

IMPROVED WYNER-ZIV VIDEO CODING EFFICIENCY USING BIT PLANE PREDICTION

Jeffrey J. Micallef, Reuben A. Farrugia and Carl J. Debono

Department of Communication and Computer Engineering, University of Malta, Msida, Malta
 jeffrey.micallef@ieee.org, reuben.farrugia@um.edu.mt, c.debono@ieee.org

ABSTRACT

Distributed Video Coding (DVC) is a coding paradigm where video statistics are exploited, partially or totally, at the decoder. The performance of such a codec depends on the accuracy of the soft-input information estimated at the decoder, which is affected by the quality of the side information (SI) and the dependency model. This paper studies the discrepancies between the bit planes of the Wyner-Ziv (WZ) frames and the corresponding bit planes of the SI. The relationship between these discrepancies is then exploited to predict the locations where the bit plane of the SI is expected to differ from that of the original WZ frame. This information is then used to derive more accurate soft-input values that achieve better compression efficiencies. Simulation results demonstrate that a WZ bit-rate reduction of 9.4% is achieved for a given video quality.

Index Terms— Distributed Video Coding, Soft-information estimation, bit plane error prediction.

1. INTRODUCTION

Traditional video coding schemes vest in the encoder to exploit the temporal and spatial redundancies in a video sequence for compression, resulting in encoders which are 5 to 10 times more complex than the decoder [1]. However, emerging video capturing devices, such as multimedia mobile phones and wireless cameras, are imposing a serious challenge to this video coding paradigm due to their limited resources. Distributed video coding (DVC) is suggested as a new video coding paradigm to minimize the encoder's complexity. DVC finds its foundations in the Slepian-Wolf theorem [2] and Wyner-Ziv (WZ) theorem [3]. Together, these theorems prove that two correlated sources can be encoded separately and jointly decoded at the same coding efficiency obtained when the sources are jointly encoded and decoded. This suggests that low-complexity encoding can be achieved by shifting the computationally-expensive tasks of exploiting the source statistics, partly or wholly, to the decoder without affecting the coding performance.

The DVC decoder generates a prediction of the current Side Information (SI) using previously decoded information. The dependency error between the bit planes of the WZ frame and those of the SI is then modeled using a bit plane-based dependency model and used to convert the SI bits into statistical soft-input values. These values are

then fed in the channel decoder where parity information is requested to correct the errors present in the SI.

The accuracy of the soft-input values, and thus the compression efficiency, depends on the quality of the SI and the bit plane-based dependency model. The latter problem was previously tackled in [4], assuming a Binary Symmetric Channel (BSC) dependency. Later on, the authors in [5-6] obtained this model from the conditional probability $P(WZ|SI)$, which is assumed to follow a Laplacian distribution. However, these techniques do not exploit the correlation that exists between the discrepancies of different bit planes. This paper considers the discrepancies between the bit planes of the WZ frames and the corresponding bit planes of the SI and shows that there exists a high correlation between the discrepancies present at consecutive bit planes. These correlations are exploited to predict most of the discrepancies expected at the bit plane being decoded, using the previously decoded bit planes. The dependency model then uses these predictions to estimate the soft-input values for the WZ video codec, obtaining a reduction of up to 9.4% in WZ bit-rate, for a given video quality.

The paper is organized as follows: Section 2 gives an overview of the Wyner-Ziv video coding architecture used. In Section 3, briefly reviews the traditional models used to estimate the soft-input information. Section 4 discusses the correlation between the discrepancies at different bit planes and introduces an algorithm to predict them, while in Section 5 this algorithm is exploited in the model. Finally, the experimental results are presented in Section 6 and Section 7 provides the final comments and conclusion.

2. WYNER-ZIV CODING FRAMEWORK

The Wyner-Ziv video codec architecture adopted in this paper is based on the IST Pixel Domain architecture proposed in [7] and is shown in Fig. 1. The incoming frame sequence is divided into key frames (odd frames) and WZ frames (even frames). The key frames (X_{2i-1} and X_{2i+1}) are assumed to be losslessly available at the decoder, while the WZ frames (X_{2i}) are divided into four quadrants which are encoded separately. All the pixels in each quadrant are uniformly quantized using 2^M levels and fed, bit plane by bit plane, to the LDPCA encoder [8]. The resulting parity bits are then stored in a buffer and transmitted upon the decoder's request. An 8-bit Cyclic Redundancy Check (CRC) is also calculated and transmitted for each bit plane.

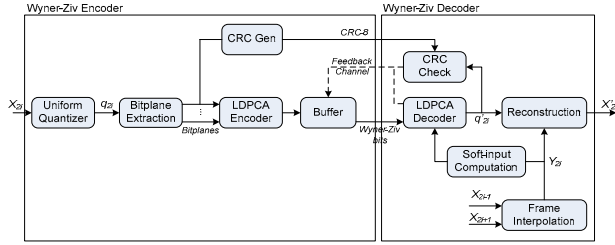


Figure 1 - Wyner-Ziv video coding architecture.

At the decoder, the SI is generated using motion compensated interpolation (MCI) techniques [7] between the two adjacent key frames X_{k-1} and X_{k+1} . The dependency error between the original WZ frame and the generated SI is assumed to follow a Laplacian distribution, whose parameters are calculated at block level in an offline process. This model is then used to convert the SI bits into soft-input values representing bit-probabilities. For each quadrant, the validity of the generated SI bits is first assessed through a CRC check. If the CRC matches, the quadrant is reconstructed using the SI bits directly without requesting any parity information. However, if it fails, the decoder requests parity information from the encoder's buffer through the feedback channel. This information is concatenated with the soft-input values and fed into the LDPCA decoder. Channel decoding is performed and if decoding fails more parity information is requested until decoding is successful. The Wyner-Ziv frame is then reconstructed with the help of the side information Y'_{2i} using a minimum-mean-square-error reconstruction.

3. SOFT INFORMATION MODULE

The soft-input values represent the level of belief in the value of a codeword's bits using the log-likelihood ratio:

$$L(x_i) = \log \left(\frac{p(x_i = 0)}{p(x_i = 1)} \right) \quad (1)$$

where $p(x_i = 0)$ and $p(x_i = 1)$ gives the probability that bit x_i is a 0 or a 1. The value of $L(x_i)$ is positive when $p(x_i = 0) > p(x_i = 1)$ and negative otherwise. The larger the magnitude of $L(x_i)$, the higher the belief in the bit value of x_i . Hence the sign of $L(x_i)$ provides the hard decision on bit x_i while the magnitude $|L(x_i)|$ represents the reliability of this decision.

The correlation noise between the SI and the original WZ frame must first be modeled to allow the evaluation of the bit-probabilities. Such model gives the confidence measure of the frame interpolation operation, resulting in large α parameters in areas with low SI reliability. Given the value of the SI and the corresponding α parameter, the conditional probability $P(WZ|SI)$ is obtained by centering a Laplacian distribution with an α parameter around SI.

The conditional bit probabilities are then evaluated by considering the integral of the probability function $P(WZ|SI)$ below the relevant areas, as indicated by the previously decoded bits [6]. A simple example of such decoding algorithm is considered in Fig. 2, using a uniform quantizer

with 8-levels. Assume that the original WZ pixel x has a value of 140 and the corresponding SI has a value of 170. The bit-probabilities $p(x_i = 0)$ and $p(x_i = 1)$ are initially obtained by evaluating the area under graph $v \in [0, 3]$ and $v \in [4, 7]$ respectively, and fed in (1) to generate the soft-information. After decoding the first bit plane, it is known that the first bit has a value of 1 ($x_0 = 1$), hence the probabilities $p(x_i = 0)$ and $p(x_i = 1)$ are evaluated by calculating the area under graph $v \in [4, 5]$ and $v \in [6, 7]$ respectively. The 2nd bit is decoded as 0, and the areas $v \in [4]$ and $v \in [5]$ are used to calculate the bit probabilities for the 3rd bit. This dependency model was adopted by the latest state-of-the-art codecs, such as the DISCOVER codec [9] and is used throughout this paper as a reference model for comparison purposes.

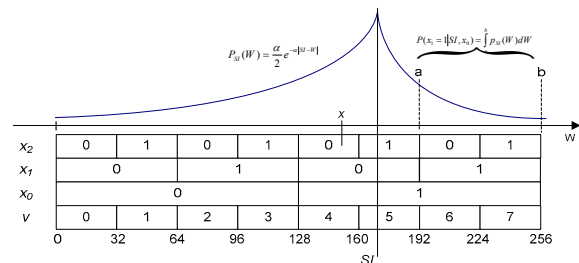
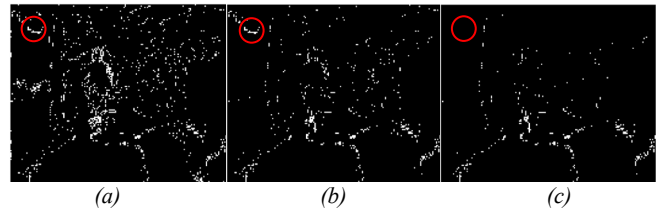


Figure 2 - Estimation of bit probabilities.

4. PREDICTION OF ERROR MAPS

The 51st frame of the Hall monitor sequence and its corresponding SI are used as an illustrative example to study the relationship between the discrepancies occurring at consecutive bit planes. The two frames were uniformly quantized using 32-levels (5-bits) and a XOR operation was performed on the 3rd, 4th and 5th bit planes of the two frames, where the 5th bit plane represents the most significant bit plane. The images produced are shown in Fig. 3. The white dots represent the locations where the bits of the two bit planes differ and hence the discrepancies occurring between the bit planes of the SI and those of the original WZ frame. For this reason these images will be called *error maps*.


 Figure 3 - Discrepancies in the (a) 3rd, (b) 4th and (c) 5th bit planes.

From the *error map*, it is clear that there exists a significant correlation between the discrepancies occurring at the different bit plane levels. It can be observed that the discrepancies occurring at the 3rd and 4th bit plane are a superset of the discrepancies occurring at the next higher significant bit plane, i.e. 4th and 5th bit plane respectively. This can be explained by looking at the binary representation of integer values. From Fig. 4, it is evident

that a small change in value will flip the value of a set of consecutive low significant bits up till a certain bit plane level. In DVC, these small changes between the values of the WZ frame and its corresponding SI are very frequent. For such locations, the *error map* for the 3rd bit plane includes the locations where bit flipping stopped at the 3rd bit plane together with the locations where bit flip propagates to higher order bit planes.

Value	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
B I T V A L U E S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1

Figure 4 – Binary representation of integer values.

The discrepancies at the 4th bit plane can therefore be predicted by traversing the locations indicated by the *error map* of the 3rd bit plane while filtering out the locations where bit flipping is not expected to propagate to the 4th bit plane. Such locations can be identified using the algorithm:

Step (1) – Search for the first value nearest to the SI, for which the 3rd bit differs from that of the SI;

Step (2) – If the 4th bit of this value differs from that of the SI as well, then it is assumed that bit flipping propagates to the 4th bit plane, and this location is included in the error map for the 4th bit plane.

The intuition is that the discrepancy at the lower significant bit plane indicates that the value of the SI is not a good predictor for the original WZ pixel value. The value nearest to the SI for which the 3rd bit agrees with that of the original WZ value is considered as a better predictor (Step 1), and is used to determine whether the 4th bit is expected to be flipped as well (Step 2). This procedure is valid even when the difference between the WZ value and SI value is greater than one, and becomes more accurate when predicting higher significant bit planes.

Consider the *error maps* in Fig. 3 as an example. Fig. 3(a) shows the discrepancies occurring at the 3rd bit plane. To predict most of the discrepancies at the 4th bit plane, these locations are scanned one by one. When considering the pixel locations marked by the red circle, it is observed that the original quantized WZ pixel value is 9 and its corresponding SI is 6. At the decoder, the value nearest to the SI which has a different bit value for the 3rd bit is the value 8 (see Fig. 5, Step 1). This value differs from the value of the SI for the next higher significant bit (step 2), hence it is expected that the 4th bit plane of the WZ frame and that of the SI differs at these locations. In fact, Fig. 3(b) indicates that the locations in the red circle differ for the 4th bit plane as well. When using the *error map* of the 4th bit plane to predict the discrepancies in the 5th bit plane, the same locations will still be considered. Once again the value 8 is chosen in step 2, however, this time the bit at the next

higher significant bit plane does not differ from that of the SI. It is therefore predicted that the 5th bit plane of the two frames does not differ at these locations, which is a correct prediction according to Fig. 3(c).

Value	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
B I T V A L U E S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1

Figure 5 – Example of error prediction for the 4th bit plane.

5. PROPOSED DEPENDENCY MODEL

This section discusses how the prediction of the *error maps* can be used by the bit plane-based dependency model. For the decoder to be able to predict such *error maps*, the bit planes are encoded starting from the least significant bit plane. The discrepancies from the previously decoded bit plane are first obtained by XORing the bit plane of the quantized SI with the bit plane obtained after successful decoding. This is then used, together with the SI, to predict the *error map* for the bit plane that is decoded next. From the predicted *error map*, a prediction of the bit plane bits is obtained by inverting the bits of the quantized SI at the locations indicated by the *error map*. The predicted bit plane bits are then used to determine the sign (hard decision) value for the soft-input, by assigning a positive value to the locations predicted to have a value of 0 and a negative value to the rest.

This algorithm was used to predict the 3rd, 4th, and 5th bit planes of frame 51, from their predicted *error maps*. The accuracy of the predicted bit plane bits was assessed by XORing the original bit plane with the predicted bit plane and the results obtained are shown in Fig. 6. Once again, the white dots indicate the location where the two bit planes differ and hence the errors in the predicted bit plane.



Figure 6 – Errors between the predicted and the original bit planes.

Fig. 6 illustrates that the proposed algorithm gives a good prediction of the required bit plane bits (hard decision values for the soft-input information). A good confidence value is however still needed. Fig. 6 shows that the errors in the predicted bit plane bits reside in those areas where the quality of the SI is very poor, such that the assumption of a small change is not valid. Hence, the confidence measure (magnitude) for the new soft-input value can still be obtained from the conditional probability $P(WZ|SI)$ as

described in section 3, since this will guarantee a low confidence value for areas having low SI reliability.

It was observed that the least significant bit plane (1st bit plane) and the next higher significance bit plane (2nd bit plane) are highly uncorrelated, and hence the proposed algorithm cannot be used to predict the discrepancies occurring at the 2nd bit plane from those in the 1st bit plane. For this reason, it is proposed that the conventional soft-input values should be used when decoding the first two bit planes, while the proposed soft-input values should be used when decoding the 3rd bit plane onwards.

6. EXPERIMENTAL RESULTS

The Pixel-Domain WZ codec described in section 2 was used to analyze the accuracy of the soft-input values estimated with the proposed model. The rate-distortion (RD) performance obtained when encoding the first 100 frames of the Foreman sequence and the Hall monitor sequence is plotted in Fig. 7 and Fig. 8 respectively. These plots take into consideration only the Luminance component of the WZ (even) frames. Both sequences have a QCIF resolution and are encoded with a GOP size of 2, at a WZ frame rate of 15 frames per second. The different RD points were obtained using the quantization levels $2^M \in \{4, 8, 16, 32\}$.

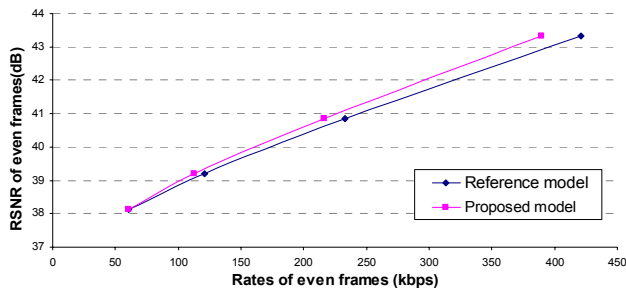


Figure 7 – Rate-Distortion performance for the Foreman sequence.

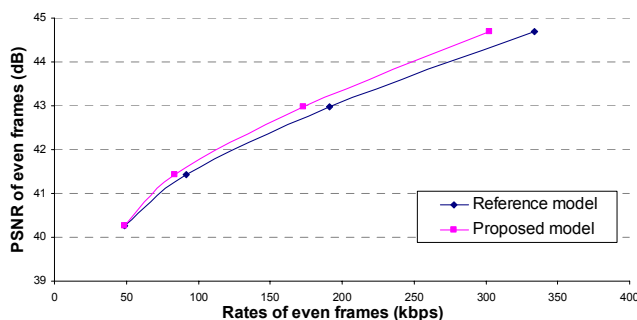


Figure 8 – Rate-Distortion performance for the Hall Monitor sequence

It can be observed that the proposed dependency model leads to an improvement in the RD performance, compared to the reference model (described in section 3). Since the side information module and the reconstruction module were not modified, the same PSNR values were obtained for the two dependency models. However, the soft-input values estimated with the proposed model were more accurate, leading to a decrease in WZ bit-rate, shifting the RD curves to the left for the same target quality. For some

quadrants the prediction of the bit plane bits was so accurate that the initial CRC was satisfied immediately, without requesting any parity information.

It was also observed that the rates required by the proposed dependency model, when correcting the errors present in every WZ frame, were always smaller than those required by the reference model. Table I summarizes the percentage reduction in Wyner-Ziv rates relative to the reference model that were achieved when using the proposed dependency model.

Table I – Percentage reduction in Wyner-Ziv rates.

No of levels	8	16	32
Foreman	6.46%	6.90%	7.52%
Hall Monitor	8.84%	9.33%	9.41%

7. CONCLUSION

This paper presented an algorithm which can predict most of the discrepancies between the bit planes of WZ frame and those of the corresponding SI, by exploiting the relationship between the discrepancies occurring at consecutive bit planes. Such prediction is then considered by the dependency model when estimating the soft-input values. Experimental results showed that, for Wyner-Ziv (WZ) video coding, the proposed soft-input estimations reduced the WZ bit-rates by up to 9.4% for a given video quality. Such gains were obtained without sacrificing the encoder's complexity and without the need of transmitting any extra information between the encoder and the decoder.

8. ACKNOWLEDGEMENT

The research work is partially funded by STEPS-Malta and partially by the European Union - ESF 1.25.

9. REFERENCES

- [1] T. Wiegand, G. Sullivan, G. Bjøntegaard and A. Luthra, "Overview of the H.264/AVC Video Coding Standard," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 13, No. 7, pp. 560-576, July 2003.
- [2] J. Slepian and J. Wolf, "Noiseless Coding of Correlated Information Sources", *IEEE Trans. Inform. Theory*, vol. 19, No. 4, pp. 471-480, July 1973.
- [3] A. Wyner and J. Ziv, "The Rate-Distortion Function for Source Coding with Side Information at the Decoder," *IEEE Trans. Inform. Theory*, vol. 22, No. 1, pp. 1-10, January 1976.
- [4] A. Aaron, S. Rane, E. Setton, and B. Girod, "Transform Domain Wyner-Ziv Codec for Video," VCIP, San Jose, CA, January 2004.
- [5] M. Dalai, R. Leonardi, and F. Pereira, "Improving Turbo Codec Integration in Pixel-Domain Distributed Video Coding," ICASSP, Toulouse, France, vol. 2, pp. 537-540, May 2006.
- [6] C. Brites, F. Pereira, "Encoder Rate Control for Transform Domain Wyner-Ziv Video Coding," ICIP, San Antonio, USA, September 2007.
- [7] J. Ascenso, C. Brites and F. Pereira, "Improving Frame Interpolation with Spatial Motion Smoothing for Pixel Domain Distributed Video Coding," *EURASIP Conference on Speech and Image Processing, Multimedia Communications and Services*, Slovak Republic, July 2005.
- [8] D. Varodayan, A. Aaron and B. Girod, "Rate-Adaptive Codes for Distributed Source Coding," *EURASIP Signal Processing Journal*, vol. 86, no. 11, November 2006.
- [9] X. Artigas, J. Ascenso, M. Dalai, S. Klomp, D. Kubasov, M. Ouaert, "The DISCOVER Codec: Architecture, Techniques and Evaluation," PCS, Lisbon, Portugal, November 2007.