Signal Processing: *Image Communication* 26 (2011) 220–235

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

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

Low delay distributed video coding with refined side information

António Tomé, Fernando Pereira *, ♣

Instituto Superior Técnico Av. Rovisco Pais, 1049-001 Lisboa, Portugal

ARTICLE INFO

Article history:
Received 6 May 2010
Accepted 3 January 2011
Available online 1 March 2011

Keywords:
Distributed video coding
Side information
Low delay
Correlation noise modeling
Iterative refinement

ABSTRACT

Distributed video coding (DVC) is a new video coding paradigm based upon two fundamental theoretical results: the Slepian–Wolf and Wyner–Ziv theorems. Among other benefits, this new coding paradigm may allow a flexible complexity allocation between the encoder and the decoder. Several DVC codecs have been developed over the years addressing the specific requirements of emerging applications such as wireless video surveillance and sensor networks. While state-of-the-art DVC codecs, such as the DISCOVER DVC codec, have shown promising RD performance, most DVC codecs in the literature do not consider low delay requirements which are relevant for some of the addressed applications. In this context, the main objective and novelty of this paper is to propose an efficient, low delay and fully practical DVC codec based on the Stanford DVC architecture adopting a side information iterative refinement approach. The obtained performance results show that the developed DVC solution fulfils the objectives regarding relevant benchmarks, notably due to the novel side information creation and correlation noise modeling tools integrated in a side information iterative refinement framework.

1. Introduction

Video coding technologies have been playing an important role in the context of audiovisual services such as digital TV, mobile video, and Internet streaming to cope with the ever growing compression requirements needs. Most available video coding standards, notably the ITU-T H.26x and ISO/IEC MPEG-x families of standards, adopt the so-called predictive video coding paradigm where the temporal and spatial correlations are exploited at the encoder by using a motion compensated prediction loop and a spatial transform, respectively. As a consequence, this video coding paradigm typically leads to rather complex encoders and much simpler decoders, with a rigid allocation of the complexity between the transmitter and the receiver. This approach fits well some application scenarios, e.g. broadcasting, where a few (complex) encoders provide coded content for millions of (simpler) decoders. However, with the explosion of mobile and wireless networks, there are a growing number of applications where many senders deliver data to a central receiver, e.g. video surveillance and sensor networks. Typically, these emerging applications require light encoding complexity, high compression efficiency, robustness to packet losses and, often, also low latency/delay. To address these emerging requirements, some research groups revisited the video coding problem at the light of an Information Theory result from the 1970s: the Slepian–Wolf theorem [1]. According to this theorem, the minimum rate needed to independently encode two statistically dependent discrete random sequences, X and Y, is the same as for joint encoding. While the Slepian–Wolf theorem deals with lossless coding, in 1976, Wyner and Ziv studied the case of lossy coding with side information (SI) at the decoder. Under some hypothesis on the joint statistics, the Wyner–Ziv theorem [2] states that, when the side information (i.e. the correlated source Y) is made available only at the decoder, there is no coding efficiency loss in encoding X, with respect to the case when joint encoding of X and Y is performed. In summary, the Slepian–Wolf and the Wyner–Ziv theorems state that it is possible to encode two statistically dependent signals independently and

---

* Corresponding author.
E-mail address: fp@lx.it.pt (F. Pereira).
♣ EURASIP member.

0923-5965/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.image.2011.01.005

decoding them jointly, while approaching the coding efficiency of conventional predictive coding schemes, which rely on joint encoding and decoding instead. The new coding paradigm, known as distributed video coding (DVC) does not rely on joint encoding and thus, when applied to video coding, it typically results on the absence of the temporal prediction loop (always used in predictive schemes) and lower complexity encoders. DVC architectures may provide the following functional benefits which are rather important for many emerging applications: (i) flexible allocation of the global video codec complexity; (ii) improved error resilience; (iii) codec independent scalability (since upper layers do not have to rely on precise lower layers); and (iv) exploitation of multiview correlation without cameras/encoders communicating among them. The functional benefits above can be relevant for a large range of emerging application scenarios such as wireless video cameras, low-power surveillance, video conferencing with mobile devices, disposable video cameras, visual sensor networks, distributed video streaming, multiview video systems, and wireless capsule endoscopy [3]. For a review on DVC basics and advances, please read [4–6].

Based on these theoretical results, the practical design of Wyner–Ziv (WZ) video codecs, a particular case of DVC, started around 2002, following important developments in channel coding technology. The first practical WZ solutions have been developed at Stanford University [4,7,8] and UC Berkeley [9,10]. As of today, the most popular WZ video codec design in the literature is clearly the Stanford architecture, which works at the frame level and is characterized by a feedback channel based decoder rate control. On the other hand, the Berkeley architecture, known as Power-efficient, Robust, hIgh compression Syndrome based Multimedia coding (PRISM), works at the block level and is characterized by an encoder side rate control approach based on the availability of a reference frame at the decoder. Regarding the Stanford DVC solutions, the side information generation process strongly impacts the overall RD performance but also the algorithmic delay, depending if an interpolation or extrapolation-based side information creation solution is used. In the context of DVC codecs, side information refers to the estimation made at the decoder, based on the already available decoded frames, for the frame to be DVC coded. In most available DVC solutions, the side information creation process is performed using an interpolation-based approach since it corresponds to estimating a frame between two available frames, one in the past and another in future; since one of the reference frames for the interpolation is in the future, this solution implies algorithmic delay, like when using B frames in predictive video coding, which may not be acceptable for some applications. For the applications requiring low delay, it is possible to adopt an extrapolation-based side information creation solution where the estimation of the side information is made by projecting decoded frames from the past to the future, without requiring the availability of future frames, and thus avoiding the algorithmic delay.

While there are in the literature many examples of practical and realistic interpolation-based SI DVC codecs, which have an associated delay penalty, the same does not happen for extrapolation-based low delay DVC codecs which are rare and typically adopt unpractical assumptions, e.g. the availability of originals at the decoder to generate the side information or to drive the request stopping criterion.

In this context, the objective of this paper is to develop a novel DVC solution based on the Stanford architecture, fulfilling two main requirements: high efficiency and low delay. To achieve this objective with a practical, realistic DVC architecture, this paper proposes efficient, extrapolation-based side information generation and adaptive correlation noise modeling solutions integrated in an iterative refinement approach. As far as the authors know, this architectural design has never been proposed and assessed in the literature. As it will be shown, the associated RD performance is promising both in comparison with standard-based solutions as well as state-of-the-art DVC interpolation-based solutions, notably for video content related to the most relevant application scenarios.

The following sections are organized as follows. Section 2 provides a classification system for the DVC solutions, and also reviews the most relevant low delay DVC codecs in the literature. Section 3 gives detailed information on the first version of the codec developed in this paper, the Advance Low Delay DVC (ALD-DVC) codec. In order to achieve a better RD performance, a side information refinement module is next integrated into the ALD-DVC codec, as described in Section 4. Section 5 presents the performance evaluation for the two proposed low delay DVC codecs. Finally, Section 6 includes the conclusions and future work.

2. Classification and background

The DVC spectrum of solutions is wide and, thus, some organization of this *landscape* would be welcome for a better understanding of the similarities, complementarities and alternatives regarding possible DVC solutions. This requires the definition of some relevant classification dimensions which do not have to be unique, but may certainly help understanding the relations between the various types of possible DVC solutions. Fig. 1 shows the DVC classification tree proposed in this paper which is based on four structuring dimensions; the presentation of the tree is simplified in order to facilitate the reading process, i.e. the classification tree under the *Multiview side* is exactly the same as under the *Monoview side*, thus only the Monoview side is showed in full detail. The proposed DVC classification dimensions are:

1. *Number of camera views*: Depending on the number of views to code, DVC solutions may be classified as *Monoview* or *Multiview*.
2. *Delay*: Regarding the coding delay, DVC solutions may fulfill or not low (algorithmic) delay requirements.
3. *Basic spatial coding support*: Regarding the type of basic spatial coding support, DVC solutions may typically be frame or block-based.
4. *Feedback channel*: Regarding the exploitation of a feedback channel when available, which is possible