

# DVC USING A HALF-FEEDBACK BASED APPROACH

J.L. Martinez<sup>1</sup>, C. Holder<sup>2</sup>, G. Fernández-Escribano<sup>1</sup>, H. Kalva<sup>2</sup>, F. Quiles<sup>1</sup>

<sup>1</sup>Albacete Research Institute of Informatics. Universidad de Castilla-La Mancha. 02071 Albacete, Spain  
{joseluismm,gerardo,paco}@dsi.uclm.es

<sup>2</sup>Department of Computer Science and Engineering, Florida Atlantic University, Boca Raton, FL, USA  
{cholder@fau.edu, hari@cse.fau.edu}

## ABSTRACT

Distributed Video Coding has become increasingly popular in recent years among the researchers in video coding due to its attractive and promising features. DVC proposed a dramatic structural change to video coding by shifting the majority of complexity conventionally residing in the encoder towards the decoder. Nevertheless, these kinds of architectures have some serious limitations that hinder its practical application. The uses of a feedback channel between the encoder and the decoder requires an interactive decoding procedure which is a limitation for certain applications such as offline processing. On the other hand, the decoder needs an efficient way to estimate the probability of error without assuming the availability of the original video at the decoder. In this paper we investigate a first approximation to solve both problems based on the use of machine learning to extract the knowledge that exists between the residual frame and the number of requests over this feedback channel. Exploiting this correlation gives us a more practical architecture without higher complexity encoders. We apply these concepts to pixel-domain Wyner-Ziv coding and the results show a loss of 0.21 dB in the rate-distortion performance.

**Index Terms**— DVC, Wyner-Ziv Coding, Machine Learning, Turbo coding, Feedback Channel.

## 1. INTRODUCTION

*Distributed Video Coding* (DVC) is a new paradigm in video compression which is the consequence of information theoretic results obtained by Slepian-Wolf (SW) [1] for lossless Distributed Source Coding (DSC), and by Wyner-Ziv (WZ) [2] for the lossy case. This mathematical background states that, under the same conditions, the RD performance achieved when performing joint encoding and decoding (i.e. like in traditional video coding scheme) of two correlated sources, can also be obtained by doing separate encoding and joint decoding. DVC enables a

flexible allocation of complexity between the encoder and the decoder mainly due to the computationally intensive motion estimation task moved to the decoder side. As a result, DVC is becoming very appealing for a wide range of real life applications where the low complexity, computational cost, power consumption and / or memory constraints dictate a low complexity encoder. Applications include multimedia wireless sensor networks, wireless video surveillance, and mobile camera phones.

One of the pioneering DVC approaches is the turbo based WZ coding scheme presented in [3], where the encoder is responsible to explore the source statistics, and therefore to achieve compression following the WZ paradigm. Most of the architectures available in the literature are based on [3] and make use of a reverse channel to the encoder requesting more information for the decoding process; this approach is referred to as *feedback based architecture* in the literature. Moreover, all architectures which are based on [3] use an ideal error estimation at the decoder which is able to measure the distortion between the original frame and the partial reconstructed frame. The original frames are used to determine when a sufficient number of parity bits are received, in order to guarantee a residual BER (bit-error rate) below a given threshold, typically set to be equal to  $10^{-3}$ . This feature is also well-known in the literature and is referred to as *rate control at the decoder*. Such feedback based architectures offers significant problems in practical scenarios where the encoded video streams need to be stored for offline processing, and also when a bidirectional communication channel is not available. Feedback mechanics also have implications in terms of delay and decoder complexity since several iterative decoding operations may be needed to decode the data to a target quality level. Moreover, the ideal capacity inherent in these architectures is impractical for real implementations due to the dependence of the decoder on the original frame which is only available at the encoder. The two main drawbacks of current DVC architectures are the dependence on a feedback channel and the original frame. Therefore a solution without the need for a reverse channel or the ideal correction detection based on original frames is necessary to develop practical solutions using DVC.

---

This work has been jointly supported by the Spanish MEC and European Commission FEDER funds under grants "Consolider Ingenio-2010 CSD2006-00046" and "TIN2006-15516-C04-02", and by JCCM funds under grant "PAI06-0106".

This paper focuses on eliminating the two main drawbacks mentioned above using a block based turbo encoder in Pixel Domain which exploits the correlation between the residual frame and the number of requests (bits parity) for each block. The proposed approach uses Machine Learning to exploit the correlation and create decision trees to be implemented in a faster encoder. Our results show that the use of Machine Learning tools is appropriate for video processing and is an acceptable and practical solution for feedback channel free DVC architectures and the performance of the proposed architecture is closer to the architectures based on [3].

This paper is organized as follows: Section 2 presents an overview of the non-feedback architectures available in the literature. In Section 3, we present the major contribution of this paper: feedback channel free solution to DVC. Section 4 presents experimental results. Finally, conclusions and some future work topics are presented in section 5.

## 2. RELATED WORK

A common rate allocation adopted in DVC algorithms is the use of a feedback channel and a Rate-Compatible Punctured Turbo Code (RCPTC) [4]. In this configuration, the turbo encoder generates all the parity bits for the bitplanes to be encoded, saves these bits in a buffer, punctures and divides them into parity bits sets (chunks). These sets are sent to the decoder in response to requests via feedback channel. The turbo decoder's task is to correct the side information mismatches / errors using these parity bits sent by the encoder. In most DVC codecs available in the literature the decoder has the ideal capacity to measure the distortion between the partial reconstructed frame and the original one. In practice, this characteristic is not possible since the original frame is only available at the encoder. To solve this problem some approaches have been recently published in the literature.

The first attempt to solve the feedback problem was proposed in 2005 by *Artigas et al* in [5] where the authors try to determine the rate at the encoder based on empirical results obtained by examining different test sequences. *Adikari et al* in [6] propose a Unidirectional DVC codec based on two parallel Wyner-Ziv encoders using different interleaves index for each one in order to scatter the original image and the bit rate is set by the user. This architecture was enhanced employing an iterative decoding with gradually updated side information using spatial-temporal predictions. *Morbée et al* in [7] propose an algorithm based on a faster estimation at the encoder side of the optimum rate calculated in the ideally turbo decoder [3] from the parameters:  $\sigma^2$  (correlation noise) and  $P_k$  (error probability) to be calculate for each bitplane. *Brites et al* in propose a transform domain architecture employing a faster method to

calculate a closer approximation of the side information at the encoder side [8].

All these approaches try to allocate, in different ways, a proper number of bits to encode each video frame in order to get the optimum rate given the side information and the correlation noise model which is achieved in the ideal feedback based approach [3]. On the other hand, all these approaches deal with the frame as a whole and they do not take into account the different features within that frame. Normally, to get an optimum approach we need to estimate in a faster way the side information which will be generated at the decoder but, always, keeping the low complexity encoder devices as DVC philosophy.

## 3. PROPOSED ARCHITECTURE

Figure 1 shows the DVC codec architecture used in this paper. The architecture is based on the feedback based Pixel Domain architecture proposed in [3]. However, some modifications have been done (see *proposed* in Figure 1) mainly in the encoder rate control which is the major contribution of this paper.

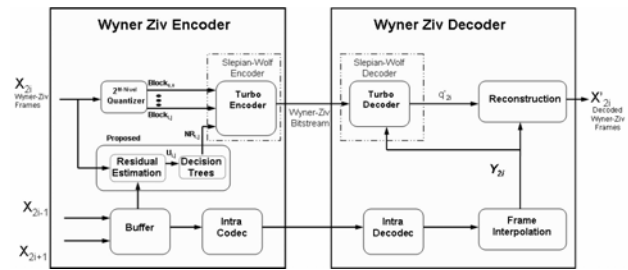


Figure 1. Video Codec Architecture Proposed

In this paper we propose a faster and efficient encoder rate control based on the information available in the residual frame (which is calculated in the Residual Estimation block, Figure 1) which may be exploited using Machine Learning Tools. Our approach is based on two important modifications to the common architecture [3]: block based input turbo encoder (see section 3.1) and the use of Machine Learning (see section 3.2) to exploit the correlation between the residual video frame and the number of requests over an imaginary feedback channel.

### 3.1. Block Based Architecture

The key to a perfect rate allocation at the encoder is to determine which parts of a frame to be encoded will have poor estimates of the side information estimated at the decoder. For these parts, the turbo decoder will spend more requests to correct the error. In this case, an intelligent encoder could spend more parity bits for these areas. On the other hand, the common turbo codes assume that the correlation noise statistics describing the relationship between the source and its side information is spatially

stationary and there is not any dependency on the spatial location into the frame. In other words, the quality of the side information is constant across the whole frame. As we know, video frames in a sequence can have areas that appear covered / uncovered, noise effects or illuminations changes which significantly affect the quality of the side information.

Because of this spatial dependency, using the whole bitplane as input to the turbo encoder (as done in [7] [8]) is not a good option. We can exploit the spatial dependency in a frame and determine an optimum encoder rate allocation for each part of a frame in the video sequence.

In order to determine the areas of a frame with poor estimate of side information we propose the use of  $n \times n$  block based turbo encoder. That is, the whole frame is divided into  $n \times n$  non-overlapping blocks, as in traditional video coding, and then the bitplane extraction is done on a block basis with blocks processed independently in a raster scan order. There are two main factors considered in determining the block size. Higher block lengths yield better turbo coding performance. On the contrary, the use of a smaller block size significantly improves the estimated side information. Moreover, smaller block sizes allow us more granularity in order to estimate different rates for different parts of the frame. The effect of the codec's overall performance with varying block size was analyzed and based on the experiments, a compromised block length of 16 ( $4 \times 4$  blocks) is used in the simulations. The block based scheme is not new in DVC community but it is novel to non-feedback based architectures.

### 3.2. Machine Learning

*Machine Learning* (ML) refers to the study of algorithms and systems that learn or acquire knowledge from experiences. Inductive ML uses the analysis of data sets for creating a set of rules to take decisions. These rules can be used, in the ML, to build a tree decision using a set of experiments or examples, named the training data set. The use of ML for video processing has primarily been in the areas of image and video understanding. The rate allocation in DVC creates an opportunity for applying ML to try to estimate in a faster way (a tree is an if-else-if operation) the rate at the encoder. So, it can be used to develop decision trees for each block trying to determine what optimum rate would be necessary in an ideal decoder scheme [3].

If we define  $U$  as the difference between the original frame which is available at the encoder and a low complexity estimation of the side information, such as the average operation between two adjacent  $K$  frames which are also available at the encoder,  $U = X_{2i} - Y'_{2i}$ . The turbo decoder will try to correct the errors in  $U$  using the parity bits sent by the encoder. From this assumption we need to extract some solid information in order to determine which parts of frame will be poorly estimated and on the other

hand we need to know how many requests will be processed for each one of these parts (blocks). Figure 2 shows the motivation behind the use of ML to try to determine the number of requests. Figure 2b shows the  $U$  image for the first frame of the foreman sequence and figure 2c shows the number of requests associated for each block. It is clear from the figure that the residual information of  $U$  is higher for the blocks which the number of requests (bits requests) is also higher and vice versa. This correlation between residual information and number of requests can be exploiting using ML tools. In our experiments, mean and variance of  $2 \times 2$  sub-blocks, non zero coefficients, the values range, the means of the variances of each  $4 \times 4$  block was used in the training stage using *Weka* [9] together the corresponding number of requests for the same block. This procedure is depicted in Figure 3. The result of this training process is a decision tree that estimates the number of requests based on the estimated side information.

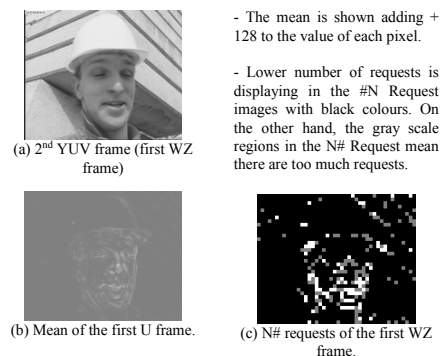


Figure 2. Exploiting the correlation with ML.

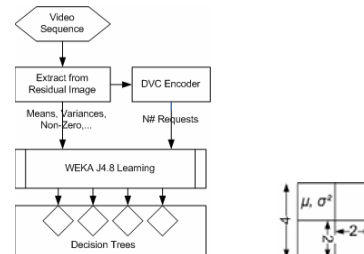


Figure 3. Training stage Procedure

## 4. EXPERIMENTAL RESULTS

In this section we present an extensive performance evaluation in terms of % of good choice of our decision trees and also we show the first approximation of a global RD performance. In the simulations, we have implemented half- feedback DVC architecture in order to get a first approximation of a feedback free architecture. Half-feedback in our approach indicates that the zero or non-zero request will be determined using a decision tree. In case of a block classified as NON-ZERO we use a feedback base DVC architecture as we shown in Figure 4. For blocks classified as ZERO no feedback is used.

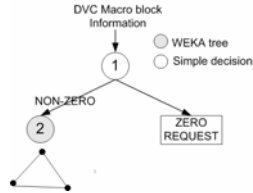


Figure 4. Half – Based DVC decision tree. First Level.

Table 1 shows the % of blocks correctly classified by our algorithm. WEKA was used to train the decision tree using the first frame of the Foreman sequence. Then, the tree was evaluated using common set of sequences used in DVC research.

Table 1. % of good choice

Sequence	1st MSB	2nd MSB	3rd MSB
CONTAINER	99.97	99.97	99.97
SALESMAN	99.82	99.55	99.05
HALL MONITOR	99.97	99.95	99.86
MOBILE	99.13	98.46	97.42
AKIYO	100.00	100.00	99.97
CARPHONE	99.43	99.00	96.52
CLAIRE	99.82	99.70	99.58
COAST	99.48	96.73	96.10
MISS	100.00	100.00	99.95
MOTHER	99.99	99.99	99.96
<i>mean</i>	99.76	99.33	98.84

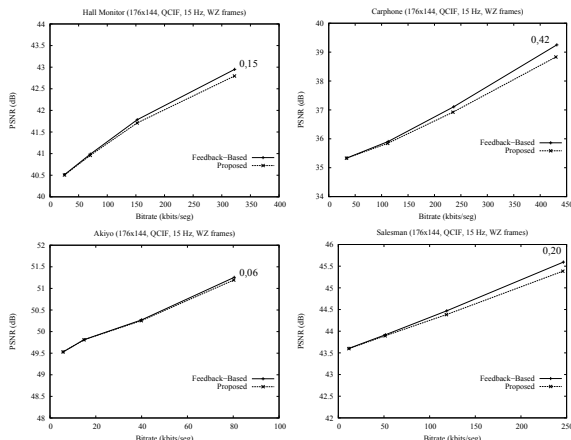


Figure 5. RD performance. Proposed Vs. Feedback-based

Figure 5 shows the RD performance of four different sequences listed in the table 1. The results show a very similar performance for higher bitplanes (lower bit rate) and a gap around 0.21 dB on average for higher rates. The drop in PSNR is because of lower accuracy of the faster estimation obtained at the encoder for higher levels.

The results are achieved in a quasi-ideal scenario that means, we save, on average, 90%, 81% and 69% the use of feedback channel as shown in Table 2.

## 5. CONCLUSIONS

In this paper we describe a practical algorithm that can be used to detect the correct number of parity bit requests in a Wyner-Ziv video coding architecture, without access to the

original frames at the decoder and without the full use of the feedback channel. We focus on the use of Machine Learning based solutions to solve these problems and we have implemented the first level of our decision tree. Results obtained demonstrate that the coding efficiency loss is negligible at low bit-rates, while it is below 0.21dB at high bit-rates compared to full-feedback based architecture. Ongoing work is extending this work into a complete tree for a feedback-free solution.

Table 2. % Saving the use of feedback

Sequence	1st MSB	2nd MSB	3rd MSB
CONTAINER	88.61	87.03	77.97
SALESMAN	93.85	81.72	70.24
HALL MONITOR	91.35	84.47	70.83
MOBILE	79.61	58.93	38.50
AKIYO	98.48	97.30	91.98
CARPHONE	84.61	66.12	50.74
CLAIRE	90.93	88.89	83.88
COAST	86.53	73.16	40.20
MISS	97.90	92.64	90.04
MOTHER	92.35	86.57	79.62
<i>mean</i>	90.42	81.68	69.40

## 6. REFERENCES

- [1] D. Slepian, J.K. Wolf, "Noiseless coding of correlated information sources," IEEE Trans. on Inform. Theory, vol. IT-19, pp. 471–480, July 1973.
- [2] Wyner and J. Ziv, "The Rate-Distortion Function for Source Coding with Side Information at the Decoder". IEEE Trans. on Inform. Theory, vol. IT-22, pp. 1–10, Jan. 1976.
- [3] B. Girod, A. Aaron, S. Rane, D.R. Monedero, "Distributed Video Coding". Proc. of IEEE Special Issue on Advances in Video Coding and Delivery, Vol.93, No. 1, pp.1-12, 2005.
- [4] D. Rowitch and L. Milstein, "On the performance of hybrid FEC/ARQ systems using rate compatible punctured turbo codes," IEEE Trans. Commun., vol. 48, no. 6, pp. 948–959, June 2000
- [5] X. Artigas, L. Torres, "Improved signal reconstruction and return channel suppression in Distributed Video Coding systems", 47th International Symposium ELMAR-2005, Multimedia Systems and Applications, June 2005.
- [6] A.B.B. Adikari, W.A.C. Fernando, W.A.R.J. Weerakkody, "Iterative Wyner-Ziv Decoding for Unidirectional Distributed Video Coding," IEE Electronics Letters. Vol. 43, Issue 2, pp. 93 – 95, January 2007.
- [7] M. Morbée, J. Prades-Nebot, A. Pizurica, and W. Philips, "Rate allocation algorithm for pixel-domain distributed video coding without feedback channel", ICASSP, April 2007, Volume 1, pp. I-521-I-524.
- [8] C. Brites, F. Pereira, "Encoder rate control for transform domain Wyner-Ziv video coding", ICIP, September, 2007.
- [9] Ian H. Witten and Eibe Frank, "Data Mining: Practical Machine Learning Tools and Techniques". 2nd Edition, Morgan Kaufmann, San Francisco, 2005.