 576 SD sequences were obtained from 1920 × 1080 HD by cropping the central region of the picture (220 pixels from right and left borders) and resampling the obtained video using a Lanczos filter. Several loss patterns were applied to 20 SD and 12 HD source sequences, thus increasing the total number of videos to 100. The sequences had either an 8-second or a 10-second duration. To ensure that the RoI extraction is faithful to the content independently of other parameters, all of the sequences were encoded such as to obtain a good video quality. Bitrates were in the range of 4 – 6 Mbs and 12 – 16 Mbs for SD and HD sequences, respectively. The video sequences were encoded in High Profile with an IBBPBBP... GOP structure of length 24. The JM 14.0 [11] encoder and decoder were used.

We formed two sub-tests of 50 videos each to avoid having a content viewed twice (in both resolutions) by the same subject during a sub-test, which could skew the RoI deriving process. We also randomized the presentation order within the sub-test.

Eye tracking data of 37 non-expert subjects with normal vision (or corrected-to-normal vision) were collected for every video sequence. The test was conducted according to the International Telecommunication Union (ITU) Recommenda-



**Fig. 2.** The eye tracker.

tion BT.500-11 [12]. Before starting the test, the subject's head was positioned so that their chin rested on the chin-rest and their forehead rested against the head-strap. Subjects were seated at a distance of  $3H$  and  $6H$  for HD and SD sequences, respectively. All sequences were viewed on an LCD display. The average sub-test completion time was 25 minutes.

### 3.2. RoI generation

A saliency map describes the spatial locations of the eye gaze over time. To compute a saliency map, the eye tracking data are first analyzed in order to separate the raw data into fixation and saccade periods. Fixation is defined as being the status of a region centered around a pixel position which was stared at for a predefined duration. Saccade corresponds to the eye movement from one fixation to another. The saliency map can be computed for each observer and each picture using two methods. The first method is based on the number of fixations (NF) for each spatial location; hence, the saliency map  $SM_{NF}^{(k)}$  for viewer  $k$  is given by:

$$SM_{NF}^{(k)}(x, y) = \sum_{j=1}^{N_{FP}} \delta(x - x_j, y - y_j)$$

where  $N_{FP}$  is the number of fixation periods and  $\delta$  is the Kronecker function. Each fixation has the same weight.

The second method is based on the fixation duration (FD) for each spatial location. The saliency map  $SM_{FD}^{(k)}$  for viewer  $k$  is given by:

$$SM_{FD}^{(k)}(x, y) = \sum_{j=1}^{N_{FP}} \delta(x - x_j, y - y_j) \cdot d(x_j, y_j)$$

where  $d(x_j, y_j)$  is the fixation duration at pixel  $(x_j, y_j)$ .

To determine the most visually important regions, all saliency maps are merged yielding an average saliency map  $SM$ . The average saliency map is given by:

$$SM(x, y) = \frac{1}{K} \sum_{k=1}^K SM^{(k)}(x, y)$$

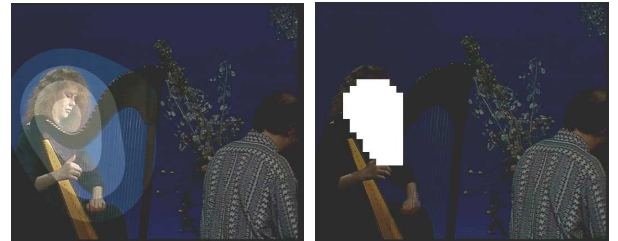
where  $K$  is the total number of viewers.

Finally, the average saliency map is smoothed with a 2D Gaussian filter which gives the density saliency map:

$$DM(x, y) = SM(x, y) * g_{\sigma}(x, y)$$

where the standard deviation  $\sigma$  depends on the accuracy of the eye tracking device.

To generate RoI maps from saliency maps, some parameters need to be set. The parameters are: *fixation duration threshold*, *fixation velocity threshold*,  $\sigma$  and the *RoI threshold*. The *fixation duration threshold* (in milliseconds) is the minimal time a region must be viewed for it to be considered as a fixation region. The *fixation velocity threshold* (in degrees per second) is the eye movement velocity threshold below which the velocity must remain for *fixation duration threshold* ms. *RoI threshold* is the minimal number of viewers who must view a region to be considered as a fixation region. The values of *fixation duration threshold*, *fixation velocity threshold* and  $\sigma$  were 200 ms, 25 °/s and 1.5, respectively, for both resolutions while the *RoI threshold* values were 4 and 2 for SD and HD sequences, respectively. Examples of saliency and RoI maps obtained with these parameter values are illustrated in Figures 3 and 4.



**Fig. 3.** Saliency map (left) and the resulting macroblock-based RoI (right) of *Harp* sequence.

### 3.3. Results and discussion

Depending on its saliency map, a video sequence can have or not an RoI. In Figure 3, it is clear that the harpist in the video attracts the visual attention of the viewers. By contrast, the sparse saliency regions in Figure 4 do not result in any RoI for this particular video sequence.

We draw two main conclusions from the eye tracking test data:



**Fig. 4.** Saliency map of *Marathon* sequence. This saliency map does not yield any RoI.

- The RoI of a video content is identical for both SD and HD resolutions.
- In the presence of a packet loss, the spatial position of the RoI can change depending on several parameters stated below.

While the first conclusion is somewhat expected, the second is worth discussing. The same packet loss pattern was applied to all the sequences to be impaired. The loss pattern consisted of losing five slices from the 6th I-picture of the sequence, two of them being the first two slices of the picture and the three others its last three. Thus, the Packet Loss Rate (PLR) in the I-picture was in the ranges 5% – 20% and 2% – 5% for SD and HD sequences, respectively.

The losses having occurred in the top and bottom regions of the picture, they were not generally in its RoI. In a video sequence having a clear RoI (*e.g.*, the ball in a football game, the face in a head-and-shoulder scene), a loss in an unimportant region of the picture might not be perceived by the user, whose attention is focused on the action in the RoI. However, when there is no clear RoI, any small loss may attract the user’s attention. The nature of the scene content also influences the perception of a loss outside the RoI. While this topic deserves to be investigated more deeply, it is not covered by the scope of this paper.

## 4. RESILIENCE MODEL

We implement an RoI-based error resilience model in the H.264/AVC encoder. The model reduces the dependencies between important and unimportant regions of the picture. To test its efficiency, we perform a controlled packet loss simulation on the encoded bitstream. We then decode the distorted bitstream and evaluate the quality of the decoded video using a perceptual quality metric.

### 4.1. Forced intra-prediction

In order to prevent error propagation from reaching the RoI in B and P-pictures, we propose to force the macroblocks in an

RoI in these pictures to be predicted in intra-prediction mode. This makes the RoI independent from past or future pictures. To this end, we implement in the JM encoder an algorithm that operates as follows: for each macroblock of a B or P-picture, it checks if the macroblock belongs to the picture’s RoI by comparing its coordinates to the coordinates given to the encoder in the form of an RoI text file. When a macroblock is flagged as being an RoI macroblock, its prediction type is forced to be intra. The selection of a macroblock’s prediction type in H.264/AVC being based on the minimization of a distortion measure between the original and the predicted pixels, we choose to force the encoding algorithm to change the prediction type of an RoI macroblock (from inter to intra) by increasing the distortion measure computed for this choice. The process is illustrated in Figure 5(a) and the pseudocode of the algorithm is given below.

---

#### Algorithm 1 Forced RoI intra-prediction.

---

```

while reading(eye tracking data)
  for all B and P-pictures
    for all MBs in a picture
      if MB ∈ RoI then
        while (predType == anyInterpredType) do
          increase cost function
        end while
      else proceed normally
      end if
    end for
  end for
end while

```

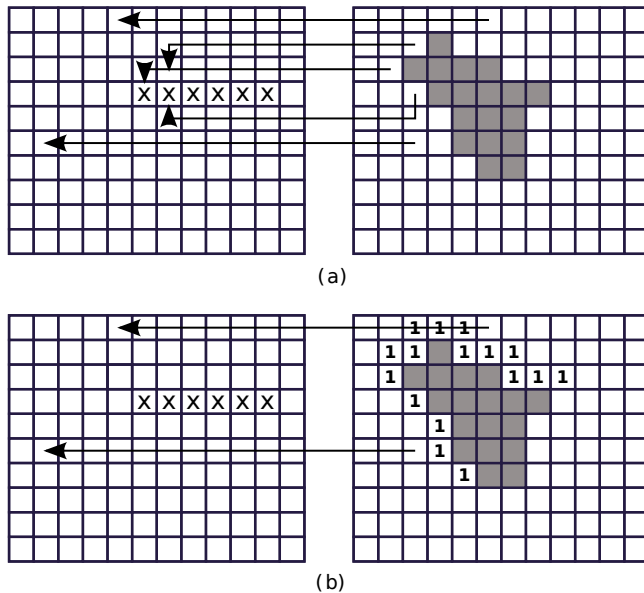
---

By forcing some macroblocks in B and P-pictures to be coded in intra mode, the quality of the video may decrease. The explanation is the following: the encoding is done at a constant bitrate, thus the number of bits to be consumed is the same for all coding schemes. The intra-coded macroblocks consume more bits than they would have if they were coded in inter-prediction mode. This results in higher quantization parameters for some macroblocks in the video.

Forcing the RoI macroblocks to be coded in intra does halt completely the temporal error propagation but not the spatial propagation. In fact, the macroblocks to the top and left of an RoI macroblock in a B or P-picture might not belong to the RoI. Thus, these macroblocks can be coded in inter-prediction mode. If the reference macroblocks of these inter-predicted macroblocks are lost and/or concealed, the error can propagate to these macroblocks. The intra-prediction being spatial, using these macroblocks to perform the intra-prediction propagates the error to the RoI.

We propose to extend the algorithm in order to cope with this situation. We establish a “security neighborhood” around each RoI macroblock to attenuate the impact of using as a reference, a macroblock using itself a lost and/or concealed reference. This security neighborhood consists of coding the

top left, top center, top right and left macroblocks surrounding an RoI macroblock in intra-prediction mode. If one of those four macroblocks is in the RoI, it is not considered as it will definitely be intra-coded. This is best illustrated in Figure 5(b) where we can see that some of the dependencies in Figure 5(a) have disappeared.



**Fig. 5.** RoI intra-coding. Gray and marked macroblocks are RoI and lost macroblocks, respectively. Arrows indicate inter-predictions. (a) Raw algorithm. (b) Algorithm with security neighborhood (“1” macroblocks).

## 4.2. Loss simulation

We use a modified version of the loss simulator in [13] to generate the transmission-distorted bitstreams. This simulator provides the possibility of finely choosing the exact slice to lose in the bitstream. At the encoding side, we take some practical considerations into account, namely we set the maximum slice size to 1450 bytes which is less than the Maximum Transmission Unit (MTU) for Ethernet (1500 bytes). The unused bytes (*i.e.*, 50) are left for the RTP/UDP/IP headers (40 bytes) and the possible additional bytes that could be used beyond the predefined threshold. In this case, every Network Abstraction Layer Unit (NALU), which contains one slice of coded data can fit in exactly one IP packet. This makes our simulation more realistic because we can map the Packet Loss Rate (PLR) at the NAL level to the PLR at the application layer (*e.g.*, RTP).

In our simulation, we never lose an entire picture; rather, we lose  $M$  slices of a picture where  $M < N$ ,  $N$  being the total number of slices in the picture. When parts of a picture are

lost due to packet loss, the error concealment algorithm implemented in the JM decoder is executed. This non-normative algorithm performs a weighted sample averaging to replace each lost macroblock in an I-picture and a temporal error concealment (based on the motion vectors) for lost macroblocks in a B or P-picture. The algorithm is described in detail in [14].

Because the macroblocks of an RoI are not confined in one slice, the RoI generally spans over three or more slices. To lose part or all of the RoI, we simulate the loss of three and five slices in the RoI of the 5th I-picture to evaluate the error propagation impact on quality. We use two loss patterns for quality evaluation: three contiguous slices and five non-contiguous slices, all containing RoI macroblocks. The goal of the loss simulation is to test the efficiency of our approach w.r.t. error propagation. Thus, we only target I-slices and look at how the algorithm copes with the spatio-temporal error propagation in subsequent B and P-pictures.

## 4.3. Quality assessment

To assess the quality of a video sequence one could either perform subjective quality tests or use an objective quality metric. During a subjective test, a group of viewers is asked to rate the quality of a series of video sequences. The quality score is chosen from a categorical (*e.g.*, bad, excellent) or numerical (*e.g.*, 1–5, 1–100) scale. An objective video quality metric evaluates the quality of a processed video sequence by performing some computations on the processed video and often on the original video too. While subjective tests are the most reliable way of assessing the quality of video sequences, they are time-consuming and require a large number of participants. Hence, a number of objective metrics providing a reliable quality assessment have been proposed to replace the subjective tests. The most widely used objective quality metric is the Peak Signal-to-Noise Ratio (PSNR). However, the performance of the PSNR metric is controversial [15]. Therefore, we propose to use in this work a perceptual video quality metric: VQM.

### 4.3.1. Video Quality Metric

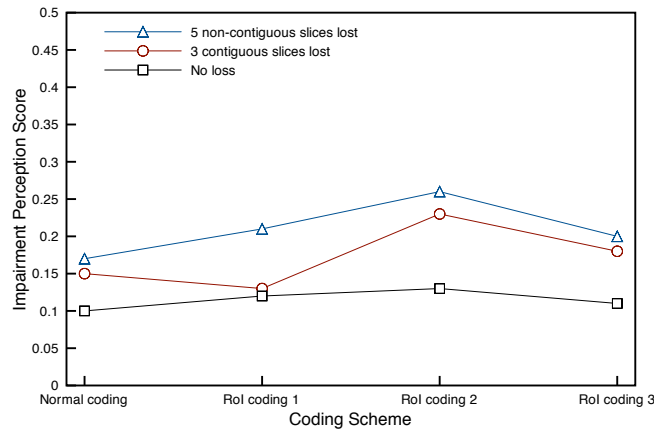
Video Quality Metric is a standardized objective video quality metric developed by the Institute for Telecommunication Sciences (ITS), the engineering branch of the National Telecommunications and Information Administration at the U.S. Department of Commerce [16]. VQM divides original and processed videos into spatio-temporal regions and extracts quality features such as spatial gradient, chrominance, contrast and temporal information. Then, the features extracted from both videos are compared, and the parameters obtained are combined yielding an impairment perception score. The impairment score is in the range 0–1 and can be mapped to the Mean Opinion Score (MOS) given by a panel of human observers during subjective quality tests. For example, 0.1 and

0.7 are mapped to MOS values of 4.6 and 2.2 on a 5-grade scale, respectively. Note that VQM can be applied over a selected spatio-temporal region of the video to assess exactly its quality. In our experiments, we used the VQM Television model which is optimized for measuring the perceptual effects of television transmission impairments such as blur and error blocks.

### 4.3.2. Results

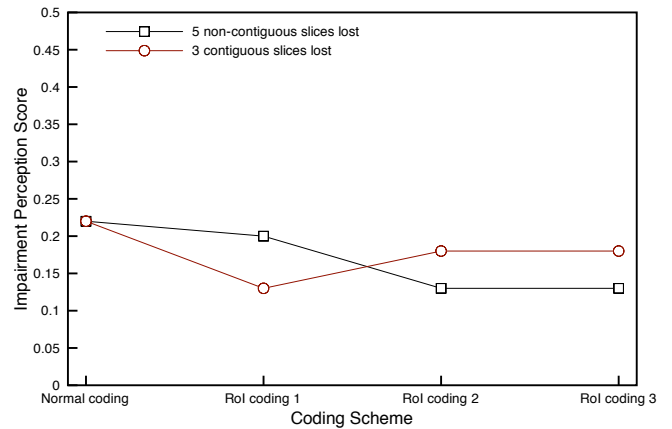
We compare the quality of the same sequence (*Harp*) encoded with the unmodified JM encoder and with three variants of our algorithm: the first one is the classical approach, denoted hereafter by *RoI coding 1*. *RoI coding 2* denotes our approach taking into account the security neighborhood. We also implemented a variant *RoI coding 3* of our approach that considers P-pictures only. The slices are lost from frame 97 which is the 5th I-picture of the sequence.

In Figure 6, the VQM impairment perception scores are plotted for each coding scheme. These scores are computed over the full spatial and temporal resolutions of the video. For the no loss case, the encoding quality of all schemes is practically the same. For the two loss patterns, the impairment generally seems to be more annoying when using the variants *RoI coding 2* and *RoI coding 3* of our algorithm.



**Fig. 6.** VQM impairment perception scores for all coding schemes. VQM is applied over the full temporal and spatial resolutions of the video.

This trend is inverted in Figure 7, where the impairment perception scores are computed only over the RoI of the picture. Only the scores of the two loss patterns are plotted in this figure. The results show that all three variants of the algorithm outperform the *Normal coding* approach, although sometimes slightly. We also note that the impairment perception score for the 3-slice loss is greater than the 5-slice loss for *RoI coding*



**Fig. 7.** VQM impairment perception scores for all coding schemes. VQM is applied over the full temporal resolution and the RoI of the video.

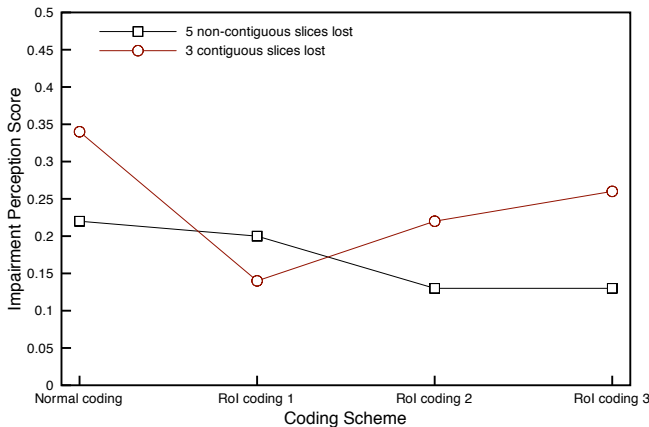
2 and *RoI coding 3*. This probably results from the content-dependency of the error concealment algorithm because the two loss patterns hit different slices.

Figure 8 depicts the impairment scores calculated over a smaller temporal region, namely 100 consecutive frames. The spatial dimensions of the evaluated region are delimited by the RoI of the picture as in the previous case. Reducing the temporal length of the region to be evaluated makes the distortion impact measure more accurate. The length is chosen such as to cover the GOP that contains the I-picture where the slices are hit and following B and P-pictures. We select 100 frames because VQM requires a temporal region of at least four seconds. The scores in Figure 8 demonstrate the efficiency of the proposed algorithm on the local region where it performs. For instance, the impairment perception score of the 3-slice loss case is 0.34 for *Normal coding* scheme while it decreases to 0.14 for *RoI coding 1*.

### 4.3.3. Discussion

The test results show that the video quality for all loss patterns was generally in the higher values of the quality scale (lower impairment perception scores). Visually, this was not always true. Some distorted frames of the *Harp* sequence given in Figure 9 illustrate this claim. This is probably due to the VQM not being perfectly adapted for RoI-based losses, i.e., no special weights are attributed to the losses in the RoI during error pooling.

Figures 7 and 8 clearly demonstrate that *RoI coding 1* is the most adapted coding scheme in the presence of bursty losses



**Fig. 8.** VQM impairment perception scores for all coding schemes. VQM is applied over 100 frames and the RoI of the video.

while *RoI coding 3* works best for single losses. The almost equal impairment scores given by VQM for the no loss case in Figure 6 show that the algorithm used does not incur a significant extra encoding cost, namely a quality decrease. Results for the 3-slice and 5-slice losses in this same figure might indicate that the algorithm fails to cope with the loss patterns. However, the error propagation for *Normal coding* and *RoI coding 1* schemes, illustrated in Figure 9, shows that our algorithm performs well in the presence of losses. Further, while the error propagation is progressively attenuated in *Normal coding* scheme, it is drastically reduced in *RoI coding 1* starting from frame 99 which is 2 frames away from the I-picture hit.

In Figure 10, two differently encoded versions of frame 103 of *Harp* sequence are depicted. The green box indicates the RoI of the picture. While the shape of the face is generally preserved when using the RoI intra-coding scheme (Figure 10(b)), we can see clearly that this is not the case with *Normal coding* (Figure 10(a)). The block effect appearing in the RoI of the picture for *RoI coding 1* scheme is due to the spatial dependency between the macroblocks adjacent to the RoI and the RoI macroblocks. When coding a macroblock in intra-prediction mode, the encoder checks if any of the upper and left macroblocks are available (*i.e.*, existing or coded in intra mode). If no macroblock is available, it uses the DC intra-prediction mode (intra-coding mode 2) which computes the average of the upper and left samples. The upper and left macroblocks being inter-predicted from a lost and/or concealed reference, a block distortion appears in the RoI. The high impairment perception scores obtained in Figure 6 could be due to the fact that VQM penalizes the block effects much more than other distortions. Note that using a security neighbor-

hood did not show a significant improvement over *RoI coding 1*.

To overcome the spatial error propagation limitation, and in an RoI-based video coding perspective, we propose to create a new RoI-based prediction type that will be applied to all RoI macroblocks which do not have an upper or left available RoI macroblock. In this case, the RoI macroblock would be intra-coded as if it were the top left macroblock of the picture. However, doing so will render the bitstream non-standard compliant. To counter this problem, we suggest to use a specific signalling for this prediction type. In the worst case, this scheme would create an overhead of one bit per intra-coded macroblock to signal this new intra-prediction type. Intra-coded macroblocks in a sequence comprise all macroblocks of I-pictures, the occasional intra-coded macroblocks and RoI macroblocks of B and P-pictures. We believe that this slight modification in the H.264/AVC can greatly improve the robustness of the bitstream against packet loss while not incurring a significant overhead. This new coding scheme can be thought of as a “cheap” FMO (in terms of overhead) because it creates totally independent regions in the picture.

On the other hand, to improve the security neighborhood variant which did not significantly increase the resilience model performance, we propose to force any RoI macroblock with at least one available RoI macroblock for intra-coding to use it as a reference to avoid that it uses a non-RoI macroblock.

## 5. CONCLUSION AND FUTURE WORK

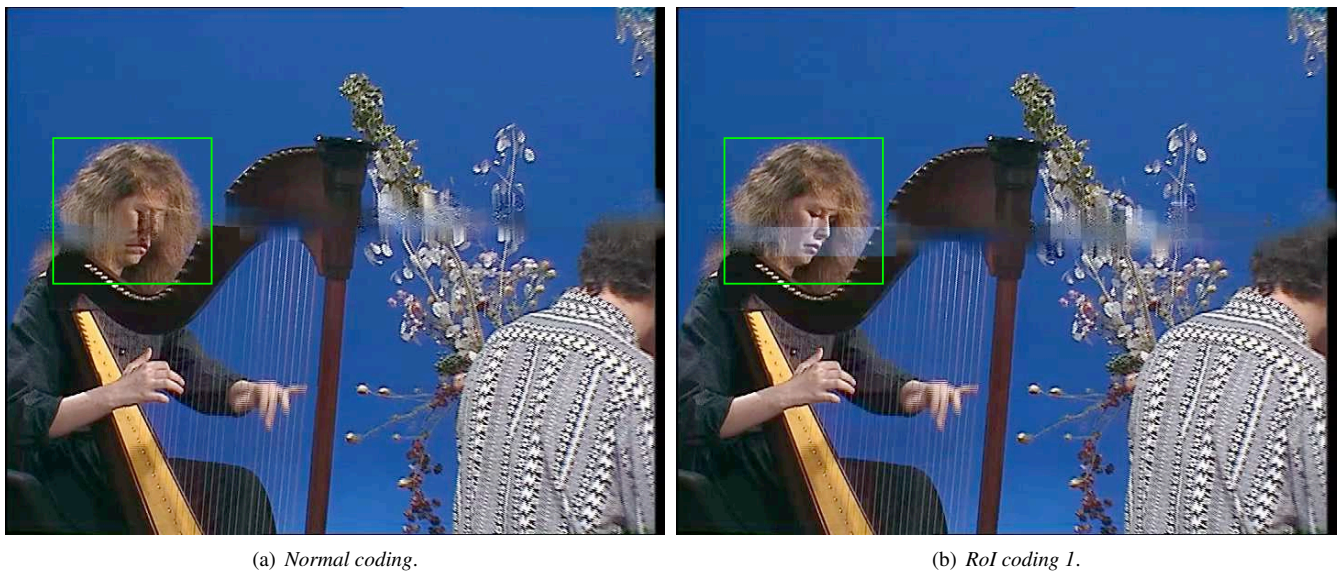
We presented in this paper a new H.264/AVC error resilience model implemented in the encoder. The model, which is based on RoI data collected through an eye tracking experiment, aims at removing the dependencies between the RoIs in B and P-pictures and the reference picture(s). We described the test procedure and the post-processing that is applied to the saliency maps in order to obtain the RoI maps. We tested the efficiency of the proposed model by simulating packet loss on RoI-encoded video sequences and then evaluating their perceived quality. Results show that the RoI intra-coding algorithm outperforms the normal encoding locally and preserves the shape of the RoI.

This work can be further improved by (1) completely removing the spatial dependency between RoI macroblocks and adjacent non-RoI macroblocks; (2) performing subjective tests for video quality assessment; and (3) incorporating an objective saliency computational model (*e.g.*, [17]) in the encoder which would steer the intra-prediction restriction algorithm. If the model chosen is reliable, it could be used as an alternative to eye tracking tests which are expensive in terms of both time and human resources. We are also working towards the development of an RoI-based UEP scheme.





**Fig. 9.** From left to right: frames 97, 99 and 116 of *Harp* sequence. Top: *Normal coding*. Bottom: *RoI coding 1*. The green box indicates the RoI.



**Fig. 10.** Frame 103 of *Harp* sequence. The shape of the RoI is better preserved in (b) than in (a).

## Acknowledgment

The authors would like to thank Romuald P epion for setting up the eye tracking tests and helping in processing their results.

## 6. REFERENCES

- [1] International Telecommunication Union-Standardization Sector. Advanced video coding for generic audiovisual services. ITU-T Recommendation H.264, November 2007.
- [2] F. Boulos, D. S. Hands, B. Parrein, and P. Le Callet. Perceptual Effects of Packet Loss on H.264/AVC Encoded Videos. In *Fourth International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, January 2009.
- [3] International Telecommunication Union. Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference. ITU-T J.144 & ITU-R BT.1683, 2004.
- [4] D. Agrafiotis, D. R. Bull, N. Canagarajah, and N. Kannonwatana. Multiple Priority Region of Interest Coding with H.264. In *Proceedings of IEEE International Conference on Image Processing, ICIP*, pages 53–56, October 2006.
- [5] Y. Dhondt, P. Lambert, S. Notebaert, and R. Van de Walle. Flexible macroblock ordering as a content adaptation tool in H.264/AVC. In *Proceedings of SPIE Multimedia Systems and Applications VIII*, volume 6015, October 2005.
- [6] S. Van Leuven, K. Van Schevensteen, T. Dams, and P. Schelkens. *An Implementation of Multiple Region-Of-Interest Models in H.264/AVC*, volume 31 of *Multimedia Systems and Applications*, pages 215–225. Springer US, 2006.
- [7] P. Lambert, W. De Neve, Y. Dhondt, and R. Van de Walle. Flexible macroblock ordering in H.264/AVC. *Elsevier Journal of Visual Communication and Image Representation*, 17(2):358–375, April 2006.
- [8] P. Baccichet, S. Rane, and B. Girod. Systematic Lossy Error Protection based on H.264/AVC Redundant Slices and Flexible Macroblock Ordering. *Journal of Zhejiang University - Science A*, 7(5):900–909, May 2006.
- [9] H. Kodikara Arachchi, W.A.C. Fernando, S. Panchadcharam, and W.A.R.J. Weerakkody. Unequal Error Protection Technique for ROI Based H.264 Video Coding. In *Canadian Conference on Electrical and Computer Engineering*, pages 2033 – 2036, May 2006.
- [10] Q. Chen, L. Song, X. Yang, and W. Zhang. Robust Region-of-Interest Scalable Coding with Leaky Prediction in H.264/AVC. In *IEEE Workshop on Signal Processing Systems*, pages 357–362, Shanghai, China, October 2007.
- [11] H.264/AVC reference software. Available at <http://iphome.hhi.de/suehring/tml/>.
- [12] International Telecommunication Union-Radiocommunication Sector. Methodology for the subjective assessment of the quality of television pictures. ITU-R BT.500-11, June 2002.
- [13] Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG. SVC/AVC Loss Simulator. JVT-Q069, October 2005. Available at [http://wftp3.itu.int/av-arch/jvt-site/2005\\_10\\_Nice/](http://wftp3.itu.int/av-arch/jvt-site/2005_10_Nice/).
- [14] International Telecommunication Union-Standardization Sector. Non-normative error concealment algorithms. VCEG-N62, September 2001. Available at [http://wftp3.itu.int/av-arch/video-site/0109\\_San/](http://wftp3.itu.int/av-arch/video-site/0109_San/).
- [15] Z. Wang, H. R. Sheikh, and A. C. Bovik. *Objective Video Quality Assessment*, pages 1041–1078. The Handbook of Video Databases: Design and Applications. CRC Press, September 2003.
- [16] Available at [http://www.its.bldrdoc.gov/n3/video/VQM\\_software.php](http://www.its.bldrdoc.gov/n3/video/VQM_software.php).
- [17] L. Itti, C. Koch, and E. Niebur. A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11):1254–1259, November 1998.