Int. J. Electron. Commun. (AEÜ) 68 (2014) 201–209

Contents lists available at ScienceDirect

International Journal of Electronics and Communications (AEÜ)

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

# An improved side information generation for distributed video coding

Suvendu Rup [a,*], Banshidhar Majhi [b,1], Sudarshan Padhy [c]

[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
[c] Institute of Mathematics and Application, Bhubaneswar, India

ARTICLE INFO

*Article history:*
Received 28 May 2012
Accepted 14 August 2013

*Keywords:*
Distributed video coding
Side information
Multilayer perceptron
Rate distortion
Decoding time requirement

ABSTRACT

This paper presents a side information (SI) scheme for distributed video coding based on multilayer perceptron. The suggested scheme predicts a Wyner–Ziv (WZ) frame from two decoded key frames. The network is trained offline using patterns from different standard video sequences with varied motion characteristics to achieve generalization. The proposed scheme is simulated along with other standard video coding schemes. Performance comparisons have been made with respect to training convergence, rate distortion (RD), peak signal to noise ratio (PSNR), number of requests per SI frame, decoding time requirement, etc. In general, it is observed that the proposed scheme has a superior SI frame generation capability as compared to its competent schemes.



## 1. Introduction

The emerging applications like mobile camera phone, video surveillance, multimedia sensor networks, wireless camera, etc. demand a complex decoder and low cost encoder with high coding efficiency. This is due to the fact that the encoder has less memory and less computational power. Distributed video coding (DVC) has emerged as a proposed solution for the purpose [1]. DVC is based on two well known information theoretic results namely, the Slepian–Wolf (SW) theorem and Wyner–Ziv (WZ) theorem [2,3]. The SW theorem suggests that it is possible to achieve the same bit rate as the joint encoding system by independent encoding and joint decoding. WZ theorem extends the SW theorem principle to a lossy case. In the latter case, side information (SI), a correlated signal of the original compressed signal plays an important role. In DVC, motion estimation is performed at the decoder to generate the SI keeping the encoder job simple. The techniques involved in SI generation at the decoder significantly influence the RD performance of the DVC. The higher the correlation between the SI at the decoder and the current WZ frame at the encoder, the better is the estimation. As a result, fewer parity bits are necessary from encoder to decoder to generate good quality WZ frame.

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]. Later, the same group have proposed a discrete cosine transform (DCT) based approach instead of pixel by pixel coding [7]. Another promising solution have been proposed by Ramchandran's group, more popularly known as power efficient robust high compression syndrome based multimedia coding (PRISM). It combines both the features of intra-frame coding with inter-frame coding compression efficiency [8,9]. This architecture uses WZ coding, but the SI generation scheme is different from Stanford-based approach.

To study the impact of any SI creation framework, it is necessary to integrate it in a practical DVC codec. The DVC codec adopted in this paper follows the Stanford-based architecture and its overall architecture is shown in Fig. 1. To understand the architecture of DVC codec, its working principles are discussed here in nut shell. Firstly, video sequence is divided into WZ frames and key frames. The even frames are the WZ frames and key frames are odd ones. For each WZ frame $X_{2i}$, a $4 \times 4$ blockwise DCT is applied. The transform coefficients of the frame $X_{2i}$ are grouped together, i.e. the coefficients from the same position of each DCT blocks are picked up to compose the 16 possible coefficient bands. After transform coding operation is over, each DCT coefficient band $X_k$ are uniformly quantized to obtain a stream $q_{2i}$. 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 plane (MSB) to least significant bit plane (LSB). 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. A pseudo random puncturing pattern is used to transmit the parity bits. Meanwhile, the key frames are encoded using the conventional intra-frame video coding. At the decoder side of DVC, the frame interpolation technique

* Corresponding author. Tel.: +91 0674 3016062; fax: +91 0674 3016009.
*E-mail address:* suvendu@iiit-bh.ac.in (S. Rup).
[1] IEEE member.

1434-8411/$ – see front matter 
http://dx.doi.org/10.1016/j.aeue.2013.08.005

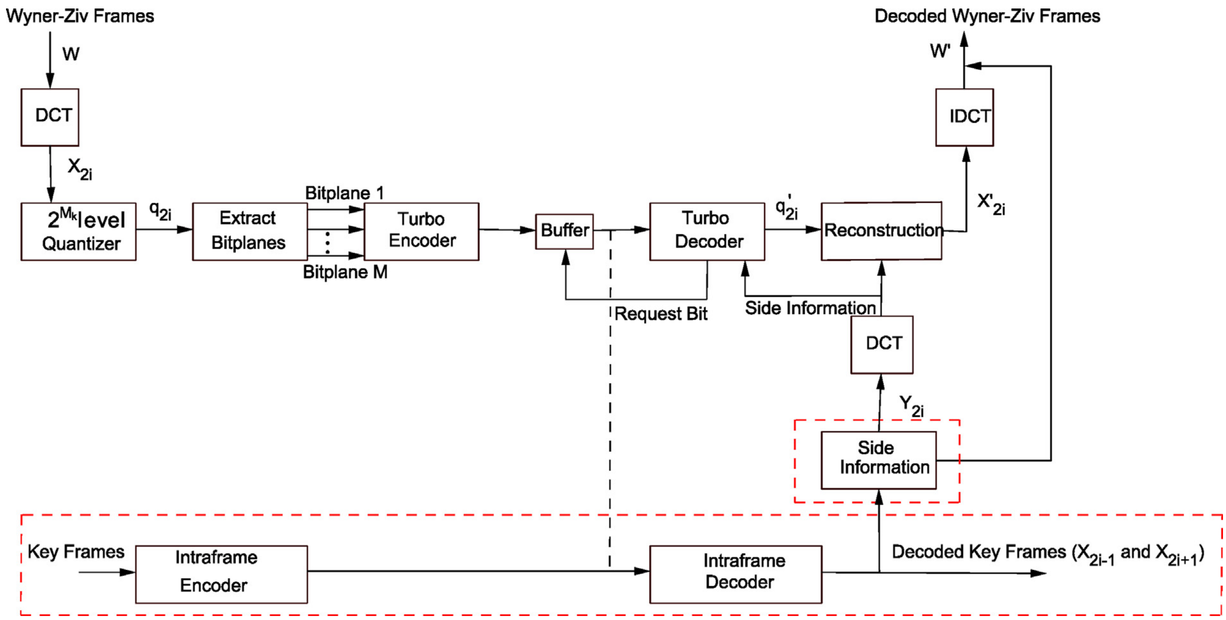


**Fig. 1.** Stanford-based transform domain DVC architecture.

is used to generate the estimation of $X_{2i}$ frame called SI denoted as $Y_{2i}$ frame from the two adjacent key frames $X_{2i-1}$ and $X_{2i+1}$. The same blockwise $4 \times 4$ DCT is applied to the interpolated frame ($Y_{2i}$) to generate an estimate of ($X_{2i}$). The coefficient band ($Y_{2i}$) and the received parity bits are used to decode the current bit plane. The decoder also determines if the current bit plane error probability i.e. $P_e > 10^{-3}$, the decoder requests more number of parity bits; otherwise, the current bit plane is executed successfully. After all bit planes are executed and the quantized symbol stream $q'_{2i}$ is obtained, then reconstruction of each DCT coefficient band is performed. When all the DCT coefficient bands are reconstructed, the inverse discrete cosine transform (IDCT) is applied to obtain the estimated WZ frame $W'$. The decoded key frames and the estimated WZ frames are sequenced together to generate the video sequence at the decoder end.

In this paper, we propose a multilayer perceptron (MLP) based SI generation scheme and integrated with the aforesaid DVC codec to study its impact on various performance issues. The rest of the paper is organized as follows. Section 2 elaborates the related work on SI generation in DVC. Section 3 presents the proposed SI generation using MLP. The overall simulation is divided into a set of experiments. The results obtained are discussed experimentwise in Section 4. Finally, the conclusion is given in Section 5.

## 2. Related work on SI generation

The RD performance of DVC strongly depends on the quality of the decoded key frames and the accuracy of the SI generation process. In the original Stanford-based architecture, SI is generated through interpolation of two decoded key frames i.e. preceding and succeeding frames of the original WZ frames. In the recent past, several related works have been reported in the literature for SI generation which involves sophisticated framework. Aaron et al. have proposed two hierarchical frame dependency arrangement in DVC [6]. In their first approach, the SI for the current WZ frame can be extrapolated from a key frame or from a WZ frame. In their second approach, a more complex arrangement has been used with an increase temporal resolution of 2:1 with bidirectional interpolation. With the above two schemes, a poor SI estimation is resulted as group of picture (GOP) size increases and motion becomes more intense and less well behaved. The same authors have proposed solutions using motion compensated interpolation (MC-I) and motion compensated extrapolation (MC-E) [7]. In MC-I, the SI for an even frame at time index $t$ is generated by performing motion compensated interpolation using the two decoded key frames at time $(t-1)$ and $(t+1)$. This interpolation technique involves symmetrical bidirectional block matching for the estimated motion. Since the next key frame is needed for interpolation, the frames have to be decoded out-of-order. This scheme is similar to the decoding of bidirectional frames in predictive coding which is a limitation of this framework. In MC-E, the SI is generated by estimating the motion between the decoded WZ frame at time $(t-2)$ and decoded key frame at time $(t-1)$. Here, the already decoded WZ frames are used for motion estimation. So the reconstruction error from the WZ frame contribute to degradation of SI quality.

Girod et al. have proposed previous extrapolation (Prev-E) and average interpolation (AV-I) techniques to generate SI for low complexity video coding solution [7]. In Prev-E scheme, the previous key frame is used directly as SI whereas, in AV-I technique, the SI for the WZ frame is generated by averaging the pixel values from the key frames at time $(t-1)$ and $(t+1)$. The limitation of Prev-E scheme is that it does not employ motion compensation to generate SI. So the Prev-E scheme is better than DCT based intra-frame coding only at lower bit rates. In AV-I scheme, the pixel values from key frames $(t-1)$ and $(t+1)$ are averaged and it is not sufficient to generate a good quality SI. A hash based motion compensation is proposed by the same authors in [10].

In 2005, Tagliasacchi et al. have presented a motion compensated temporal filtering technique [11]. This scheme is based on pixel domain coding solution. Natario et al. have proposed a motion field smoothing algorithm to generate SI in [12]. Artigas and Torres have proposed an iterative motion compensated interpolation technique where the turbo decoder runs several times for decoding the WZ frame to estimate SI and as a result a significant delay is associated [13].

Adikari et al. have proposed a multiple SI stream for DVC [14]. It uses two SI streams which are generated using motion extrapolation and compensation (ME-C). The first SI stream (SS-1) is predicted by extrapolating the motion from the previous two closest key frames. The second SI stream (SS-2) is predicted using the immediate key frame and the closest WZ frame. Fernando et al. have proposed a SI scheme using sequential motion compensation,