# ADAPTIVE KEY FRAME RATE ALLOCATION FOR DISTRIBUTED VIDEO CODING

*Ugo Ciracì, Marco Dalai and Riccardo Leonardi*

Department of Electronics for Automation
University of Brescia, Italy

## ABSTRACT

In the context of Distributed Video Coding (DVC), rate allocation among sources represents an important problem to solve. While in the information theoretical setting of Distributed Source Coding (DSC) the statistical correlation between sources is usually assumed to be known, in practical DVC systems there is no a priori knowledge of the underlying statistics of visual data. This lack of information makes it difficult to deal with the problem of rate allocation in practical DVC codecs. In this paper we focus on the problem of how to distribute the rate between differently encoded parts of the video sequence in a DVC system. Namely, we propose an adaptive rate allocation scheme for the encoding of the key frames depending on an estimation of the local motion activity of the sequence.

***Index Terms***— *Distributed video coding, rate allocation, motion compensated interpolation.*

## 1. INTRODUCTION

Distributed Video Coding (DVC) systems have been studied intensively by many research groups in recent years (see [3,4,5] and references therein). DVC is a video coding paradigm inspired by the theory of Distributed Source Coding (DSC), initiated in the ‘70s with the famous work by Slepian and Wolf [1] and with the following work by Wyner and Ziv on the rate distortion function with side information available at the decoder [2]. The most significant difference between the practice of DVC and the theory of DSC is the knowledge of the joint statistics of the involved sources. In particular, in DSC it is always assumed that the statistical properties of the sources at hand are completely known, and consequently the amount of rate required to compress a source within some given distortion can be perfectly computed by an encoder. On the contrary, in practical DVC systems, the statistics of the sources are not usually known a priori and it is then difficult to estimate the rate required in order to have a successful decoding of the compressed data (see [4,5]; see also [6]). In particular, one of the most crucial tasks for DVC systems is the process of estimation of the correlation between different frames that is needed in order to properly allocate rate for the Wyner-Ziv encoding of data. This problem arises in all major schemes for DVC in the literature [4,5]. In this paper we concentrate on the scheme proposed by Stanford’s group[1] [5].

In the cited DVC system, the video sequence frames are separated in two groups, namely a group of *key-frames*, which are encoded in a classical fashion using a standard picture or video codec and a group of so called *Wyner-Ziv frames*, which are encoded with a distributed approach. An effective encoding of the WZ frames basically consists in applying a DCT-like transform, then quantizing the obtained coefficients, and then computing a *syndrome* of these coefficients to be sent to the decoder. A *syndrome* is actually constructed by applying a channel encoding process (with a systematic code) to a bitplane representation of the transformation coefficients and then transmitting a fraction of the obtained parity bits. These bits represent the Wyner-Ziv code. At the decoder, a Motion Compensated Interpolation (MCI) between the key frames is performed in order to construct an approximation of the WZ frames, which represents the Side Information (SI). The parity bits received from the encoder are then used in order to “correct” the transformed SI and recover at the decoder a better approximation of WZ frame.

In this context the problem of rate allocation is twofold. The first problem encountered is that the rate allocation for the syndrome encoding of a WZ frame is a very difficult task. In general, it is not easy to establish the number of parity bits needed by the decoder to “correct” the SI. For this reason the scheme proposed in [5] used a feedback channel in order to allow the decoder to ask the encoder for more and more parity bits as long as an acceptable decoding can be achieved. On another hand, in a practical implementation of such a codec, one is faced with a second problem of allocation. Namely, it is necessary to decide how to distribute the rate between key frames and WZ frames. In the literature, this problem has not been explicitly discussed up to now, and sequence-dependent ad-hoc rate allocation has been used for the keyframes.

In this paper we propose a first approach to the design of an adaptive scheme of rate allocation to the keyframes. Based on the analysis of the performance of different blocks of a DVC scheme we propose an adaptive rate allocation for the key frames depending on the motion activity and on the desired average distortion of the reconstructed frames.

---

[1] The work presented was developed within DISCOVER project and the respective software started from IST_WZ software developed at the Image Group from Instituto Superior Tecnico (IST).

## 2. PROBLEM DESCRIPTION

In this section we give a more detailed description of the considered system and we explain in detail how the problem of rate allocation is distributed on each processing block. In Figure 1 a conceptual representation of the DVC codec is shown.

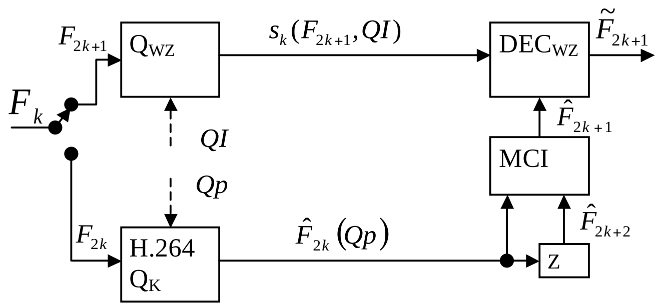


**Figure 1. The DVC system architecture.**

As already anticipated, input frames $F_k$ are divided in two subsets. Odd indexed frames $F_{2k+1}$ are encoded using Wyner-Ziv techniques, while even indexed frames $F_{2k}$, i.e. keyframes, are encoded in a traditional way using a standard video codec. Both key frames and WZ frames need to be quantized in order to allow lossy compression, and two parameters, respectively Qp (quantization increases as Qp increases) and QI (quantization increases as QI decreases), are thus used for the tuning of the quantizers.

The problem we are interested in here is how to set the values of the two quantization indices in order to effectively distribute the rate between WZ frames and keyframes. This problem has not yet been studied in the literature. What is usually implicitely assumed is that for a given target quality of the reconstructed video sequence one is able to establish a pair of fixed values of QI and Qp which are to be used for the encoding of the entire sequence. It is however important to point out that the choice of QI and Qp are not sequence-independent. More specifically, for a given quantization level QI on the Wyner-Ziv frames, a different value of Qp may be needed depending on the type of sequence at hand. This is due to the fact that, depending on the type of sequence, the quality of the key frames directly impact the quality of the reconstructed WZ frames. The dependence between the values of QI and Qp is not known and ad-hoc values are used for every tested sequence. In this paper we provide a first approach to an adaptive bitrate allocation between key frames and WZ frames.

## 3. PERFORMANCE MEASURES

### 3.1. Side information and key frames

As we have already made clear the reconstruction quality of the WZ frame at the decoder side is strongly affected by the quality of the SI frame, which is itself obtained by motion compensated interpolation between key-frame pairs. Since correlation between two consecutive key frames affects the quality of the SI, it is important to estimate the quality of the SI generated at the decoder, in order to better allocate the available bandwidth at the encoder side. An estimate of the motion content can thus be used in order to guess the SI quality. The typical low-complexity requirement of DVC systems [3] imposes to design motion low complexity indicators. We simply adopte a frame difference measure $d_M = \|F_{2k} - F_{2k+2}\|_2^2$.

Figure 3 collects a statistics about the quality of the SI frames versus the average quality of keyframe pairs that produce this SI. The motion content range can be partitioned achieving several "*channels*". Each channel contains all curves whose keyframe pairs have a confined motion content: a mean and a standard deviation curve can be associated to each channel. It is interesting to note that there is a saturation of the SI quality for a certain key frame quality, and that the associated threshold tends to decrease as $d_M$ increases.

A typical performance objective for video coding system is an almost constant, or smoothly varying, quality along the sequence. The described saturation effect has thus an important impact on the coding efficiency. For fast motion frames, for example, it is not possible to generate a high quality SI at the decoder, no matter how best the key frame quality. Furthermore, it can be observed that if the quality of the SI is much lower than the target quality for the WZ frame, the WZ encoding becomes far less efficient than any low complexity non predictive encoding procedure. This implies that for high motion frames, when the quality of the key frames increases, in order to achieve the same quality on the WZ frame, a huge amount of rate must be allocated for the WZ syndrome encoding.

Based on the previous considerations, we propose to adaptively adjust the Qp and QI values, in order to properly allocate the rate between key-frames and WZ frames, based on the motion content in the sequence.

### 3.2. MCI and quantization

As already anticipated, the SI is generated at the decoder by applying a process of MCI to a pair of key frames. In this section we are interested to characterize the error between the SI and the original WZ frame. We decide to model the "noise" affecting the SI as composed of two components, namely the one resulting from an inexact compensated interpolation on the key frames, and the one resulting from the quantization of the key frames. In order to better understand what happens, it is useful to consider how MCI operates. The MCI function is composed of two simpler operations: a) the motion estimation and b) the proper interpolation along the motion vectors. The quantization on the key frames clearly affects both steps of the procedure. However, it can be shown here that the effect of the quantization on the first step, i.e. the motion estimation, can be neglected as long as the compensation is performed on quantized keyframes.

Let thus $V$ be the motion field calculated by block matching on the quantized key frames at the decoder side. Let then $V^*$ be the motion field estimated on the lossless keyframes. This motion field represents the result of an ideal motion estimation, and can be thus used as a reference for comparing it with the motion estimation performed on the quantized key frames.

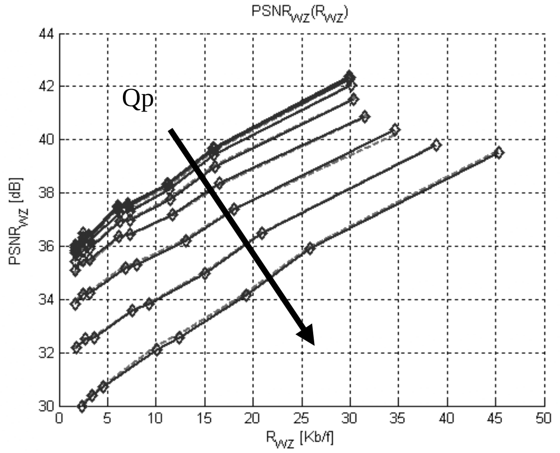


**Figure 2. Performance comparison between motion compensated interpolation using ideal (dashed lines) or non ideal (continuous lines) motion estimation, for varying key frame quality levels.**

We can then try to motion-compensate the lossy key frames using $V^*$ in place of $V$, i.e, we generate an SI by compensating the lossy keyframes along the ideal motion vectors specified by $V^*$. Doing so, the quantization adds distortion to the SI only through the compensation step, which uses quantized information. It is interesting to note in Figure 2 that there is no significant difference in using the motion field $V^*$ instead of $V$ for the compensation. The SI generated by motion estimation on lossy key frames obtains practically the same quality in both cases for the reconstructed WZ frames. This result may seem counterintuitive. However, it is important to consider that when the frames used for the compensation are lossy, most of the details which would need fine motion accuracy are not present. This gives a practical evidence in support of the fact that we can indeed consider the correlation noise between the SI and the original WZ frames as composed of two mainly independent components: the one resulting from motion and the one resulting from quantization.

## 4. ADAPTIVE RATE ALLOCATION SCHEMES

In this section we start with a brief analysis of the SI quality with respect to the key frames quality. We describe the criteria can be used to exploit motion content information in a practical DVC system to adaptively allocate the rate between keyframes and WZ frames.

### 4.1. Motion content channel separation

We have already pointed out that, for a fixed WZ frames quality, motion content will affect the achievable SI quality. Referring to Figure 3, if we properly define the number and extent of the motion content channels, different almost non overlapping (in terms of dispersion) regions exist, so that, for a given motion, the decoder can determine the maximum quality improvement it can achieve for the SI frames. This way, it can guess how many parity bits (i.e. the QI parameter) are required for a correct recovery of the current WZ frame.

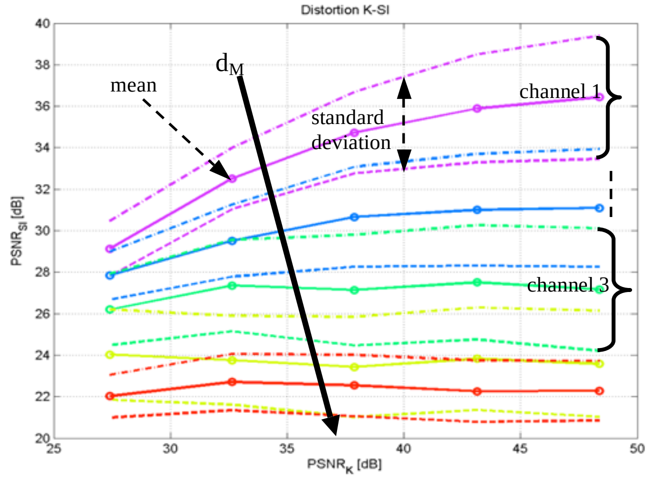


**Figure 3. Statistical characterization of the SI quality with respect to the quality of the keyframe pairs used in MCI.**

By setting a global quality level QL for the coded sequence, we have to set the values of Qp, and thus the values of QI, that lead to this quality depending on the moton content of the interested keyframes.

### 4.2 Bitrate switching schemes

We propose now a simple bitrate allocation switching scheme that chooses what $(Qp, QI)$ pair the encoder has to employ to properly allocate the rate between key frames and WZ frames once the motion content channel is known. For example, three uniform motion content channels $ch_M$ could be enough to represent High, Medium and Low motion range, so that we have the following bitrate allocation strategy:

| $ch_M$ | $QL_1$ | … | $QL_N$ |
|---|---|---|---|
| H | $(Qp,QI)_1^H$ | … | $(Qp,QI)_N^H$ |
| M | $(Qp,QI)_1^M$ | … | $(Qp,QI)_N^M$ |
| L | $(Qp,QI)_1^L$ | … | $(Qp,QI)_N^L$ |

where $QL_i$ establishes a global Quality Level for the output sequence and $(Qp,QI)_i^{ch_M}$ is the quantization parameters pair chosen by the encoder when the motion content is classified as belonging to the $ch_M$ channel.

We can resume the procedure in few steps: a) QL is defined first; b) the encoder determines then the motion content between two adjacent key frames at any given time; c) consequently a proper motion content channel is selected; d) the respective quantization parameters are identified leading to the proper bitrate allocation between WZ frames and key frames.

The choice of $(Qp,QI)_i^{ch_M}$ pairs to fill the bitrate switching scheme depends on what requirements are specified for the output sequence. For example, if we would like a temporally smoothly varying quality for the output sequence, the $(Qp,QI)_i^{ch_M}$ pairs have to be set accordingly, possibly, by lowering both key frame quality in presence of high motion and conversely when there is less motion, nonetheless targeting an overall QL quality level.

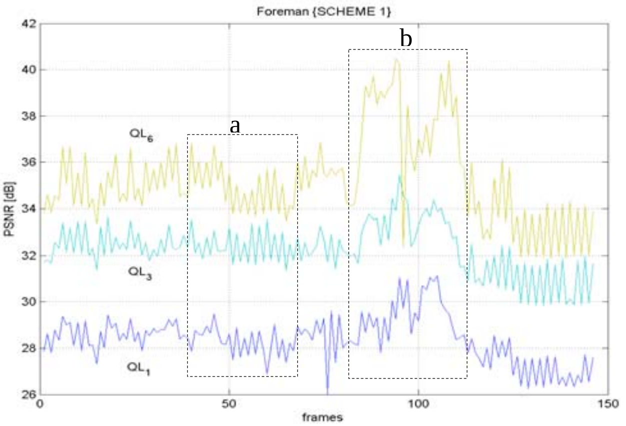


**Figure 4. Example of the quality of the decoded frames as a function of time: a) low/medium motion content; b) high motion content.**

In Figure 4 the temporal behaviour of a decided sequence is shown. We observe a quite constant quality when low/medium motion content occurs and instability in case of high motion content. In this case the SI frames achieve a too low quality to obtain a reliable adaptation of the WZ rate.

## 5. PRELIMINARY EXPERIMENTAL RESULTS

In this section some preliminary experimental results are presented in order to show the potential advantages a DVC architecture can achieve using an adaptive bitrate allocation scheme.

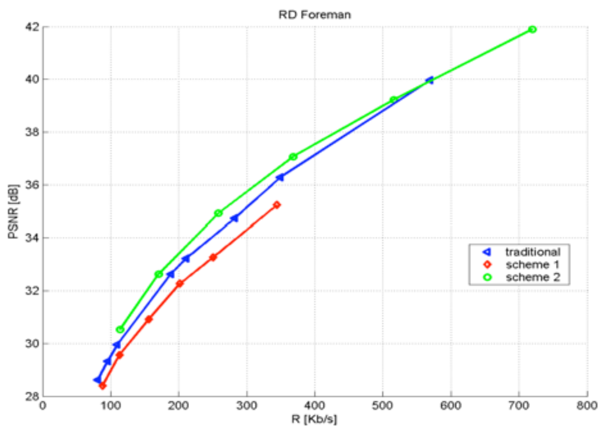


**Figure 5. Performance comparison between the traditional architecture and the udgraded architecture tuned with two different bitrate allocation schemes.**

In Figure 5 we plot the rate-distortion performances with two different schemes tested on the *foreman* sequence. Both schemes uniformely split the range of achievable motion contents in three channels (H, M and L channels) but different quantization parameter pairs are selected:

| Scheme 1 | | | | | | |
|---|---|---|---|---|---|---|
| $ch_M$ | $QL_1$ | $QL_2$ | $QL_3$ | $QL_4$ | $QL_5$ | $QL_6$ |
| H | (42,1) | (40,2) | (39,3) | (36,4) | (35,5) | (33,6) |
| M | (40,2) | (39,3) | (36,4) | (35,5) | (33,6) | (31,7) |
| L | (39,3) | (36,4) | (35,5) | (33,6) | (31,7) | (26,8) |

In this scheme the $(Qp, QI)_i^{ch_M}$ pairs select key frame rates to achieve the maximum SI quality just before it saturates (Figure 3). WZ parity bits are chosen to achieve a WZ frame quality similar to the key frame quality (Figure 4).

| Scheme 2 | | | | | | |
|---|---|---|---|---|---|---|
| $ch_M$ | $QL_1$ | $QL_2$ | $QL_3$ | $QL_4$ | $QL_5$ | $QL_6$ |
| H | (36,1) | (32,2) | (28,3) | (24,4) | (20,5) | (16,6) |
| M | (36,2) | (32,3) | (28,4) | (24,5) | (20,6) | (16,7) |
| L | (36,3) | (32,4) | (28,5) | (24,6) | (20,7) | (16,8) |

In scheme 2 for each QL a certain Qp determines the target quality and QI is chosen to achieve this quality.
We can observe that varying the used scheme we can achieve higher or lower rate-distortion results than using ad-hoc selected parameters which are constant along the sequence. Thus strategy is needed an accurate design of the bitrate switching to successfully exploit the SI quality statistics with respect to the key frames quality variations.

## 6. CONCLUSION AND FUTURE WORK

In this paper we have presented some statistical properties of a DVC system architecture. Motion content together with the side information, the WZ frames quality and the key frames quality have been examined in detail to characterize the codec behavior.
The result is that we can find a number of non overlapping motion content channels such that, with a good confidence, the encoder can dynamically - i.e. during the encoding process - selects $(Qp, QI)$ pairs that ensure a correct bitrate allocation between WZ frames and key frames.
In the future, more efficient schemes will be explored to better estimate statistics of visual content. Furthermore, we will propose other strategies for a better design of the bitrate allocation schemes using nonuniform motion content partitioning.

## REFERENCES

[1] D. Slepian, J. Wolf. Noiseless Coding of Correlated Information Sources. *IEEE Transaction On Information Theory, Vol.IT-19, NO.4*, July 1973.
[2] D. Wyner, J. Ziv. The Rate-Distortion Function for Source Coding with Side Information at the Decoder. *IEEE Transaction On Information Theory, Vol.IT-22, NO.1*, January 1976.
[3] Girod, A. Aaron, S. Rane, D. Rebollo-Monedero. Distributed Video Coding.
[4] R. Puri, and K. Ramchandran. PRISM: A new robust video coding architecture based on distributed compression principles. In *Proc.Of 40th Allerton Conf. on Comm., Controland Comp*., Monticello, 2002.
[5] Aaron, R. Zhangand B. Girod. Wyner-Ziv coding for motion video. *Asilomar Conference on Signals, Systems and Computers, Pacific Groove*, USA, 2002.
[6] João Ascenso, Catarina Brites, Fernando Pereira. Content Adaptive Wyner-Ziv Video Coding Driven By Motion Activity.