IWSSIP 2010 - 17th International Conference on Systems, Signals and Image Processing

A Proposed Framework for Implementation of Distributed Video Coding

Marco Antonio Grivet M. Maia, Rodrigo S. Mello
Centro de Estudos em Telecomunicações (CETUC)
Pontifícia Universidade Católica do Rio de Janeiro
Rio de Janeiro, Brasil
mgrivet@cetuc.puc-rio.br, rodrigosm@cetuc.puc-rio.br

Clayton Escouper das Chagas
Grupo de Guerra Eletrônica (GGE)
Centro Tecnológico do Exército (CTEx)
Rio de Janeiro, Brasil
escouper@ctex.eb.br

Abstract – Distributed Video Coding (DVC) is a coding paradigm based on the concepts of Distributed Source Coding (DSC), which is based on Information Theory developed by Shannon, for a scenario where we have a distributed encoding of information and a joint decoding. Implementations with different architectures have been presented over the past years, but due to various reasons such as unavailability of detailed documentation, lack of concern of the projects in relation to software engineering, not uniform and heterogeneous implementation technologies, among others, the development of this area ended up being hampered by a lack of tools and documentation more applied, preventing the researcher from focusing their studies and research only on the party seeking to enhance or supplement the project. This paper applies concepts and theories of software engineering, object oriented, components, frameworks and other, in order to design, implement, document and test an open framework, incremental and reusable implementation of the boxes and pieces of a codec architecture for Distributed Video Coding, submitted in a collaborative environment so that it can be used for studies and the contributions developed in the future can be aggregated to the framework with little coding effort.

Keywords – Distributed Video Coding (DVC); Distributed Source Coding (DSC); Information Theory; Slepian-Wolf Theorem; Wyner-Ziv Theorem; Side Information.

I. INTRODUCTION

With the development of new devices that take on board modern technology and infrastructure communications that support these innovations, the information passed by the ubiquity as one of its main features, as it can be accessed and transmitted from anywhere. This made the traditional data transmission/reception architectures search for new patterns so that their operations adapt to this scenario.

Regarding the transmission of video, due to the inherent characteristics of this type of information, most traditional architectures are asymmetric, with video encoding performed by a machine of greater complexity and, after the transmission and reception, decoding is performed on a client device with less complex operations. This is reasonable from the moment that you have a scenario in which the production of content, encoding and transmission are made by companies with the ability to buy equipment with great processing power and ensure their customer service at low cost access to this content, through a device with a built cheaply and with little processing power. This scenario can be exemplified by the relationship that has a system for transmission of digital terrestrial television, considering the technical characteristics of the television companies and viewers.

Unlike what happens in the scenario described, new technologies focused on mobility and transmission primarily through the air, meant that there was need for a paradigm shift in relation to the technical complexity and cost of the devices used. The hardware has become simpler, since we now have devices with different requirements on the production of content (where we have the encoder) in relation to mobility, weight, processing power, battery life and other important characteristics.

However, the complexity of the side that makes the decoding was increased, requiring more expensive equipment and more processing power, able to cope with higher algorithmic complexity.

This change in the architecture of video systems has brought the need to develop new models of codec, since the best-known models were not suitable for the new topology. This new family of implementations is included in the concept of DVC.

This paper presents a mathematical theory that opened the possibility to achieve the performance target and the design of this new architecture, which is done in sec. II. The classical architectures that have implemented this new paradigm are presented in sec. III. Sec. IV is proposed modifications to the classical schemes, giving rise to a new architecture for the DVC and showing the topology and organization of an open and collaborative framework, object-oriented and its computational implementation. Sec. V describes some applications of DVC. Sec. VI presents the simulation results with the proposed architecture and comparisons are made with other codecs. Finally, conclusions and suggestions for future works are discussed in sec. VII.

II. THE THEORY BEHIND THE ARCHITECTURE

The architecture of some modern communication systems that deal with video data requires a paradigm shift in traditional coding, moving the side with greater computational complexity of the encoder to the decoder. This result can be achieved with the implementation of two important results of Information Theory, which operate in the decoder the statistical dependence of sources independently coded. The first study that gave rise to this new branch of Information Theory, called Distributed Source Coding (DSC) was done by Slepian and Wolf and published in 1973 [1] leading to the theorem of Slepian-Wolf. This goes beyond the basic theorem of Information Theory of Shannon [2] for encoding two related sources, independently coded and exploits this correlation in the decoder, allowing reconstruction without loss if they met the following restrictions:

- The statistical correlation between the sources must be known at the decoder
- $R_X \geq H(X|Y)$
- $R_Y \geq H(Y|X)$
- $(R_X + R_Y) \geq H(X,Y)$

Where X and Y are discrete sources correlated, R_X and R_Y are the transmission rates of sources X and Y, $H(X|Y)$ is the entropy of X given knowledge of Y, $H(Y|X)$ is the entropy of Y given knowledge of X and $H(X,Y)$ is the joint entropy of sources X and Y.

In 1976, Wyner and Ziv expanded the results of the theorem of Slepian-Wolf for lossy coding [3], so the result was widespread, being the theorem of Slepian-Wolf a particular case of the theorem of Wyner-Ziv where the distortion is zero.

Wyner and Ziv showed that, given a rate-distortion function $R_X(D)$ associated with the encoding and decoding of the source X, and a function $R_{X|Y}(D)$ associated with encoding and decoding of X given side information Y, there is a rate-distortion function $R_{WZ}(D)$, called rate-distortion function of Wyner-Ziv, such that:

- $R_{WZ}(D) \geq R_{X|Y}(D)$ and
- $R_{WZ}(D) \leq R_X(D)$

III. STATE OF THE ART IN ARCHITECTURE DVC

The first two DVC architectures, which were the basis for all research and developments, were the frameworks developed by Stanford University and the University of California Berkeley.

Researchers at Stanford University presented a first draft in 2002 [4] which was extended and improved in 2004 [5] and 2005 [6]. In this approach, frames coded according to the method of Wyner-Ziv, called Wyner-Ziv frames, undergo a transformation and its result is quantized and divided into units called bitplanes. One of the peculiar characteristics of this implementation is the use of Turbo codes in channel coding. The so-called key frames, i.e. those who are not Wyner-Ziv, are encoded and decoded by a conventional intraframe codec, and used in the DVC decoder to generate the side information used in the reconstruction of the Wyner-Ziv frames, as shown in Fig 1, which presents the details and information flow of this architecture.

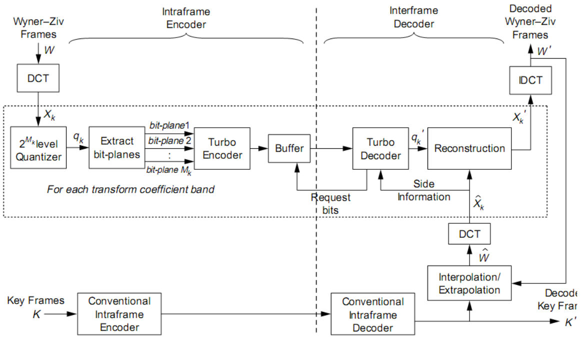
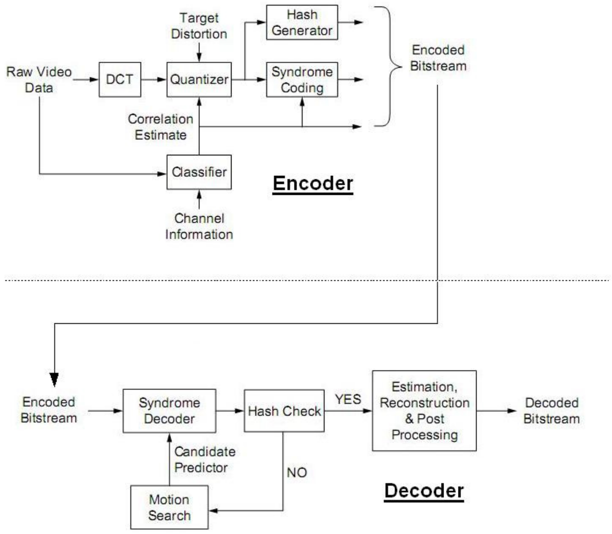


Figure 1. Architecture of DVC codec Stanford University

The University of California Berkeley proposed an architecture known as PRISM [7, 8], acronym of Power-efficient, Robust, High-compression, Syndrome-basing Multimedia coding. One of the peculiarities of PRISM is that it does not require a return channel to reconstruct the Wyner-Ziv frames, as shown in Fig 2.

Figure 2. Architecture of DVC codec PRISM



From these two architectures, new projects with different techniques optimized were published, however, the DVC is still facing the problem of having a lower performance than the conventional encoders, therefore, applicability being open to new ideas, which still promotes a lot of research in the area of video encoding.

IV. ARCHITECTURE OF THE PROPOSED FRAMEWORK

In this paper, we present an architecture of a distributed encoder WZ-LI, i.e., based on Wyner-Ziv encoder with channel coding based on irregular LDPC (Low Density Parity-Check Codes) technique for the generation of optimized side information, as shown in Fig 3.

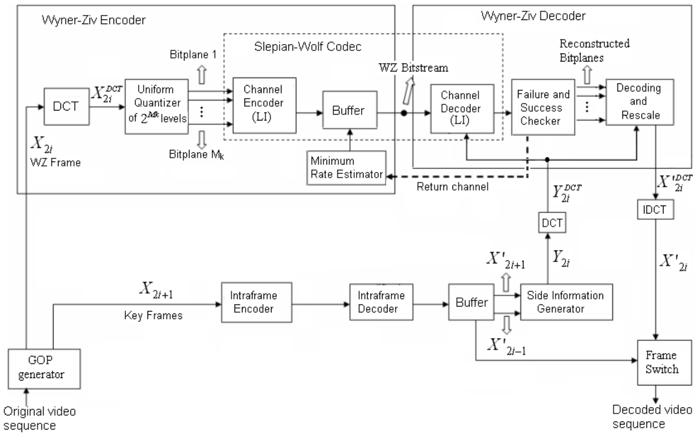


Figure 3. Architecture of the proposed codec DVC WZ-LI

The distributed codec scheme described in this project was based on [9], with the addition of new techniques and improvements.

The process of encoding and decoding begins with the division of the video sequence in Wyner-Ziv frames (even frames) and key frames (odd frames). Then, apply the DCT transform in the 4x4 blocks of the Wyner-Ziv frames. The resulting transform coefficients are grouped according to their positions in the 4x4 block, resulting in 4^2 bands of DCT coefficients. After that, each band k is uniformly quantized with 2^{M_k} levels, where M_k is the number of bits needed to encode the band k. From these quantized coefficients are extracted the bitplanes that feed the encoder LIA, generating their syndromes, which are sent on demand to the decoder as a request made through the return channel.

As shown in Fig 3, the encoder is simple, with low computational complexity, requiring the most basic requirements in relation to energy consumption and autonomy of the equipment.

In the decoder, the decoding process starts running the generation process of side information, as shown in figure 4, which is composed by a bidirectional motion estimation and compensation from the key frames (the adjacent frames before and after the WZ current frame), followed by two interpolation steps: a weighted interpolation of the motion compensated frames based on their correlation measure; and a weighted interpolation based on the sequence movement level. Finally, it is executed a pixel extrapolation of the resulting frame, in order to fill the zeroed pixels (black pixels), producing the estimate of the WZ frame, that after going through a DCT transform, helps in the reconstruction of the original Wyner-Ziv frame. To achieve this goal, the bitplane is decoded by the decoder LI, receiving as input data the syndrome sent by the encoder, the side information (estimate of WZ frame) and the virtual channel characteristics, i.e., the correlation between the estimate and WZ frame, which performs an iterative process of error checking. Then, from the bitplane rebuilt, it leads to the bands of DCT quantized coefficients, and after the rescale step, it provides the DCT coefficients. Hence, performing an inverse DCT, we obtain the rebuilt WZ frame.

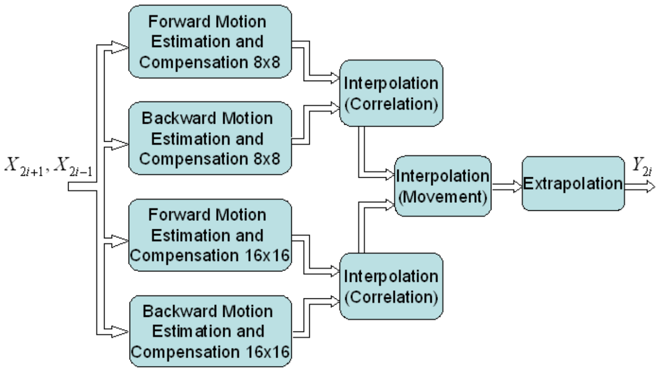


Figure 4. Proposed side information generation scheme

V. APPLICATIONS

The topologies of equipment, mobility and ubiquity of new networks were the main instigators of the paradigm shift in relation to the architecture of video codecs. New requirements in relation to these aspects with the needs of autonomy and low power consumption have made new surveys were developed to seek more efficient solutions that met this demand. In this scenario, the DVC comes as a solution, whose applications include those described below.

A. *Sensor networks with video support*

With the advancement of research in sensor networks, many of its areas can be improved if the network elements aggregate the capacity of visual monitoring. These desirable features should be aggregated, but cannot conflict with the basic requirements of a network that can have hundreds or thousands of elements, all with low complexity of construction, low cost and low power consumption. The shift of the computational complexity of the encoder to the decoder in the DVC architecture makes it a great solution to the problem, because the data acquired by sensors are processed by an apparatus with high computing power in the central of processing, consolidating data from network elements.

B. *Video calling in mobile service:*

Video calling in mobile service is other branch which is interesting that you can enjoy the possibilities of DVC, because each of the terminals normally have a device with low processing power, as a mobile phone, and so, this asymmetric system with low complexity in the encoder becomes interesting for who is doing the video calling, because the encoding algorithm is low complexity. Then the signal is sent to a central, which has more computing power and, after identifying the destination, you can send the call with a traditional encoding scheme as the MPEG-x or H.26x, where we have a complex encoder that will work in the plant, which has no problem, since our central support such complexity, and in the other device client end, which receive the video calling, the video is decoded with a simple algorithm of the same family MPEG-x or H.26x.

C. *Distributed cameras in vehicles and weapons on the battlefield*

Distributed cameras in vehicles and weapons on the battlefield a very interesting application in the military field is the use of DVC codec in military systems for use in command and control platform, where cameras with low encoder complexity should be used, since the topology of a field of combat is highly complex given the diversity and quantity of sources, coupled with the diffuse mobility of the elements of this network and the requirement of low power consumption, as the soldier can not carry a large amount of reserve batteries due to the extra weight this will have. Thus, all network components, such as a soldier with his rifle or camera on the helmet, tanks by ground and surveillance aircraft send images to be merged and processed in a central command and control with high processing power at the rear of the battalions, where strategists use the images from different sources and many different angles to make decisions in the battlefield.

VI. RESULTS

The quality of the side information influences significantly the performance of Wyner-Ziv codec. With more sophisticated techniques than the copy of the previous frame or just an interpolation of adjacent frames, the proposed method makes use of tools for interpolation and extrapolation of motion compensated, to generate a side information for high quality, i.e., making the estimate of the WZ frame the closest possible to the original frame, achieving thus a higher efficiency of the process. Fig. 5 (a) shows the estimate of a WZ frame sequence of Foreman, obtained by the proposed method, the PSNR is 37.6 dB. Fig. 5 (b) shows the estimate using only the interpolation of adjacent frames, with PSNR of 35.1 dB. Fig. 5 (c) shows the original frame. Besides the large difference in PSNR, observing the images, there are also great distortions in the image using only the interpolation, which is not in the picture as to the method proposed.

(a) (b) (c)

Figure 5. (a) Estimates obtained by the proposed method. (b) Estimates obtained by simple interpolation. (c) Original frame.

Besides, the fig. 6 presents the PSNR along the 100 frames of the sequence Foreman, QCIF@15Hz, with GOP of 2 and a rate of 400 kbps. As observed in fig. 6, there are peaks in regular intervals that correspond to intra-coded frames (key frames), which have more PSNR values, as expected. The average PSNR is 40.6 dB with a standard deviation of 1.8 dB.

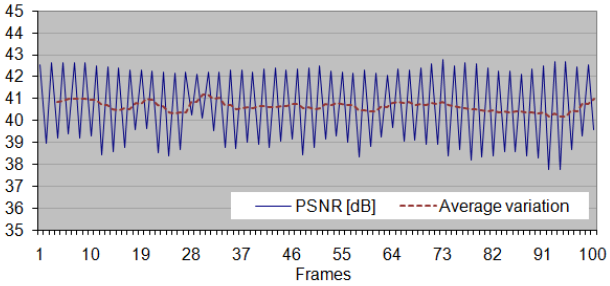


Figure 6. PSNR along the Foreman sequence

VII. CONCLUSIONS AND FUTURE WORKS

With the need for paradigm shift in traditional video codecs, where the computational complexity was on the side of the encoder, the Distributed Video Coding (DVC) came to attend a number of emerging applications that have special needs, spending most computational load to the decoder.

Thus, this paper presented a framework based on Wyner-Ziv codec using an irregular LDPC in channel coding and optimizing the technique of generating the side information, implemented it in an open and collaborative framework, incremental through new features planned that can be easily aggregated to the current library, given its object-oriented architecture based on components.

It aims to provide the future framework, which will be published in the academic and scientific for students and researchers can easily explore and implement new functionality to it, making it a reference to teaching how to do DVC, with comprehensive documentation, facilitating early access to technology.

REFERENCES

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

[2] C.E. Shannon, “A Mathematical Theory of Communication”, Bell System Technical Journal, vol. 27, pp. 379-423, 623-656, July, October, 1948.

[3] 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.

[4] A. Aaron, R. Zhang and B. Girod, “Wyner-Ziv Coding of Motion Video”, in Proc. Asilomar Conference on Signals and Systems, Nov. 2002.

[5] A. Aaron, S. Rane, E. Setton, and B. Girod, “Transform-Domain Wyner-Ziv codec for video,” in Proc. SPIE Visual Communications and Image Processing, Jan. 2004.

[6] A. Aaron, D. Varodayan, B. Girod, “Rate-adaptative Distributed Source Coding using low-density parity-checks codes”, in Proc. Asilomar Conference on Signals, Systems and Computting, Nov 2005.

[7] R. Puri and K. Ramchandran, “PRISM: A New Robust Video Coding Architecture Based on Distributed Compression Principles”, in Proc. Allerton Conference on Communication, Control and Computing, Oct 2002.

[8] R. Puri, A. Majumdar and K. Ramchandran, “PRISM: A reversed multimedia coding paradigm”, in Proc. IEEE International Conference on Image Processing, Set 2003.

[9] B. Girod, A. Aaron, S. Rane and D. Rebollo-Monedero, “Distributed Video Coding”, in Proc. Of the IEEE, vol. 93, no. 1, Jan 2005.