A COMPARISON OF DIFFERENT SIDE INFORMATION GENERATION METHODS FOR MULTIVIEW DISTRIBUTED VIDEO CODING

Xavi Artigas, Francesc Tarrés and Luis Torres
Department of Signal Theory & Communications, Universitat Politècnica de Catalunya (UPC)
Campus Nord, D-5. Jordi Girona 1-3, 08034 Barcelona

Keywords: Distributed Video Coding, Multiview Side Information Generation.

Abstract: This paper presents a comparison of the performance of different side information generation methods for multiview distributed video coding scenarios. Existing literature on this topic relies on pure temporal interpolation (by means of motion compensation), pure inter-camera interpolation (by means of disparity compensation) or a combination of both. In this work a different approach is used by calculating the motion vectors on an available conventionally-encoded camera (one that is not encoded using distributed principles) and then using the obtained motion vectors to generate the side information for the camera that requires it. Variations of this technique are also presented, along with a mechanism to merge all of them together. Finally, simulation results comparing with other techniques and conclusions are given.

1 INTRODUCTION

Video coding research and standardization have been adopting until now a video coding paradigm where it is the task of the encoder to explore the source statistics, leading to a complexity balance where complex encoders interact with simpler decoders. Distributed Video Coding (a particularization of Distributed Source Coding) adopts a completely different coding paradigm by giving the decoder the task to exploit the source statistics to achieve efficient compression. This coding paradigm is particularly adequate to emerging applications such as wireless video cameras and wireless low-power surveillance networks, disposable video cameras, medical applications, sensor networks, multi-view image acquisition, networked camcorders, etc., where low complexity encoders are a must because memory, computation, and energy are scarce.

However, even though the theoretical bases for Distributed Source Coding (DSC) were set thirty years ago with the work by Slepian & Wolf (Slepian, 1973) (for the lossless case) and Wyner & Ziv (Wyner, 1976) (for the lossy case), it has been only recently that research on the topic has taken a new momentum. This research has been encouraged by the rise of some new applications, and has been leaded mainly by Ramchandran *et al*. (Puri, 2002) and Girod *et al*. (Girod, 2005). A good review of other works can be found in (Girod, 2005).

On the other hand, Multiview techniques have been researched in the past, both for coding (Ohm, 1999) and for camera interpolation, since they allow creating views from virtual (non-existent) cameras, or what is called Free Viewpoint Navigation of scenes given only recordings from a few cameras (Shum, 2000).

The objective of Multiview DVC is to efficiently encode different video streams, but exploiting the possible redundancies at the decoder, thus obtaining benefits inherent to DVC like lower encoding complexity, embedded error resilience or the fact that no connection is necessary between the different cameras. This paper compares the performance of a number of methods that particularize Distributed Video Coding to the Multiview scenario.

Multiview DVC has only recently received attention from the scientific community. Ramchandran *et al.* (Toffetti, 2005) and Girod *et al.* (Zhu, 2003) worked with static images. Guo *et al.* (Guo, 2006) and Ouaret *et al.* (Ouaret, 2006) worked with video sequences using a homography approach to perform the inter-camera interpolation and fuse it with conventional intra-camera interpolation. In

Artigas X., Tarrés F. and Torres L. (2007).
A COMPARISON OF DIFFERENT SIDE INFORMATION GENERATION METHODS FOR MULTIVIEW DISTRIBUTED VIDEO CODING.
In *Proceedings of the Second International Conference on Signal Processing and Multimedia Applications*, pages 440-445
DOI: 10.5220/0002142304400445

(Artigas, 2006) additional depth information was used in order to help the process.

Section 2 quickly describes the used DVC codec and Section 3 depicts the selected multiview scenario. Then, Section 4 details the different techniques being compared and Section 5 gives the results of the comparison. Section 6 finally extracts some conclusions.

2 DISTRIBUTED VIDEO CODING

The process followed to turn Wyner and Ziv's theoretical principles (Wyner, 1976) into a practical codec is summarized next. More details can be found in (Girod, 2005) and the general scheme is depicted in Figure 1.

Some of the input frames (the Intra frames) are independently coded and decoded. Furthermore, the receiver uses them to generate an estimate for the rest of the frames (the WZ frames). This estimate is called side information. Then, for the WZ frames, the emitter only needs to transmit the necessary bits to correct possible estimation errors. This is why the better the side information (the more correlated it is with the frame being estimated), the less bits will be required to encode that image (there will be less errors to correct).

The distributed encoder uses systematic turbo codes to generate parity bits for the WZ frames. The systematic part of the codes is discarded, since it will be replaced by the side information at the decoder, and only the parity bits are kept.

The parity bits are punctured out to achieve compression and transmitted (this information is called Main Signal in Figure 1). The receiver then uses the popular MAP (or BCJR) algorithm (Bahl, 1974) to turbo decode the received parity bits using its side information as systematic bits. If decoding is not reliable enough after a given number of turbo iterations, more parity bits are requested through the return channel. Ideal error detection is used to assess the reliability of the decoded bits. More details can be found in (Girod, 2005).

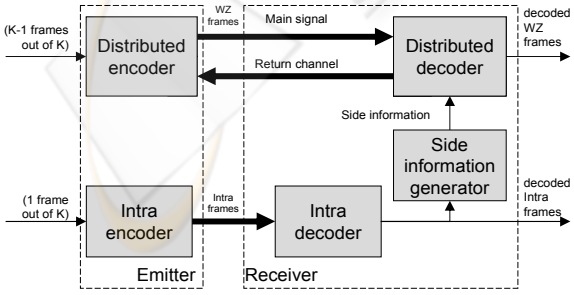

Figure 1: General scheme of the DVC-based system. 1 out of each K input frames is intra encoded and decoded, while the rest of the frames uses distributed principles.

The techniques presented later in Section 4 deal only with the generation of the side information and no more remarks will be made regarding the rest of the distributed video codec.

3 PROBLEM STATEMENT

The setup depicted in Figure 2 is proposed. It can be further augmented by adding more cameras or changing their configuration, but this structure is sufficient to describe the proposed techniques. Three cameras are used, which do not communicate among them. Two of them are called **Intra Cameras** and work in a non-distributed fashion, i.e., their video stream is encoded and decoded independently of the other cameras (using H.264, for example). The third camera, called **Wyner-Ziv camera** (or WZ camera), independently encodes its video sequence but requires the video streams from the other cameras to decode it (Figure 2). This joint decoding allows the WZ camera to transmit at a lower rate than if it was decoded on its own, as stated by the Slepian-Wolf theorem.

The Wyner-Ziv camera transmits some frames in Intra mode, as in (Girod, 2005); this is, coded independently of the other frames (for example using JPEG or H.264 I-frames coding). The rest of the frames are called Wyner-Ziv frames and are the ones that will benefit from the joint decoding performed at the receiver (Figure 3 depicts the three kinds of frames and their relationship). In the following explanations it will be assumed that only one WZ frame is present between every two Intra frames.

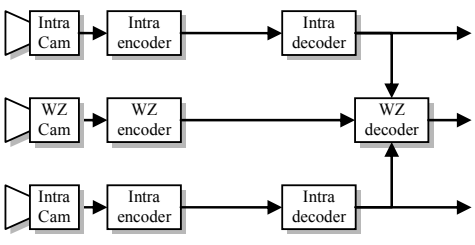

Figure 2: General setup. Intra cameras operate in a conventional fashion while the Wyner-Ziv camera requires joint decoding.

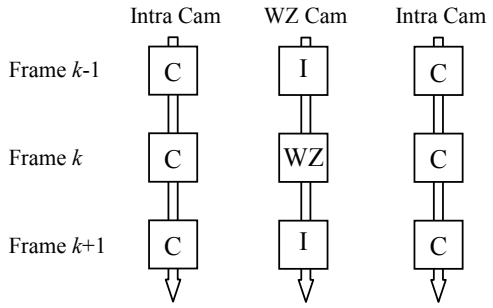

Figure 3: Spatiotemporal frame structure. "C" frames belong to a conventionally encoded video sequence. "I" frames are encoded as single images. "WZ" frames use the distributed coding scheme.

4 SIDE INFORMATION GENERATION

As can be seen in Figure 3, a WZ frame has a number of nearby frames that can be used to generate its side information. There are a number of methods which only use information from the WZ camera (**Intra-Camera** methods) like Motion-Compensated Temporal Interpolation (MCTI) (Lee, 2003), and also a number of methods which use information from the Intra cameras but only at time instant *k* (**Inter-Camera** methods) like Image-Based Rendering methods (Shum, 2000) or Disparity-Compensated View Prediction (DCVP, described next).

These two classes of methods estimate correctly some parts of the WZ frame, but fail in other parts: the Intra-Camera class has problems with high motion areas and the Inter-Camera class cannot easily deal with scene occlusions and reflections.

To overcome these difficulties a number of proposals have recently appeared (Guo, 2006), (Ouaret, 2006), (Artigas, 2006) that merge the correctly predicted parts of each estimation and discard the rest. The approach followed in this work is different in that it does not try to merge independently-obtained intra-camera and inter-camera estimates, but it directly uses all available information, as described next.

4.1 Basic Side Information Generation Techniques

Two of the simplest side information generation techniques are MCTI and DCVP, shown in Figure 4. MCTI obtains its motion vectors by means of block-matching, and uses temporally adjacent frames from

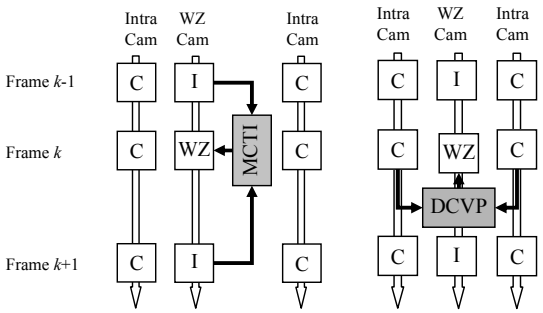

Figure 4: Basic techniques.

the WZ camera as references. DCVP employs the same method as MCTI, but uses frames from the Intra cameras, at the same time instant, as references. In order to avoid the limitations stated above for these techniques, the following method was researched.

4.2 Multiview Motion Estimation

The main idea behind Multiview Motion Estimation (MVME) is depicted in Figure 5: the motion vectors are first found on an Intra camera and then used on the WZ camera to estimate the WZ frames.

Details can be found in Figure 6. Firstly, the relationship between the two cameras is examined by finding the disparity vectors. These vectors play the same role as the motion vectors: they relate each block in the WZ camera with the most similar block in the Intra camera. This is similar to what DCVP presented before does.

Secondly, each matched block in the intra camera is again searched for, but in a temporally adjacent frame (as one would do in conventional motion compensation). In this step, the motion vectors are found.

Finally, the motion vectors obtained in the Intra camera are applied to the WZ camera to generate the estimation.

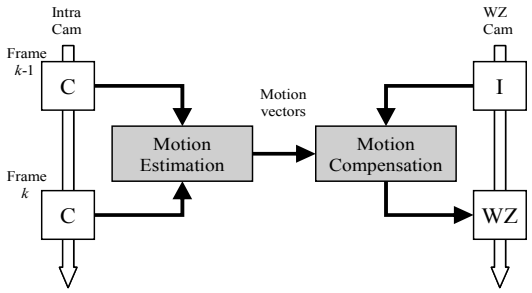

Figure 5: General scheme for the MVME technique.

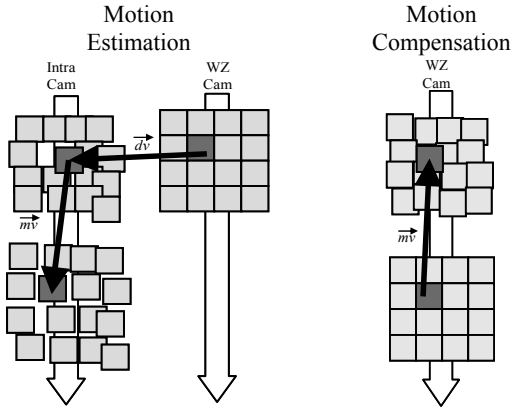

Figure 6: Detailed view of the MVME technique. $\vec{mv}$ and $\vec{dv}$ are the motion and disparity vectors respectively.

This technique works as long as all the cameras lie in the same plane and point in the same direction. If this is not the case, then the motion vectors need to be transformed before applying them to a different camera, which requires that the calibration matrix of each Intra and WZ camera is available.

The technique has been described so far for only one intra camera (left or right) and one reference frame (previous or next) in that intra camera. The particular frames that have been used (and the order in which they have been used) are called the *path*, and with the technique described so far 4 different paths are possible (depicted in Figure 7).

It is straightforward to increase the number of paths by taking the "orthogonal" ones, this is, 4 new estimates are generated by finding the disparity vectors in the previous time instant and applying them to the current time instant to estimate the WZ frame. The new paths are the disparity paths, and are obtained using the exact same process as the previous 4 paths (the motion paths), but use the

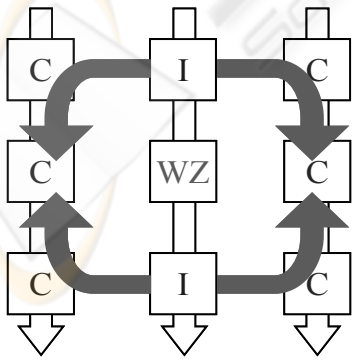

Figure 7: The four different paths obtained with two Intra cameras and two reference frames in each Intra camera.

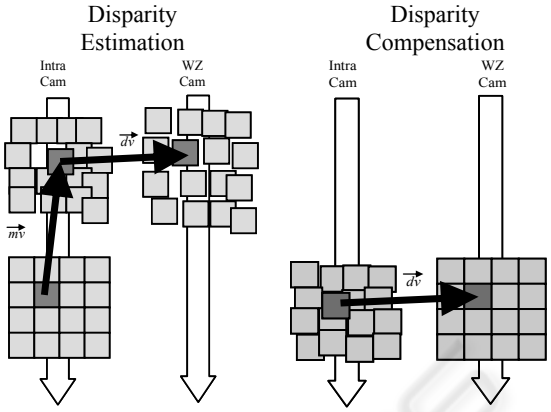

Figure 8: One disparity path for MVME. Note that the same frames are being used as the motion path in Figure 6, but in different order.

reference frames in a different order (Figure 8).

This technique doubles the amount of motion and disparity search that the decoder must carry out, but also doubles the quantity of available estimates, increasing the total performance as shown later.

The (different) estimates produced by each path need now to be merged, and the simplest mechanism is to average all of them. This is the technique called MVME. A more evolved method could calculate, for each block, a measure of the reliability of each path, and then build the final estimate by weighting each path by its reliability and adding all of them together. This technique is called MVME with Weighted Average (MVME-WA) and allows the most reliable paths to have a higher contribution to the final estimate.

The reliability measure used in this work to weight the different paths is based on the local variance of the motion (or disparity) field around each block. I.e., it is a measure on how uniform the field is. The rationale behind this choice is that real fields are usually uniform, except at object's boundaries, while incorrectly calculated fields are usually very noisy. Other reliability measures are currently being researched.

The side information generation techniques described in this section have been integrated in a distributed video codec to verify their validity. The outcome of the simulations is given next.

5 RESULTS

Results have been obtained by simulating a number of side information generation techniques for the three multiview test sequences listed in Table 1. The tested techniques are listed in Table 2.

Table 1: Simulated multiview test sequences and their characteristics.

Sequence	Frame size (pixels)	Number of frames	Frame rate (Hz)	Content
Breakdancers	256 x 192	100	15	Extremely fast movement.
Exit	192 x 144	250	30	Slow movement and small disparities.

Table 2: List of examined techniques.

Name	Description
MCTI	Pure temporal interpolation
MVME-4m	MVME using the 4 motion paths
MVME-WA-4m	MVME-WA using the 4 motion paths
MVME-WA-4d	MVME-WA using the 4 disparity paths
MVME-WA-8	MVME-WA using all 8 paths
H.264 Intra	Intra coding of all frames (no DVC)

A complete DVC codec[1] has been used to assess the performance of the methods presented. Results for H.264 in intra mode have also been added so the DVC results can be compared with a state-of-the-art codec with no motion search at the encoder.

The results of all the conducted experiments can be found in Figure 9. It can readily be seen that no single technique works best for all test sequences; For Breakdancers the best technique is H.264 Intra, and for Exit it is DVC with MCTI-generated side information. This is due to the different motion content of the sequences: The motion in Breakdancers is too high for the motion estimation algorithm to track, so independent encoding of every frame delivers the best performance.

This is also the reason why, for the Breakdancers sequence, even when all MVME variants are worse than H.264 Intra, they still outperform pure temporal interpolation (MCTI), the most complex one (MVME-WA-8) giving a gain of up to 1dB over MCTI.

However, this is not the case for the other sequence, where MCTI gave consistently better results than MVME. These results indicate that the MVME technique has yet to be improved to be useful for every type of sequence.

Also, the disparity path usually brought the worst results among the MVME-WA techniques. This is expected since correlation is normally higher between frames of the same camera than between different cameras.

[1] This software is called DISCOVER-codec and is the copyrighted work of the research project "Distributed coding for video services" (DISCOVER), FP6-2002-IST-C contract no.: 015314 of the European Commission. It cannot be copied, reproduced nor distributed without the consent of the project consortium.
The DISCOVER software started from the so-called IST-WZ software developed at the Image Group from Instituto Superior Técnico (IST), Lisbon-Portugal (amalia.img.lx.it.pt), by Catarina Brites, João Ascenso, and Fernando Pereira.

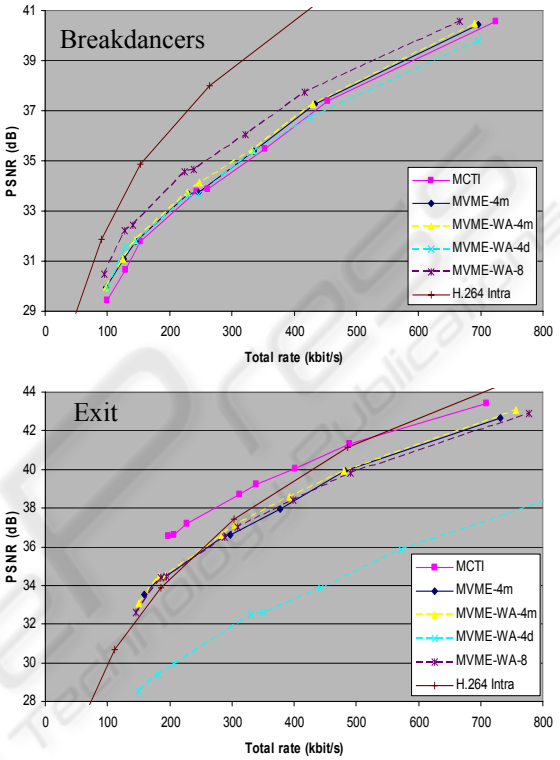

Figure 9: PSNR versus bit-rate for different side information generation techniques. All curves using weighted average are dashed for convenience.

6 CONCLUSIONS

Comparisons among side information generation approaches have been presented. In addition, a technique has been introduced to generate side information for multiview DVC scenarios, based on finding motion (or disparity) vectors on a different camera (or time instant) and apply them to the present camera (or time instant). Research is still ongoing, but it already presents promising results for some test sequences.

ACKNOWLEDGEMENTS

The work presented was developed within DISCOVER (www.discoverdvc.org), a European Project funded under the European Commission IST FP6 programme, and by grant TEC2005-07751-C02-02 of the Spanish Government.

REFERENCES

Artigas, X., Angeli, E. and Torres, L., 2006, "Side Information Generation for Multiview Distributed Video Coding Using a Fusion Approach", 7th Nordic Signal Processing Symposium, NORSIG'06, Reykjavik, Iceland, June 7 - 9.

Bahl, L.R., Cocke, J., Jelinek, F., Raviv, J., 1974, "Optimal Decoding of Linear Codes for minimising Symbol Error Rate," IEEE Transactions on Information Theory, vol.20, pp.284-287, March.

Girod, B., Aaron, A., Rane, S. and D. Rebollo-Monedero, 2005, "Distributed video coding", Proc. of the IEEE, vol. 93, no. 1, January.

Guo, X., Lu, Y., Wu, F., Gao, W. and Li, S., 2006, "Distributed multiview video coding", Proceedings of SPIE, January, San Jose, California, USA, Vol. #6077.

Lee, S.H., Kwon, O. and Park, R. H., 2003, "Weighted-Adaptive Motion-Compensated Frame Rate Up-Conversion", IEEE Trans. on Consumer Electronics, Vol. 49, No. 3.

Ohm, J.R., 1999, "Stereo/Multiview Video Encoding Using the MPEG Family of Standards", Invited Paper, Electronic Imaging '99, San Diego, January.

Ouaret, M., Dufaux, F. and Ebrahimi, T., 2006, "Fusion-based multiview distributed video coding", ACM International Workshop on Video Surveillance and Sensor Networks, Santa Barbara, CA, USA October 27.

Puri, R. and Ramchandran, K., 2002, "PRISM: A new robust video coding architecture based on distributed compression principles". Proc. of 40th Allerton Conf. on Comm., Control, and Computing, Allerton, IL, October.

Shum, H.Y. and Kang, S.B., 2000, "A Review of Image-based Rendering Techniques", IEEE/SPIE Visual Communications and Image Processing (VCIP) 2000, pp. 2-13, Perth, June.

Slepian, D. and Wolf, J., 1973, "Noiseless coding of correlated information sources", IEEE Trans. Inform. Theory, vol. 19 pp. 471-480, July.

Toffetti, G., Tagliasacchi, M., Marcon, M., Sarti, A., Tubaro, S. and Ramchandran, K., 2005, "Image Compression in a Multi-camera System based on a Distributed Source Coding Approach". European Signal Processing Conference, Antalya, September.

Wyner, A. and Ziv, J., 1976, "The rate-distortion function for source coding with side information at the decoder", IEEE Trans. Inform. Theory, vol. 22, pp. 1-10, January.

Zhu, X., Aaron, A. and Girod, B., 2003, "Distributed compression for large camera arrays", Proc. IEEE Workshop on Statistical Signal Processing, SSP-2003, St Louis, Missouri, September.