

RESEARCH

Open Access

# Temporal scalable mobile video communications based on an improved WZ-to-SVC transcoder

Alberto Corrales-García<sup>\*</sup>, José Luis Martínez, Gerardo Fernández-Escribano and Francisco José Quiles

## Abstract

Wyner–Ziv (WZ) to scalable video coding (SVC) transcoding can offer a suitable framework to support scalable video communications between low-cost devices. In addition, the video delivery provided by SVC covers the needs of a wide range of homogeneous networks and different devices. Despite the advantages of the video transcoding framework, the transcoder accumulates high complexity and it must be reduced in order to avoid excessive delays in communication. In this article, an approach for WZ to SVC transcoding is presented. The information generated during the first stage is reused during the second one, and as a consequence the time taken by the transcoding is reduced by around 77.77%, with a negligible rate-distortion penalty.

**Keywords:** Transcoding, Scalable video coding, Wyner–Ziv coding, Temporal scalability

## 1. Introduction

Traditionally, video communications, such as television broadcasting, have been based on a down-link model, where the information is encoded once and then transmitted to many terminal devices. These applications are supported by traditional video codecs, such as those adopted by all MPEG and ITU-T video coding standards [1], which are characterized by architectures where most of the complexity is present in the encoding part. However, in the last few years, with the ever-increasing development of mobile devices and wireless networks, a growing number of emerging multimedia applications have required an up-link model. These end-user devices are able to capture, record, encode, and transmit video with low-constraint requirements. Applications, such as low-power sensor networks, video surveillance cameras, or mobile communications, present a different framework in which low-cost senders transmit video bitstreams to a central receiver. In order to manage this kind of applications efficiently, distributed video coding [2], and specially Wyner–Ziv coding (WZ) [3], proposed a solution in which most of the complexity is moved from the encoder to the decoder. In particular, the WZ codec used in this study is based on the architecture proposed by VISNET-II [4], which is an improved

version of the architecture proposed by the project DISCOVER [5]. More detail about WZ coding can be found in [3].

However, as existing networks are heterogeneous and the receivers have different features and limitations (such as power consumption, available memory, display size, etc.), a new scalable standard has recently been proposed in order to support this variety of networks and devices. In particular, scalable video coding (SVC) [6] has been standardized as a scalable extension of the H.264/AVC standard [1]. SVC supports three main types of scalability: (1) temporal scalability; (2) spatial scalability; and (3) quality (SNR) scalability. For a comprehensive overview of the scalable extension of H.264/AVC, the reader is referred to [6].

In particular, this article proposes a WZ-to-SVC transcoder with temporal capabilities; a bitstream provides temporal scalability when it can be divided into a temporal base layer (with an identifier equal to 0) and one or more temporal enhancement layers (with identifiers that increase by 1 in every layer), so that if all the enhancement temporal layers with an identifier greater than one specific temporal layer are removed, the remaining temporal layers form another valid bitstream for the decoder. In this way, to achieve temporal scalability, SVC links its reference and predicted frames using hierarchical prediction structures [7] which define the temporal layering of the final structure.

<sup>\*</sup> Correspondence: albertocorrales@dsi.uclm.es  
Instituto de Investigación en Informática de Albacete, University of Castilla-La Mancha, Avenida de España s/n, 02071, Albacete, Spain

At this point, this article presents a straightforward step in the framework of WZ-based transcoders towards a scalable WZ-based transcoder (Figure 1). In this way, WZ encoding provides low complexity on the encoder side and by including the SVC paradigm, different receivers can satisfy their requirements and the video can also be delivered over a variety of networks. In addition, the main idea of this article is to perform this process as efficiently and as fast as possible by using information gathered in the first part in order to reduce the delay caused by the conversion. With this aim, the improved WZ-to-SVC transcoder with temporal scalability presented in this article reuses the motion vectors (MVs) generated during WZ decoding, because they can give us an idea about the quantity of movement in the current frame and this information will be used to accelerate the motion estimation (ME) stage as part of the SVC encoding algorithm.

The remainder of this article is organized as follows. In Section 2, the several works related to this topic are analyzed. Then, in Section 3 our improved approach is presented, and some implementation results are shown in Section 4. Finally, in Section 5 conclusions are presented.

## 2. Related work

Taking into account the advantages of WZ coding for low-cost video encoding, in the literature several WZ to traditional video coding transcoders have been proposed in order to support video communications between portable low-cost devices, such as one based on H.263 [8] and another on H.264/AVC [9], although these approaches did not concern about heterogeneous network or devices.

Regarding the transcoding based on SVC with temporal scalability, several approaches have been proposed

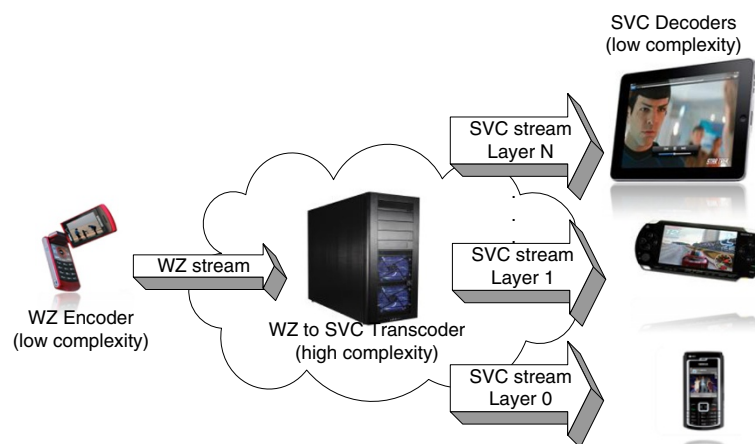
recently. Al-Muscatti and Labeau [10] proposed a technique for transcoding that provided temporal scalability. The method presented was applied in the baseline profile and reused information from the mode decision and ME processes from the H.264/AVC stream. In the same year, Garrido-Cantos et al. [11] presented an H.264/AVC to SVC video transcoder that efficiently reuses some motion information from the H.264/AVC decoding process in order to reduce the time consumption of the SVC encoding algorithm by reducing the ME process time. The approach was developed for main profile and dynamically adapted for several temporal layers [11].

In our previous work [12], we propose a WZ-to-SVC transcoding approach, which only works with baseline profile (or  $P$  frames). However, this article proposes a flexible transcoding architecture which supports both baseline and main profiles and whatever combination of  $P$  and  $B$  frames. As a consequence, it can support any WZ to SVC conversion providing flexibility, and then the transcoding framework is more realistic. In addition, this proposal has deeply been evaluated by using different WZ GOPs, frame rates, and sequence resolutions, as can be seen in Section 4.

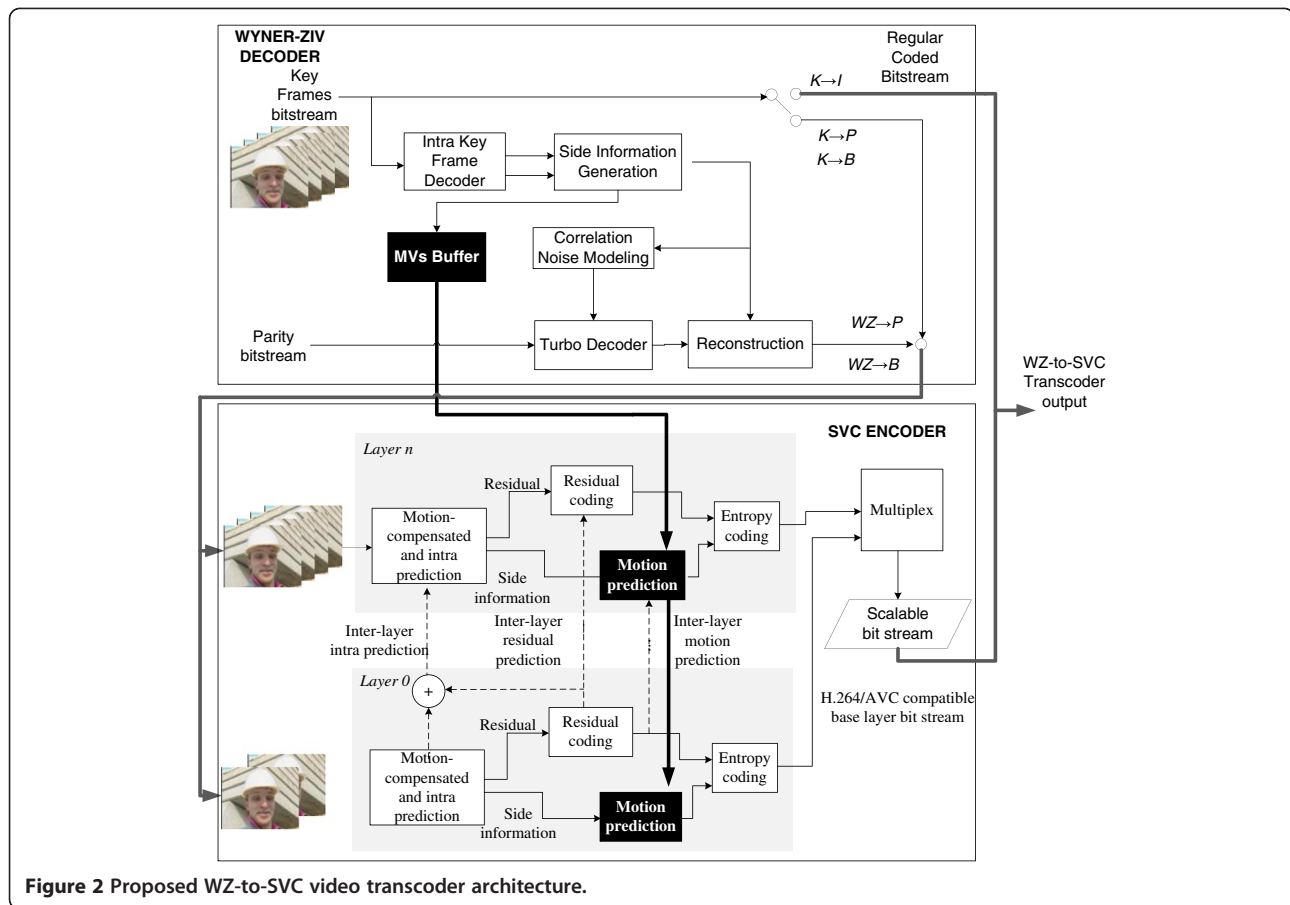
Concerning WZ-to-SVC transcoding, there are several approaches focused on reducing the WZ decoding stage (e.g., by using parallel computing [13,14] or avoiding the use of a feedback channel [15]); something which could also be employed with the proposed dynamic search is ME in SVC, but this study wishes to isolate the second part of the transcoder and thus is focused only on accelerating the SVC encoding.

## 3. WZ-to-SVC transcoding

The main aim of a transcoder is to convert from one source video coding format to another. Taking into account a framework where the delay of the video



**Figure 1** WZ-to-SVC transcoder framework.

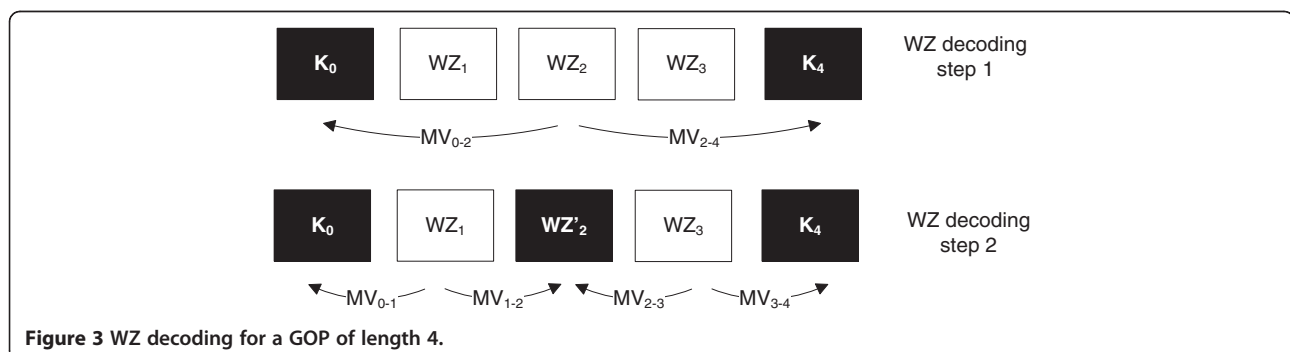


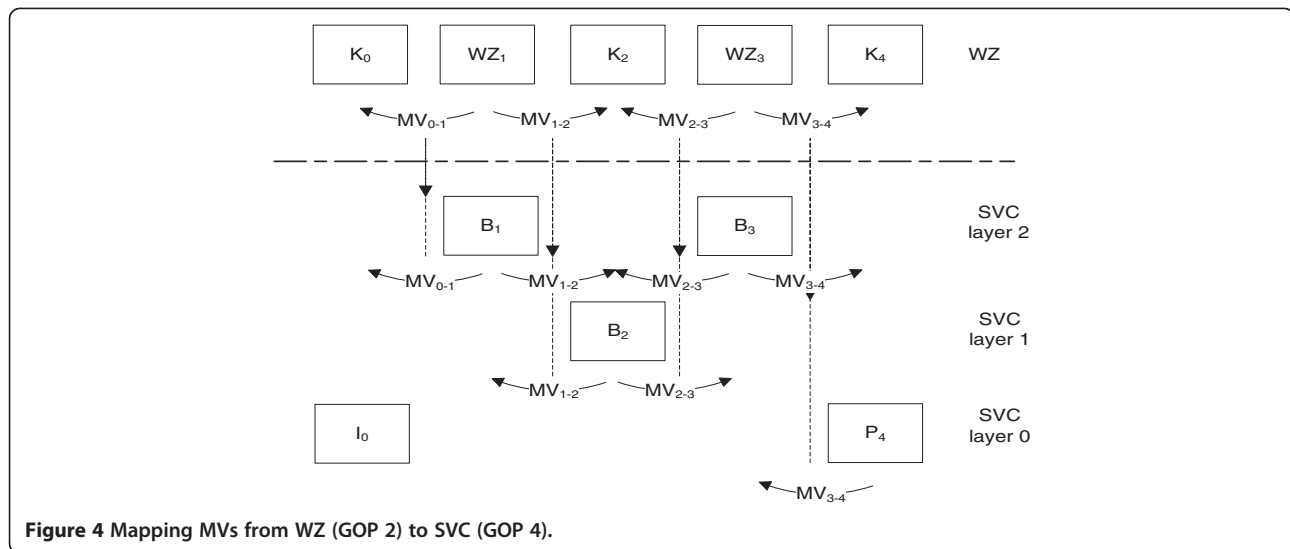
conversion could lead to problems with communication stability (e.g., with a video streaming broadcast), the time spent by the transcoding process is of utmost importance. The most complex task in the SVC architecture is the ME process. In particular, the idea behind this study consists of analyzing the quantity of movement information during WZ decoding to accelerate SVC encoding. This movement information is contained in the MVs generated during the SI process. In other works, the complexity of SVC encoding is reduced without increasing the complexity of the WZ decoding stage. As is

shown in Figure 2, the WZ-to-SVC transcoder is composed of a WZ decoder concatenated with an SVC encoder. In the proposed architecture, the MVs are temporally stored in a buffer and sent to the motion prediction module of SVC, where they are processed as described in the following sections.

### 3.1. MV extraction

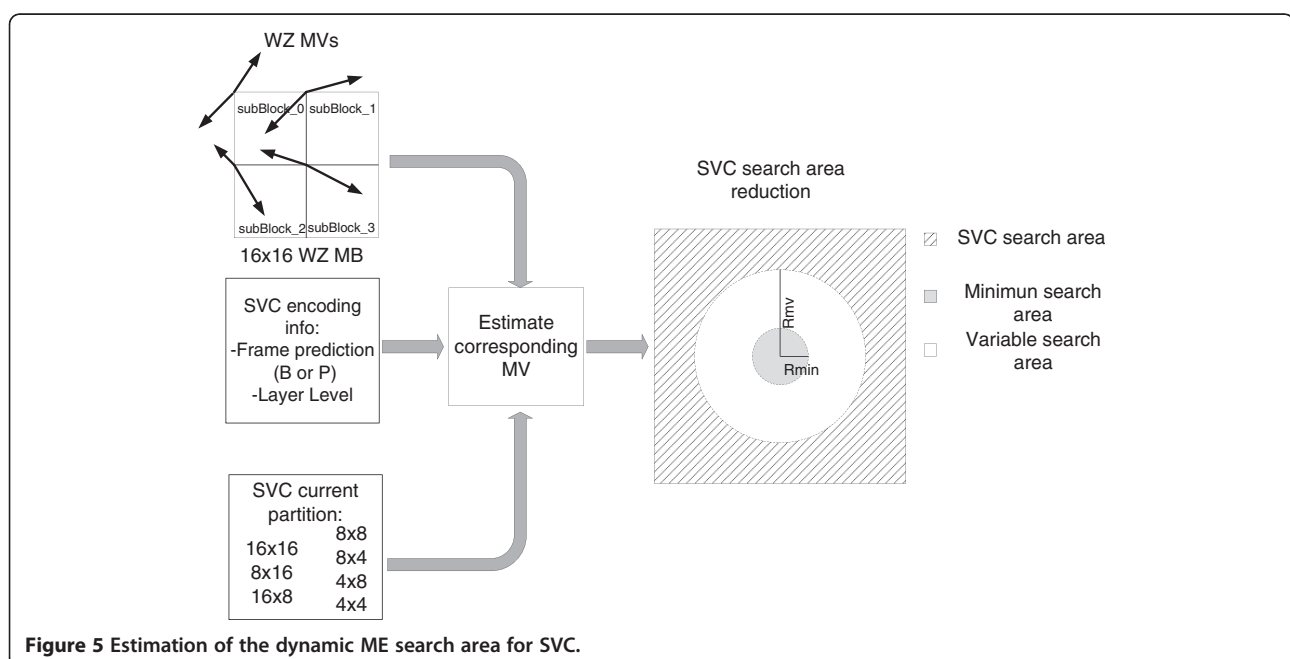
In the WZ decoding process, side information generation is one of the most important tasks, because WZ frames are decoded by reconstructing the information





provided by the SI. There are many studies on this topic but there are mainly two major approaches: hash-based ME and motion compensated temporal interpolation (MCTI). In particular, the SI generated by the VISNET-II codec is based on MCTI [16] with the following steps: first, a forward ME is performed between the two  $K$ -frames. In this step, each  $16 \times 16$  macro block (MB) of the backward frame looks for the MB which generates the lowest residual inside the forward frame. This searching is carried out within a fixed search range of  $32 \times 32$ . Subsequently, the bidirectional ME calculates two MVs from the MV generated during the previous

step. In order to improve the accuracy of the MVs generated, they are up-sampled for  $8 \times 8$  blocks and a new bidirectional ME is performed. Once the bidirectional motion field is obtained, it can be observed that the MVs sometimes have low spatial coherence, so the MVs are improved by a spatial smoothing algorithm targeting the reduction in the number of false MVs. This is based on weighted vector median filters. Finally, bidirectional motion compensation is performed again. Since the SI is estimated from the  $K$ -frames or decoded WZ-frames, MV extraction is a process that is independent of the domain used during encoding of the sequence.



**Table 1 Performance for 15 fps QCIF sequences and two temporal layers**

QCIF sequences	WZ GOP	One layer (7.5 fps)		Two layers (15 fps)		TR (%)
		$\Delta$ Bitrate (%)	$\Delta$ PSNR (dB)	$\Delta$ Bitrate (%)	$\Delta$ PSNR (dB)	
Foreman	2	-0.25	-0.01	0.39	-0.01	-71.41
	4	-0.06	0.00	0.71	-0.03	-73.08
	8	-0.23	-0.01	0.43	-0.02	-73.46
Hall	2	-0.07	0.00	0.19	-0.01	-74.67
	4	0.06	0.00	-0.13	0.00	-74.70
	8	-0.01	0.00	-0.02	0.00	-74.74
CoastGuard	2	0.01	-0.01	-0.07	0.00	-74.33
	4	0.09	0.00	-0.02	0.00	-74.52
	8	-0.22	-0.01	0.28	-0.01	-74.58
Soccer	2	-0.56	-0.03	1.05	-0.04	-68.76
	4	-0.47	-0.02	1.56	-0.06	-71.67
	8	-0.82	-0.03	2.04	-0.07	-72.21
Mean		-0.21	-0.01	0.53	-0.02	-73.18

For more details, the steps of the MCTI process are also described in [16].

As a result of this process, although at the beginning the side information starts with  $16 \times 6$  MBs, at the end, for each  $8 \times 8$  sub-block, two MVs (forward and backward) are calculated and stored. These MVs can help us to estimate the quantity of movement during the SVC stage. For GOP lengths longer than 2, only the MVs generated during the last step are reused by the SVC encoder, because they are more accurate. For example, Figure 3 shows the WZ decoding of a GOP of length 4. For this case, during the first step frame 2 is decoded by using frames 0 and 4 as references, but these MVs are

not considered. In the second step, frames 1 and 3 are decoded by using frames 0, 2, and 4 as references, and these MVs are stored.

### 3.2. WZ-to-SVC transcoding and MV mapping

Once the MVs are obtained from the SI process, we must decide how we can use them to accelerate the SVC encoding. WZ and SVC are quite different video codecs. Thus, the first step is to decide the mapping needed to assign the MVs to the predicted (P) and bidirectional (B) frames of SVC, and then reduce the ME process. Figure 4 represents the transcoding from a WZ GOP 2 to a SVC GOP 4. The first  $K$ -frame is transcoded to an  $I$ -frame

**Table 2 Performance for 30 fps QCIF sequences and three temporal layers**

QCIF sequences	WZ GOP	One layer (7.5 fps)		Two layers (15 fps)		Three layers (30 fps)		TR (%)
		$\Delta$ Bitrate (%)	$\Delta$ PSNR (dB)	$\Delta$ Bitrate (%)	$\Delta$ PSNR (dB)	$\Delta$ Bitrate (%)	$\Delta$ PSNR (dB)	
Foreman	2	0.88	-0.05	1.03	-0.06	1.32	-0.05	-77.61
	4	0.86	-0.05	1.54	-0.07	1.68	-0.07	-77.78
	8	0.54	-0.03	0.90	-0.04	1.01	-0.04	-78.00
Hall	2	-0.01	0.00	0.08	0.00	-0.04	0.00	-78.34
	4	-0.01	0.00	-0.01	0.00	-0.02	0.00	-78.37
	8	-0.06	0.00	0.01	0.00	0.10	0.00	-78.39
CoastGuard	2	0.14	-0.01	0.16	-0.01	0.25	-0.01	-78.10
	4	0.14	-0.01	0.24	-0.01	0.23	-0.01	-78.17
	8	-0.03	0.00	0.01	0.00	-0.05	0.00	-78.18
Soccer	2	0.41	-0.02	1.12	-0.05	0.98	-0.03	-75.68
	4	0.48	-0.02	1.05	-0.03	1.20	-0.03	-77.19
	8	0.28	-0.01	1.23	-0.04	1.41	-0.04	-77.45
Mean		0.30	-0.02	0.61	-0.03	0.67	-0.02	-77.77

**Table 3 Performance for 30 fps CIF sequences and three temporal layers**

CIF sequences	WZ GOP	One layer (7.5 fps)		Two layers (15 fps)		Three layers (30 fps)		TR (%)
		$\Delta$ Bitrate (%)	$\Delta$ PSNR (dB)	$\Delta$ Bitrate (%)	$\Delta$ PSNR (dB)	$\Delta$ Bitrate (%)	$\Delta$ PSNR (dB)	
Foreman	2	1.31	-0.08	1.48	-0.07	1.75	-0.06	-79.30
	4	1.32	-0.08	1.90	-0.07	2.09	-0.07	-79.99
	8	0.82	-0.04	1.42	-0.05	1.78	-0.05	-80.17
Hall	2	0.10	0.00	0.08	0.00	0.08	0.00	-81.26
	4	0.12	0.00	0.09	0.00	0.13	0.00	-81.44
	8	0.05	0.00	0.17	-0.01	0.15	-0.01	-81.45
CoastGuard	2	1.47	-0.07	1.44	-0.06	1.43	-0.05	-80.98
	4	1.42	-0.07	1.56	-0.06	1.46	-0.05	-81.03
	8	1.63	-0.07	1.51	-0.05	1.58	-0.05	-81.06
Soccer	2	1.54	-0.06	2.46	-0.10	2.86	-0.10	-75.90
	4	1.50	-0.06	2.15	-0.08	3.39	-0.09	-78.13
	8	0.98	-0.04	2.44	-0.09	3.91	-0.11	-78.53
Mean		1.02	-0.05	1.39	-0.05	1.72	-0.05	-79.94

without any conversion, as shown in Figure 4. On the other hand, for every WZ frame we have two MVs (forward and backward predicted). Then, they are mapped considering the position of each WZ frame. For B frames, both MVs are considered, but the orientation is changed when necessary (as in frame 2). For P frames, just the backward MV is mapped.

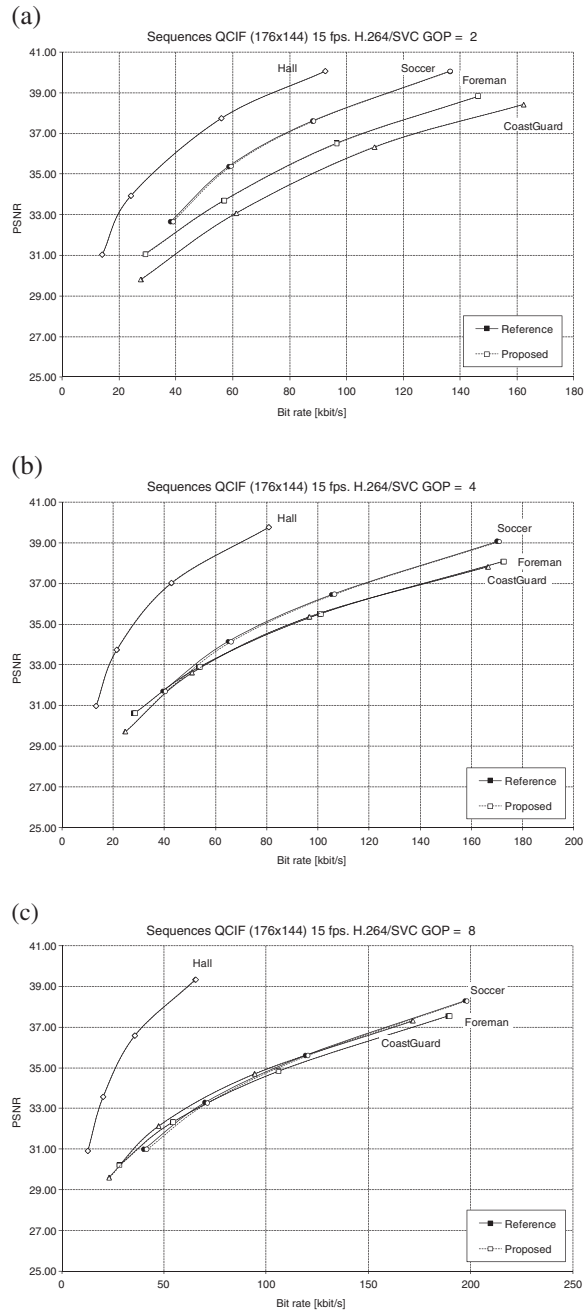
### 3.3. Fast SVC encoding

As is well known, this complexity depends largely on the search range used in the SVC ME process, as the range is linked to the number of positions checked. However, this process may be accelerated because the search range can be reduced by avoiding unnecessary checking without a significant impact on quality or bit rate. In this way, we can use the MVs generated by WZ decoding (which contain information about the quantity of movement per MB) to reduce the search area used by the ME of the SVC encoder. In Section 3.1, we explained how the MVs are calculated and extracted, and in Section 3.2 we described how we map the MVs from GOPs of WZ to SVC. Once the MVs are mapped between frames, depending on the partition checked in the SVC encoding algorithm, we can use a different group of these  $8 \times 8$  MVs. As is well known, in SVC (as in the H.264/AVC standard) inter prediction is carried out by means of the process of variable block size ME. This approach supports motion compensation block sizes ranging between  $16 \times 16$ ,  $16 \times 8$ ,  $8 \times 16$ , and  $8 \times 8$ , where each of the sub-divided regions is an MB partition. If the  $8 \times 8$  mode is chosen, each of the four  $8 \times 8$  block partitions within the MB may be further split in four ways:  $8 \times 8$ ,

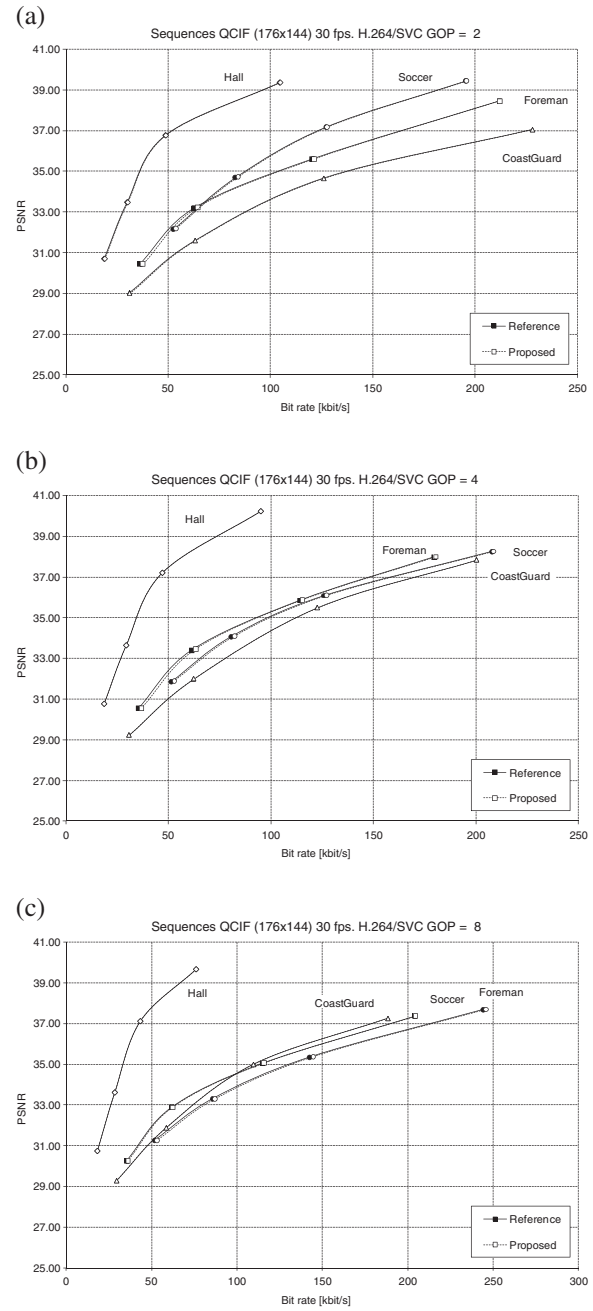
$8 \times 4$ ,  $4 \times 8$ , or  $4 \times 4$ , which are known as sub-MB partitions. For all these partitions, ME is carried out and a separate MV is generated. As is shown in Figure 5, WZ provides one backward and one forward MV for every  $8 \times 8$  sub-partition in each MB. The backward ones are used for P-frame decoding in SVC, and B-frame decoding uses both MVs. Then, depending on the sub-partition to be checked by SVC, the final MV predicted is calculated as follows: if the sub-partition is bigger than  $8 \times 8$  ( $16 \times 16$ ,  $8 \times 16$ ,  $16 \times 8$ ), the predicted MV is calculated by taking the average of the MVs included in the sub-partition. For example, for the  $8 \times 16$  MB-partition, only the MVs allocated at *Block\_0* and *Block\_1* will be used, since the MVs of *Block\_0* and *Block\_1* contain the information about the displacement of the  $8 \times 16$  MB-partition. If the sub-partition is equal to, or smaller than,  $8 \times 8$ , the corresponding MV is applied directly. Then, for each MB and sub-partition, we have one MV for P-frames (previous reference) and two MVs for B-frames (previous and future references). Each MV is composed of two components:  $MV_x$  and  $MV_y$ .

As is shown by Equations (1) and (2), the components  $MV_x$  and  $MV_y$  are multiplied by a factor, which is directly dependent on the number of layers used,  $n$ , and the layer number of the current frame,  $m$ . This means that the MV length of frames from less significant layers will be shorter than the MV length of frames from more significant layers. This is because frames from more significant layers have their reference frames farther than frames from less significant layers. In other words, in WZ decoding the MVs extracted from the SI generation process are always calculated in a frame distance of two, whereas in SVC with temporal scalability these MVs can





**Figure 6** PSNR/bitrate results for transcoding from WZ GOP = 2, 4, and 8 to SVC GOP 2 (15 fps and two layers) with QCIF sequences. Reference symbols: ■ Foreman, ♦ Hall, ▲ CoastGuard, and ● Soccer.



**Figure 7** PSNR/bitrate results for transcoding from WZ GOP = 2, 4, and 8 to SVC GOP 4 (30 fps and three layers) with QCIF sequences. Reference symbols: ■ Foreman, ♦ Hall, ▲ CoastGuard, and ● Soccer.

be mapped for a longer frame distance.

$$MV'_x = MV_x * (n - m) \quad (1)$$

$$MV'_y = MV_y * (n - m) \quad (2)$$

With the scaled MVs, the dynamic area can be defined. In Figure 5, the default search area used by SVC is

defined by Equation (3) and labeled as  $S$ . In Equation (3),  $(x, y)$  represents the coordinates to check in the area.  $A$  is the search range used by SVC which represents the square of the search range  $d$ , as shown in Equation (4). Finally,  $C$  is a circumference which restricts the search with centre on the upper left corner of the MB.  $C$  is

defined by Equation (5).

$$S = \{(x, y) / (x, y) \in (A \cap C)\} \quad (3)$$

$$A = d * d \quad (4)$$

$$C = \pi * R_{mv}^2 \quad (5)$$

$R_{mv}$  represents the radius of the dynamic search area calculated from the MVs calculated during the WZ decoding stage, as Equation (6) shows. In addition, the components of  $R_{mv}$  labeled as  $r_x$  and  $r_y$  are calculated from Equations (7) and (8), which take the maximum between the MVs estimated in Equations (1) and (2) and a minimum radius,  $R_{min}$ , defined as a quarter of the search range (labeled as  $d$ ), as shown by Equation (9). This minimum area is considered to avoid applying search ranges that are too small.

$$R_{mv}^2 = r_x^2 + r_y^2 \quad (6)$$

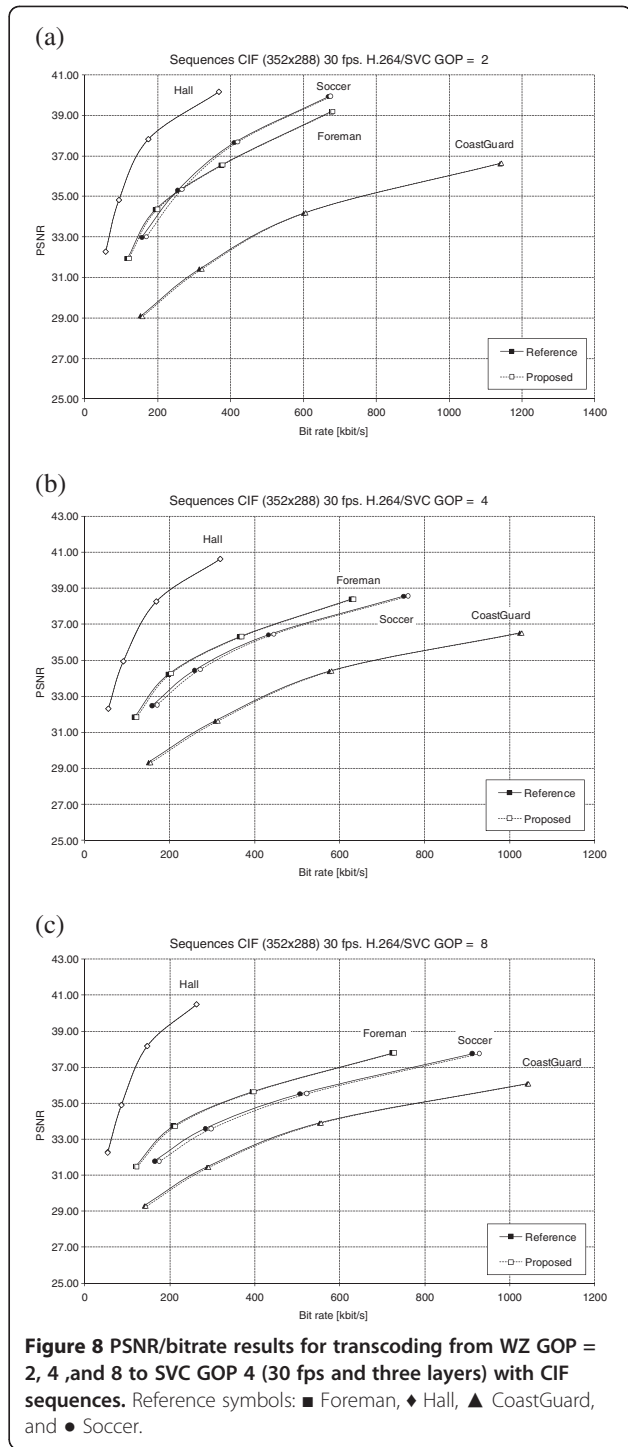
$$r_x = \max(MV'_x, R_{min}) \quad (7)$$

$$r_y = \max(MV'_y, R_{min}) \quad (8)$$

$$R_{min} = \frac{d}{4} \quad (9)$$

#### 4. Experimental results

The source WZ video was generated by the VISNET II codec using a fixed matrix  $QP = 3$  in pixel domain (which means that there are three bitplanes processed) and a GOP length of 2, 4, and 8. While sequences are being decoded, the MVs are passed to the SVC encoder without any increase in complexity. Afterwards, the transcoder converts this WZ video input into an SVC video stream using  $QP = 28, 32, 36$ , and  $40$ , as specified in Sullivan and Bjøntegaard's common test rule [17]. The simulations were run using version JSVM 9.19.14 [18] and the main profile with the default configuration file. The baseline profile was selected because it is the most widely used profile in real-time applications due to its low complexity. In order to check our proposal, we have chosen four representative sequences in QCIF and CIF resolutions with different motion levels at 15 and 30 fps, coding 150 and 300 frames, respectively; these are the same sequences that were selected in the DISCOVER codec evaluation [5]. Furthermore, the percentage of Time Reduction (%TR) is calculated as is specified by Equation (10). In Tables 1 and 2, the reported TR (%) displays the average time reduction for the four SVC QP points under study, compared with the reference transcoder, known as cascade pixel domain video transcoder (CPDT), which is composed of the full WZ decoding and SVC encoding algorithms.



$$TR(\%) = 100 * \frac{(Time_{proposed} - Time_{reference})}{Time_{reference}} \quad (10)$$

Table 1 shows the RD penalty measured for 15 fps sequences encoded using two temporal layers and the TR of the proposed transcoder. Concerning the RD



results, we can observe that the drop penalty is negligible for every layer; even for low-complexity sequences (such as the Hall or CoastGuard sequences) the coding efficiency is better than that of the reference transcoder because the MVs stored are shorter. The TR achieved is -73.18% on average, so the time consumed by SVC encoding is greatly reduced by using the MVs generated in the WZ decoding stage. For different WZ GOP lengths, the results are similar since the proposed MV extraction method is similar for every WZ GOP length.

In addition, Table 2 includes the results for the 30 fps sequences, and then three layers of temporal scalability. The results reported are similar to those of the 15 fps sequences, achieving a TR of -77.77% on average without a significant RD drop penalty. The proposed transcoder avoids a lot of unnecessary checking, but the number of MVs selected by the reference transcoder is the same as for the proposed one of around 92%. As a consequence, the encoding time is greatly reduced with a negligible RD drop penalty.

Taking into account the equations described in Section 3.3, we can observe that the performance is maintained for any layer by using a dynamic search area, which is calculated from the number of layers ( $n$ ) and the current layer ( $m$ ). In particular, frames from more significant layers have more distant reference frames (as can be seen in Figure 4), so it is compensated by using bigger factors in the equations of the algorithm, and as a consequence, bigger search areas for those frames. Keeping the quality in more significant layers is an important issue in order to maintain the performance of the rest of the layers, because in temporal scalability frames from enhancement layers use previous frames as reference frames.

Furthermore, Table 3 shows the performance of the proposed transcoding algorithm using CIF sequences. As for the case of QCIF sequences, proposed algorithm does not involve significant RD penalty for bigger resolution sequences, since it is not dependent on a particular resolution. In addition, it achieves better time reduction (-79.94 on average), because CIF sequences involves more a more complex process for the CPDT transcoder.

Finally, Figures 6 and 7 show the RD obtained for QCIF sequences by using a QP = 28, 32, 36, and 40, respectively, for a transcoding from WZ GOP 2 to SVC GOP 2 (15 fps with two temporal layers) and SVC GOP 4 (30 fps with three temporal layers). As can be seen, there are no significant differences in quality or the bit rate obtained by the SVC reference codec and our proposed one. There are only tiny bitrate increases for several QP points and above all in high movement/complexity sequences. Similar RD results are obtained when comparing with PSNR, as is shown for 15 fps in Figure 6

and for 30 fps in Figure 7. For longer GOP lengths, the results are similar. For CIF sequences RD results (Figure 8), we obtain similar conclusions as for QCIF sequences.

## 5. Conclusions

In this article, a novel WZ-to-SVC transcoding framework is proposed to support scalable video communications between a wide range of mobile devices and over heterogeneous networks. Our proposed transcoder with temporal scalability analyzes and adapts the motion information generated during WZ decoding to accelerate the ME process of the SVC encoder with a main profile. In order to manage this approach, several sequences were transcoded by using different scalability layers for 15 and 30 fps frame rates. In our results, the SVC encoding time is reduced by around 77.77% whilst maintaining the efficiency in RD terms.

### Competing interests

The authors declare that they have no competing interests.

### Acknowledgments

This study was supported by the Spanish MEC and MICINN, as well as European Commission FEDER funds, under Grant nos. CSD2006-00046 and TIN2009-14475-C04. It was also partly supported by the JCCM funds under Grant nos. PEI09-0037-2328 and PII209-0045-9916, and the University of Castilla-La Mancha under Project AT20101802. This study was performed by using the VISNET2-WZ-IST software developed in the framework of the VISNET II project.

Received: 27 September 2012 Accepted: 11 February 2013

Published: 26 February 2013

### References

1. ISO/IEC International Standard 14496-10, "Information Technology - Coding of Audio - Visual Objects - Part 10: Advanced Video Coding", 2003
2. B Girod, AM Aaron, S Rane, D Rebollo-Monedero, Distributed video coding. *Proc. IEEE* **93**, 71-83 (2005)
3. A Aaron, Z Rui, B Girod, Wyner-Ziv coding of motion video, in *Proceedings of 36th Asilomar Conference on Signals, Systems and Computers*, Vol. 1 Pacific Grove, CA, USA, 2002), pp. 240-244
4. J Ascenso, C Brites, F Dufaux, A Fernando, T Ebrahimi, F Pereira, S Tubaro, The VISNET II DVC codec: architecture, tools and performance, in *Proceedings of European Signal Processing Conference (EUSIPCO)* (Aalborg, Denmark, 2010)
5. X Artigas, J Ascenso, M Dalai, S Klomp, D Kubasov, M Ouaret, The DISCOVER codec: architecture, techniques and evaluation, in *Proceedings of Picture Coding Symposium (PCS)* (Lisbon, Portugal, 2007), pp. 1-4
6. ITU-T and ISO/IEC JTC 1, *Advanced Video Coding for Generic Audiovisual Services*. ITU-T Rec. H.264/AVC and ISO/IEC 14496-10 (including SVC extension), 2009
7. H Schwarz, D Marpe, T Wiegand, Overview of the scalable video coding extension of the H.264/AVC standard. *IEEE Trans. Circuits Syst. Video Technol.* **17**, 1103-1120 (2007)
8. E Peixoto, RL Queiroz, D Mukherjee, A Wyner-Ziv video transcoder. *IEEE Trans. Circuits Syst. Video Technol.* **20**, 189-200 (2010)
9. JL Martínez, G Fernández-Escribano, H Kalva, WAC Fernando, P Cuenca, Wyner-Ziv to H.264 video transcoder for low cost video encoding. *IEEE Trans. Consum. Electron.* **55**, 1453-1461 (2009)
10. H Al-Muscati, F Labeau, Temporal transcoding of H.264/AVC video to the scalable format, in *Proceeding of 2nd International Conference on Image Processing Theory Tools and Applications (IPTA)* (Paris, France, 2010), pp. 138-143
11. R Garrido-Cantos, J De Cock, JL Martínez, S Van Leuven, P Cuenca, A Garrido, R Van de Walle, Video adaptation for mobile digital television, in

*Proceedings of 3rd joint IFIP Wireless and Mobile Networking Conference (WMNC)* (Budapest, Hungary, 2010), pp. 1–6

12. A Corrales-García, JL Martínez, G Fernández-Escribano, FJ Quiles, Scalable mobile-to-mobile video communications based on an improved WZ-to-SVC transcoder, in *Proceedings of International Conference on MultiMedia Modeling (MMM), Lectures Notes in Computer Science, Vol. 7131* (Klagenfurt, Austria, 2012), pp. 233–243
13. A Corrales García, JL Martínez Martínez, G Fernandez Escribano, Reducing DVC decoder complexity in a multicore system, in *Proceedings of IEEE International Workshop on Multimedia Signal Processing (MMSP)* (Saint-Malo, France, 2010), pp. 315–320
14. A Corrales García, JL Martínez Martínez, G Fernandez Escribano, FJ Quiles Flor, WAC Fernando, Wyner-Ziv frame parallel decoding based on multicore processors, in *Proceedings of 13th IEEE International Workshop on Multimedia Signal Processing (MMSP)* (Hangzhou, China, 2011), pp. 1–6
15. J Areia, J Ascenso, C Brites, F Pereira, Low complexity hybrid rate control for lower complexity Wyner-Ziv video decoding, in *Proceedings of 16th European Signal Processing Conference (EUSIPCO)* (Lausanne, Switzerland, 2008)
16. J Ascenso, C Brites, F Pereira, Improving frame interpolation with spatial motion smoothing for pixel domain distributed video coding, in *Proceedings of 5th EURASIP Conference on Speech and Image Processing, Multimedia Communications and Services (EC-SIP-M)* (Smolenice Castle, Slovak Republic, 2005)
17. G Sullivan, G Bjøntegaard, Recommended simulation common conditions for H.26L coding efficiency experiments on low-resolution progressive-scan source material, in *ITU-T VCEG, Doc. VCEG-N81*, 2001
18. *Joint Scalable Video Model (JSVM) Reference Software, Version 9.19.3*. Available at: <http://www.hhi.fraunhofer.de/en/fields-of-competence/image-processing/research-groups/image-video-coding/svc-extension-of-h264avc/jsvm-reference-software.html>

doi:10.1186/1687-6180-2013-35

**Cite this article as:** Corrales-García et al.: Temporal scalable mobile video communications based on an improved WZ-to-SVC transcoder. *EURASIP Journal on Advances in Signal Processing* 2013 **2013**:35.

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)