Optik 124 (2013) 4929–4938

Contents lists available at ScienceDirect

Optik

journal homepage: www.elsevier.de/ijleo

A mixed framework for transform domain Wyner–Ziv video coding

Suvendu Rup[a,*], Banshidhar Majhi[b,1]

[a] Department of Computer Science and Engineering, International Institute of Information Technology, Bhubaneswar, India
[b] Department of Computer Science and Engineering, National Institute of Technology, Rourkela, India

ARTICLE INFO

Article history:
Received 12 October 2012
Accepted 2 March 2013

Keywords:
Distributed video coding
Wyner–Ziv video coding
Side information
Burrows–Wheeler transform
Multilayer perceptron

ABSTRACT

This paper presents a mixed framework based on an efficient intra key frame coding and an improved side information (SI) generation scheme in transform domain Wyner–Ziv (WZ) video coding. The performance of the WZ video coding strongly depends on the quality of the SI. The SI can be generated from the decoded key frames resulted from intra key frame video coding. The better the decoded key frames are the better would be the SI generation. In this paper, a Burrows–Wheeler transform (BWT) based intra-frame video coding is proposed to generate improved decoded key frames. Furthermore, an improved SI generation scheme with multilayer perceptron (MLP) is proposed. Comparative analysis with other standard techniques of WZ video coding reveals that the proposed scheme has better standing as compared to its counterparts in terms of both coding efficiency and improved perceptual quality.

1. Introduction

In recent years, distributed video coding (DVC) is becoming more popular due to the emerging applications like mobile camera phone, video surveillance, wireless PC-camera, visual sensor network etc. The traditional video coding architecture has complex encoder and a simple decoder and it is a challenge for the traditional video coding to fulfill the requirement of the above mentioned applications [1]. In traditional video coding, the video encoder is typically five to ten times more complex than the decoder due to the motion estimation task at the encoder. So DVC is an ultimate choice over traditional video coding where the computational complexity has been shifted from encoder to decoder.

The fundamental basis of DVC is mainly based on two major information theoretic results namely: Slepian–Wolf (SW) and WZ theorem. The SW theorem suggests, it is possible to achieve the same bit rate as the joint encoding system by independent encoding and joint decoding [2]. The WZ theorem extends the SW theorem to a lossy case [3].

Around 2002, most of the research communities have emerged in the field of DVC. Recently, major practical solutions of the DVC have been proposed by two groups Bernd Girod's group at Stanford University and Ramchandran's group at the University of California, Berkley. The first practical solution toward DVC was pixel domain coding solution proposed by Girod's group [4–6]. In pixel domain coding solution the video frames are divided into key frames and WZ frames. The WZ frames are encoded independently and decoded jointly. These schemes are the simplest ones, because neither discrete cosine transform (DCT) nor motion estimation is required.

Another relevant WZ video coding solution proposed by Girod and co-workers where, a blockwise DCT is performed instead of pixel by pixel [7]. The DCT coefficients are independently quantized and compressed by SW encoder. The SI can be generated from previously reconstructed frame with or without motion compensation. This scheme has a higher encoder complexity than pixel domain solution but shows improved performance results.

Another promising solution have been proposed by Ramchandran's group named as PRISM (power efficient robust high compression syndrome based multimedia coding) [8,9]. In this solution, they combine both the feature of intra-frame coding with inter-frame coding compression efficiency. This architecture uses WZ coding, but the SI generation scheme is different from other schemes.

Till now, the Stanford architecture and its development have been popular among various researchers. So far, the best rate distortion (RD) performance has been achieved by Stanford based architecture is the DISCOVER (distributed coding for video services) [10]. DISCOVER has a strong potential for new applications targeting new advances in coding efficiency, error resilience, and scalability.

So from the different DVC solutions available till date, the RD performance of DVC did not beat the conventional H.264/AVC (Inter) due to more intense and irregular motion. The degradation is mainly due to the poor quality of SI resulting from intra key frame video coding.

The intra key frame coding plays an important role in DVC architecture. The better are the decoded key frames the better would be the SI generation. The SI is treated as the noisy version of the

---

[*] Corresponding author. Tel.: +91 674 3016062; fax: +91 674 3016009.
E-mail addresses: suvendu@iiit-bh.ac.in (S. Rup), bmajhi@nitrkl.ac.in (B. Majhi).
[1] IEEE Member.

0030-4026/$ – see front matter © 2013 Elsevier GmbH. All rights reserved.
http://dx.doi.org/10.1016/j.ijleo.2013.03.152

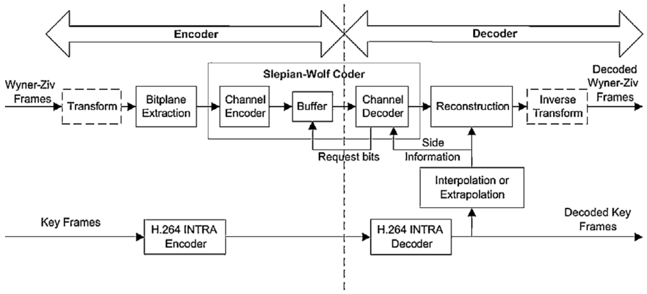

Fig. 1. Architecture of DVC.

WZ frame passing through the virtual correlation channel [7,11]. In transform domain DVC architecture the encoder has no knowledge about the SI generation. However, the SI generation process has been adopted at the decoder side. As the SI is more accurate, the encoder needs to send fewer number of parity bits to minimize the error between original WZ frame and the SI.

In this paper, a mixed framework based on an efficient intra key frame video coding using BWT and an improved SI generation scheme using MLP is proposed. The suggested schemes are applied to transform domain DVC architecture to improve both coding performance and perceptual visual quality.

The rest of the paper is organized as follows. Section 2 gives the architectural description of DVC. Section 3 elaborates the related works on intra-frame video coding in DVC. Section 4 presents the theoretical foundations of BWT. In Section 5, the proposed BWT based intra key frame coding in DVC is introduced. Section 6 illustrates the related work on SI generation in DVC. Section 7 presents the proposed SI generation framework using MLP. The results obtained are discussed experimentwise in Section 8. Finally, Section 9 provides the concluding remarks.

2. Architectural scheme of DVC

The architecture of the transform domain DVC is depicted in Fig. 1. Normally in DVC, the frames are divided into two groups WZ frames (even frames) and key frames (odd frames). Then a 4×4 blockwise DCT is applied to all WZ frames. The DCT coefficient bands are organized by picking the DCT coefficients from the same position [1,11]. The transform coefficients within a given band X_i are grouped together and each transform coefficients are uniformly quantized. The coefficients are binarized after quantization and the different quantized coefficients of the same band are grouped together and different bit planes are extracted. The bit planes are organized from the most significant bit (MSB) plane to least significant bit (LSB) plane. Next, turbo encoding is applied to each bit plane. The turbo encoder generates the parity bits for each bit plane which is saved in a buffer and sent to the decoder upon request. At the decoder side of the DVC, the SI (Y_i) can be generated using two closest key frames, one temporally in the past X_b' and the other in the future X_f'. For group of picture (GOP) size 2, the two closest neighboring key frames within the WZ frame to be decoded. For longer GOP sizes, previously decoded WZ frames may also play the role of reference frames for the decoding of other WZ frames. The same blockwise DCT is applied to generate side information Y_i and the coefficients bands are grouped together in the same way. The turbo decoder does not use the SI (Y_i) DCT band directly. They are first converted into soft-input information through a correlation channel model that exploits the statistics between X_i (at the encoder) and Y_i (at the decoder). The WZ video coding efficiency strongly depends on the virtual channel model. In most of the DVC solution, a Laplacian distribution model is used. This model computes the residual error, (X_i–Y_i) which represents the confidence or matching success of the frame interpolation operation [11]. The noise or error can be mathematically described as,

$$r = X_i - Y_i, \quad f(r) = \frac{\alpha}{2}e^{-\alpha|r|}, \quad \alpha = \frac{\sqrt{2}}{\sigma} \tag{1}$$

where r refers to the difference between Y_i and X_i and σ^2 denotes the variance of r. The residual between X_i and Y_i is modeled by Laplacian distribution as shown in (1). Firstly, we need to get the parameter α by calculating the variance parameter of the residual between X_i and Y_i. Then calculate the probability of each pixel in Y_i. The probability of a pixel y_k in Y_i equals to each pixel in X_i can be computed as follows,

$$P(x_i' = x_m|y_k) = \frac{\alpha}{2}e^{-\alpha|y_k - x_m|} \quad x_m \epsilon X_i \tag{2}$$

where y_k and x_m are pixels and x_i' denotes the prediction of input symbol. The turbo decoder decodes the current bit plane using the Logarithmic Maximum A Posteriori (Log-MAP) algorithm. Then the turbo decoder generates the decoded quantized symbol stream by using received parity bits and soft-input information provided by the correlation model. If the current bit plane error probability P_e exceeds 10^{-3}, the decoder requests more number of parity bits from the encoder; otherwise the current bit plane is executed successfully. After all the bit planes are decoded, the quantized symbol stream can be rebuilt as q'. The reconstruction of WZ frame is defined in [12] as follows,

$$X_i' = E(X_i|q', Y) \tag{3}$$

where X_i' is the reconstructed WZ frame, Y is the SI, $E(\cdot)$ is the expectation operator, and X_i is the original WZ frame. After reconstruction is over finally, the decoded key frames and the estimated WZ frames are sequenced together to generate the video sequence at the decoder end.

3. Related works on intra-frame coding in DVC

The coding performance of the WZ video coding strongly depends on the quality of the SI. The SI in DVC can be generated from the decoded key frames resulting from the intra frame DVC coder. So it is necessary to have good decoded key frames at the decoder side of the DVC. Several related works have been reported in literature for intra-frame coding in DVC. Girod et al. have proposed two hierarchical frame dependency arrangement [7]. In this framework, the frames are encoded as intra (I) frame with a fixed quantization parameter using H.263 (Intra) coding. The scheme utilizes motion compensation interpolation which is less accurate and thus degrades SI quality. Brites et al. have proposed an improved transform domain WZ video coding architecture [13]. Here, the key frames are sent directly to the decoder without any compression.

In [14,15], the authors have claimed high RD performance, but a high amount of bits are resulted from encoder to decoder due to lossless key frames without any compression. All the above mentioned reported works are claimed about the improvement of DVC codec design. But none of the schemes are addressed toward the improvement of intra key frame coding in DVC.

However, for the first time Adikari et al. have proposed an independent key frame coding using correlated pixels in DVC [16]. In this framework, a novel intra coding technique is proposed that eliminates the requirement of a secondary coding scheme for coding the key frames in DVC. The authors have claimed a peak signal to noise ratio (PSNR) improvement, but this scheme was mainly designed for pixel domain coding solution. Recently, most schemes in DVC deals with transform domain coding solution. The latest frameworks adopted recently in [10,17] are popular among DVC researchers, however, they do not address the improvement in