Signal Processing: *Image Communication* 28 (2013) 553–568

Contents lists available at SciVerse ScienceDirect

Signal Processing: *Image Communication*

journal homepage: www.elsevier.com/locate/image

Distributed video coding based on lossy syndromes generated in hybrid pixel/transform domain

Simone Milani *, Giancarlo Calvagno

Department of Information Engineering, University of Padova, via Gradenigo 6/B, 35131 Padova, Italy

ARTICLE INFO

Article history:
Received 16 July 2012
Received in revised form
21 March 2013
Accepted 25 March 2013
Available online 10 April 2013

Keywords:
Distributed video coding
Lossy syndromes
H.264/AVC
Hybrid domain
Low-complexity

ABSTRACT

A recently proposed class of distributed source coding based video coders enables low-complexity compression and robust transmission over unreliable channels. These architectures process the video signal either in the pixel or in the transform domain generating some side information that permits a correct decoding of the coded image from a set of possible correlated sources. The approach proposed in this paper processes the video sequence both in the pixel and in the transform domain exploiting the advantages of both schemes and generating a set of lossy syndromes. The resulting video coding scheme requires a lower computational complexity at the decoder with respect to their transform-domain counterparts (like DISCOVER or PRISM) and provides a high compression gain and an increased robustness against channel losses.

1. Introduction

The technical literature of the last years has assisted the proposals of new video coding schemes which aim at satisfying the requirements of low encoding complexity and robustness to channel losses [1]. Many of them adopt the Distributed Video Coding (DVC) framework, which relies on the information-theoretic concept of Distributed Source Coding (DSC) [2,3]. In particular, they use the principles of lossy distributed compression (also known as source coding with side-information or Wyner–Ziv coding), which imply the assumption of perfect knowledge of the statistical correlation structure between the source and side-information. In most of the video coding approaches, the source is a block of pixels from the frame to be coded, the side-information is a predictor block from a previously decoded frame, and the knowledge of correlation corresponds to knowledge of statistics for the Displaced Frame Difference (DFD) between the current block and its possible predictors. Under this assumption, motion search can be theoretically transferred from encoder to decoder, reducing the encoding complexity without any loss in compression performance [4].

In Distributed Video Coding, the correlation statistics between the source and the side information are used to determine the amount of information about the source that must be encoded to guarantee successful decoding [1]. This information can be conceptualized as a “hash value”, “checksum”, or “error correcting information” of the source, and it can be represented via an integer value that identifies a set of source samples whose mutual correlation is strictly related to the strength of specified error correcting information. This strength can be measured via the amount of bits that characterize the checksum information, which are computed according to the correlation between the actual source and the related side information. Following the terminology used in some previous works (which was borrowed from linear codes theory), we will refer to this information with the term “syndrome”.

In this paper, we present a hybrid pixel/transform domain DVC scheme which characterizes the signal to be decoded via “lossy” syndromes. The rationale behind this idea is that correlation can be exploited more effectively in

* Corresponding author. Tel.: +39 049 8277641; fax: +39 049 8277699.
E-mail addresses: simone.milani@dei.unipd.it (S. Milani), calvagno@dei.unipd.it (G. Calvagno).

0923-5965/$ - see front matter
http://dx.doi.org/10.1016/j.image.2013.03.007

the pixel domain, while transform domain can be useful to achieve high compression gains. Throughout the paper, the adjective “lossy” refers to coding distortion derived from quantization whenever it is used in relation to syndromes, while it refers to channel distortion produced by packet losses whenever it is applied to wireless channels and transmission.

A second innovative element in the proposed architecture is the adoption of a low-cost Fast Motion Estimation (FME) strategy at the encoder to estimate the correlation level between temporally adjacent blocks and to compute syndrome strength. This choice makes the proposed codec different from a traditional pure DVC architecture and permits estimating correlation accurately without increasing significantly the computational complexity at the encoder (with respect to traditional DVC encoder).

It will be shown that the hybrid DVC approach permits obtaining an increased robustness against losses (typical of pixel domain DVC schemes) and a high compression gain (typical of transform domain DVC schemes) with the additional advantage of requiring a lower decoding complexity compared with other schemes existing in literature. The proposed coder permits improving PSNR of 1 dB (on average) with respect to the DISCOVER architecture [5] while reducing the computational complexity at the decoder by a factor of 10 with an increase of complexity at the encoder around 5%. In this way it is possible to mitigate the decoding complexity, which is still nowadays one of the major drawbacks concerning Distributed Video Coding.

In the following, Section 2 overviews some of the previous works proposing DVC approaches. Section 3 presents the general structure of the proposed scheme describing the coding and decoding processes, while Section 4 underlines the differences between our solution and the previous ones based on nested scalar quantization in the transform domain. Experimental results are reported in Section 5, and the conclusions are drawn in Section 6.

2. Related works

Different kinds of approaches have been proposed in technical literature, which process the video signals and compute syndromes in different domains [1]. In [6] Aaron et al. propose a coding scheme where the image blocks are transformed using a Discrete Cosine Transform (DCT), and parity bits are computed from the resulting coefficients using a Turbo encoder. From this scheme, several solutions have followed based on the same architecture (see [7]). In a similar way, the PRISM coder [8] analyzes the video signal in the DCT domain and generates the syndromes from the output levels of the adopted quantizers for transformed coefficients. In [9] Fowler et al. present an approach that generates hash information processing the image in the Discrete Wavelet Transform (DWT) domain, alike approaches [10,11]. In [12], this hash function is adaptively selected according to the characteristics of the input signal. A simpler scheme is proposed by Brites et al. in [13], where the error correcting information is computed in the pixel domain avoiding the huge number of transform operations required by the other solutions. Moreover, the signal in the pixel domain proves to be highly spatially correlated and is not affected by the chosen transform block size. These facts permit increasing the robustness of the whole scheme. Most of the proposed DVC schemes rely on the temporal correlation. However, some schemes exploits spatial correlation as well (see [14,15] as examples), while others exploit the availability of multiple side information in order to obtain a better rate-distortion performance or a more accurate reconstruction of the coded signal [16].

In addition, it is possible to classify the different approaches according to the method adopted for the generation of syndromes from the input source given the amount of correlation with the associated side information. Many of the works presented in literature use capacity-achieving channel codes to approach the Wyner–Ziv bound [17–20]. In these approaches, transform coefficients (obtained via 4×4 or 8×8 block transform) for a whole or a part of a frame are grouped together according to their spatial frequencies. Then, selecting the bit values at the same positions for all the coefficients within the same set, the coders creates different bitplanes (one for each bit position within the current coefficient group). A channel code is then applied to generate parity bits. At the decoder, the correlation noise model is used in order to reconstruct the coded data from the transmitted parity information. Such long block-length typically entails high decoding complexity. A lighter approach was presented in [8], where the coset partitioning is performed using a nested scalar quantizer. In this way it is possible to achieve high coding gains with a limited complexity in syndrome decoding (see [21]).

In recent works, the DVC paradigm has been applied to different kinds of video encoders and applications [22]. As an example, lossy syndromes have been proposed as a robust solution to characterize the prediction residual error after Motion Estimation (ME) in Multiple Description schemes (see [23]) since they permit mitigating the distortion propagation. In the following, the paper presents a novel DVC coder based on lossy syndromes (together with a detailed analysis of the lossy syndrome technique) and compares its performance with some of the previous state-of-the-art coders.

3. Structure of the proposed video architecture

Like the solution in [21], the proposed coder inherits the building blocks of the H.264/AVC architecture [24] in order to exploit the enhanced coding solutions of the standard. However, the DVC coder in [21] processes the signal in the transform domain, while the scheme proposed here presents some differences. Fig. 1(a) shows the block diagram of the proposed Lossy Syndromes DVC (LSDVC) coding architecture. The input signal is divided into 16×16 macroblocks, which are then segmented into 4×4 blocks. Each 4×4 block $\mathbf{x}$ is classified using the motion-search unit. More precisely, the ME block computes the best predictor $\mathbf{x}_p$ from the previous frame, which is then employed to estimate the correlation level (modelled by the parameter n in Section 3.2). The amount of calculation for ME at the encoder is limited by using a reduced search window and employing a fast motion estimation algorithm. In this way, it is possible to constrain