



Region-of-Interest Intra Prediction for H.264/AVC Error Resilience

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

► To cite this version:

Fadi Boulos, Wei Chen, Benoît Parrein, Patrick Le Callet. Region-of-Interest Intra Prediction for H.264/AVC Error Resilience. IEEE International Conference on Image Processing, Nov 2009, Cairo, Egypt. pp.3109 - 3112, 10.1109/ICIP.2009.5414458 . hal-00458957

HAL Id: hal-00458957

<https://hal.science/hal-00458957>

Submitted on 22 Feb 2010

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.

REGION-OF-INTEREST INTRA PREDICTION FOR H.264/AVC ERROR RESILIENCE

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

Nantes Atlantique Universités
IRCCyN CNRS UMR 6597, Polytech’Nantes
Rue Christian Pauc, 44306 Nantes, France
firstname.lastname@univ-nantes.fr

ABSTRACT

Packets in a video bitstream contain data with different levels of importance that yield unequal amounts of quality distortion when lost. In order to avoid sharp quality degradation due to packet loss, we propose in this paper an error resilience method that is applied to the Region of Interest (RoI) of the picture. This method protects the RoI while not yielding significant overhead. We perform an eye tracking test to determine the RoIs of a video sequence and we assess the performance of the proposed model in error-prone environments by means of a subjective quality test. Loss simulation results show that stopping the temporal error propagation in the RoIs of the pictures helps preserving an acceptable visual quality in the presence of packet loss.

Index Terms— Eye tracking, region of interest, error resilience, subjective quality.

1. INTRODUCTION

The Quality of Experience (QoE) associated with video services over IP is not always satisfactory. This is mainly due to the loss of video packets during transmission. These packets contain different parts of the bitstream which generally yield unequal amounts of distortion when lost. To attenuate the quality degradation resulting from packet loss, several techniques have been proposed to improve the important packets’ resilience against losses. In the context of the H.264/AVC video coding standard, error resilience features such as Flexible Macroblock Ordering (FMO) and Data Partitioning (DP) have been extensively studied. DP partitions a coded slice into three separate NAL units each of which contains a different part of the slice. FMO allows the ordering of macroblocks in slices according to a predefined map rather than using the raster scan order. When coupled with RoI-based coding, FMO is generally used to assemble the RoI macroblocks into a single slice.

In H.264/AVC, there is a high dependency between many parts of the coded video sequence which allows better compression ratios. However, this dependency has the drawback of allowing a spatio-temporal propagation of the error resulting from a packet loss. Some research works have looked

into attenuating the effect of this spatio-temporal dependency when important parts of the bitstream are lost.

In [1], an error resilience scheme based on intra-refresh in the RoI is proposed. The RoI is determined by means of an objective attention model [2]. Restrictions are applied over inter-prediction between reference frames and intra-prediction within the same image in order to halt error propagation. The intra-refresh period is set to 4 frames for a GOP size of 30 and the size of the RoI is fixed at 25% of the image size. Significant quality enhancement is noted in the RoI compared to random intra-refresh for various loss rates.

In [3], an error resilience scheme based on FMO is presented. It first computes the importance of each macroblock by considering the distortion resulting from the loss of every pixel in it and the number of times this pixel is used as a reference for other pixels within and beyond the picture. The macroblocks are then grouped into slices depending on their importance. Results for packet loss rates in the range 5% – 20% show that video quality when using the proposed approach is significantly better than the quality obtained with traditional FMO-based coding schemes.

A robustness model for RoI-based scalable video coding is proposed in [4]. 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.

We propose in this work to use ground-truth RoI data to enhance a video bitstream’s robustness to packet losses. This is done by first confining the RoI in separate slices to cut the error propagation within a picture and then constraining the prediction process in the RoIs to prevent the loss distortion from reaching other pictures’ RoIs.

This paper is organized as follows: in Section 2, we propose our method which forces the RoI macroblocks to be coded in intra prediction mode. In Section 3, we briefly describe the eye tracking tests we performed in order to deter-

mine the saliency maps of the video sequences. Then, we present the video quality subjective tests which aimed at assessing the quality of sequences encoded with the resilience model and affected by packet loss. We finally provide the results and we discuss them.

2. INTRA PREDICTION RESILIENCE MODEL

The distortion caused by a lost slice located in an unimportant region of a picture can propagate to its RoI and to subsequent pictures'. The first situation can be caused by the spatial intra prediction within a slice when it contains macroblocks from both regions. To prevent this, we propose to group RoI macroblocks into one or more slices independently of other slices in the picture. This can be done in the H.264/AVC encoder by using the FMO feature. There are several ways of ordering the macroblocks when using FMO. We choose FMO type 6 which provides full control over the assignment of macroblocks to slices.

However, the macroblocks belonging to the RoI should be known to both the encoder and the decoder. We use another feature of H.264/AVC, namely Picture Parameter Set (PPS) to include this information in the bitstream. A PPS is a Network Abstraction Layer Unit (NALU) that contains header data useful to the decoding and display of a series of NALUs belonging to one or more pictures. Generally, one PPS NALU is sent for a large number of pictures to avoid creating an overhead. But with the RoI changing from one picture to another, we choose to include one PPS NALU for every picture in the video bitstream. This PPS NALU will contain the RoI map of each picture. The RoIs are derived from eye tracking subjective tests described in Section 3.1.

With the spatial error propagation problem solved, we tackle the more difficult problem of temporal error propagation. The use of inter picture prediction allows a packet loss in a reference picture to drift to the corresponding inter-coded pictures. To overcome this problem, we propose to force the RoI macroblocks to be coded in intra prediction mode. The forced intra prediction algorithm 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 RoI coordinates fed into the encoder. When a macroblock is flagged as being an RoI macroblock, its prediction type is forced to be intra. Given that in H.264/AVC the selection of a macroblock's prediction type is 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 algorithm's pseudocode is given in Algorithm 1.

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 constant bitrate, thus the amount of bits allowed to be used is the

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  $\in$  RoI then
        while (predType == anyInterpredType) do
          increase cost function
        end while
      else proceed normally
    end if
  end for
end for
end while

```

same for all coding schemes. The intra-coded macroblocks consume more bits than what they would have if they were coded in inter prediction mode. This results in higher quantization parameters for some macroblocks in the video and hence lower quality. The quality decrease depends also on the size of the RoI.

3. SUBJECTIVE TESTS

We describe in this section the subjective tests that aimed at (1) extracting the RoIs of the videos and; (2) assessing the efficiency of the proposed algorithm. Both tests were conducted in a standardized environment according to ITU Recommendation BT.500-11 [5]. We also present the results of the video quality tests and compare the performance of the resilience algorithm against classical coding schemes.

3.1. Eye tracking test

The goal of performing eye tracking tests is to track the eye movement of the viewers in order to extract the RoIs of a video sequence. We used a dual-Purkinje eye tracker for the experiment. The video testbed included 30 SD sequences (720×576). The sequences had either 8-second or 10-second duration. 37 non-expert subjects participated in the test.

To extract an RoI truly faithful to the content, all of the sequences were encoded such as to obtain a good video quality. Bitrates were in the range of 4 – 7 Mbs. The video sequences were encoded in High Profile with an IBBPBBP... GOP structure of length 24. All I-pictures were Instantaneous Decoding Refresh (IDR) pictures which forbid the use of references for inter prediction from a previous GOP. This helps in attenuating the temporal error drift by limiting it to at most one GOP ($\simeq 1$ s). The JM v14.0 reference software was used for encoding and decoding video sequences.

To translate the eye tracker saliency maps into RoI maps, we follow a computational procedure involving psycho-visual parameters. The values of these parameters are set empirically. The method is explained in detail in [6]. An example of saliency and RoI maps is given in Figure 1.

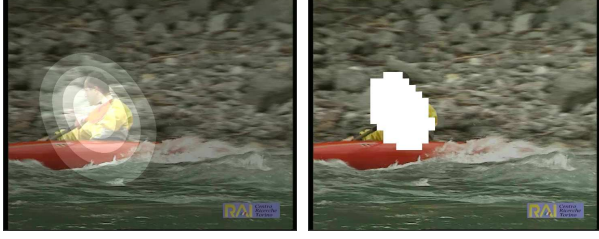


Fig. 1. Saliency map (left) and the resulting macroblock-based RoI (right) of frame 65 of *Canoe* sequence.

3.2. Video quality test

A subjective video quality test is the most reliable method to evaluate perceptual video quality. It consists in asking a group of viewers to rate a series of video sequences displayed on a monitor. We used the Absolute Category Rating (ACR) method with a five-level quality scale. 25 non-expert subjects participated in the test. We chose 3 source sequences out of the 30 included in the eye tracking test and we encoded each of them using three encoding configurations at the same bitrate: (1) *Normal coding*: default encoding scheme; (2) *RoI coding*: grouping the RoI macroblocks into independent slices using FMO; and (3) *RoI intra coding*: same as *RoI coding* but with the error resilience model.

We also introduced some slice losses in the RoI and outside of it to test the efficiency of the proposed models w.r.t. to both error propagation and the spatial location of the loss. When the losses were outside the RoI, they occurred in the slices adjacent to the RoI. All losses were in a single I-picture to allow a longer temporal propagation. We used the error concealment algorithm implemented in the JM decoder.

3.3. Results and discussion

The first experiment consisted in comparing the quality degradation caused by the loss of slices inside and outside the RoI. In an earlier work [7], we performed subjective tests which results proved that the MOS of a sequence strongly depends on the spatial position of the loss in the picture. That is, losing less slices in the RoI can sometimes yield worst quality than losing more slices outside of it. Therefore, we tested two loss scenarios: in the first, the whole RoI was lost (“RoI lost”) and in the second the loss occurred in the slices adjacent to the RoI (“Around RoI lost”). Note that the size of the RoI (in slices) is not the same for all contents.

Figure 2 gives the Mean Opinion Score (MOS) for three sources with Loss Ratios (LR) smaller than 1. LR is the ratio of the number of slices lost in the RoI to the number of slices lost outside the RoI. The three sequences were encoded using *RoI coding* scheme which groups the RoI macroblocks into separate slices without applying any further prediction constraint. For *Captain* sequence, losing in the RoI

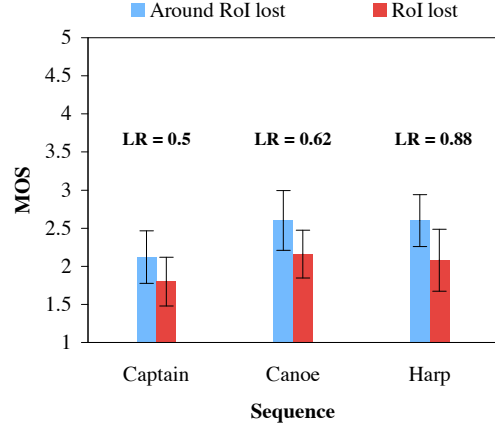


Fig. 2. MOS values for *RoI coding* with losses inside and outside the RoI. The three sequences have different LR values.

half the number of slices lost outside of it yields worst quality ($\Delta_{MOS} = 0.3$). For *Canoe* and *Harp* sequences, we can see that quality difference is slightly higher for an increasing number of RoI slices lost: $\Delta_{MOS} = 0.45$ and 0.5 , respectively. For all three contents, the loss affected more slices outside the RoI yet the quality was better than with smaller losses in the RoI. This stresses the strong correlation between saliency and perceptual quality.

In Figure 3, the results of the second experiment are plotted. We compare the three coding schemes in both error-free and error-prone environments. Two conclusions can be drawn: (1) the RoI-based model does not yield significant overhead compared to classical coding schemes; and (2) *RoI intra coding* outperforms other schemes when the RoI is lost.

The first conclusion is inferred by the almost equal MOS values of video sequences which do not have any of their packets lost. This shows that no extra cost is mandated by the use of the RoI-based error resilience model because the quality is sustained at a good to excellent level ($MOS = 4.5$).

The right part of the plot in Figure 3 indicates a large quality difference when the RoI of an I-picture is lost ($\Delta_{MOS} \simeq 1$) between *RoI intra coding* scheme which yields a fair to good quality ($MOS = 3.4$) and the other two schemes which yield a poor quality on average ($MOS = 2.25$). This is mainly due to the cut of temporal error propagation in the RoI which makes the distortion appear in the I-picture’s RoI only. Although there are no restrictions on the inter prediction process outside the RoI and hence the error can propagate to that region, this case did not seem to happen very often. Note that *Normal coding*, wherein RoI macroblocks are not grouped into particular slices yields a slightly better quality because the loss cannot occur in the RoI exactly.

The *RoI intra coding* scheme’s performance indicates that applying this error resilience model to a video bitstream can make all of its packets equally robust to packet loss: losses in the RoI no longer yield more distortion than losses in other

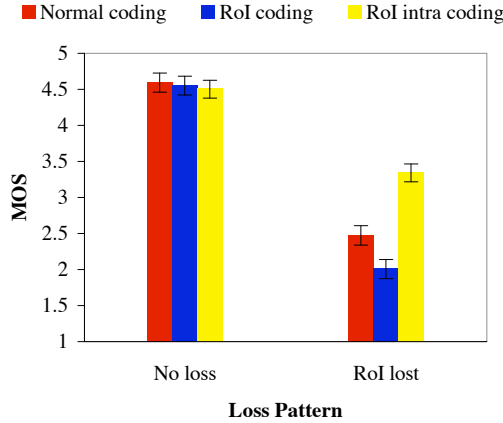


Fig. 3. MOS values for all coding schemes without and with losses in the RoI.

parts of the picture. Therefore, we can consider our model as a layering technique where the priority of a slice depends on its position w.r.t. the RoI's. This new form of bitstream prioritization, which is obtained with no significant complexity increase, can be compared to common layering strategies such as temporal and quality scalability.

In Figure 4, we illustrate how the *RoI intra coding* scheme stops the temporal error propagation by comparing it to the *RoI coding* scheme. Figures 4(a) and 4(b) depict frame 97 of *Captain* sequence without and with losses, respectively. Given that frame 97 is an I-picture, we expect that the distortion propagates to subsequent pictures. This is indeed the case in Figure 4(c) where the sequence is coded using the simple *RoI coding* mode and therefore the B-picture suffers from severe degradation in quality. However, no distortion appears in Figure 4(d) because the intra-coded RoI is completely independent of the previous reference frame.

4. CONCLUSION AND FUTURE WORK

We presented in this paper an RoI-based H.264/AVC error resilience model that copes with spatio-temporal error propagation. We derived RoIs from ground-truth data collected through an eye tracking experiment. We also performed a subjective quality test to evaluate the efficiency of the algorithms. Results showed that in the presence of packet loss, forcing the macroblocks of all RoIs in a video sequence to be coded in intra prediction mode improves its perceptual error robustness leading to an increase of 1 MOS value on a scale of 5 compared to normal coding.

We envision investigating more deeply the influence on visual quality of both the RoI size and the distance between the RoI and the less important region of the picture where the loss occurs. We also plan to test the efficiency of our method when using RoI data obtained from computational models such as [2].

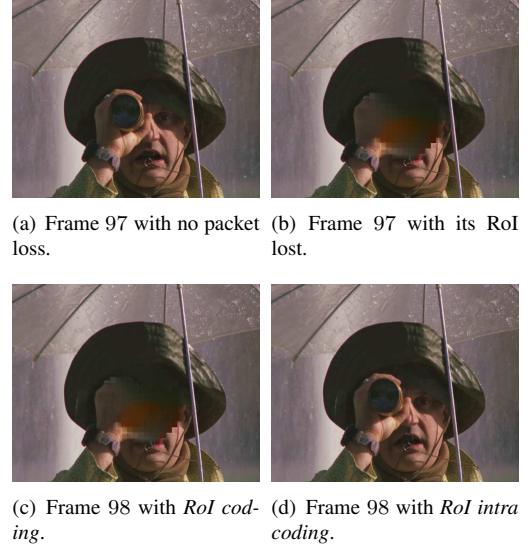


Fig. 4. Frames 97 and 98 of *Captain* sequence.

5. REFERENCES

- [1] Q. Chen, Z. Chen, X. Gu, and C. Wang. Attention-Based Adaptive Intra Refresh for Error-Prone Video Transmission. *IEEE Communications Magazine*, 45(1):52–60, Jan. 2007.
- [2] L. Itti, C. Koch, and E. Niebur. A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. *IEEE Trans. on PAMI*, 20(11):1254–1259, Nov. 1998.
- [3] Y. Dhondt, P. Lambert, and R. Van de Walle. A Flexible Macroblock Scheme for Unequal Error Protection. In *Proc. of IEEE Int. Conf. on Image Processing*, pages 829–832, Oct. 2006.
- [4] 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, Oct. 2007.
- [5] International Telecommunication Union-Radiocommunication Sector. Methodology for the subjective assessment of the quality of television pictures. ITU-R BT.500-11, Jun. 2002.
- [6] F. Boulos, W. Chen, B. Parrein, and P. Le Callet. A new H.264/AVC Error Resilience Model Based on Regions of Interest. In *17th Int. Packet Video Workshop*, May 2009.
- [7] F. Boulos, D. S. Hands, B. Parrein, and P. Le Callet. Perceptual Effects of Packet Loss on H.264/AVC Encoded Videos. In *4th Int. Workshop on Video Processing and Quality Metrics*, Jan. 2009.