

WYNER-ZIV TO H.264 VIDEO TRANSCODER

J. L. Martínez¹, H. Kalva², G. Fernández-Escribano¹, W.A.C Fernando³, and P. Cuenca¹

¹Universidad de Castilla-La Mancha, Albacete, Spain. [joseluismm,gerardo,pcuenca]@dsi.uclm.es

²Dep. Computer Science and Engineering, Florida Atlantic University, USA. hari@cse.fau.edu

³Center for Communications Research, University of Surrey, UK. W.Fernando@surrey.ac.uk

ABSTRACT

This paper proposes an improved Wyner-Ziv to H.264 video transcoder as part of a framework for mobile to mobile video communications. In this scheme, the user devices keep the low complexity constraints by using the Wyner-Ziv encoding and H.264 decoding algorithms. They shift their complexities to the network where the proposed transcoder is allocated. The main goal of the transcoder is to convert the bitstream and reduce the delay efficiently. The results show that the proposed transcoder reduces the complexity by a factor of 95% with a negligible rate-distortion loss.

Index Terms— DVC, Wyner-Ziv, H.264, AVC, transcoding

1. INTRODUCTION

Today digital video coding relies on the so-called motion compensated hybrid scheme, adopted by all MPEG and ITU-T video coding standards. In this configuration the encoders may become rather complex in comparison with decoders [1]. On the other hand, *Wyner-Ziv* (WZ) video coding [2], a particular case of *Distributed Video Coding* [3] (DVC), is known to be a technique used to reduce the processing complexity of the encoder, leading to a low-cost implementation, while the majority of the computations are taken over by the decoder.

Mobile-to-mobile video communication such as video telephony is becoming increasingly important. Mobile phones, PDA and the new portable devices have enabled the use of this kind of communications where, as a main characteristic, both the transmitter and receiver devices may not have the necessary computing power, resources or complexity constraints to perform complex video algorithms. Nevertheless, the requirements to have low complexity at both encoder and decoder side have not been met using traditional video codecs such as [1]. This scenario is depicted in Figure 1.

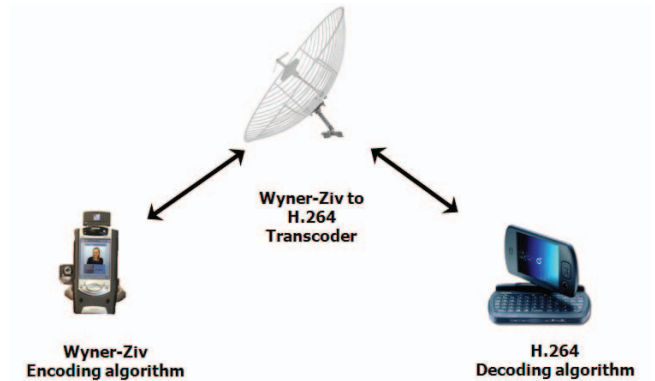


Figure 1. System using a WZ/H.264 transcoder.

Therefore, in order to efficiently exploit the advantages that these two video coding paradigms can offer in terms of low complexity encoding (using WZ coders [2]) and decoding (using traditional video decoders such as [1]), this paper proposes the use of an improved WZ to H.264 video transcoder device regarding to the reference one (formed by the concatenation of the full WZ decoding and H.264 encoding algorithms) which is located in the network that converts the WZ video from the lower complexity sender to H.264 video to the lower complexity receiver in order to efficiently support these mobile-to-mobile video communications. Moreover, the proposed transcoder has been improved in two steps: 1) by reusing the *Motion Vectors* (MVs) of the incoming side information generation process (the motion estimation done at the WZ decoding) in order to reduce and focalize the search area and window size done in the H.264 encoding algorithm and, 2) by reusing some correlated information from the WZ decoding, so that the transcoder selects only a sub-set of the different partitions of the structure motion compensation done in the H.264 inter prediction algorithm.

The paper is organized as follows: Section 2 identifies the state-of-the-art in WZ based transcoders. Section 3 shows our improved WZ/H.264 video transcoder which is evaluated in Section 4 with some simulation results. Finally, in Section 5, conclusions are presented.

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 "PEII09-0037-2328" and "PII2I09-0045-9916", and by UCLM funds under "TC20091095". The work presented was developed as a part of VISNET II, a European Network of Excellence (<http://www.visnet-noe.org>), funded under the European Commission IST FP6 programme.

2. RELATED WORK

One of the most referenced overviews of DVC [3], mentioned as one of the benefits of this new video coding paradigm, is the support of mobile-to-mobile communications using a transcoder device. This work was presented in 2005 but so far not much work has been done on the transcoding problem. This first WZ video transcoder approach was presented by Peixoto *et al* in [4] in 2008, which is based on WZ/H.263 to support this kind of communications. The authors propose a mapping between the WZ *Group of Pictures* (GOP) and the traditional GOP and, moreover, for the P or B slices in H.263 some ME refinement is also proposed. However, they did not exploit the correlation between the WZ MVs and the traditional ME successfully and only used them to determine the starting centre of the ME process. In our approach, both the search range and search window are reduced based on the length and the orientation of the MV estimated at the WZ decoder. Moreover, our proposal also accelerates the inter prediction done at H.264. Furthermore, it is well known that the rate distortion performance of the H.264/AVC video standard is much superior to the performance of H.263 standard; therefore the overall performance of the transcoder will be increased. As previous work, two independent approaches have been proposed in the framework of WZ to H.264 transcoder [5,6]. The work presented in this paper is a combination of these approaches working together.

3. PROPOSED VIDEO TRANSCODER

The key insight of the proposal consists of perfectly exploiting the information that the WZ decoding algorithm offers in order to reuse it within the H.264 encoding algorithm because to estimate some information again is a waste of time.

The end-user devices of the proposed scenario employ a WZ encoder and a H.264 decoder; the lower complexity parts in both paradigms (as Figure 1 shows). The proposed transcoder is depicted in Figure 2. The WZ encoder used in the sender device is our previous WZ architecture based on *Transform Domain* (TD) and *Turbo Trellis Coded Modulation* (TTCM) that was proposed in [7]. TTCM is a well-known channel coding technique used to optimize the bandwidth requirements while protecting the information bits by increasing the size of the symbol constellation. The WZ encoding algorithm works as follows: some frames (called key frames) are coded with a regular intra-frame encoder (in this case H.264/AVC Intra). The other frames (called WZ frames) are encoded using the WZ paradigm [2]. At the WZ sender, key frames are encoded using H.264 Intra; these frames are passed through the receiver without any transcoding process as I frame in the transcoder device. All the computations are taken over by the transcoder which is the major contribution of this paper.

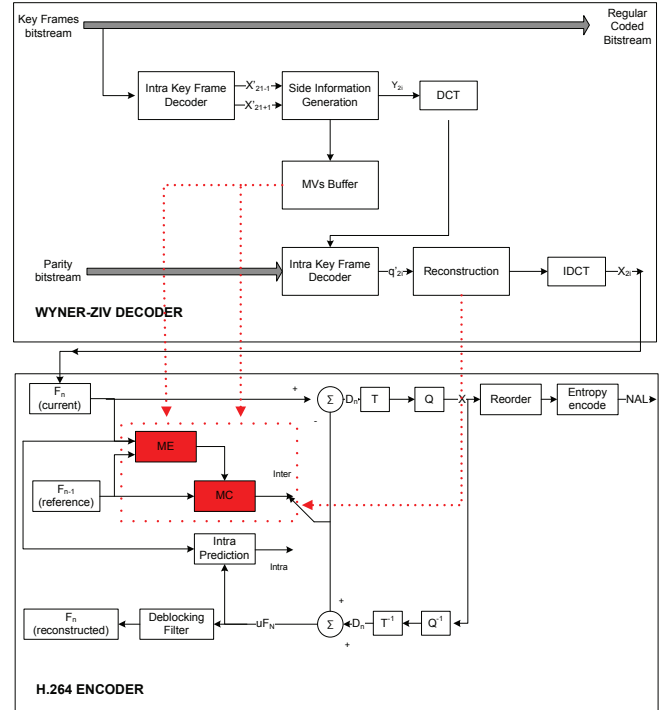


Figure 2. Proposed WZ to H.264 Video Transcoder

It is well-known that the motion estimation and inter prediction done at H.264 encoding are the most complex tasks and, therefore, they are the most suitable processes to be accelerated. In the inter frame coding of H.264, seven different block division modes (16x16, 16x8, 8x16, 8x8, 8x4, 4x8 and 4x4) can be selected for the ME prediction. Moreover, H.264 adopts the spatial domain intra prediction in the block sizes 16x16 and 4x4, which include four and nine directional predictions, respectively. For each of the latter sub-partition, the ME is performed using each partition as original data and trying to find the closest one. Therefore the ME is performed many times per MB. In this paper the two processes have been accelerated by re-using information incoming from WZ decoding stage. Section 3.1 depicts the motion estimation reduction and Section 3.2 shows a decision tree algorithm developed to reduce the different MB partition that can be checked.

3.1. Motion Estimation Reduction

The communication delay and the transcoder complexity depend largely on the search range used in the H.264 ME process. Therefore, adapting the search range based on the *Motion Vectors* (MVs) of the incoming MB could accelerate the system without severely impacting the PSNR-bitrate.

In the proposal, the search range for every H.264 MB is adaptively determined; and it is recalculated for every MB (or sub-MB partition) that can occur in the MB mode coded decision and reduced depending on the length and the orientation of the incoming side information MVs. The ME complexity can be further reduced by exploiting the side information MVs orientation. It is used to determine a

dynamic search range area around the side information MVs orientation. The description of the approach is depicted in Figure 3. The MB with MVs equal to (0,0) (the null MV) in the side information generation represents very simple MBs, so the search range area is limited to 1 pixel. On the other hand, in the case of incoming MB modes that the MV is higher than (0,0), the search range is determined by the area created by the circumference equation, centered in the (0,0) point for each H.264 mode or sub-mode, with the length of the incoming vector such as the radius of the circumference (see Figure 3). In this way, the length of the side information MV will limit the search range area. The same process is developed with the rest of the H.264 MB modes, generating a set of circle areas where the original side information MV is refined. This procedure is depicted in Figure 2 by a line connected from MVs buffer to ME module.

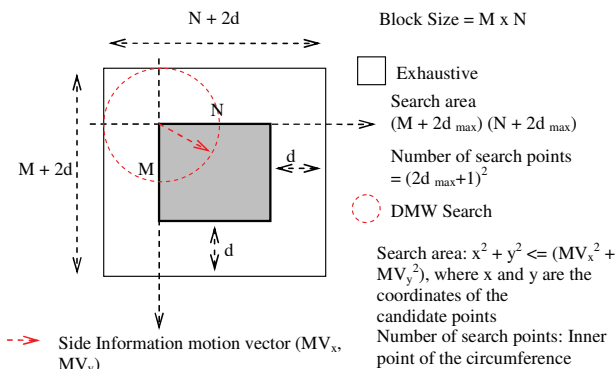


Figure 3. Proposed reduced-ME algorithm

3.2. Inter Prediction Reduction

The approach depicted here to accelerate the MB mode coded partition used in H.264 is based on a *Machine Learning* (ML) process that generates a decision tree that selects only a sub-set of partition modes based on the correlation of three parameters extracted from WZ decoding algorithm. Subsequently, the correlation of these parameters according to the different partitions is depicted.

In WZ video decoding, the side information generation process is the procedure where an estimation of the current frame (available only at the encoder) is constructed at the decoder. We found, in some experimental observation, that the stationary areas or object with slow motion or with slow camera motion are often coded in inter mode for inter-frame coding with larger block partition (such as 16x16, 16x8 or 8x16) or even as *Skipped*. On the other hand, the regions with scene change, with light change or with the object which appears suddenly are coded in inter-mode with lower MB mode partition (such as 4x4) or even to *Intra* mode for the inter-frame coding. In fact, this is one of the motivations for our approach, the SAD calculation derived for side information generation of the motion compensation backward prediction frame could be used to determine the similarity between the current block and the corresponding block in the previous frame (in a P frame in H.264).

Moreover, in some situations, although SAD is large, the blocks are still chosen as inter mode blocks. For example, the objects with uniform and fast motion in the block cause the SAD to be large. However, in this case, inter mode coding performs better than *Intra* mode. To avoid this problem, the proposed algorithm adds another factor to prevent the temporal correlation of the MB in two adjacent frames: the length of the MVs.

Finally, the turbo decoder task is to try to correct the side information mismatches using the parity bit sent by the encoder. Once the turbo decoding algorithm is successful, the reconstruction is performed; taking the corresponding side information sample if it falls into the quantization interval [2]. Based on the observations, we found that the number of reconstructed pixel index that differs from the side information quantization bin is higher for the blocks that are poorly estimated by the side information and these blocks are then mapped into more complex MB partitions such as 4x4 or *Intra*. Based on this observation, the parity bit information sent by the encoder is also taken into account by our algorithm in order to determine the final MB partition decision.

In a nutshell, there are three types of information extracted from the WZ decoding algorithm that we found much correlated regarding the MB coding mode partition performed in a P frame in H.264. They can be used in a ML process in order to convert this knowledge into faster rules that can replace the most complex MB coding mode decision procedure. These are the value of the SAD, the length of the MVs and the pixels that fall off to the quantization interval bin that have been described above.

We extracted the latter information (correlated to the class variable) per each MB and we call the ML learning algorithm. In this work the well-known C4.5 [8] algorithm proposed by *Ross Quinlan* has been called ML algorithm. The training file was generated using 10 frames of QCIF *flower and garden* sequence, 30 fps. The ML algorithm gives us a decision tree formed by MB statistics variables and classifies each MB into a set of different MB mode partitions. Basically, the final decision follows a binary-decision tree where it has been created per level according to the relations between MB partitions. In this way the training process was developed at the following levels: The first level is divided into *LOW COMPLEXITY* and *HIGH COMPLEXITY* modes. The first one is formed by {SKIP, 16x16, 16x8, 8x16} and the second one by {8x8, 8x4, 4x8, 4x4 and INTRA}. Then, within the *LOW COMPLEXITY* tree, we distinguish between two more branches. {SKIP, 16x16} and {16x8, 8x16}. In terms of the *HIGH COMPLEXITY* tree, it is divided into other two branches: {8x8, 8x4 and 4x8} and {4x4 and INTRA}. Finally, the fourth level continues with the tree formed by {8x8, 8x4 and 4x8}, it splits up into two branches: {8x8-4x4 DCT, 8x8-8x8 DCT} and {8x4, 4x8}. All this process is depicted in Figure 4. The gray circles mean that we force the decision

by a tree and in the white ones we leave freedom to H.264 to select the better option.

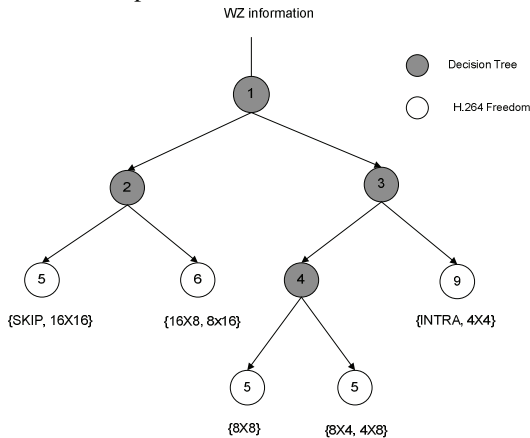


Figure 4. The proposed decision tree algorithm.

We emphasize that the information required by the tree are not new calculations, and they are not generated in order to be passed through the H.264 encoder part (the second half of the transcoder), yet this information is generated in the backward motion estimation in the side information generation and in the turbo decoding processes in order to find the best MV per block in the WZ decoder and to recover the mismatches in the side information respectively. Therefore, our approach only has to store this information which can be used to determine the optimum partition for this MB block.

4. PERFORMANCE EVALUATION

The incoming video from WZ sender was generated using a fixed QP matrix generation as a trade-off among quality, bit rate and encoder complexity. Then, the transcoder converts this WZ video input into a H.264 video stream with QP ranging from 28 to 40. In order to show transcoding results, the experiments were carried out on the test sequences with four different quantization parameters, i.e., QP = 28, 32, 36 and 40 as specified in *Bjontegaard and Sullivan's* common test rule [9]. The H.264 version used in the simulations was JM 14.1 and the parameter configuration was the baseline profile with all parameters as they appear in the configuration file by default, RD optimization was turned off. The H.264 baseline profile is chosen because it is a lower complexity profile and is the common profile used in real-time applications. In the simulations, the reference transcoder is comprised of a full WZ decoder followed by a full H.264 encoder. Finally, the % of *Time Reduction* (%TR) reported was the average of the times reduction of the four H.264 QP points under study. The results are reported in Table 1. We have preferred to show the results in this system instead of the traditional RD figures because of the following reasons: 1) with tabular results, we can include more sequences (only one more row in the table instead of a full figure) and, 2) due to very close RD performance, using

traditional figures both curves (the proposed approach and the reference transcoder) look identical on the plot and it is difficult to appreciate the differences. As Table 1 shows the TR of the proposed transcoder is remarkably over 95% with negligible rate-distortion loss compared to the full complex reference transcoder.

Table 1. Performance of the proposed transcoder.

RD performance of the WZ/H.264 video transcoder – GOP 2			
Sequence	Δ PSNR (dB)	Δ Bitrate (%)	TR (%)
Akiyo	-0,004	0,10	95,23
Coastguard	-0,036	0,95	94,99
Container	-0,001	0,03	96,00
Hall Monitor	-0,006	0,14	95,43
Mother	-0,005	0,14	95,32
mean	-0.011	0.27	95.39

5. CONCLUSIONS

This work presents a WZ to H.264 transcoder which has been improved by reducing the motion estimation and the inter prediction done in the H.264 encoding. Reusing the MVs generated at the WZ decoding algorithm, the proposed transcoder reduced dynamically both the search area and the search window of the motion estimation. Moreover, based on some correlated information, a decision tree algorithm is introduced; it reduces the sub-MB partitions that the H.264 encoding algorithm can check and, therefore the overall time is further reduced.

6. REFERENCES

- [1] ISO/IEC International Standard 14496-10:2003, "Information Technology – Coding of Audio – Visual Objects – Part 10: Advanced Video Coding".
- [2] A. Aaron, R. Zhang, and B. Girod, "Wyner-Ziv Coding for Motion Video," Asilomar Conference on Signals, Systems and Computers, Pacific Grove, USA, November 2002.
- [3] B. Girod, A. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed Video Coding," Proceedings of the IEEE, vol. 93, no. 1, pp. 71-83, Jan. 2005.
- [4] E. Peixoto, R. L. de Queiroz, and D. Mukherjee, "Mobile video communications using a Wyner-Ziv transcoder", Symposium on Electronic Imaging, Visual Communications and Image Processing (SPIE), San Jose, CA, USA, Jan. 2008.
- [5] J.L. Martinez, G. Fernández-Escribano, H. Kalva, W.A.C. Fernand and A. Garrido, "Wyner-Ziv to H.264 video transcoder for mobile telephony", IEEE ICCE, Las Vegas, USA, January 2009.
- [6] J. L. Martinez, H. Kalva, W.A.C Fernando, P. Cuenca and F. J. Quiles, "Efficient WZ-to-H.264 Transcoding Using Motion Vector Information Sharing", ICME, June 2009, New York, USA.
- [7] J. L. Martinez, W.A.R.J. Weerakkody, W.A.C. Fernando, G. Fernandez-Escribano, H. Kalva and A. Garrido, "Distributed Video Coding using Turbo Trellis Coded Modulation", The Visual Computer, Vol. 25. No 1, pp. 69-82, January 2009.
- [8] I. H. Witten and E. Frank, "Data Mining: Practical Machine Learning Tools and Techniques," 2nd Edition, Morgan Kaufmann, San Francisco, 2005.
- [9] G. Sullivan and G. Bjontegaard, "Recommended Simulation Common Conditions for H.26L Coding Efficiency Experiments on Low-Resolution Progressive-Scan Source Material". ITU-T VCEG, Doc. VCEG-N81. September 2001.