Distributed Video Coding: Bringing New Applications to Life

Catarina Brites[1], Fernando Pereira[2]

Instituto Superior Técnico - Instituto de Telecomunicações, Av. Rovisco Pais, 1049-001 Lisboa, Portugal
[1]catarina.brites@lx.it.pt, [2]fp@lx.it.pt

Abstract — **Distributed source coding (DSC) is a new coding paradigm based on two Information Theory results: Slepian-Wolf and Wyner-Ziv theorems. DSC theory relies on the coding of two or more dependent random sequences in an independent way, i.e. associating an independent encoder to each sequence. A single decoder is used to perform joint decoding of all encoded sequences exploiting the statistical dependencies between them. Based on the DSC independent encoding-joint decoding configuration, a new video coding paradigm, called distributed video coding (DVC), allows to shift complexity from the encoder to the decoder. Typically, this is difficult to achieve with current hybrid video coding where encoders are more complex and should enable a whole new set of applications. The major goals of this paper is to review the principles, advantages and recent developments in DVC and discuss the set of applications for which DVC seems to be well-suited, typically when low-complexity or low-power consumption encoders are required.**

I. INTRODUCTION

Today's digital video coding paradigm, represented by the ITU-T VCEG and ISO/IEC MPEG standardization efforts, relies on hybrid block-based transform and interframe predictive coding. In this coding framework, the encoder architecture is based on the combination of motion compensated predictions with the DCT transform in order to exploit both the temporal and spatial redundancy present in the video sequence. In this type of framework, the encoder has a higher computational complexity than the decoder (typically 5 to 10 times more complex [1]), mainly due to the tools that explore the temporal correlation, i.e. the motion estimation tool. After all, it is the encoder that is responsible for all coding decisions and to work in order to attain maximum rate-distortion performance, while the decoder remains a pure executer of the encoder "orders". This type of architecture is well-suited for applications where the video is encoded once and decoded many times, i.e. one-to-many topologies, such as broadcasting or video-on-demand, where the cost of the decoder is more critical than the cost of the encoder.

In recent years, with emerging applications such as wireless low-power surveillance, multimedia sensor networks, wireless PC cameras and mobile camera phones, the traditional video coding architecture is being challenged. These applications have very different requirements than those of traditional video delivery systems. For some applications, it is essential to have a low power consumption both at the encoder and decoder, e.g. in mobile camera phones. In other types of applications, notably when there are a high number of encoders and only one decoder, e.g. surveillance, low cost encoder devices are needed. To fulfil these new requirements, it is essential to have a coding configuration with a low-power and low-complexity encoder device, possibly at the expense of a high-complexity decoder (see *Fig. 1*). In this configuration, the goal in terms of compression efficiency would be to achieve a coding efficiency similar to the best hybrid video coding schemes (e.g. the recent H.264/AVC standard [2]). That is, the shift of complexity from the encoder to the decoder should ideally not compromise the coding efficiency; this is currently rather far from happening and much research needs to happen in this area.

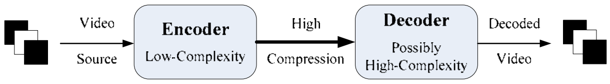

Fig. 1– Ideal coding configuration for some emerging video applications.

In this context, the major goals of this paper are: i) present a brief review of the DVC theory, its implications and advantages in terms of video coding solutions, and check some practical coding solutions recently developed, and ii) discuss some emerging applications which can strongly benefit from the DVC paradigm. This paper is organised as follows: Section II presents the theoretical background that supports the DVC paradigm; Section III gives an overview about currently available distributed video coding solutions; finally, several application scenarios that can benefit from DVC paradigm are discussed in Section IV, and some final remarks are presented in Section V.

II. THEORETICAL BACKGROUND

For some visual communication systems, having an efficient and low-complexity video encoder is a strong demand. Some results from Information Theory suggest that it is possible to achieve this goal by exploiting source statistics, partially or totally, at the decoder. These results can be used in the design of a new type of video coding algorithms, the so-called distributed video coding solutions.

A. Slepian-Wolf Theorem

The first study in distributed source coding was made in the 1970s when Slepian and Wolf studied the situation described in *Fig. 2*. In this context, the term "distributed" refers to the encoding operation.

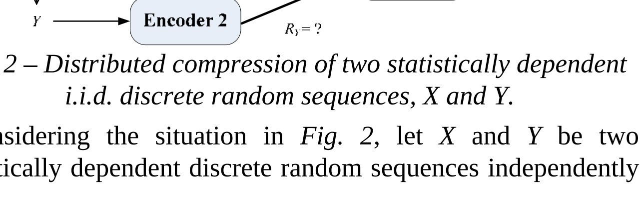

Fig. 2 – Distributed compression of two statistically dependent i.i.d. discrete random sequences, X and Y.

Considering the situation in *Fig. 2*, let X and Y be two statistically dependent discrete random sequences independently

---

The work presented was developed within VISNET, a European Network of Excellence (http://www.visnet-noe.org).

and identically distributed (i.i.d.). These sequences are separately encoded with rates R_X and R_Y, respectively, but are jointly decoded, exploiting the correlation between them. The possible rate combinations of R_X and R_Y for a reconstruction of X and Y with an arbitrarily small error probability, were obtained by Slepian-Wolf in [3]. The possible rate combinations are expressed by:

$$R_X \geq H(X \mid Y) \tag{1}$$

$$R_Y \geq H(Y \mid X) \tag{2}$$

$$(R_X + R_Y) \geq H(X, Y) \tag{3}$$

where $H(X|Y)$ is the conditional entropy of X given Y and $H(Y|X)$ is the conditional entropy of Y given X. Equation (3) shows that even when the encoding of correlated sources is performed independently, a total bit rate,

$$R = R_X + R_Y, \tag{4}$$

equal to the joint entropy is enough. So, separate encoding in distributed video coding schemes does not (theoretically) need to have any compression efficiency loss when compared to the joint encoding used in the traditional video coding paradigm. This is exactly what the Slepian-Wolf theorem states [3]. Fig. 3 illustrates the achievable rate region for which the distributed compression of two statistically dependent i.i.d. sources, X and Y, allows recovery with an arbitrarily small error probability according to the Slepian-Wolf theorem. In Fig. 3, the vertical, horizontal and oblique blue dashed lines, corresponding to the conditions in (1), (2) and (3), respectively, represent the inferior bounds for the achievable rate combinations of R_X and R_Y.

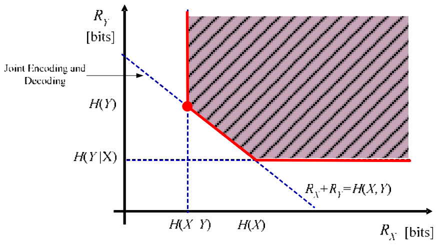

Fig. 3 – Achievable rate region by Slepian-Wolf theorem [3].

Slepian-Wolf coding is the term generally used to characterize coding architectures that follow the scenario described in Fig. 2. Slepian-Wolf coding is also referred in the literature as lossless distributed source coding since it considers that the two statistically dependent sequences independently encoded are reconstructed with an arbitrarily small error probability at a joint decoder (approaching the lossless case). Notice that in this context, lossless is different from mathematically lossless since a controlled amount of error is allowed. One interesting feature of Slepian-Wolf coding is the relationship that it has with channel coding. Due to sequence X being correlated with sequence Y, it can be considered a virtual "dependence channel" between sequences X and Y. The Y sequence is therefore a "noisy" or "erroneous" version of the X sequence. The "errors" between the X and Y sequences can be corrected applying a channel coding techniques to encode the X sequence. This relationship was studied in the 1970s by Wyner [4].

B. Wyner-Ziv Theorem

In 1976, A. Wyner and J. Ziv [5] studied a particular case of Slepian-Wolf coding corresponding to the rate point ($H(X|Y)$, $H(Y)$) identified in *Fig. 3* by the red dot. This particular case deals with the source coding of the X sequence considering that the Y sequence, known as side information, is available at the decoder. *Fig. 4* illustrates such scenario; in the literature, this case is known as lossy compression with decoder side information. The designation of lossy compression is due to Wyner and Ziv having considered an average, acceptable distortion d, between the sequence to be encoded, X, and its decoded version, X'.

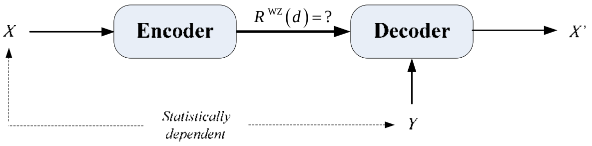

Fig. 4 – Lossy compression with decoder side information.

Wyner-Ziv coding is another designation used in the literature to characterize the situation depicted in *Fig. 4*. Considering *Fig. 4*, let X and Y be two statistically dependent i.i.d. random sequences where X is the sequence to be encoded, the so-called main information, and Y, the so-called side information, is considered available at the decoder; for the case, it is not relevant how this side information is made available to the decoder. As illustrated, there is no exploitation of the statistical dependency between X and Y at the encoder. Wyner and Ziv work established the minimum rate $R^{WZ}(d)$ necessary to encode X guaranteeing its reconstruction with an average distortion below d, assuming that the decoder has the side information Y available. The results obtained by Wyner and Ziv indicate that when the statistical dependency between X and Y is exploited only at the decoder, the transmission rate increases comparing to the case where the correlation is exploited both at the encoder and the decoder, for the same average distortion, d. This is precisely what the Wyner-Ziv theorem states [5]. To decode X in an efficient way, the decoder needs to be aware and exploit the correlation model between X and the side information Y.

Further, Zamir [6] has demonstrated that the increase in the rate is smaller than 0.5 bit/source symbol when compared to the rate in a joint encoding and decoding situation (encode of the sequence X exploiting the correlation with Y both at the encoder and decoder). This result was obtained for general statistics using a mean-squared error (MSE) to measure the reconstruction error at the decoder.

Wyner and Ziv showed, however, that there is no rate increase, for all $d > 0$, when X and Y are jointly Gaussian sequences and a mean-squared error distortion measure is considered [5]. This means that, in these conditions, there is no increase in the transmission rate if the statistical dependency between X and Y is only explored at the decoder compared to the case where it is explored both at the decoder and the encoder.

In conclusion, the Slepian-Wolf and the Wyner-Ziv theorems for long well-known in Information Theory suggest that it is possible to compress two statistically dependent signals in a distributed way (separate encoding, jointly decoding) using a rate similar to the one used in a system where the signals are encoded and decoded together, i.e. like in traditional video coding schemes. In traditional video coding schemes those signals may correspond, for instance, to the odd and even frames of a video sequence. This conclusion opens the doors for coding solutions where the encoder share in the overall complexity budget may be significantly reduced at the cost of increasing the decoder complexity share which is currently not possible.

III. DVC STATE-OF-THE-ART REVIEW

Theoretical foundations for distributed source coding have been established in the 1970s, as mentioned in Section II. However, practical efforts with the aim of approaching the Slepian-Wolf and Wyner-Ziv bounds started more recently driven by emerging applications where low encoding complexity is a major requirement.

Considering *Fig. 4*, it was assumed in Section II.*B* that the side information Y is available at the decoder; in this context, available means that a reliable reconstruction of Y is accessible to the decoder. From the Information Theory, encoding Y with a rate $R_Y \geq H(Y)$ provides such a reliable reconstruction. For example, Y may be provided to the decoder using a traditional coding solution for some of the frames in the video sequence such as the MPEG-x or H.26x standards. The dilemma would be then how to encode X in order to reach optimal performance in terms of rate, i.e. how to built a system capable of approaching the rate point $(H(X|Y), H(Y))$ (see Section II.*B*). To achieve this goal, well-suited techniques to encode X have been studied. Firstly, it was considered in the study of distributed source coding the use of quantization at the encoder. A source data observation is simply quantization-based encoded, and transmitted to a decoder. The decoder, with the help of an available uncoded source data observation correlated to the one encoded (side information Y), attempts to obtain the original source data observation, i.e. reconstructs the data.

The relationship between Slepian-Wolf coding and channel coding, previously studied in [4], has then stimulated the usage of channel coding techniques to encode the sequence X. Several channel codes were therefore tested in the context of DSC such as turbo codes, low density parity check codes (LDPC codes) as well as syndromes. The test results show that performing channel coding after quantization allows approaching the theoretical bounds (see Section II).

In a non-distributed coding scenario, transform coding is a widely used source coding technique along with quantization and entropy coding to reduce the transmission rate. Relying on this idea, the rate-distortion performance of DSC schemes that perform transform coding followed by quantization and channel coding has been evaluated; the Discrete Cosine Transform (DCT) was one of the transforms used. Generally, the results obtained corroborate the idea that performing transform coding before the quantization and the channel coding stages allows a reduction of the X sequence transmission rate [7]. Notice that the major part of the approaches that have been developed considered the scenario illustrated in *Fig. 4*, i.e. Wyner-Ziv coding.

The most important distributed video coding approaches recently present in the literature are described in the following.

- In 2002, an approach known as “Power-efficient, Robust, hIgh-compression, Syndrome-based Multimedia coding” or PRISM was proposed by Puri and Ramchandran [8] for multimedia transmissions on wireless networks using syndrome encoding. The PRISM solution aims to combine intraframe coding features (low-complexity encoding and robustness to transmission errors) with interframe coding compression efficiency. In [8], PRISM and the H.263+ standard are compared in terms of compression efficiency and robustness performance. The results show that PRISM solution outperforms H.263+ in robustness to transmission errors; however PRISM has still compression efficiency inferior to H.263+ when interframe encoding is performed.
- In the same year, making use of turbo codes, Aaron, Zhang and Girod [9] have shown video coding results using a pixel-domain intraframe encoding-interframe decoding scheme. The Wyner-Ziv solution proposed in [9] presents a rate-distortion performance better than H.263+ intraframe coding; however, comparing to H.263+ interframe coding the rate-distortion performance is still inferior. This solution was later improved by adding a DCT transform to the encoder and transmitting additional information to help the decoder in the motion estimation task. The most recent solution is presented in [10] and outperforms H.263+ intraframe coding up to 8 dB; however, a gap in performance is still observed in comparison to H.263+ interframe coding.

IV. APPLICATIONS FOR DVC

As was seen is previous sections, distributed video coding is a new coding paradigm described by a configuration where the encoder has low-complexity at the expense of a higher decoder complexity. This configuration makes DVC a promising coding solution for emerging applications, where the encoder complexity or the power consumption are scarce resources, such as those described in the following:

- **Wireless low-power surveillance network:** Nowadays, events sensing is almost present everywhere. One example of event sensing is the video surveillance scenario, illustrated in *Fig. 5*, where multiple cameras are sensing the same event from different locations.

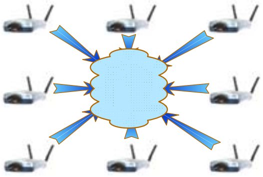

Fig. 5 – Video surveillance scenario.

Neighbouring cameras typically sense partially overlapped geographical areas and therefore their associated video sequences are correlated. Since the number of encoders is usually much higher than the number of existing decoders (typically one), it is possible to reduce the cost of the system if low-complexity encoders are used and if only one or a few more complex decoders are necessary. Wyner-Ziv coding is well-suited for this scenario, since enables to explore the correlation between the multiple encoded sequences just at the decoder, providing a low encoding complexity.

- **Wireless mobile video:** Another application that can benefit from the DVC paradigm is wireless video, e.g. wireless video communication between a pair of camera phones as illustrated in *Fig. 6*.

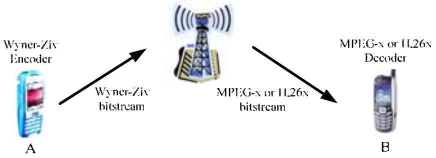

Fig. 6 – Wireless mobile video scenario.

The major requirement in this application is to have a low-complexity encoder and decoder in each terminal, since power consumption and battery life is closely related to the complexity of the encoder/decoder pair. However, to take advantage of Wyner-Ziv coding in a wireless mobile video scenario, it is necessary to have a high-complexity decoder device, as was seen previously. For this application, the high-complexity decoder is collocated at a base station along with a transcoder. This base station will be responsible for receiving the low-complexity encoded bitstream (known as Wyner-Ziv bitstream), transcode it to a MPEG-x or a H.26x bitstream and transmit it to another terminal with a low complexity decoder (terminal B in *Fig. 6*). This will enable to have a low-complexity encoder and decoder at each terminal.

- **Multi-view acquisition:** In several applications, a scene or an object is acquired from many cameras located at fixed spatial positions, e.g. for special movie effects, image-based rendering (3D reconstruction with texture mapping). One of the research challenges is to find the best algorithms and sensors to acquire the video sequences especially because a large number of cameras are necessary to fulfil the requirements, e.g. to have photo-realistic scenes or immersive 3D models. *Fig. 7* illustrates a large camera array scenario to acquire these scenes.

Fig. 7 – Large camera array scenario.

In such scenario, neighbouring cameras of a large camera array capture overlapped views and therefore views that are correlated. With Wyner-Ziv coding, independent encoding of each view can be performed in each camera while a central station has to perform joint decoding, in order to exploit the correlation between views. This will enable to have low-complexity encoders, and thus to use low-cost cameras, minimizing the total cost of the camera array.

- **Video-based sensor networks:** Sensor networks [11] are becoming a new field of research driven by advances in microelectronics and communications networks. The main goal is the development of technologies that would enable to employ thousand of sensors in a chosen environment to accomplish a certain task. If these sensors have video acquisition capabilities, several applications are possible such as tracking of persons throughout an environment, monitoring of activities, tracking events and signalling any alarms if necessary. Also by having a large number of sensors, multiple camera angles are available, making some computer vision tasks (e.g. gesture recognition) much easier than using a single view. Wyner-Ziv coding can help the construction of such video-based sensor networks, since it allows the construction of low-complexity, low-cost and low power consumption encoder devices. In this type of networks, the decoder is a central processing unit with high computational capabilities responsible to collect and process all the information received (namely to explore the redundancy between all the received video signals).

V. FINAL REMARKS

This paper presents a review of distributed video coding and discusses its importance for the emergence of several applications such as wireless low-power surveillance networks, wireless video, multi-view acquisition and video-based sensor networks. These applications are still largely conditioned on their deployment by the current high encoder complexity, low decoder complexity paradigm. DVC may allow getting out of this “trap”, bringing new video applications to life.

REFERENCES

[1] T. Wiegand, G. Sullivan, G. Bjøntegaard and A. Luthra, “Overview of the H.264/AVC Video Coding Standard”, *IEEE Trans. on CSVT*, vol. 13, no. 7, July 2003.

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

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

[4] A. Wyner, “Recent Results in the Shannon Theory”, *IEEE Trans. on Information Theory*, vol. 20, no. 1, January 1974.

[5] 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.

[6] R. Zamir, “The Rate Loss in the Wyner-Ziv Problem”, *IEEE Trans. on Information Theory*, vol. 42, no. 6, November 1996.

[7] A. Aaron, S. Rane, E. Setton and B. Girod, “Transform-Domain Wyner-Ziv Codec for Video”, VCIP, San Jose, USA, January 2004.

[8] R. Puri and K. Ramchandran, “PRISM: A New Robust Video Coding Architecture Based on Distributed Compression Principles”, *40th Allerton Conference on Communication, Control and Computing*, Allerton, USA, October 2002.

[9] A. Aaron, R. Zhang and B. Girod, “Wyner-Ziv Coding for Motion Video”, *Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, USA, November 2002.

[10] A. Aaron, S. Rane and B. Girod, “Wyner-Ziv Video Coding with Hash-Based Motion Compensation at the Receiver”, *IEEE ICIP*, Singapore, October 2004.

[11] Z. Xiong, A. Liveris and S. Cheng, “Distributed Source Coding for Sensor Networks”, *IEEE Signal Processing Magazine*, vol. 21, no. 5, September 2004.