AN APPROACH TO DISTRIBUTED VIDEO CODING USING 3D FACE MODELS[1]

Xavi Artigas, Luis Torres

Technical University of Catalonia, Barcelona, Spain
{xavi, luis}@gps.tsc.upc.edu

ABSTRACT

Distributed Source Coding (DSC), formulated thirty years ago, is lately witnessing a great research effort. This effort is encouraged by emerging new applications that could greatly benefit from the properties inherent to DSC, like low complexity encoders and embedded error resilience.

Among the many challenging topics related to DSC there is the generation of the Side Information, an estimation made by the decoder of the information being decoded.

This paper quickly summarizes the theoretical bases of DSC and then focuses on the essential task of generating the Side Information, by developing the model-based method drafted by the authors on a previous publication. It is shown that the application of model-based Side Information generation techniques on top of current state of the art DSC-based video codecs, provides an improvement of 0.5 to 0.75dB over a wide range of bit-rates. Comparisons against H.264 are also provided.

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. This paradigm is strongly dominated and determined by applications such as broadcasting, video on demand, and video streaming. 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 change of paradigm also moves the encoder-decoder complexity balance, theoretically allowing the provision of efficient compression solutions with simple encoders and complex decoders. 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 were set thirty years ago with the work by Slepian & Wolf [1] (for the lossless case) and Wyner & Ziv [2] (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. [3] and Girod et al. [4]. A good review of other works can be found in [4].

Although Distributed Source Coding can be used in other areas, like Robust Channel Transmission, this paper focuses purely on the aspects related to compression using low-complexity encoders.

The novelty presented here is the combination of model-based and block-based side information generation techniques, and the iterative refinement of them both. A predefined 3D model of a human head is used, therefore restricting the area of application of this codec, but also, by focusing on a particular application (like video-conferencing), the decoder has extra knowledge that can be used to help the decoding process. Previous work on theses issues was presented in [5] and [6].

This paper is organized as follows. Section **2** presents a quick review of the Slepian and Wolf's and Wyner and Ziv's theorems for Distributed Source Coding. Section **3** then summarizes the approach followed in [4], for the particular case of Distributed Video Coding as it provides the basis of our codec. Next, Section **4** introduces a recent block-based motion-compensated interpolation technique that uses refinement of the motion vectors. Section **5** proceeds to present the model-based approach. Simulation results and comparison to current state-of-the-art are given in Section **6**, and, finally, Section **7** draws some conclusions.

[1] The work presented was developed within DISCOVER, a European Project (http://www.discoverdvc.org), funded under the European Commission IST FP6 programme.

2. THEORETICAL FOUNDATIONS

It is a well known fact that the minimum lossless rate at which a signal X can be transmitted is $H(X)$, the signal's entropy. It is also well known that if two statistically dependent signals X and Y are to be transmitted, the best thing that can be done is to encode them together, in order to exploit the statistical dependencies, and that the minimum lossless rate is then $H(X, Y)$, their joint entropy. Slepian and Wolf showed in 1973 [1] that this lower bound for the lossless joint transmission rate is also achievable when the signals X and Y are encoded *separately*, provided that some conditions are fulfilled. That is, when the encoder for X does not have access to Y, and vice versa. A codec that exploits this theorem is called a *Slepian-Wolf codec*.

It is the ability of encoding X and Y separately that makes Distributed Source Coding so attractive, because encoders for separate signals do not have to search for inter-correlations among signals, and therefore require fewer computations. Note that, in order to correctly decode the transmitted signals, these inter-correlations still have to be found, but this is now done in the decoder. On an implementation context, this means that the complexity of the coder is transferred to the decoder.

Three years later, in 1976, the work from Wyner and Ziv [2] extended the work by Slepian and Wolf [1] by studying the lossy case in the same scenario.

3. BASE WYNER-ZIV CODEC

The proofs of the above theorems are asymptotical and non-constructive, meaning that the implementation of a codec based on them is not straightforward, and, as a consequence, different approaches are currently being explored. The two approaches that have had more continued effort in the field of video coding are the one by Ramchandran et al., [3], and the one by Girod et al., [4]. The Wyner-Ziv codec chosen as the base implementation for this paper follows the approach in [4] and is briefly described next.

The pixel-domain codec described in [4] separates the input video sequence in Intra frames and Wyner-Ziv frames. Intra frames are independently encoded and decoded using a conventional codec. The decoder uses neighboring Intra frames to build a motion-compensated estimation of the Wyner-Ziv frames which is called Side Information. The encoder calculates parity bits for the original Wyner-Ziv frames using a Turbo Codec which are then used by the decoder to correct the possible prediction errors present in the Side Information. It is important to generate as accurate Side Information as possible since this directly affects the rate-distortion performance of this architecture. The parity bits generated by the encoder for the Wyner-Ziv frames are called Wyner-Ziv data in this paper. While still not reaching the performance of state-of-the-art inter-frame coding, the system is reported to perform 10-12 dB better than H.263+ intra-frame coding [4].

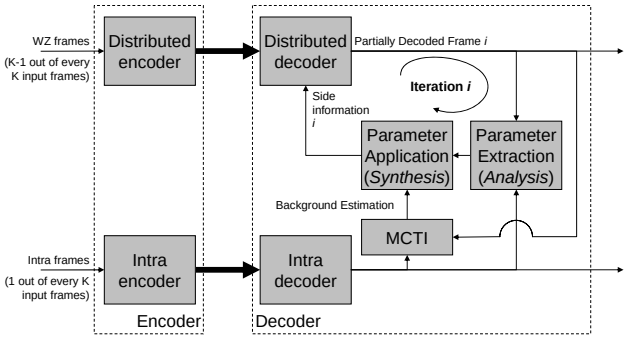

Figure 1 Proposed model-based distributed video codec.

4. MOTION COMPENSATED RESTORATION

Since Side Information plays such an important role in DVC codecs, improving its quality (correlation with the original frame) is a very good method to improve the overall codec performance. One particular approach to accomplish this goal is to continuously refine the Side Information as Wyner-Ziv data is being received [5][7].

The basic idea behind the Motion Compensated Restoration technique introduced in [5] is to refine the motion vectors used to generate the Side Information every time a bitplane is decoded. Before decoding starts, an initial Side Information is created using motion compensated interpolation. Once the first bitplane is decoded, it is used to help adjust the motion vectors and a second Side Information is produced, which can then be used in the reconstruction phase and as initial Side Information for the subsequent bitplanes.

5. MODEL-BASED SIDE INFORMATION

The main objective of this paper is to increase the quality of the Side Information in a DVC scheme by mixing model-based and block-based approaches. A 3D model will be used to estimate the face present in the foreground in videoconference sequences, and the more general block-based approach will be used for the background.

In addition, both approaches allow progressive refinement. The motion vectors of the blocks in one case (as seen in the previous section), and the 3D model parameters in the other case can be both refined as more Wyner-Ziv data is received.

The block diagram of the proposed model-based scheme appears in Figure 1, and was drafted in a previous publication by the authors [6], built on top of the codec described in [4]. The main difference of our proposal and that of [4] described in Section 3 is the generation of the side information as described next.

The block-based motion-compensated temporal interpolation (MCTI) is created by using information only from neighboring Intra frames. At the same time, the Parameter Extraction module "fits" a deformable 3D model of a human head [8] to the same neighboring Intra frames. This means that all the parameters that control the deformation of the model, like position, rotation, mouth opening, etc. have to be found. These parameters are then interpolated and passed to the Parameter Application module, which creates a synthetic image to be used as Side Information. The parts of the frame not covered by the 3D model are filled with the MCTI estimation. At this point an initial Side Information has been created and is ready to be used by the Wyner-Ziv decoder.

In the Parameter Application stage, the interpolated parameters only allow the creation of "wireframe" images like the ones in Figure 2, therefore, to give the synthetic images a more realistic look, they have to be textured. For example, the textures can be taken from the neighboring Intra frames and averaged. Taken as a whole, the parameter analysis and synthesis stages perform image warping, in a way similar to mesh-based interpolation, displacing all the pixels inside a reference triangle into the destination triangle. The main advantage of this method over block-based interpolation is that contiguous triangles in the reference frame continue to be contiguous after warping, so no blocking artifacts are created. Figure 2 shows how the pixels inside the reference triangle adapt to the shape of the destination triangle by shrinking or expanding as necessary.

Once an initial Side Information has been created, it is combined with the Wyner-Ziv data in the distributed decoder, and a *Partially Decoded Frame* is produced. DVC schemes not implementing Side Information refinement directly regard this frame as the output frame. However, since this partially decoded frame contains information that was not present in the reference frames used to create the Side Information, it is worth using it to improve the Side Information. It can be used in the MCTI module to refine the motion vectors, as explained in Section 4, leading to a more accurate estimation of the background. The partially decoded frame can also be used to directly estimate the model parameters, instead of interpolating them from neighboring Intra frames. This means fitting the 3D model to the partially decoded frame, which will lead to a more accurate estimation of the face in the foreground.

This process can be clearly iterated, with each iteration giving more accurate estimations for both the background and the foreground of the frame. These more accurate estimations mean more accurate Side Information that will in turn generate better quality output frames.

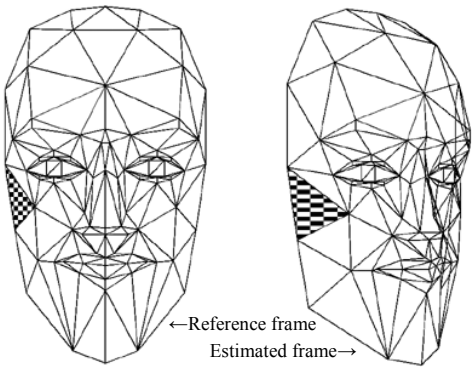

Figure 2 Pixel triangles in the reference frame are warped to build the estimated frame.

6. SIMULATION RESULTS

At the current stage of this research, the model is being adjusted manually. Every frame is fitted to the partially decoded frame and requires user intervention. It is clear that an automatic fitting of the model is needed in real scenarios, but the objective here is to prove the validity of the 3D model-based approach.

The proposed model-based system has been built on top of a DVC codec following the approach in [4]. A return channel is used, and the error probability required to decide if more parity bits are needed is assumed available at the decoder (ideal error detection). The DVC codec used[2] incorporates a block-based motion refinement mechanism, described in [7] and [9]. 100 frames of the Foreman sequence in QCIF format at 30 Hz. have been coded. Even frames were intra coded and decoded while odd frames, the Wyner-Ziv frames, used the proposed codec. The 3D model used is a slightly modified version of the Candide model [8], using two reference frames for the texturing process.

Rate-distortion plots are shown in Figure 3, along with H.264 in intra mode, for comparison purposes with a codec of similar *encoder* complexity. The rate axis is the rate of the Wyner-Ziv frames, that is, the frames that are not intra coded. It can be seen that there are 0.5 to 0.75dB of improvement when the model-based side information is added on top of the existing codec[2].

Intermediate steps in the decoding of a particularly difficult frame (number 193) are shown in Figure 4 . For this frame, the MCTI gives a very low PSNR for the area of the face, which is partially corrected by the model. It can be seen that the initial Side Information (images **b)** and **f)**) has already better quality when using the model-based approach. After two bitplanes have been decoded (images **c)** and **g)**) Side Information refinement has fixed most of the errors in the initial estimation. In the final results (images **d)** and **h)**) the difference between the two approaches lies only

[2] 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 (http://amalia.img.lx.it.pt/), by Catarina Brites, João Ascenso, and Fernando Pereira.

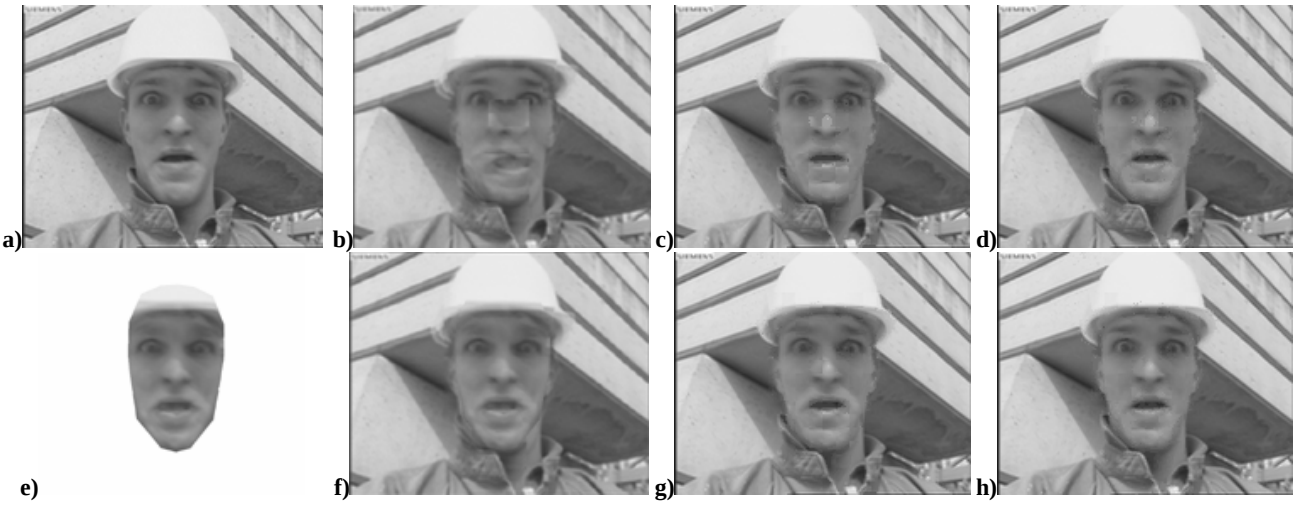

Figure 4 Frame 193 of the QCIF Foreman sequence. **a)** original frame. **e)** image generated by rendering only the 3D face model. **b)**, **c)** and **d)** use the base Wyner-Ziv codec. **f)**, **g)** and **h)** use the proposed Model-based codec. **b)** and **f)** initial Side Information. **c)** and **g)** results after decoding two bitplanes. **d)** and **h)** final results (decoding four bitplanes).

in the last four bitplanes, which have not been transmitted and therefore rely solely on the Side Information.

It is interesting to note that the model does not suffer (and will never suffer) from the triple-eye effect, although it comes with its own artifacts, like the seams at the perimeter of the face. It is also worth mentioning, that the approach is still very far (more than 10dB) from current H.264 interframe coding results, so more work is needed to decrease the current gap of DVC systems and current state of the art of hybrid codecs.

7. CONCLUSION

This paper has shown the importance of generating a good Side Information in order to increase the performance of a Distributed Video Coding scheme. A scheme using model-based Side Information has been proposed which improves the performance of an existing state-of-the-art DVC codec by 0.5 to 0.75dB depending on the bit-rate.

Moreover, the proposed Side Information generation method allows iterative refinement of its parameters (the 3D face model deformation parameters) which can greatly increase its performance.

8. REFERENCES

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

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

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

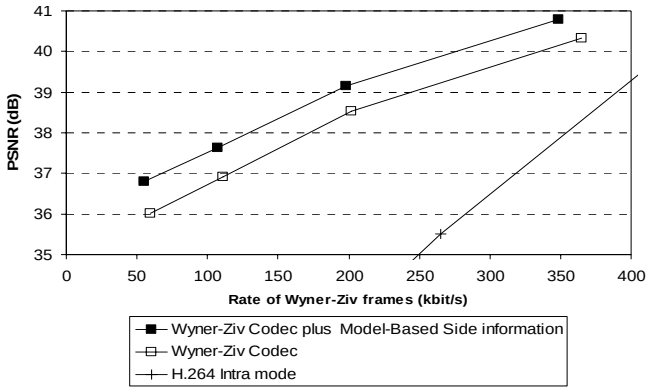

Figure 3 Rate-distortion comparison of the model-based approach with the base Wyner-Ziv codec and H.264 in intra mode. Foreman QCIF @ 30 Hz (rate of the Wyner-Ziv frames 15 Hz)

[4] B. Girod, A. Aaron, S. Rane and D. Rebollo-Monedero, "Distributed video coding", P*roc. of the IEEE*, vol. 93, no. 1, January 2005

[5] X. Artigas, L. Torres, "Iterative Generation of Motion-Compensated Side Information for Distributed Video Coding*", Proceedings of the IEEE International Conference on Image Processing*, Genova September 11-14 2005

[6] X. Artigas, L. Torres, "A model-based enhanced approach to distributed video coding", Image Analysis for Multimedia Interactive Services, WIAMIS, Montreux, Switzerland, April 13-15 2005.

[7] J. Ascenso, C. Brites, F. Pereira, "Motion Compensated Refinement for Low Complexity Pixel Based Distributed Video Coding" , *IEEE International Conference on Advanced Video and Signal-Based Surveillance*, Como – Italy, September 2005

[8] http://www.bk.isy.liu.se/candide/

[9] J. Ascenso, C. Brites, F. Pereira, "Improving frame interpolation with spatial motion smoothing for pixel domain distributed video coding" , 5th EURASIP Conference on Speech and Image Processing, Multimedia Communications and Services, Smolenice – Slovak Republic, June 2005