J. Vis. Commun. Image R. 25 (2014) 361–372

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Low complexity distributed video coding

Krishna Rao Vijayanagar [a,*], Joohee Kim [a], Yunsik Lee [b], Jong-bok Kim [c]

[a] Department of Electrical and Computer Engineering, Illinois Institute of Technology, 3301 South Dearborn Street, Chicago, IL 60616, USA
[b] Korea Electronics Technology Institute (KETI), Bundang-gu, Sungnam-si, Gyungki-Do, South Korea
[c] Sane System, 439 DongAhn-Gu, AnYang-Si, Gyungki-Do, South Korea

ARTICLE INFO

Article history:
Received 5 April 2013
Accepted 3 December 2013
Available online 11 December 2013

Keywords:
Distributed video coding
Wyner–Ziv
GOP size control
Video surveillance
Encoder rate control
BCH code
Unidirectional
Low-complexity

ABSTRACT

Context: Conventional video encoding is a computationally intensive process that requires a lot of computing resources, power and memory. Such codecs cannot be deployed in remote sensors that are constrained in terms of power, memory and computational capabilities. For such applications, distributed video coding might hold the answer.
Objective: In this paper, we propose a distributed video coding (DVC) architecture that adheres to the principles of DVC by shifting the computational complexity from the encoder to the decoder and caters to low-motion scenarios like video conferencing and surveillance of hallways and buildings.
Method: The architecture presented is block-based and introduces a simple yet effective classification scheme that aims at maximizing the use of skip blocks to exploit temporal correlation between consecutive frames. In addition to the skip blocks, a dynamic GOP size control algorithm is proposed that instantaneously alters the GOP size in response to the video statistics without causing any latency and without the need to buffer additional frames at the encoder. To facilitate real-time video delivery and consumption, iterative channel codes like low density parity check codes and turbo codes are not used and in their place a Bose–Chaudhuri–Hocquenghem (BCH) code with encoder rate control is used.
Results: In spite of reducing the complexity and eliminating the feedback channel, the proposed architecture can match and even surpass the performance of current DVC systems making it a viable solution as a codec for low-motion scenarios.
Conclusion: We conclude that the proposed architecture is a suitable solution for applications that require real-time, low bit rate video transmission but have constrained resources and cannot support the complex conventional video encoding solutions.
Practical implications: The practical implications of the proposed DVC architecture include deployment in remote video sensors like hallway and building surveillance, video conferencing, video sensors that are deployed in remote regions (wildlife surveillance applications), and capsule endoscopy.

1. Introduction

Conventional video compression techniques have been around for a long time and have been used successfully in various applications with a wide range of bitrate restrictions. In order to reduce the bitrates and improve compression efficiency, conventional video codecs like H.263 and H.264/AVC [1] employ motion prediction techniques in the encoder to take advantage of temporal redundancy between neighboring frames. The motion estimation techniques employed have high complexity but lead to better video compression and rate distortion (RD) performance. The decoding scheme for conventional video codecs is highly optimized and is geared towards high frame rates to enable real-time decoding and viewing of the coded information. Such codecs are also referred to as asymmetric codecs in which the encoding is highly complex and the decoding is very simple. The use of such codecs is ideal in scenarios where a video is encoded once but decoded several times (e.g., broadcasting a video over the Internet). However, such an encoder requires a lot of computational resources and computing power and can be quite a burden for applications with limited resources.

There are many applications which need efficient video encoders but have constrained computing capabilities, battery and memory resources. In such situations, one cannot perform complex motion estimation which might drain the battery source or it might be too time-consuming considering the computing capabilities onboard. A simple example of such a situation can be found in wireless video sensors or wireless surveillance systems. Such applications require real-time transmission and possibly cannot handle the latency induced by motion estimation. Also, it is possible that a lot of computing power cannot be packed into small

* Corresponding author.
E-mail address: kvijayan@hawk.iit.edu (K.R. Vijayanagar).

1047-3203/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jvcir.2013.12.006

video sensors and thus complex encoding is not a viable option. Another upcoming application that requires real-time video transmission is in capsule endoscopy [2]. This requires a video sensor and a transmission device to be fitted into a small capsule that can be ingested by a human. It also requires that the data be transmitted in real-time and all of this requires a low complexity encoding solution.

For all such scenarios, distributed video coding (DVC) can prove to be a promising solution. DVC's principles were established in [3] [4] and can be concisely described as the reverse paradigm of conventional video coding. DVC is a promising solution for the above-mentioned scenario because it aims at shifting the coding complexity from the encoder to the decoder. Recently, several practical implementations have been proposed including the Stanford codec [5], DISCOVER codec [6], PRISM codec [7] and SPI–DVC [8].

DVC follows the principles of distributed source coding [3,4], where two correlated sources are independently encoded but jointly decoded. In other words, the correlation or redundancy between the two sources is exploited at the decoder and not at the encoder like in conventional codecs. In a frame-based DVC architecture [6], adjacent frames of a video sequence are treated as the correlated sources and are divided into two categories – key and Wyner–Ziv (WZ) frames. Key frames are encoded using a conventional encoder like H.264/AVC in the intra mode. For a WZ frame, the side information (SI) which is an estimate of the original frame is generated at the decoder by exploiting the interframe correlation. WZ frames are iteratively refined using the parity bits transmitted from the encoder.

Though it is theoretically proved in [3,4] that the same coding efficiency can be achieved whether the correlation between the sources is exploited at the encoder or decoder, practically such results have not been achieved yet. State-of-the-art DVC architectures like PRISM [7], DISCOVER [6] and MLWZ [9] all lag behind the rate-distortion (RD) performance of H.264/AVC in the No-Motion mode. Apart from coding inefficiency, many of the existing DVC architectures [7,6,9] use iteratively decodable channel codes like turbo and low density parity check accumulate (LDPCA) codes [10] to generate parity bits for the WZ frames. Results from [6] show that several seconds are needed to decode a single frame when turbo coding or LDPCA is used. Such latencies are not acceptable for applications that require real-time encoding/decoding.

Another factor that is specific to DVC is the use of neighboring frames for side information generation. Most DVC schemes like [6,7] use the previous and the future frame for generating the side information and this can cause delays. Another cause of latency in current DVC architectures is the use of a feedback channel. The feedback channel is required because the decoder needs to inform the encoder when it needs more parity bits in iterative channel decoding for WZ frames.

Considering the strengths and limitations of DVC, the major applications areas of DVC will be situations where simple but resource-limited encoders needed such as wireless surveillance or wireless video sensor networks. Both these scenarios need simple encoding devices that can effectively capture, compress and transmit video without a lot of latency. What is common to both scenarios is that the input to the systems will be low-motion sequences. When we use the term "low-motion sequences", we refer to situations that have a fairly static background with some activity in the foreground. In addition to a fairly static background, there is a high possibility of large groups of frames being identical to each other when there are no subjects in the camera's path. Such large redundancies should be exploited to achieve a good RD performance.

In this paper, we propose a low-delay, unidirectional DVC codec applicable to resource–constrained systems. The codec is block-based so that it can isolate blocks that can be skipped in order to reduce the bitrate while not affecting the video quality. A block classification scheme is proposed that aims at improving the reconstructed video quality by shifting focus of the classification on key blocks. Also, a dynamic group of pictures (GOP) size control algorithm is proposed that is simple and highly effective in comparison to existing GOP size control schemes. In order to reduce the latency, the proposed system uses only the previous frame for SI generation and BCH coding with encoder rate control instead of turbo codes or LDPCA codes. Simulation results show that the proposed architecture is much faster than current DVC codecs and despite being simpler it can achieve up to 2–2.5 dB gain in RD performance for video surveillance applications and up to 4 dB gain in video conferencing applications.

The rest of the paper is organized as follows. We review related works in Section 2 and the architectural details of the proposed codec are presented in Section 3. In Section 4 we explain the details of the proposed GOP size control algorithm and in Section 5 we show the simulation results of the codec in comparison with other DVC implementations and the paper is concluded in Section 6.

2. Related work

In the past few years, several attempts have been made to improve the performance of DVC. As mentioned in the introduction, the main issues restricting the use of current DVC architectures in practical wireless applications is its relatively low coding efficiency, high latency and the presence of a backward channel.

In order to improve the coding efficiency of DVC, several methods have been proposed to reduce the number of parity bits required to refine the WZ frames or blocks by using a highly efficient channel code. Most of these involve either turbo codes [11,8] or LDPC codes [12] or high rate LDPCA codes [6,10]. In [13], rateless Slepian–Wolf codes called Matrioshka codes are proposed which are designed to adapt to changing source statistics. In [14], rate-adaptive Slepian–Wolf codes based on rate-adaptive LDPC codes are proposed. These iterative channel coding methods cause a lot of latency in the system and are not suitable for real-time communication. Since the number of parity bits needed to be transmitted is proportional to the number of WZ blocks that have to be refined and the quality of the SI, the amount to which each block has to be refined will vary. With this as an intuition, in [15,16] the authors propose a scheme in which only the DC coefficients of the WZ blocks are compressed and sent to the decoder and they are compared with the DC coefficients of the SI. If there is sufficient matching between the two values for a block, then it is not refined in the next stage. This scheme improves the coding efficiency but has the disadvantage of having to send a map of all the blocks needing refinement back to the encoder using a backward channel. In [17], the iterative channel codes are replaced with a high speed entropy coding scheme based on data hiding. Since minimal effort is needed in encoding and decoding such a code, the system is low in terms of latency.

Another method to improve the coding efficiency is to increase the efficiency of the SI generation scheme. In [18], a novel SI generation technique is proposed for block-based DVC architectures. To take advantage of spatial and temporal redundancies, in [19] a hybrid error concealment scheme is proposed. Several other SI generation schemes [20–23] have been proposed for improving the quality of the side information generated.

In an effort to reduce latency, the authors of [24] propose a scheme in which only the previous frame is used in SI generation. This is termed as a "low-delay DVC" architecture. This is in contrast to the SI generation schemes used in [6,8,9] which involve bi-directional motion estimation and compensation that adds to the complexity and latency of the system.

In [25], the authors propose a scheme where motion estimation is performed at the decoder and the motion vectors are sent back