CODEC-INDEPENDENT SCALABLE DISTRIBUTED VIDEO CODING

Mourad Ouaret, Frederic Dufaux, Touradj Ebrahimi
Institut de Traitement des Signaux
Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland.

ABSTRACT

In this paper, we introduce novel schemes for scalable Distributed Video Coding (DVC), dealing with temporal, spatial and quality scalabilities. More specifically, conventional coding is used to obtain a base layer. DVC is then applied to generate enhancement layers. The side information is generated either temporally by motion compensated interpolation, or spatially by a spatial bi-cubic interpolation. Note that this scalable DVC approach is independent from the codec used to encode or decode the base layer. Simulation results show that most of the proposed schemes outperform non-scalable DVC, in addition to enabling the scalability features.

Index Terms—Video codecs, Source coding.

1. INTRODUCTION

Scalable coding is becoming important nowadays in heterogeneous multimedia networks. Different clients on a network might require decoding the same video at different resolutions and qualities. For this purpose, scalable coding encodes the video once and enables decoding at different qualities (or Signal to Noise Ratio (SNR)), spatial and temporal resolutions. This makes scalable coding attractive for several applications such as video surveillance and video browsing.

The Moving Picture Experts Group (MPEG) has recently introduced the Scalable Video Coding (SVC) [1], which is an extension of the H.264/MPEG-4 Advanced Video Coding (AVC) standard [2]. SVC achieves very good compression performance. On the other hand, SVC entails a high complexity at the encoder side.

Simultaneously, work is conducted in the field of Distributed Video Coding (DVC) [3] as new paradigm in compression. Using DVC, the source statistics can be exploited at the decoder side. For instance, the burden of motion estimation and compensation can be shifted from the encoder to the decoder. In a practical scenario, this implies low power / low complexity encoders. In addition, DVC is attractive for error resilience since the absence of a prediction loop prevents drifts, as well as the capability of the Wyner-Ziv bits to correct errors due to lost or corrupted data. Thus, combining scalability and DVC would be beneficial for compression in heterogeneous video networks and video surveillance systems.

Tagliasacchi et al. [4] implemented a scalable version of PRISM (Power-efficient, Robust, hIgh compression, Syndrome-based Multimedia coding) [5]. The approach enhances an H.264/AVC base layer with a PRISM refinement bitstream resulting in a spatio-temporal scalable video codec. It focuses on the case where estimation and most of the motion compensation task is performed at the decoder. Results show that scalable PRISM outperforms non-scalable PRISM and H.263+ Intra, but has a poorer performance when compared to motion compensated H.263+. In fact, since the base layer used H.264/AVC, comparison should have been made with respect to the latter.

A solution to the problem of scalable predictive video coding is introduced in [6] by posing it as a variant of the Wyner-Ziv side information problem. It discusses mainly quality scalability. Results show that the proposed codec is approximately 4.0 dB superior to a naive scalable codec. However, it does not address temporal and spatial scalabilities. In addition, motion compensation is performed at the encoder which increases its complexity.

In this work, we introduce scalable video coding schemes based on DVC, enabling temporal, spatial and quality scalabilities. For temporal scalability, side information generated by motion compensated interpolation is used to enhance the temporal resolution. Spatial downsampling and upsamling are used along with Wyner-Ziv bits to ensure spatial scalability. Furthermore, the amount of parity bits sent defines the refinement quality of decoded video. Finally, we show that the approach is codec independent. In other words, it is independent of the coding scheme used for the base layer. For this purpose, H.264/AVC Intra and JPEG2000 [7] are used as illustration to encode the base layer.

This paper is structured as follows. First, the paradigm of DVC is outlined in section 2. Then, the proposed scalable DVC schemes are introduced in section 3. Experimental results are presented in section 4, comparing the rate-distortion performance with respect to non-scalable DVC. Finally, some concluding remarks are drawn in section 5.

2. DISTRIBUTED VIDEO CODING

DVC is the consequence of information-theoretic bounds established by Slepian and Wolf [8] for distributed lossless coding, and by Wyner and Ziv [9] for lossy coding with decoder side information. In a practical scenario, lossy coding is used. In this paper, we consider the DVC architecture from [10] as illustrated in Fig. 1.
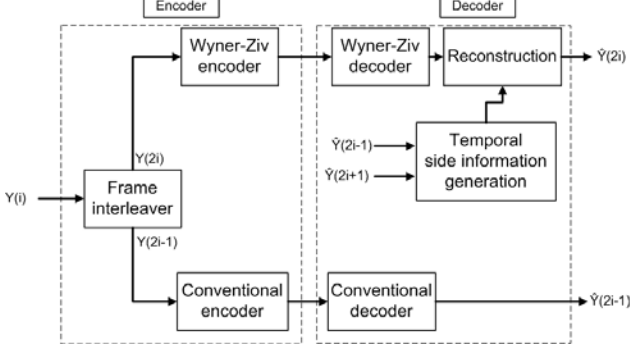

Figure 1. Non-scalable DVC scheme (GOP =2).

The Wyner-Ziv encoder operates in the DCT domain. In other words, an interleaved turbo encoder is used to generate parity bits for the quantized DCT coefficients. For a Group Of Pictures (GOP) size equal to two, the conventionally decoded previous and next key frames are used to generate side information by motion compensated interpolation. To exploit the side information, the decoder assumes a statistical model, which is a Laplacian distribution of the difference between the individual DCT coefficients of the original frame and the side information. The decoder combines the side information and the received parity bits to recover the original frame. A feed back channel is used to request the parity bits from the encoder until a small error probability is reached. For more details on the used DVC scheme, see [3] and [10].

3. SCALABLE DISTRIBUTED VIDEO CODING

In this section, we introduce novel DVC schemes for temporal, quality and spatial scalabilities. For simplicity, we consider the case where the GOP size and the spatial downsampling/upsampling factor are equal to two.

3.1. Temporal scalability
The DVC scheme depicted in Fig. 1 ensures temporal scalability. By decoding only key frames, the decoded video has half the original temporal resolution (GOP=2). This makes temporal scalability straight forward. Furthermore, the scheme can be easily extended to n temporal enhancement layers. In this case, a GOP of 2^n is chosen. This results in a base layer with $1/2^n$ the full temporal resolution, whereas the decoding of each enhancement layer doubles the temporal resolution. Fig. 2 illustrates the decoding process for two enhancement layers (GOP=4).
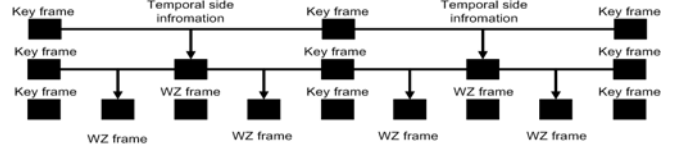

Figure 2. Temporal scalability for two enhancement layers (GOP=4).

3.2. Quality scalability
In Fig. 3, a DVC scheme for quality scalability is introduced. Each frame is simultaneously encoded by Wyner-Ziv and conventional coding. The generated parity bits are used to enhance the quality of the conventionally decoded frames. As mentioned previously, the DVC encoder operates in the DCT domain. Moreover, the quantized DCT coefficients are organized in bands, which are separated into bitplanes. The latter are organized from most to least significant. The parity bits are generated for a certain number of bitplanes. As the number of bitplanes increases, the quality of the refined frame improves.
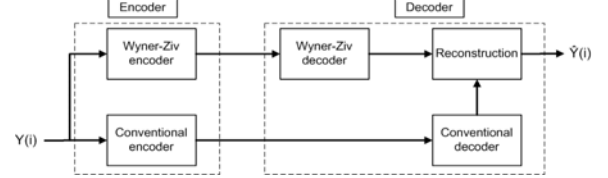

Figure 3. Quality scalability scheme.

3.3. Spatial and Quality scalability
We now introduce a DVC scheme for spatial and quality scalability as an extension of the previous one. A spatial downsampling/upsampling is introduced prior/after the conventional encoding/decoding in the lower branch as illustrated in Fig. 4. In parallel, a quality refinement stream is sent for the enhancement of the base layer $\hat{Y}_B(i)$ to produce $\hat{Y}_L(i)$. One of the latter is used as side information to decode the video, $\hat{Y}(i)$, at the full spatial resolution. The downsampling and upsampling (Bi-cubic interpolation) are performed in both vertical and horizontal directions. Moreover, the downsampling is preceded by a convolution with a Gaussian kernel to reduce the effect of aliasing.

In the case of n enhancement layers, n successive downsampling/upsampling stages are required. At each, a quality refinement layer for the current spatial resolution is sent to the decoder.
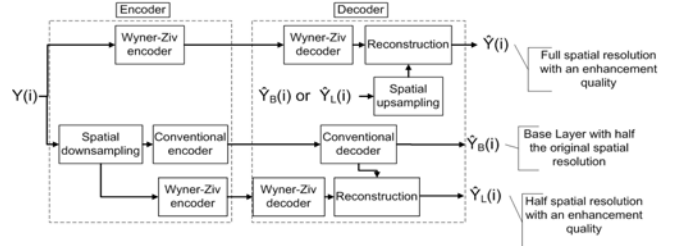

Figure 4. Quality and spatial scalability scheme for two enhancement layers.

3.4. Temporal, Spatial and Quality scalability

Finally, we introduce a DVC scheme for temporal, quality and spatial scalability in Fig. 5, by combining the three schemes previously described. First, the odd frames are spatially downsampled and then conventionally encoded in block A. When decoded, the latter produce the base layer, $\hat{Y}_B(2i-1)$ and $\hat{Y}_B(2i+1)$, which has half the spatio-temporal resolution of the original video. At the same time, Wyner-Ziv bits are sent in parallel to enable the quality enhancement of the base layer to generate $\hat{Y}_L(2i-1)$ and $\hat{Y}_L(2i+1)$.

Then, temporal side information and DVC decoding are used to generate video at full frame rate and half spatial resolution in block B, i.e $\hat{Y}_L(2i)$.

Furthermore, $\hat{Y}_L(2i-1)$, $\hat{Y}_L(2i+1)$ (or $\hat{Y}_B(2i-1)$, $\hat{Y}_B(2i+1)$) and $\hat{Y}_L(2i)$ are spatially interpolated and used as side information to generate video at the full spatio-temporal resolution in blocks C and D. Note that the video can be decoded at half its original temporal resolution and full spatial resolution if blocks B and D are skipped.

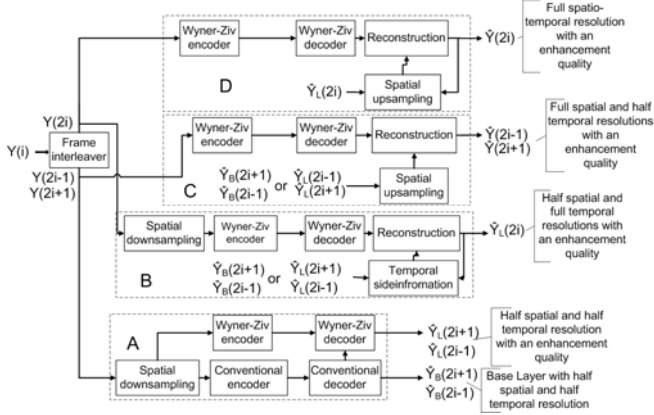This scheme can be easily extended to *n* layers, since it is a combination of the previous ones.

Figure 5. DVC scheme for Temporal, Spatial and Quality scalability.

4. EXPERIMENTAL RESULTS

4.1. Video sequences

The four sequences, *Soccer*, *Crew, Foreman* and *Football,* are used to evaluate the rate-distortion performance of the proposed schemes with respect to non-scalable DVC. The sequences *Football* and *Soccer* contain high motion video content. The sequence *Crew* contains more or less uniform texture with sudden illumination changes due to camera flashes. Finally, the sequence *Foreman* contains a camera pan. The spatio-temporal resolution 352x288@15fps is used.

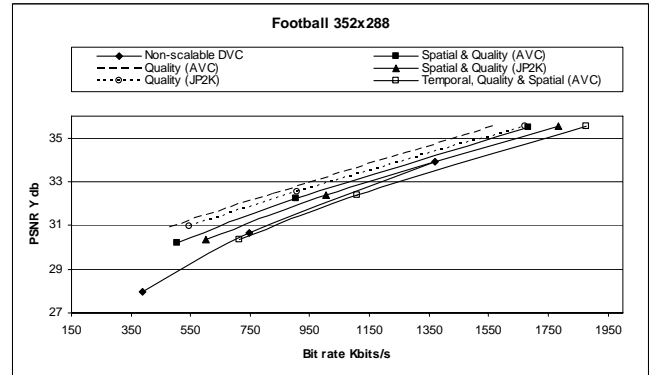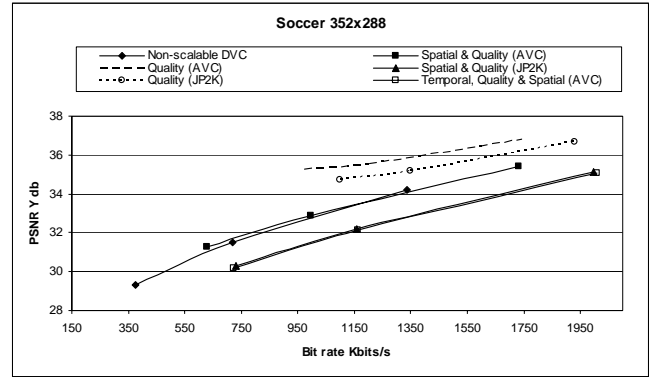### 4.2. Rate-distortion performance

The rate-distortion performance is evaluated for non-scalable DVC (GOP=2) and the introduced scalability schemes. The base layer is encoded using either JPEG2000 or H.264/AVC Intra with the same refinement stream. For example, "Quality(AVC)" in the rate-distortion figures

means the quality scalability scheme with an H.264/AVC Intra encoded base layer.

For the Sequences *Soccer*, *Crew* and *Football*, the quality scalable scheme is superior to non-scalable DVC with around 1.0~2.0 dB. Moreover, the spatial and quality scheme outperforms the non-scalable scheme by around 1.0 dB while it has the same performance as the spatial and quality scheme when H.264/AVC Intra is used as the base layer codec for *Soccer* and *Crew*. When JPEG2000 is used instead of H.264/AVC Intra, the proposed scheme is outperformed by non-scalable DVC by around 1.0 dB. For the scheme enabling all three scalabilities, a 1.0 dB performance gap separates the scheme enabling all the three scalabilities from the non-scalable one for *Soccer* and *Crew*, while they have similar performance for *Football*.

For the sequence *Foreman*, the non-scalable scheme outperforms the quality scheme at high bit rates, while the latter has a better performance at lower bit rates when the H.264/AVC Intra is used as the base layer codec. For the rest of the schemes, they are outperformed by the non-scalable scheme. This is due to the fact that this sequence contains less motion when compared to the other sequences.

For the introduced schemes, the quality scheme has the best performance followed by the spatial and quality scalability scheme. Finally, the worst performance is for the scheme enabling all three scalabilities. This makes sense since as more scalability layers are added, the poorer is the performance.
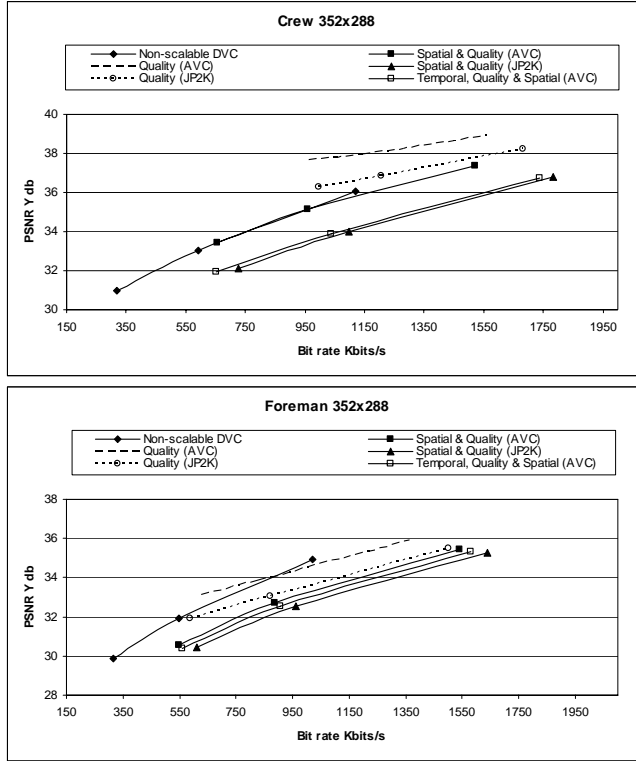
Figure 7. Soccer, Football, Crew and Formane sequence rate-distortion performance.

Table 1 shows the percentage of Wyner-Ziv bits out of the total bit stream. The quality scheme contains the least amount of the Wyner-Ziv data. Thus, the quality scheme contains the largest amount of AVC Intra stream when compared to the other schemes. This explains why this scheme has the best performance among the introduced schemes.

	Quality	Spatial & Quality	Spatial, Quality & Temporal	Non-scalable
Soccer	30.86%	53.43%	88.31%	73.06%
Crew	27.56%	48.42%	89.01%	70.55%
Football	27.60%	73.33%	88.76%	78.50%
Foreman	21.47%	59.66%	83.10%	63.67%

Table 1. Wyner-Ziv bits average percentage out of the total bit stream.

5. CONCLUSIONS

In this paper, schemes for scalable Distributed Video Coding (DVC), dealing with temporal, spatial and quality scalabilities are introduced. Simulation results show that most of the proposed schemes have a better performance, or similar to, non-scalable DVC, in addition to enabling scalability. Furthermore, our approach implies low complexity encoders since all the motion compensation is performed at the decoder. Finally, the approach is codec-independent, which means that base-layer could use any conventional encoder.

This work can be extended by studying the behaviour of the different schemes when errors occur in the base layer, the refinement stream or both. The introduced schemes should be quite robust with respect to errors since parity bits are used to produce the higher layers. In addition, the base layer is encoded in the Intra mode. The latter is attractive for error resilience since each frame is encoded on its own without using information from neighbouring frames. In addition, the DVC decoding uses a feedback channel. A more practical solution would be to estimate the total amount of Wyner-Ziv bits correctly at the encoder.

7. REFERENCES

[1] International Organization for Standardization, "Introduction to SVC Extension of Advanced Video Coding", ISO/IEC JTC1/SC29/WG11, International Organization for Standardization, Coding of Moving Pictures and Audio, Poznań, Poland, July 2005. URL: http://www.chiariglione.org/mpeg/technologies/mp04-svc/.

[2] Thomas Wiegand, Gary J. Sullivan, Gisle Bjøntegaard, and Ajay Luthra, "Overview of the H.264/AVC Video Coding Standard", IEEE Trans. on Circuits and Systems for Video Technology, vol. 13, no. 7, July 2003.

[3] Bernd Girod, Anne Aaron, Shantanu Rane and David Rebollo-Monedero, "Distributed Video Coding". Proceedings of the IEEE, vol. 93, no. 1, pp. 71-83, January 2005.

[4] M. Tagliasacchi, A. Majumdar, K. Ramchandran: "A Distributed Source Coding based Spatio-Temporal Scalable Video Codec". Picture Coding Symposium 2004, San Francisco, December 2004.

[5] R. Puri and K. Ramchandran, "PRISM: A New Robust Video Coding Architecture Based on Distributed Compression Principles". Allerton Conference on Communication, Control and Computing, October 2002.

[6] A. Sehgal, A. Jagmohan, N. Ahuja, "Scalable Video Coding Using Wyner-Ziv Codes", Proc. International Picture Coding Symposium, PCS'04, San Francisco, CA, December 2004.

[7] Skodras, A., Christopoulos, C., Ebrahimi, T., "The JPEG 2000 still image compression standard", Signal Processing Magazine, IEEE vol. 18, Issue 5, pp. 36 – 58, September 2001.

[8] J. Slepian and J. Wolf, "Noiseless Coding of Correlated Information Sources", IEEE Trans. on Information Theory, vol. 19, no. 4, July 1973.

[9] A. Wyner and J. Ziv, "The Rate-Distortion Function for Source Coding with Side Information at the Decoder", IEEE Trans. on Information Theory, vol. 22, no. 1, January 1976.

[10] C. Brites, J. Ascenso, F. Pereira, "Improving Transform Domain Wyner-Ziv Video Coding Performance", International Conference on Acoustics, Speech and Signal Processing, Toulouse, France, May 2006.