A ROBUST H.264 DECODER WITH ERROR CONCEALMENT CAPABILITIES

†*Gianluca Gennari,* †*Gian Antonio Mian and* ‡*Luca Celetto*

†Department of Information Engineering
University of Padova
Via Gradenigo 6/B, 35131 Padova - Italy
email: jena@dei.unipd.it, mian@dei.unipd.it

‡ST Microelectronics Srl
Advanced System Technology
Via C. Olivetti 2, 20041 Agrate Brianza - Italy
email: luca.celetto@st.com

ABSTRACT

Video transmission on unreliable networks can produce artifacts on received images due to either the loss of a part of the bit-stream or its corruption. In this paper we describe a H.264/AVC decoder robust to transmission errors, which contains an innovative algorithm, based on temporal correlation between frames and scene change detection, that greatly improves the objective and subjective quality of incorrectly decoded images, with a computational complexity allowing real-time implementation.

1. INTRODUCTION

Motion video is the most demanding mobile multimedia application in terms of bit-rate and computational complexity. Dealing with unreliable channels adds several challenges to this difficult task. In a real-world situation, a network can not provide a guaranteed quality of service due to issues like interfering traffic, signal to noise fluctuations etc. Transmission glitches range from single bit errors to burst errors, or even temporary loss of connection, causing a wide range of different conditions. The compressed video is, in general, extremely vulnerable to transmission errors. In order to obtain the low coding bit rate imposed by channel constraints, inter-frame coding is heavily used. Error propagation from a corrupted reference image to the following ones lowers the quality of the reconstructed video sequence. Traditional techniques for low bit-rate video encoding in error-free channels can lead to severe image corruption without adoption of special countermeasures.

H.264/AVC is the newest video coding standard that enhances compression performance and provides a "network-friendly" video representation [1, 2]. Slice partitioning, data partitioning, Intra refreshing, separation of picture parameters from slice data are all important features of the H.264 standard that improve the bit-stream robustness.

In order to be able to decode a corrupted bit-stream, a robust H.264 decoder must detect syntax errors, discarding the rest of the corrupted data unit and recovering the synchronization with the encoder as described in Fig. 1. Frame loss detection is also important to keep synchronization with different streams (like audio channels, etc.). Concealment methods are necessary to reconstruct the missing parts of a damaged image, interpolating the lost information from the spatially or temporally neighboring macro-blocks. All the discussed concealment techniques are based on image post-processing done at the decoder side and do not include feedback sent to the encoder.

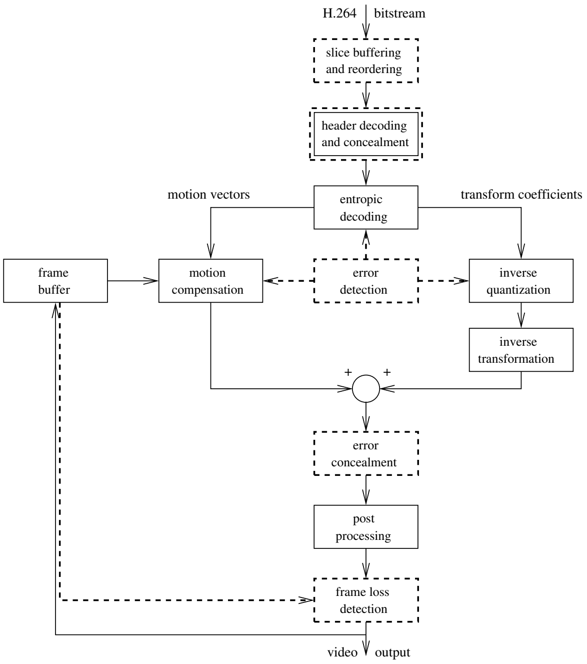

Figure 1: Scheme of the modified H.264 decoder. The dotted blocks represent the added features.

The article is organized as follows. Section 2 describes the algorithm implemented in the reference H.264/AVC software model. Section 3 shows our innovative solution, which solves the problems arising during scene changes by means of a novel approach based on temporal correlation between images. Finally, section 4 shows the quality improvement and section 5 concludes the paper.

2. THE JVT CONCEALMENT ALGORITHM

We based our work on the H.264 reference decoder version 6.0*a* [3] supplied by the JVT committee (Joint Video Team of ISO/IEC MPEG and ITU-T VCEG), which implements state-of-the-art concealment algorithms to reconstruct the missing data from the received information. The decoding status is collected for each macro-block, signalling the incorrectly decoded image regions for a later concealment. The JVT algorithm treats differently Inter and Intra images, following the modalities described in the next paragraphs.

2.1 Intra concealment

In Intra images, the JVT algorithm uses information from surrounding blocks to reconstruct the missing data. A bi-

---

This work was carried out within the Italian Ministry of Education, University and Research (MIUR) Project "FIRB PRIMO".

linear interpolation estimates original values from the available blocks using the samples on their borders. The already concealed blocks can be used for the interpolation of the remaining damaged blocks, when there are not enough correct neighbors to ensure a good prediction.

The adopted interpolation technique follows the criterion of maximally smooth recovery in the spatial domain exposed in [4]. The reconstruction formula is the following:

$$p_0 = \frac{\sum_{i=1}^{4} d_i \, p_i}{\sum_{i=1}^{4} d_i} \tag{1}$$

where p_0 is the value of the interpolated pixel, p_i are the values of the pixels on the border of the surrounding blocks and d_i are the distances of the p_i pixels from p_0. Therefore, the interpolation is a simple weighted average. The algorithm leads to a visually pleasant reconstruction unless the image losses do extend to a great part of the image.

2.2 Inter concealment

JVT concealment for Inter images performs a motion estimation for temporal prediction in order to reconstruct the missing data. Since a good motion estimation leads to small prediction errors, the algorithm assumes that this signal is null. The hypothesis leads to an acceptable image quality in most situations, especially with low-motion content.

The algorithm tries to estimate the motion vector associated to the damaged or lost blocks from the vectors in the spatially adjacent blocks, which are stored into a data structure that is updated each time the decoding process of a macro-block ends correctly. For each damaged block, the method cycles on the four spatially adjacent macro-blocks, using their motion information as a predictor. As in the data structure there is a motion vector for every 8×8 block, every macro-block is tested against two different motion predictors for each available neighbor. The simple copy from previous frames is also tested as it frequently leads to good results.

The JVT algorithm selects the predictor that minimizes a particular distortion function measuring the uniformity of the predictor block in respect to the border pixels of the surrounding blocks. This is performed computing the average of the absolute value of the difference between the pixels lying on the edges of the predictor and the pixels pertaining to the corrupted image. This technique is equivalent to the principle of the maximally smooth recovery described in [4], while the adopted spatial distortion measure corresponds to the one illustrated in [5].

3. ERROR DETECTION AND RECOVERY

In [6], the error concealment problem is divided in two steps: locating the error position and recovering the missing data. The first part of the problem can be easily solved when the underlying network stack handles this aspect in a transparent way. In order to be more general, we implemented error detection algorithms in our solution.

The H.264 standard allows using two different entropy coding algorithms (called CABAC and CAVLC) to encode the transmitted symbols. With both algorithms, a transmission error causes an incorrect decoding of the corrupted symbol, which propagates the error in the following values, until a resynchronization point is reached. Errors are detected when the syntax of the stream is violated. Experimental results show that error detection is harder with CABAC, as syntax errors are detected well behind the point where they actually occurred. Some incorrectly decoded macro-blocks are not reported, so they can not be concealed, driving to a lower overall quality (see [7]). For this reason, all the results shown in this paper were obtained with CAVLC encoding.

In order to detect transmission errors, several checks were introduced. Every syntax exception causes the immediate interruption of the decoding process, given the impossibility to recover the remaining information in the current slice. Due to the data structure introduced by the H.264 standard, every slice is independent from the others and therefore the resynchronization with the data flow can happen in correspondence of every new slice.

3.1 Slice header correction

Image slice partitioning is an effective method to reduce error propagation. All the slices can be independently decoded as they have no reference to other parts of the image. A part of the information is repeated in each slice header, producing a *bit rate overhead*. This information redundancy allows implementing an error correction strategy, which becomes more effective increasing the number of slices per frame. A histogram structure stores the number of times a certain value is received, so that when an error is detected the most frequent value replaces the incorrect one. Slices with too many detected header errors are discarded, in order to eliminate heavily damaged data units.

This algorithm can only work when there are three or more slices in the packets buffer pertaining to the current image. Under this hypothesis, it allows performing a slice header concealment that is much more effective than a simple syntactical control, correcting the errors in nearly all the realistically possible situations.

3.2 JVT algorithm enhancements

Through an extensive analysis, we detected some weak points of the original algorithm: Intra images use spatial prediction for the concealment of the damaged blocks that proved to be effective only when recovering small areas in the image. In some situations, it could cause a blurring effect on the concealed blocks that is easily perceptible especially in presence of highly detailed sequences. In section 3.2.2 we describe an algorithm that uses temporal prediction to improve the concealment.

The Inter images algorithm supplies good performance in case of static or low-motion sequences, but in case of scene changes it extracts the predictor blocks from an uncorrelated frame, producing unacceptable artifacts. In section 3.2.3 we describe the modifications applied to the reference algorithm for Inter image concealment, while in the paragraph 3.2.1 we introduce some simple algorithms for scene change detection.

3.2.1 Scene change detection

In the H.264 standard, the encoder is free to choose the best encoding strategy on a per-block basis. For example, it could encode a macro-block in Intra mode even in a Inter slice, if the lack of correlation between the frames prevents finding a good temporal predictor. The experimental results have shown that when a scene change occurs in correspondence of an Inter image, most macro-blocks are encoded in Intra mode. This observation inspired a simple technique for scene change detection.

Our proposal computes the percentage of Intra blocks in each predictive frame, then the obtained value is compared against a threshold to decide whether there is a scene change. However, such technique is not applicable to Intra images, hence we implemented a different approach: we choose 4 correctly decoded reference macro-blocks in the Intra image, checking also that they refer to a correctly decoded part of the reference image. Then a modified motion estimation – as explained in 3.2.2 – is performed to find the best predictor, using the motion vectors pertaining to the latest Inter frame.

The 4 reference macro-blocks and their predictors are compared computing the sum of the absolute differences (*SAD*) between their pixels. A mean of the values found is then compared against a fixed threshold. Higher values indicate that motion estimation failed detecting good predictors and there is a high probability of a scene change, while lower values show a good correlation between the compared blocks and therefore a high probability they belong to the same scene.

3.2.2 Intra concealment enhancements

The main changes in Intra image concealment regard the way it is applied using the scene change algorithm described in paragraph 3.2.1. In order to preserve image details, we use a temporal interpolation, unless a scene change is detected. The behavior is similar to the original algorithm for Inter images, but the last available reference frame is used. The two algorithms differ in the estimation technique of the more probable motion vector for the reconstructed block. As Intra images does not have any motion information, we need to extrapolate them from available data.

Due to the real-time requirements, we based our approach on the hypothesis of slow variations in the motion vector fields, associating to every Intra 8×8 block the motion vector pertaining to the homologous block in the latest Inter reference frame. This motion vectors are then compared through the spatial distortion measure described in 2.2, and the macroblock with the lower cost is chosen as predictor.

With fast moving images, our technique could produce visible artifacts on the reconstructed frame. However, in this situation the scene change algorithm would detect a low correlation between frames selecting the spatial concealment.

3.2.3 Inter concealment enhancements

Modifications to Inter concealment follow the same strategy applied to Intra images and are based on the scene change technique introduced in 3.2.1. We apply the original JVT temporal concealment (described in 2.2) between highly correlated frames, but whenever a scene change is detected we use the spatial interpolation described in 2.1, in order to avoid prediction from a bad reference frame. This comes at the expense of the typical interpolation *blurring* effect, but it's acceptable because, in presence of a scene change, the natural inertia of the human eye masks the undesired effect.

Bidirectional predictive pictures require a special treatment, because they can belong to both past and future reference frames. The temporal concealment is therefore constrained to the most frequently accessed reference frame, which has the higher correlation with the damaged image.

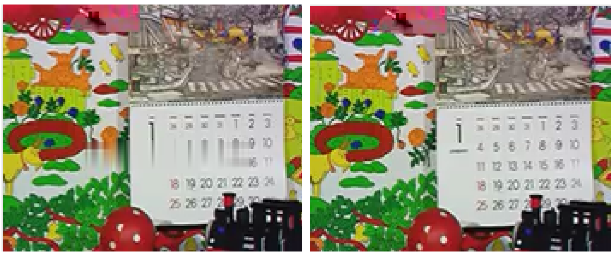

Figure 2: Left and right pictures reconstructed respectively by JVT and our proposed algorithms in case of a corrupted Intra frame.

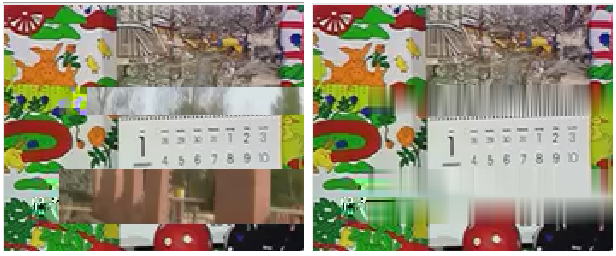

Figure 3: Left and right pictures reconstructed respectively by JVT and our proposed algorithms in case of a scene change.

4. EXPERIMENTAL RESULTS

In our experiments we used several QCIF sequences, but for conciseness we show only the results related to a 100 frames sequence created by assembling 5 fragments that were alternatively extracted from "Foreman" and "Mobile" QCIF sequences. Results from the "Foreman" QCIF sequence are also reported in order to evaluate the performance of the modified decoder in a scene-changes free scenario.

Fig. 2 shows the effects of the JVT concealment algorithm and our proposed one on a corrupted Intra image. A central slice was reconstructed by the JVT algorithm on the left image, interpolating data from the surrounding blocks. On the right image, the proposed algorithm improves the reconstruction, inferring information from past frames. In the central zone of the calendar (right image) we can see readable numbers while they are deleted in the spatially interpolated frame (left image).

Fig. 3 shows the improvements deriving from scene change detection. On the left image we can notice that some macro-blocks were copied from an unrelated previous image, while the proposed algorithm detects the scene change and prefers a spatial interpolation. Image quality is still corrupted but with a less unpleasant effect.

Table 1 summarizes the parameters used for simulations. As the image quality can greatly change depending on where the error occurs in the bit-stream, we repeated simulations many times, inserting errors in different positions and reporting a "mean" result.

After extensive simulations with BERs between $10^{-2.5}$ and 10^{-5}, we decided to adopt a partitioning of the picture in slices of 500 bytes – a good compromise between bit-rate overhead and slice size. This kind of partitioning gives higher performance for wide bit-rate fluctuations than the one with a fixed number of macro-blocks per slice.

The three error concealment methods are measured in Fig. 5 with a GOP length of 6 frames and an error rate of

Parameter	Value
Sequences	Foreman, Foreman + Mobile (QCIF, 100 frames)
Slice partitioning	500 byte per slice
Entropic coding	CAVLC
GOP Structure	IBBPBBP, 6, 12, 24, 48 and 96 frames
Quantization pass	From 6 to 46 by steps of 10
Bit Error Rate	From $10^{-2.5}$ to 10^{-5}

Table 1: Encoding parameters used for simulations.

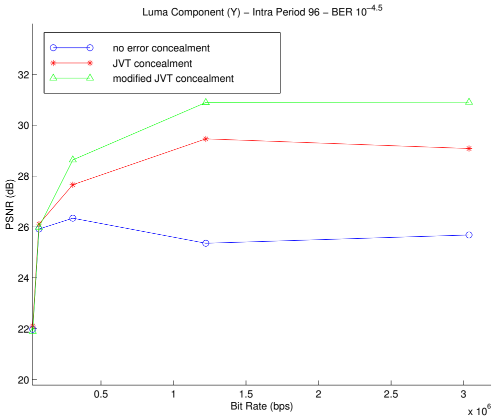

Figure 4: Comparison between the concealment methods with scene changes, with a GOP lengths of 96 frames and a $10^{-4.5}$ BER.

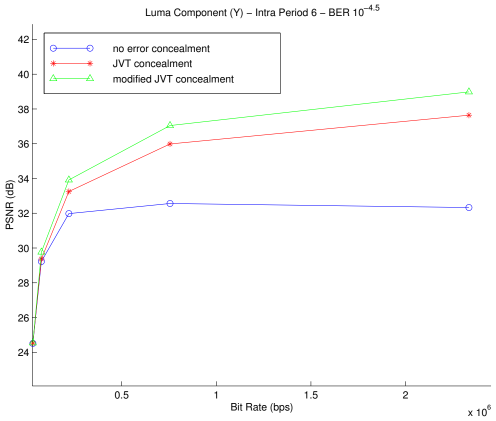

Figure 5: Comparison between the concealment methods without scene changes, with a GOP lengths of 6 frames and a $10^{-4.5}$ BER.

$10^{-4.5}$ on the "Foreman" sequence, and in Fig. 4 with a GOP length of 96 frames and the same error rate on the composite sequence. The *rate-distortion* diagrams demonstrate the improvement over the JVT algorithm, in the first case due to the enhanced Intra concealment and in the second case due to the scene change detection. Temporal concealment for Intra pictures allows a mean advantage of more than 1 dB, while scene change detection provides an improvement of about 2 dB at high bit rates.

In Fig. 6, we report the experimental results in terms of *rate-distortion* diagrams, showing how PSNR scales with the error rate. It reaches a minimum of approximately 12 dB with a Bit Error Rate of $10^{-2.5}$, while it tends to the error-free curve for BER lower than 10^{-5}.

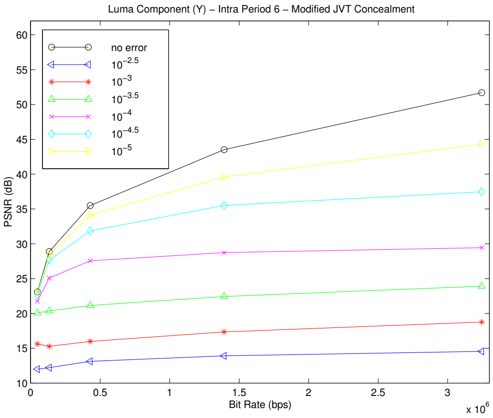

Figure 6: Visual quality as a function of Bit Error Rate, measured on the composite sequence containing scene changes.

5. CONCLUSIONS

The realization of a H.264 decoder that is robust to single bits errors allowed testing the effectiveness of the methods implemented in the standard in order to improve data robustness and comparing various image concealment techniques.

The experimental results demonstrate the effectiveness of the modifications applied to the concealment algorithm. The new algorithm allows taking advantage of the temporal prediction between subsequent frames and detecting the scene changes in order to avoid predicting a block from a reference frame pertaining to an uncorrelated context.

Future developements will focus on feed-back based methods. We will try to improve the image quality, tuning the encoder parameters to the current network conditions.

REFERENCES

[1] T. Wiegand, G. Bjøntegaard, G. J. Sullivan and A. Luthra *Overview of the H.264/AVC Video Coding Standard*, IEEE Trans. on Circuits and Systems for Video Technology, Vol. 13, No. 7, July 2003, pp. 560–576.

[2] Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG *Joint Final Committee Draft (JFCD) of Joint Video Specification (ITU-T Rec. H.264 — ISO/IEC 14496-10 AVC), JVT-D157*, 4th Meeting: Klagenfurt, Jul. 2002.

[3] http://bs.hhi.de/ suehring/tml/jm60a.zip

[4] Y. Wang, Q.-F. Zhu, L. Shaw *Maximally smooth image recovery in transform coding*, IEEE Trans. on Communications, Vol. 41, Oct. 1993, pp. 1544–1551.

[5] Q.-F. Zhu, Y. Wang *A comparison of smoothness measures for error concealment in transform coding*, Proc. SPIE Conference on Visual Communications, Taipei, Taiwan, 1995, Vol. 2, pp. 1205–1214.

[6] Y. Wang and Q.-F. Zhu *Error Control and Concealment for Video Communication: A Review*, Proc. IEEE, Vol. 86, No. 5, May 1998, pp. 974–995.

[7] G. Gennari *Decodificatore H.264 robusto nei confronti degli errori di trasmissione*, Laurea Degree Thesis, Department of Information Engineering, University of Padova, Italy, July 2003.