

ENHANCING ERROR RESILIENCE IN WIRELESS TRANSMITTED COMPRESSED VIDEO SEQUENCES THROUGH A PROBABILISTIC NEURAL NETWORK CORE

Reuben A. Farrugia and Carl J. Debono

Department of Communications and Computer Engineering, University of Malta, Msida, Malta
{rrfarr, cjdebo}@eng.um.edu.mt

ABSTRACT

Video compression standards commonly employed in the delivery of real-time wireless multimedia services regularly adopt variable length codes (VLCs) for efficient transmission. This coding technique achieves the necessary high compression ratios at the expense of an increased system's vulnerability to transmission errors. The more frequent presence of transmission errors in wireless channels requires video compression standards to accurately detect, localize and conceal any corrupted macroblocks (MBs) present in the video sequence. Unfortunately, standard decoders offer limited error detection and localization capabilities posing a bound on the perceived video quality of the reconstructed video sequence.

This paper presents a novel solution which enhances the error detection and localization capabilities of standard decoders through the application of a Probabilistic Neural Network (PNN). The proposed solution generally outperforms other error detection mechanisms present in literature, as it manages to improve the standard decoder's error detection rate by up to 95.74%.

Index Terms— Error detection coding, learning systems, multimedia communications, video coding, wireless networks.

1. INTRODUCTION

Numerous wireless multimedia services, including Digital TV broadcasting, video-telephony and videoconferencing applications, have recently emerged. The video compression standards adopted by these systems [1] – [5] usually employ variable length codes (VLC) to reduce the transmission bit-rate and storage space requirements. However, when VLC codes are transmitted over an error-prone channel, a single corrupted bit will desynchronize the bitstream until the next synchronization marker. This results in a number of corrupted macroblocks (MBs), causing annoying visual artifacts which propagate both in the temporal and spatial domains.

The video compression standards generally adopt syntax and semantic violation detection to identify errors in the transmitted bitstream. However, some of the corrupted bitstreams will still form valid entries in the VLC table resulting in only 40% - 60% of the corrupted MBs being detected [7]. This is a rather low error detection rate

which impacts negatively on the overall experience of the end user.

A number of dissimilarity metrics were proposed in literature which exploit the inherent redundancies present in the neighboring MBs and the corresponding MBs in the previous frame [6], [10] – [17]. While enhancing the error detection capability of standard decoders, these dissimilarity metrics only manage to detect 20% – 40% of the corrupted MBs which were not detected by the syntax and semantic violation test (residual corrupted MBs).

An iterative error detection and concealment approach that manages to detect up to 93% of the residual corrupted MBs was proposed in [9], where a combination of dissimilarity metrics was considered. However, this solution modifies the syntax within the bitstream making it incompatible with standard decoders and was tested on two video sequences which contain limited movement, thus the results may be sequence dependent. Moreover, this algorithm enhances the quality of the reconstructed video sequence through the utilization of an iterative algorithm, inducing a significant increase in computational time making it unsuitable for real-time applications.

This paper presents a novel error detection algorithm that can be employed in real-time applications. The proposed algorithm adopts a Probabilistic Neural Network (PNN) and manages to detect on average 93.93% of the residual corrupted macroblocks. The algorithm was tested on a wide range of video sequences in order to assess the validity of the proposed solution and the results obtained confirm that this new technique outperforms other error detection mechanisms employed in literature.

This paper is organized as follows; an overview of the distortions caused by transmission errors is presented in Section 2, followed by a detailed description of the proposed error detection algorithm. The simulation results are delivered in Section 4 while the final comments and conclusions are given in Section 5.

2. DISTORTIONS CAUSED BY TRANSMISSION ERRORS

The perceived quality of a reconstructed video sequence is significantly degraded when the compressed bitstream is corrupted by transmission errors, as shown in Fig. 1. This results in a number of corrupted macroblocks with different levels of visual distortion. Some of these artifacts are very annoying, others are acceptable while

the rest have no impact on the perceptual quality of the reconstructed video sequence.



Figure 1. Perceived quality of a typical reconstructed video sequence

In this paper, different distortion levels (DLs), inspired by the Mean Option Score (MOS), were scaled according to Table 1. The features adopted by the PNN to enhance the robustness of the decoder were extracted at image level and the set of features which best described the problem was selected. This data was then used to optimize the detection of corrupted MBs which provide distorted frames while ignoring others which do not reduce the perceptual quality of the reconstructed video sequence. This ensures that resources are not wasted in concealing MBs which do not provide annoying visual artifacts.

Table 1. Rating Scale used for level of visual distortions of the corrupted MBs

Distortion Level (DL)	Description
0	Uncorrupted/Imperceptible
1	Perceptible but not annoying
2	Slightly Annoying
3	Annoying
4	Very Annoying

3. ERROR DETECTION AND LOCALIZATION ALGORITHM

The proposed error detection mechanism, illustrated in Figure 2, extracts ten features which provide an intuitive description of the problem. These features form the vector which describes the condition of the considered MB and is used as input to a classifier which uses a supervised machine learning approach. Any supervised learning classifier can be employed as classification module but in this work, only three different classifiers were considered, namely; (1) Fisher Discriminant Analysis (FDA) [18], Backpropagation Neural Networks (BPNN) [20] and Probabilistic Neural Networks (PNN) [19]. A brief description of each of these modules is provided in the following sub-sections.

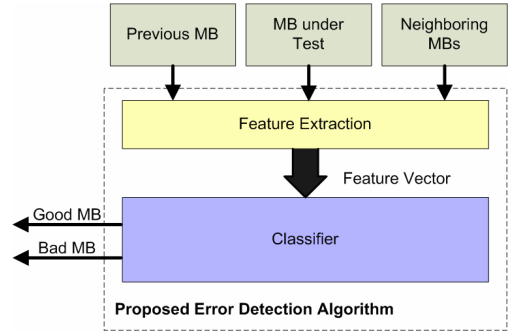


Figure 2. Proposed Error Detection Algorithm

3.1. Feature Extraction Module

The feature extraction module extracts ten features which exploit both the color and the texture consistency in both the temporal and spatial domains. These can be divided in two different classes; 1) Pixel domain features, which exploit the color consistency and are computed in the perceptually uniform CIE LUV color space model, and 2) DCT domain features, which exploit both the color and the textural consistency and are computed in the DCT compressed domain. The following subsections summarize the concepts behind these features, while more information can be found in [16], [17].

3.1.1. Spatial Domain Features

The *Average Inter-Sample Difference across boundaries* (AIDB) computes the Euclidean distance across the boundaries of the neighboring MBs. On the other hand, the *Internal AIDB* (IAIDB) and *Internal AIDB per block* (IAIDB_{block}) features exploit the fact that the pixel transition from adjacent 8x8 blocks within an MB is smooth, and thus can be used to detect the presence of a single corrupted 8x8 block.

The *Sum of Euclidean Pixel difference* (SED) is based on the fact that for an uncorrupted MB the pixel transition of an MB and the corresponding MB in the previous frame varies smoothly, and thus the Euclidean distances between corresponding pixels are computed. The *Internal SED* (ISED) was designed to detect the presence of a single corrupted 8x8 block, and therefore the Euclidean distances between the corresponding pixels of each 8x8 block are computed and the largest difference is used as a feature.

3.1.2. DCT Domain Features

Figure 3 illustrates the method used to extract the metrics. Each metric contains information about the color, together with 1) vertical, 2) diagonal, and 3) horizontal edge information. These metrics were computed by summing the coefficients along the respective direction.

The spatial DCT metrics ($d_{1\text{spat}}$ and $d_{3\text{spat}}$) exploit the fact that there is a correlation between the DCT coefficients of the current block and its neighboring blocks. On the other hand, the temporal dissimilarity metrics ($d_{1\text{temporal}}$, $d_{2\text{temporal}}$ and $d_{3\text{temporal}}$) are computed using the Euclidean distance between the current MB and the analogous metrics in the corresponding MB in the previous frame.

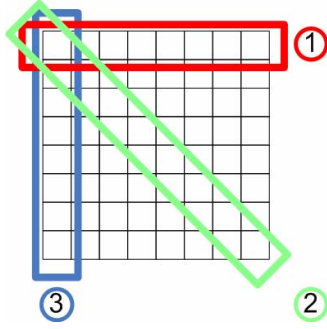


Figure 3. DCT metrics containing color and textural data

3.2. Classifier Module

The classifier module applied in this context solves a binary classification problem. It adopts the supervised learning approach, where training is accomplished by presenting 500 training vectors per class, each with its associated target output vector. The weights are then adjusted according to the learning process. The FDA, BPNN and PNN were implemented to verify the validity of the proposed approach.

3.2.1. Fisher Linear Discriminant Analysis

The FDA classifier [18] moves a straight line around the space to find an orientation for which the projected samples are well separated. This is achieved by computing the scalar dot product between a linear combination of the input data \mathbf{x} and weight vectors \mathbf{w} . The corresponding classes \mathbf{y} are then found by shifting the result with bias \mathbf{b} as follows:

$$\mathbf{y} = \langle \mathbf{w} \cdot \mathbf{x} \rangle + \mathbf{b} \quad (1)$$

3.2.2. Backpropagation Neural Network

BPNN is one of the simplest and most effective supervised machine learning algorithms, which employs hidden neurons to achieve nonlinear separation of the data set. The BPNN algorithm was trained using different architectures adopting the conjugate gradient method for learning with the Polack-Ribiere formulation [20].

3.2.3. Probabilistic Neural Network

In the PNN [19], shown in Figure 4, each pattern \mathbf{x} of the training set is normalized and placed on the input units. The modifiable weights linking the input units and the corresponding pattern units are set such that $\mathbf{w}_k = \mathbf{x}_k$ for $x = 1, 2, \dots, n$. Then, a single connection between each pattern unit is made to the category unit corresponding to the known class of that pattern.

The classification of the PNN normalizes the input vector \mathbf{x} , and computes the inner product between \mathbf{x} and \mathbf{w} . The non-linear function which computes the activation function of each pattern unit is given by:

$$\mathbf{z} = \exp\left(\frac{\mathbf{z}_{in} - 1}{\sigma^2}\right) \quad (2)$$

where σ is the smoothing parameter. The summation units then accumulate the values of the activation functions of a given class. The decision unit assigns the class according to the magnitude of the accumulated activation functions.

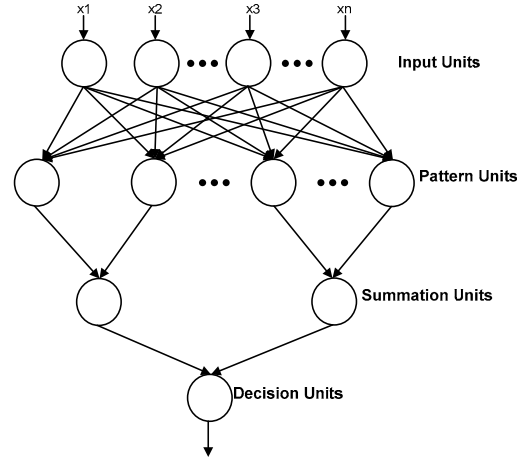


Figure 4. PNN Classification Architecture

4. SIMULATION RESULTS

The proposed error detection and localization algorithm was tested using a wide range of video sequences; “Akiyo”, “Erik”, and “Silent” which contain moderate movements, together with “Football” and “Tennis” which contain fast moving objects. The H.263++ codec was adopted as a testbed to compress the raw video sequences at CIF resolution. This compressed sequence was transmitted through a wireless noisy channel [21] with a bit-error rate (BER) of 1.31E-03.

The classifiers were trained as described above, and the algorithms were tested on a set containing 1000 testing vectors. Table 2 summarizes the performance of the proposed algorithm using the three different classification methods compared to the AIDB [9], [10] and the Spatial Feature [16] results available in literature. It can be noticed that the proposed method outperforms other error detection algorithms, and its gain in performance becomes more significant with the application of NNs. The PNN presents the best performance with an overall error detection rate of 92.16% and manages to detect most of the corrupted MBs which provide major visual distortions, as indicated in Table 3.

Table 2. Error Detection rate using different Classifiers

Classifier	Average Class ⁿ Rate
AIDB	0.4413
Spatial Feature	0.6100
Fisher LDA	0.7480
Probabilistic NN	0.9216
Backpropagation NN	0.9080

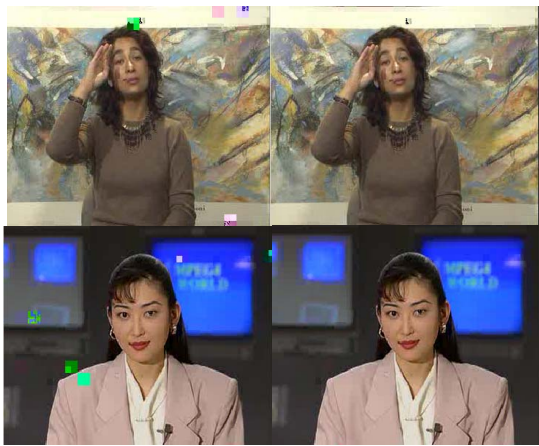
The proposed algorithm was included within the standard decoder, and its error detection capabilities were tested on the five video sequences mentioned above. The enhanced decoder manages to detect and accurately localize on average 93.93% of the corrupted MBs as shown in Table 4, outperforming all other error detection mechanisms present in literature, including [9]. Figure 5 illustrates the enhanced subjective quality of the reconstructed video sequence when applying the proposed algorithm.

Table 3. Error Detection rate of different distortion levels

DL	PNN	BPNN
1	0.5517	0.7241
2	0.8454	0.7938
3	0.9091	0.8751
4	0.9336	0.9231

Table 4. Error Detection rate of different video sequences

Video Sequence	PNN
Erik	0.9394
Silent	0.9574
Akiyo	0.9360
Football	0.9302
Tennis	0.9333

**Figure 5.** Reconstructed frames of some of the considered video sequences when decoded by the standard decoder (Left) and by the proposed algorithm using PNN (Right)

5. COMMENTS AND CONCLUSION

This paper has presented a novel solution which enhances the error detection capabilities of standard video decoders by applying a PNN at its core. Results have shown that the proposed solution manages to detect on average 93.93% of the corrupted MBs which were undetected by the tested standard H.263++ decoder. These also show that the proposed solution significantly outperforms all the other methods that are published in literature [6], [10]-[17]. The extracted features are easily computed and the PNN can be implemented in parallel, thus minimizing the overheads making it applicable for real-time applications. Moreover, this is achieved without any increase in data-rate.

The proposed solution manages to locate the corrupted MBs more accurately presenting a significant gain in the quality of the reconstructed video sequences when compared to the standard decoder and hence gives room for the channel BER requirements to be relaxed.

6. REFERENCES

[1] ISO/IEC 10918, "Information technology – Digital compression and coding of continuous tone still images," 1994.

- [2] ITU-T Rec. H.264, "Advanced video coding for generic audiovisual services," 2005.
- [3] ITU-T Rec. H.263, "Video Coding for Low Bit-Rate Communication," 2005.
- [4] ISO/IEC 11172, "Information technology – Coding of moving pictures and associated audio for digital storage media at up to 1.5 MBit/s," 1993.
- [5] ISO/IEC 13818, "Information technology – Generic coding of moving pictures and associated audio information," 2000.
- [6] M.R. Pichering, M.R. Frater, and J.F. Arnold, "A Statistical Error Detection Technique for Low Bit-Rate Video," *IEEE Proc. of TENCON Conf.*, vol. 2, pp. 773-776, Dec. 1997.
- [7] W. Park, and B. Jeon, "Error Detection and recovery by Hiding Information into Video Bitstream using Fragile Watermarking," *Proc. SPIE Visual Communications and Image Processing 2002*, vol. 4671, pp. 1-10, Jan. 2002.
- [8] J. Wen, J.D. Villasenor, "Reversible Variable Length Codes for Robust Image and Video Transmission," *Proc. 31st Asilomar Conf. on Signals, Systems and Computers*, vol. 2, pp. 973-979, Nov. 1997.
- [9] E. Khan, S. Lehmann, H. Gargi, and M. Ghanbari, "Iterative Error Detection and Correction of H.263 Coded Video for Wireless Networks," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 12, pp. 1294-1307, Dec. 2004.
- [10] W. Chu, and J. Leou, "Detection and Concealment of Transmission Errors in H.261 Images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, no. 1, pp. 74-84, Feb. 1998.
- [11] H. Shyu, and J. Leou, "Detection and Concealment of Transmission Errors in MPEG-2 Images – A Genetic Algorithm Approach," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 6, pp. 937-948, Sept. 1999.
- [12] Y. Han and J. Leou, "Detection and Correction of Transmission Errors in JPEG Images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, no. 2, pp. 221-231, Apr. 1998.
- [13] S. Ye, X. Lin, and Q. Sun, "Content Based Error Detection and Concealment for image Transmission over Wireless Channel," *IEEE Proc. of ISCAS Conf.*, vol. 2, pp. 368-371, May 2003.
- [14] O. Lehtoranta, T.D. Hamalainen, and V. Lappalainen, "Detecting Corrupted Intra Macroblocks in H.263 Video," *IEEE Workshop on Multimedia Signal Processing*, pp. 33-36, Dec. 2002.
- [15] K. Bhattacharyya, and H.S. Jamadagni, "DCT Coefficient-Based Error Detection Techniques for Compressed Video," *IEEE Proc. of ICME Conf.*, vol. 3, pp. 1483-1486, Aug. 2000.
- [16] R.A. Farrugia, and C.J. Debono, "Enhancing the Error Detection Capabilities of the Standard Video Decoder using Pixel Domain Dissimilarity Metrics," *IEEE Proc. of EUROCON Conf. 2007*, accepted for publication.
- [17] R.A. Farrugia, and C.J. Debono, "Enhancing the Error Detection Capabilities of DCT Based Codecs using Compressed Domain Dissimilarity Metrics," *IEEE Proc. of EUROCON Conf. 2007*, accepted for publication.
- [18] R.O. Duda, P.E. Hart, and D.G. Stork, *Pattern Classification (Second Edition)*, pp. 161-350, John Wiley & Sons, USA, 2000.
- [19] D.F. Specht, "Probabilistic Neural Networks for Classification, Mapping or Associative Memory," *IEEE Int. Conf on Neural Networks*, vol. 1, pp. 525-532, Jul. 1988.
- [20] M.M. Gupta, L. Jin, and N. Homma, *Static and Dynamic Neural Networks from Fundamental to Advanced Theory*, pp. 105-215, John Wiley & Sons, USA, 2003.
- [21] R.A. Farrugia and C.J. Debono, "A Statistical Bit Error Generator for Emulation of Complex Forward Error Correction Schemes," *IEEE Proc. of ICC 2007*, Jun. 2007.