# EFFICIENT WZ-TO-H.264 TRANSCODING USING MOTION VECTOR INFORMATION SHARING


*J. L. Martínez[1], H. Kalva[2], W.A.C Fernando[3], P. Cuenca[1]and F. J. Quiles[1]*

[1]Instituto de Investigación en Informática de Albacete. Universidad de Castilla-La Mancha. Spain
[joseluismm, pcuenca,paco]@dsi.uclm.es
[2]Department of Computer Science and Engineering. Florida Atlantic University. USA
hari@cse.fau.edu
3 Center for Communications Research. University of Surrey, UK
W.Fernando@surrey.ac.uk



## ABSTRACT

In mobile-to-mobile video communications, both the sender and the receiver devices should not have higher complexity requirements to perform complex video compression tasks. The traditional video coding solutions are not suitable to support this communications due to its extremely complex encoding algorithm. On the other hand, the new Wyner-Ziv video coding paradigm reduces the complexity of the encoder at the expenses of a more complex decoder. In this paper, we propose an improved WZ/H.264 video transcoder to support this mobile-to-mobile communications, using the low complexity Wyner-Ziv encoding and the traditional H.264 decoding to be implemented in the end-user devices. The improved transcoder converts the video from the Wyner-Ziv to H.264 and reuses the motion vectors generated in the Wyner-Ziv decoding, in order to reduce the computational complexity of the motion estimation process in the H.264 encoding. Simulations results show a complexity reduction up to 55% with negligible rate – distortion drop.



***Index Terms* —** Wyner-Ziv, DVC, Transcoding, H.264/AVC.


## 1. INTRODUCTION

Mobile-to-mobile video communication such as video telephony is increasingly becoming important. To effectively use mobile video services we need low complexity, yet efficient, video encoding and decoding solutions. Traditional hybrid video encoder such as H.264/AVC has asymmetric complexity; i.e., encoder is significantly more complex compared to decoders [1]. Implementation of such complex encoders on mobile devices sacrifice quality by avoiding complex encoding modes in order keep encoding complexity low.

There has been increasing interest in Wyner-Ziv (WZ) video coding [2,3], a particular case of *Distributed Video Coding* (DVC) [2]. A key characteristic of WZ encoders is that they reduce the processing complexity of the encoder, leading to a low-cost implementation, while increasing the complexity of the decoder.

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 in [1]) this paper proposes the use of an improved WZ/H.264 video transcoder. Such a transcoder can reside in the network and enable communication between devices that use WZ encoding and H.264 decoding by transcoding WZ encoded video to H.264. This scenario is depicted in Figure 1.

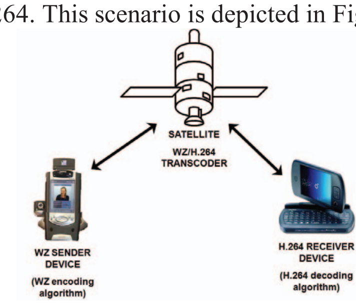


Figure 1. Mobile-to-mobile communications system using WZ/H.264 transcoder.

In this paper, moreover, the proposed transcoder has been improved by reusing the *Motion Vectors* (MV) estimated at the WZ decoding state in order to speed up the *Motion Estimation* (ME) done at the H.264 encoder.

The paper is organized as follows: Section 2 identifies the state-of-the-art in WZ based transcoders and the problem statement. Section 3 shows our improved WZ/H.264 video transcoder which is evaluated in Section 4 with some


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". 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.

simulation results. Finally, in Section 5 the conclusions are presented.

## 2. RELATED WORK

One of the more referenced overviews of DVC [2], 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 video transcoder approach was presented by *Peixoto et al* in [4] in 2008, which is based on a WZ/H.263 video transcoder to support this kind of communications. They 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 well the correlation between the WZ MV and the traditional ME and only used them to determine the starting center of the ME process. In our approach, both the search range and search window is reduced based on the length and the orientation of the MV estimated at the WZ decoder. Moreover, 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 better using the proposed combination.

The authors of this paper have recently proposed an improved WZ/H.264 transcoder [5] that outperforms the results of [4]. In the approach reported in [5], the proposed transcoder is focused on the *Macro Block* (MB) coding mode decision done in H.264 instead of the ME itself. Reusing the *Sum of Absolute Differences* (SAD) and the length of the MV generated at the *side information* process (the ME done in WZ video coding), the proposed transcoder selected a candidate set of MB mode decisions and reduce, therefore, the amount of complexity by more than 86%. As far as the authors, these are the two approaches available in the literature focusing on WZ video transcoder. In this paper, we propose a new approach to reducing the complexity of WZ to H.264 transcoders. This approach, based on reducing the ME complexity, reduces transcoder complexity by about 55%. This approach can also be combined with the transcoding techniques reported in [5] to further reduce the complexity.

## 3. PROPOSED SCENARIO

The end-user devices of the proposed scenario use a WZ encoder and a H.264 decoder; the lower complexity parts in both paradigms. The WZ encoder used in the sender device is our previous WZ architecture based on *Turbo Trellis Coded Modulation* (TTCM) that was proposed in [3]. This architecture is based on *Transform Domain* (TD) and TTCM codes. 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. For H.264 [1] decoding, the standard reference decoder (JM version 14.0) has been used.

The WZ encoding (performed at the sender) 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 as follows: the pixels values are converted to a transform domain using the integer Discrete Cosine Transform (DCT) as proposed in H.264 [1]. The coefficients values are quantized using a quantizer matrix (different RD points are achieved simply varying this matrix). Then, the quantized symbols are encoded by a TTCM code as in [3]. The WZ encoder may have a GOP length of 2 (one key frame followed by 1 WZ frame), 3 (one key frame followed by 2 WZ frames), or even more. On the other hand, the traditional video decoder just performs entropy decoding, motion compensation and inverse transform. We emphasize that only these are the algorithms that have been implemented in the end devices: i.e. WZ encoding and H.264 video decoding. On contrary, all the computations are taken over by the transcoder which is the major contribution of this paper and is depicted in the next section.

The proposed WZ/H.264 video transcoder is depicted in Figure 2. The transcoder is comprised of a WZ decoder [3] as well as the H.264 encoder [1]. Instead of re-encoding the sequence, the transcoder uses information that was calculated in the WZ decoding process for speeding up the transcoding (that is depicted in Figure 2 by a red points line connected both ME).

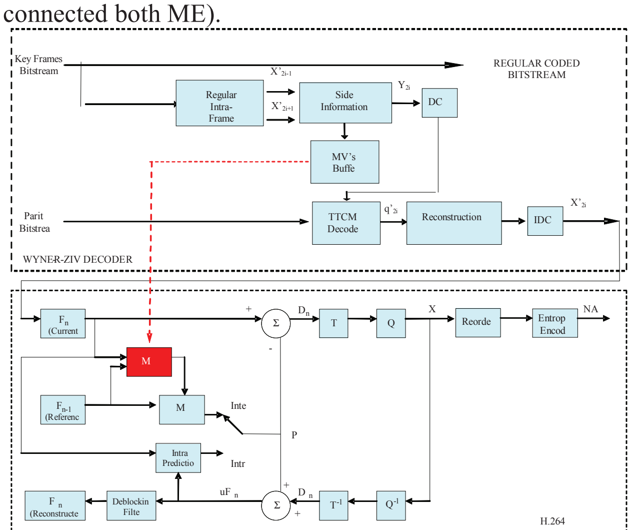


Figure 2. The proposed WZ/H.264 video transcoder

The WZ decoder is based on our TTCM approach [3] and it works as follows: firstly, the WZ decoder generates the side information (the estimation at the decoder of the original frame available at the encoder) using motion-compensated interpolation between past and previous key frames (In the proposed transcoder we employ the approach based on [2]). This side information scheme [2] can be used to generate more than one frame between two key frames. The side information jointly together with the TTCM symbols sent by the encoder are fed to the TTCM decoder to generate the reconstructed coefficients. The final steps are the reconstruction process and the inverse DCT in order to recover the pixel values.

The inter prediction in H.264 supports motion compensation block sizes ranging form 16x16 to 4x4 with many option available between them; this method is know as tree structured motion compensation. Then, for each partition and sub-partition the ME process is carried out. The final decision is done based on the SAD between the current block and previous /past ones inside the search range window and the length of the MV. Therefore the ME process is called too many times per block ensuring an optimum MB selection. In this paper, all MB partition selections are kept untouched but, for each MB mode partition the ME process is reduced by reusing the information gathered at the WZ decoder side information generation. Basically, the MVs side information give us some information about the movement of the content in the image and therefore, it does not require to make the full ME process again in H.264 encoding algorithm. In other words, the main contribution of this paper is a dynamic motion window / search range scheme in the framework of WZ/H.264 mobile-to-mobile video communications. The improvements to the proposed transcoder are presented in the following sub sections.

***Dynamic Motion Estimation Approach***

Our approach can reduce the search range adaptively based on the length of the MVs estimated at the WZ decoder. In the proposed transcoder, the MVs search range for every H.264 MB is adaptively determined and reduced depending on the length of the incoming MVs. Moreover it is recalculated for every MB (or sub-MB partition) than can occur in the MB mode coded decision. The ME complexity can be further reduced by exploiting the orientation of the MV side information. The proposed approach is depicted in Figure 3. The MB with zero MVs in the side information generation represents very simple MB, so the search range is limited to 1 pixel. On the other hand, in the case of estimated (at the WZ decoder) MB modes that have MV information the search range is determined by the area of the circle, centered in the (0,0) point for each H.264 mode or sub-mode (top-left corner of the MB). The length of the incoming vector determines the radius of the circle (see Figure 3). In this way, the length of the MV side information will reduce the search area. An example of the calculation of the search range for the 16x16 H.264 MB mode is shown in Figure 3, when the MVs in the side information are bigger than 1. Same process is developed with the rest of the H.264 MB sub partition modes, generating a set of circle areas where the original MVs side information is refined.

In a nutshell, the main ideas of the transcoding process are: 1) copying the bitstream of the intra frames since that will not be changed; they are directly passed to the receiver without any processing (Intra H.264 is used for key frames) and, 2) reusing the MVs of the side information generation process instead of re-calculating them at the H.264 encoder using the proposed Dynamic Motion Estimation approach that has been depicted in this section.

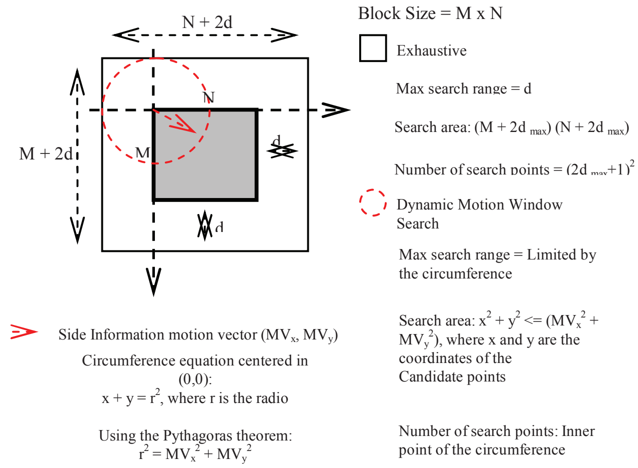


Figure 3. The proposed dynamic motion estimation

## 4. PERFORMANCE EVALUATION

To investigate the performance of the WZ/H.264 transcoder proposed in this paper we implemented it in the H.264/AVC reference software 14.0 version [1]. Basically we have concatenated our TTCM based WZ decoder [3] to the H.264 encoder reference software and, information has been mutually used at the H.264 ME process from the WZ decoder stage (as Figure 2-3 illustrate). The tests were carried out with the popular sequences whose main characteristics are shown in Table 1.

Table 1. Main characteristics of video sequence test. Frame rate 30fps

| Sequence characteristics | | |
|---|---|---|
| Sequence name | Format | N# frames |
| Akiyo | QCIF | 300 |
| Carphone | QCIF | 382 |
| Coastguard | QCIF | 300 |
| Foreman | QCIF | 300 |
| Mobile | QCIF | 300 |
| Paris | QCIF | 449 |
| Salesman | QCIF | 449 |
| Soccer | QCIF | 150 |

Basically, the incoming video from WZ sender was generated using a fixed QP matrix generation as a trade-off between 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. For showing 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 *Bjøntegaard and Sullivan´s* common test rule [6]. The PSNR of the decoded frames and bit rates are averaged over the sequence only for the luminance component at the receiving end. In this performance evaluation two GOP sizes have been evaluated, GOP 2 (K-WZ-K) and GOP 3 (K-WZ-WZ-K) that are converted to H.264 GOP size (I-P-I) and (I-P-P-I) respectively. The H.264 was running with RD-off because this is the most low complexity mode that is more suitable for mobile-to-mobile communications. The H.264 parameter configuration used in the simulation was the baseline profile with all parameters as appeared into the configuration file by default. Only three parameters have been modified:

- *NumberReferenceFrames*. By default it is 5 but it is set to 1 since our main goal is real time applications, so we reduced the complexity to get it.
- *SearchMode*. It is fixed to -1. *Full Search* (FS) mode.
- *SearchRange*. It is defined to 16. By default it is set to 32. 16 pixels are enough for transcoding the sequences. Moreover, it is closer to the search range used for side information generation process.

The H.264 baseline profile is chosen because it is a lower complexity profile and is the common profile used in real-time applications, such as mobile TV and video conference.

Note that bit rate and PSNR differences should be regarded as equivalent, i.e., there is either the decrease in PSNR or the increase in bit rate, but not both at the same time. The detail procedures in calculating these differences can be found from a JVT document authored by *Bjøntegaard* [6], which is recommended by JVT Test Model Ad Hoc Group. This mechanism is proposed for finding numerical averages between RD-curves as part of the presentation of results. This is a more compact and in some senses more accurate way to present the data and is used instead of the RD-plots. We have preferred this way to show the results instead of the traditional RD figures because 1) with tabular results we can include more sequences (only one row more 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.

The reference transcoder is comprised of a full WZ decoder followed by a full H.264 encoder. Finally, the % of time reduction reported was the average of the times reduction of the four H.264 QP points under study.

As Table 2 and Table 3 show that the RD performance is practically the same. For some sequences in both GOP formats, the RD performance is better than the reference encoder. This happens because the best rate distortion solution is obtained by enabling the RD optimization, and in the experiments reported in the tables we have disabled RD optimization n the H.264 encoder. We are working in a real-time scenario and are trying to see how can reduce the time of the faster H.264 encoder (with the SAE cost as evaluation function), without loss in PSNR. To summarize the results, on average, for all sequences the PSNR penalty is less than 0.01 dB with a maximum bitrate increase of 0.127%.

Table 2. Performance of the proposed WZ/H.264 transcoder with GOP 2.

| RD performance of the WZ/H.264 video transcoder – GOP 2 | | | |
|---|---|---|---|
| Sequence | ΔBitrate (%) | ΔPSNR (dB) | TR (%) |
| Akiyo | 0.000 | 0.000 | 51.93% |
| Carphone | -0.120 | 0.005 | 53.97% |
| Coastguard | 0.020 | -0.001 | 62.09% |
| Foreman | -0.510 | 0.016 | 54.07% |
| Mobile | 0.008 | -0.035 | 70.83% |
| Paris | 0.006 | -0.021 | 59.28% |
| Salesman | -0.010 | 0.000 | 59.06% |
| Soccer | 0.470 | -0.014 | 42.38% |
| ***mean*** | **-0.017** | **-0.006** | **56.70%** |

Table 3. Performance of the proposed WZ/H.264 transcoder with GOP 3.

| RD performance of the WZ/H.264 video transcoder – GOP 3 | | | |
|---|---|---|---|
| Sequence | ΔBitrate (%) | ΔPSNR (dB) | Time Reduction (%) |
| Akiyo | -0.030 | 0.001 | 49.76% |
| Carphone | -0.020 | 0.001 | 53.95% |
| Coastguard | 0.130 | 0.005 | 63.43% |
| Foreman | 0.060 | -0.002 | 51.80% |
| Mobile | -0.010 | 0.001 | 69.03% |
| Paris | -0.010 | 0.001 | 58.18% |
| Salesman | -0.020 | 0.001 | 59.91% |
| Soccer | 0.920 | -0.028 | 40.91% |
| ***mean*** | **0.127** | **-0.003** | **55.87%** |

## 5. CONCLUSIONS

In this paper a novel WZ/H.264 video transcoder to efficiently support mobile-to-mobile video communications has been introduced. Moreover the proposed transcoder has been improved by reusing the MVs generated at the side information generation process at the WZ decoder in order to speed up the ME done at the H.264 encoder. This improved WZ/H.264 offers a considerable decrease of complexity up to 55% compared to the conventional back to back configuration without any RD penalty. On average, for all sequences the PSNR penalty is less than 0.01 dB with a maximum bitrate increase of 0.127%.

## 6. REFERENCES


[1] ISO/IEC International Standard 14496-10:2003, “Information Technology – Coding of Audio – Visual Objects – Part 10: Advanced Video Coding”.

[2] B. Girod, A. Aaron, S. Rane, and D. Rebollo-Monedero, “Distributed Video Coding,” Proceedings of the IEEE, vol. 93, pp. 71-83, Jan. 2005.

[3] 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.

[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. Martínez, G. Fernandez-Escribano, H. Kalva, W.A.C. Fernando and A. Garrido, “Wyner-Ziv to H.264 Video Transcoder for Mobile Telephony”, International Conference on Consumer Electronics (ICCE), Las Vegas, NV, USA, January, 2009.

[6] G. Sullivan and G. Bjøntegaard, “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.