Journal of Network Intelligence
Taiwan Ubiquitous Information
 ISSN 2414-8105 (Online)
Volume **2**, Number **4**, November **2017**

JND-based Wyner-Ziv Video Coding

Jie Cheng, Lili Meng, Jia Zhang, Yanyan Tan, Yuwei Ren, Li Liu

Department of Information Science and Engineering
Shandong Normal University, Jinan, China
Shandong Provincial Key Laboratory for Novel Distributed Computer Software Technology
mengll_83@hotmail.com
Corresponding author: Lili Meng

Abstract. *Distributed video coding (DVC) is the implementation of distributed source coding in video coding. Its core idea abide by the fundamental of distributed source coding, where correlated sources are encoded independently but decoded jointly. Video coding is meant to achieve the best possible reconstruction quality for a given bit rate. In recent years, research community think highly of DVC and make their work focus on how to attain a better DVC system. In this paper, we introduce the concept of Just-noticeable distortion (JND) in distributed video coding system, a measure of maximum image distortion that the human eye cannot detect, which can effectively improve the encoding efficiency and not affect perceptual quality. Meanwhile, the quality of the side information (SI) plays an extremely significant role in the performance of distributed video coding system. The better the quality of side information is, the higher the system performance will be. Also, experimental results in this paper illustrate that by using one side information distributed video coding system with JND model can achieve the same effect as multiple SIs system without JND model.*
Keywords: Wyner-ziv video coding, Just-noticeable distortion

1. **Introduction.** Compared with the traditional video coding, distributed video coding (DVC) is a totally advanced coding technique. As we all know, distributed video coding is the application of distributed source coding in video coding domain, thus its central idea relies on the criterion of distributed source coding. In the wake of developments in science and technology recent years, the application such as digital television and network video call is popular increasingly in our daily life. That makes a series of meaningful progress on video compression research. But burgeoning wireless video surveillance networks, wireless camera, mobile video telephone are now facing the new challenge. Then, how to achieve a lighter encoding complexity and a higher compression efficiency becomes particularly urgent. Faced with such problems, a new method which called distributed video coding was born. The theoretical basis of DVC is Slepian-Wolf (SW) theorem [1] and Wyner-Ziv (WZ) theorem [2]. The SW theorem indicates that we may get the same bit rate by using the way of independently encoding and jointly decoding, which states that when two or more correlated sources are respectively encoded and jointly decoded, the same compression ratio as joint encoding can be achieved. WZ theorem enlarges the former theorem to a lossy case. In this latter case, the task of the decoder is to exploit the correlation between the sources of code, meaning the complexity balance between encoder and decoder can be potentially reversed with regard to traditional coding methods.

First practical implementations of DVC systems were made in [3] and [4]. In [3], the PRISM codec is introduced, which is based on independent syndrome coding of pixel

blocks. In [4], a codec based on turbo codes operating on the whole frame is proposed. The innovation of this page is applying just noticeable distortion (JND) model to distributed video coding system. Human visual system is not sensitive to the information below JND threshold so that they can be regarded as visual redundancy which can be deserted. That means useless information which the human eye cannot detect should not be encoded. The advantage of this application is enhancing encoding efficiency and system performance. Transmission of the reconstructed image can use less bit rates. Thus, JND is a wonderful application in distributed video coding system. Later, we give a definite scheme and detailed description of the application.

The structure of the paper is reproduced below. A detailed description of JND in transform domain is given in section 2. In section 3, detailed explanation of DVC and the proposed scheme will be shown. Section 4 mainly talks about experiments and analysis for data. Section 5 summarizes the paper.

2. JND models.

2.1. Just-noticeable distortion.

JND is the abbreviation of Just-noticeable distortion. It is a threshold in substance as well as a measurement of maximum distortion of the human eye. It embodies the tolerance of the human eye for an image change and is employed to measure sensitivity of the human eye for image distortion in image processing field. Human eye cannot distinguish the image changing because of some physiological limitation. Depending upon this characteristic, we can reduce the redundancy of vision, which means decreasing useless information that the human eye cannot detect and dealing with important information for the human eye only. In this way can improve video compression performance effectively.

Pixel domain JND model and DCT domain JND model are two methods of JND model. In the practical application process, most compression of image and video is conducted in discrete cosine transform (DCT) domain what makes the DCT domain JND model the major research direction. In this paper, we apply just noticeable distortion model to distributed video coding system, meaning that we use JND model to deal with coefficient in the DCT domain. As for JND value in Pixel domain JND, it is used to compute the PSPNR for the sake of making comparisons between the former scheme and our proposed scheme. Computational formula will be determined by following section.

2.2. DCT domain JND model.

We can depict the DCT domain JND model according to equation in [5], as follows.

$$JND\_DCT(i,j) = J_{CSF}(i,j) \times F_{lum} \times F_{contrast}(i,j) \tag{1}$$

where $J_{CFS}(i,j)$ and F_{lum} are spatial contrast sensitivity function and luminance adaption factor respectively. Meanwhile, $F_{contrast}(i,j)$ stand for contrast masking weighting factor. We can obtain CSF factor via equation (2).

$$J_{CSF}(i,j) = \frac{s}{\phi_i \phi_j} \times \frac{\frac{exp(c\omega_{ij})}{a+b\omega_{ij}}}{r + (1 - r\cos\psi_{ij})} \tag{2}$$

In this equation, s is regarded as a measurement of spatial influence and its empirical value is 0.25. ϕ_i and ϕ_j are normalization coefficients of DCT. ω_{ij} is spatial frequency of DCT coefficients. $r + (1 - rcos\psi_{ij})$ means tilting effect of the human visual system. ψ_{ij} means direction angle of corresponding DCT component. In this paper,the value of a and b are 1.33 and 0.11, respectively. Experiential value of r is 0.6.

Human eye has different sensitivity to different brightness areas of an image because of adaptive luminance masking effect. Thus the value of F_{lum} is related to average brightness value of the block. And then we can get it from equation (3).

$$F_{lum} = \begin{cases} 1 + \frac{60-\hat{I}}{150} & \hat{I} \leq 60 \\ 1 & 60 < \hat{I} < 170 \\ 1 + \frac{\hat{I}-170}{425} & \hat{I} \geq 170 \end{cases} \tag{3}$$

where $\hat{I}$ represent average brightness value. We use equation (4) to compute the value of $\hat{I}$.

$$\hat{I} = \frac{DC}{N} + 128 \tag{4}$$

where N represents the size of image block and N equals 4 in this paper. DC is direct current coefficient in DCT domain. Then we can compute the value of contrast masking weighting factor according to equation (5).

$$F_{contrast}(i,j) = \begin{cases} \Psi, & \text{for } (i^2+j^2) \leq 16 \text{ in Plane or Edge block} \\ \Psi \times \min(4, \max(1, (\frac{C(i,j)}{J_{CSF}(i,j) \times F_lum})^{0.36})), & \text{others} \end{cases} \tag{5}$$

Before we compute the value of $F_{contrast}(i,j)$, we should make an edge detection on the image in order to get edge pixel density. Then classify the image block into three areas: plane, edge, texture. Canny operator is a very popular and effective detector. For a given image, it can detect the edge pixel to a crumb. If a block contains a lot of edge pixels, the block can be regarded as a texture block. Otherwise, the block can be smooth region if it contains less edge pixels. The classification principle is determined via (6).

$$type = \begin{cases} Plane & \rho_{edgel} \leq \alpha \\ Edge & \alpha < \rho_{edgel} \leq \beta \\ Texture & \rho_{edgel} > \beta \end{cases} \tag{6}$$

Here $\alpha = 0.1$ and $\beta = 0.2$.

Ψ represent weighted value. Human eye has different sensitivity to different areas of an image so that different regions should be weighted in different values. In general, human visual system is sensitive to the plane region and edge region. In contrast, human eye is not sensitive to the texture region. The weighting principle is depicted by equation (7).

$$\Psi = \begin{cases} 1, & \text{for Plane and Edge block} \\ 2.25, & \text{for } (i^2+j^2) \leq 16 \text{ in Texture block} \\ 1.25, & \text{for } (i^2+j^2) > 16 \text{ in Texture block} \end{cases} \tag{7}$$

By all the above equations, we can finally obtain the JND threshold in DCTdomain. Compared JND threshold in DCT domain, DCT coefficients should be abandoned if they are not greater than JND threshold. By doing this comparison process, preconditioning for DCT coefficients has been finished ultimately.

3. **Proposed JND system.**

3.1. **Overview of DVC system.**

There are two kinds of frames in DVC system. They are so-called key frames (K frames) and WZ frames. K frames are encoded by traditional intra-frame encoder and decoded by traditional intra-frame decoder. The frames between Key frames are WZ frames, for whose process mode is encoded by intra-frame encoder and decoded by inter-frame decoder. For every WZ frame, the decoder generates side information (SI) as an evaluation of WZ frame from the previous K frames. After that,

the channel decoder makes a combination of side information and the parity bit to rebuild the W frame. Thus, it indicates that the accuracy of side information plays a significant role on compression performance of the coding system. If we want to obtain a higher performance of rate-distoration.

In sake of enhancing the side information scheme there are several schemes. Aaron in [6] refer to send to a Hash of the WZ frames to be decoded in order to promote interpolation in SI. And in [7], Fan. et. al indicate approach of transform domain DVC. Then in [8], an improvement of transform-domain DVC was raised. Although it refines the quality of the SI after all the DCT bands are decoded, it raises decoding complexity. Up to now, an effective coding approach is the transform domain Wyner-Ziv video codec. Using DCT can decrease the spatial redundancy. Firstly, in the system, SI of W frames are arose by decoder from previously decoded Key frame. After that, a combination of the parity bit and the side information will be taken to reconstruct the video stream. The transform-domain Wyner-ziv video coding system takes advantage of fewer parity to decode W frame.

One side information in reconstruction is developed in [9] under normal conditions. In [10], it is shown from the information theory perspective that the DSC coding efficiency can be enhanced by multiple SIs, because of their reduction of the conditional entropy of the source. In the typical setting of the DVC, two SIs for each WZ frame can be readily obtained from the neighboring key frames, using forward and backward motion estimation respectively. In [11], it is present an effective method for getting better quality of two side informations by using Bayesian conditional pdf which could outperform other ways.

In this proposedscheme, after decoding K frames, the decoder need two side informations what can obtained from forward and backward prediction and here we draw lessons from [11] to acquire conditional pdf of $f(x|y_1, y_2)$.

$$f(x|y_1, y_2) = \frac{f(x|y_1)f(x|y_2)f(y_1)f(y_2)}{f(x)f(y_1, y_2)}. \tag{8}$$

In the experimental part, we will make a comparison between the proposed scheme in this paper and scheme proposed in [11] for the sake of testing the performance of our method and the experimental result will be given.

3.2. **Overview of proposed system.** The JND-based distributed video coding framework is shown in Fig.1.

In order to better understand the structure, now we give some detailed explanations about the operation process of the system. Firstly, test video sequences are divided into two kinds of frames. The odd ones are called K frame and W frame is the even ones. The input W frame is made a 4×4 block wise DCT. Then, we compute the JND threshold in DCT domain of W frame to make a pretreatment of DCT coefficient. DCT coefficients will be abandoned when the value is below JND threshold and be retained when the value is greater than JND threshold by using JND threshold to make filtration operation. After that, the treated DCT coefficients are grouped together to form coefficients bands and then quantized by quantizer. Bit planes will be abstracted from different quantized coefficients of the same band after that operation. Next bit planes are arranged in a specific order and encoded by LDPCA encoder. Encoded information of W frame will be sent to decoder.

Meanwhile, K frames are encoded by traditional intra-frame encoder and decoded by conventional intra-frame decoder to generate side information which is the estimate of W frame. Similarly, generated side informations are manufactured the same 4×4 block wise DCT. JND threshold of side information in DCT domain is calculated. If DCT coefficient

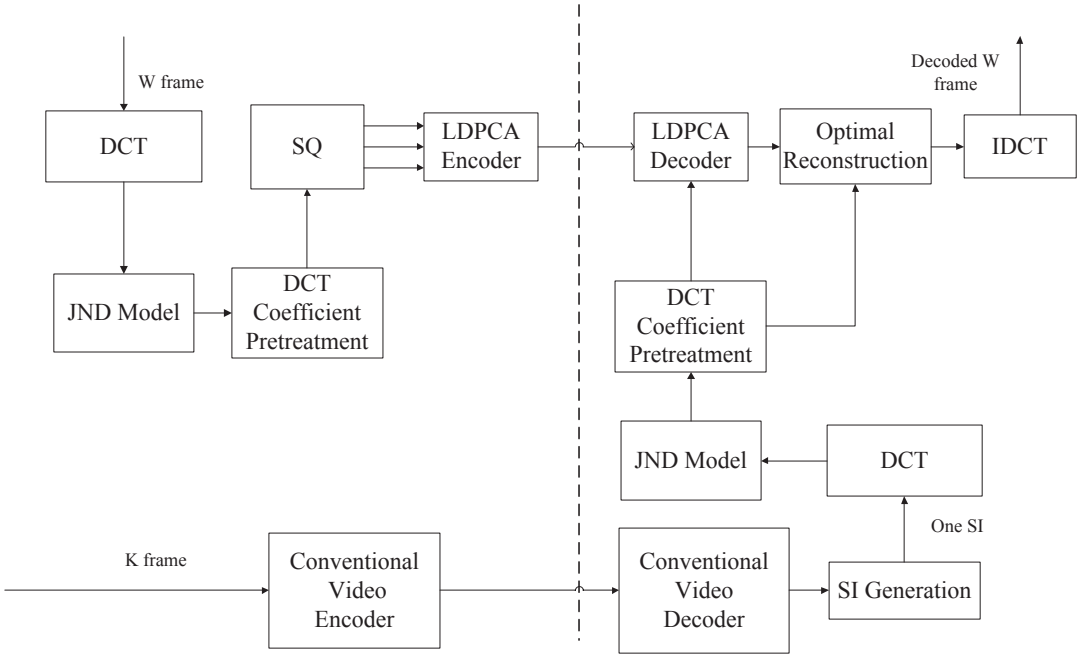

FIGURE 1. proposedscheme for one side information.

of side information is no greater than JND threshold of side information in DCT domain, it would be discarded. Otherwise it would be conserved. New side information will be obtained after that comparison.

The decoder combines additional side information with encoded information of W frame to make the reconstruction. After IDCT process, optimal reconstruction will be obtained finally. Up to now, JND-based DVC system finishes the encode and decode process.

As we know above, at the decoder, the W frames are decoded by using both encoded information of W and side information produced from decoded K frames, which means the decoder need to know some model of statistical correlation between side information and W frames to use side information. The statistical correlation can be modeled by $Y = X + Z$ where Y is called side information and X stands for W frames. Z denotes the correlation noise. The conditional probability density function of $f_{Y|X}(y|x)$ could be found equal to $f_Z(y - x)$.

The decoder assumes that Z has Laplacian distribution which can be expressed below.

$$f(z) = \frac{\alpha}{2} e^{-\alpha|z|} \tag{9}$$

α is Laplacian parameter of each DCT band and can be estimated by plotting the residual histogram of several sequences for the side information. In our experiment we use six sequences to train it and estimate it from the variance σ^2 of SI by using expression of $\alpha^2 = \frac{2}{\sigma^2}$. The model of Z plays a significant part in reconstruction of W frames. For the sake of testing performance of the proposed system, a series of experiments are carried out. In the next portion, experimental results will be displayed.

4. **Experiment and Results.** In this section, we will give some results of experiments. The test sequences whose R-D performance will be demonstrated are Foreman, Mother, Highway and Salesman. In order to better evaluate performance of the proposed scheme, a definition of PSPNR is introduced, which is widespread in perceptual distortion metric. The PSPNR is obtained in [12] via (8).

$$PSPNR = 10\log_{10}\frac{255\times 255}{\frac{1}{WH}\cdot\sum_{i=1}^{W}\sum_{j=1}^{H}(err(i,j)^2)\delta(i,j)} \tag{10}$$

where W denote width of the image and H means height. Here:

$$err(i,j) = |P(i,j) - \bar{P}(i,j)| - JND_S(i,j) \tag{11}$$

$JND_S(i,j)$ represent the JND threshold of coordinate value in pixel domain of reconstructed frame. When the formula $|P(i,j) - \bar{P}(i,j)| \geq JND_S(i,j)$ is correct, the value of $\delta(i,j)$ is 1. Otherwise the value of $\delta(i,j)$ equal 0.

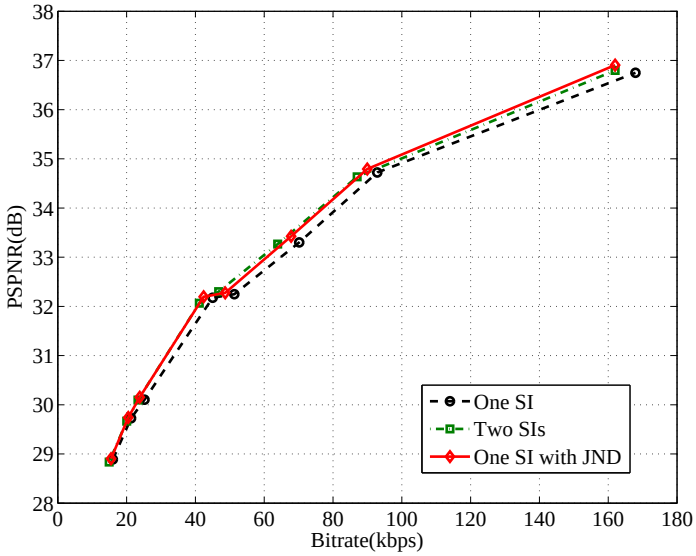

FIGURE 2. Rate-Distoration result for Foreman.

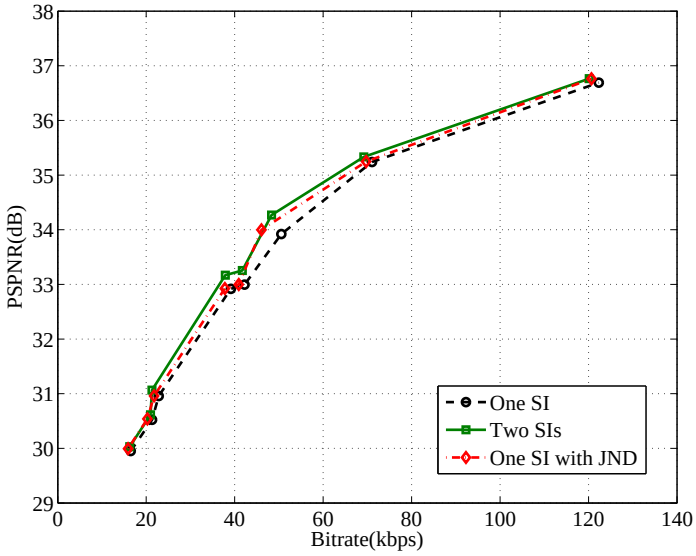

FIGURE 3. Rate-Distoration result for highway.

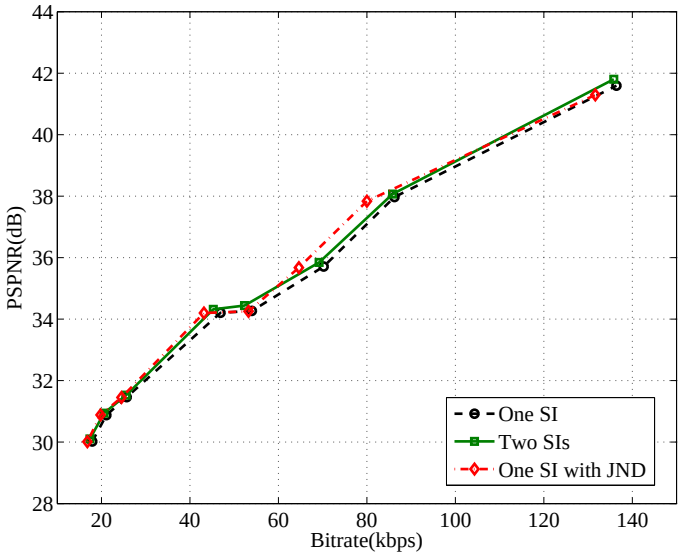

FIGURE 4. Rate-Distoration result for Mother.

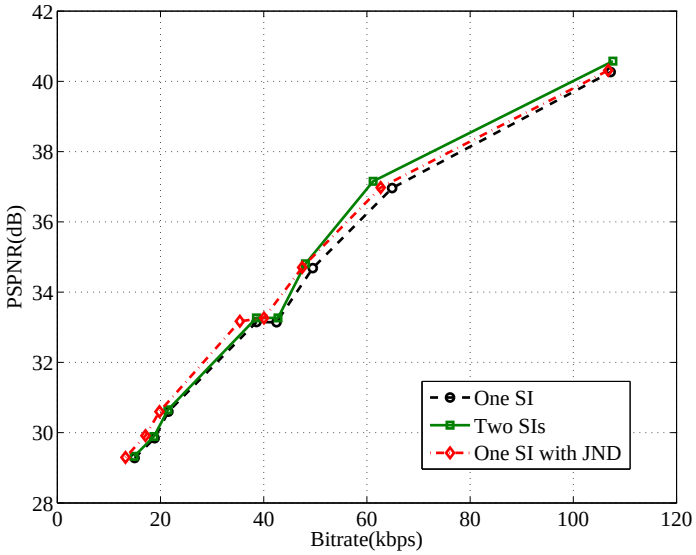

FIGURE 5. Rate-Distoration result for Salesman.

According to our proposed system, we provide three experimental schemes and implement. They are one side information scheme, two side information scheme and one side information with JND model. We give one chart for every test sequence for the sake of making comparisons between different schemes and making readers observe the experimental result clearly. The test sequences whose R-D performance will be shown are Foreman, Highway, Mother and Salesman. And the results are shown in Fig.2, Fig.3, Fig.4 and Fig.5.

Among these results, the system performance with two SIs scheme is superior to the system with one SI method which both of them are not applied JND method which means the better side information are available, the better the system performance can be improved. Also, the charts present that our proposed method that using JND model

in one SI system make a better Rate-Distortion performance than the system with one SI method without JND model. It is demonstrated in the experiment that when applying JND method in system, performance will be better compared with no JND method system which means the scheme we put forward is feasible. Simultaneously, compared with two SIs scheme, one SI scheme with JND model can achieve the same effect. In summary, Rate-Distortion performance of proposed scheme is better than original scheme which means our scheme has reached the preset effect.

5. **Conclusions.** In this paper, distributed video coding system has been improved by introducing JND model. According to the experimental results, JND model has similar performance to previous methods. For different kind of side information, R-D performance of DVC system is also different. Furthermore, we can also draw from the experimental results that distributed video coding system based on JND has a better performance when compared with the DVC system without JND model. What we will discuss next is to study how to make the performance DVC system based on JND better than the original system, as well as how to obtain better quality of side information.

Acknowledgment. The work is partially supported by the National Natural Science Foundation of China (No. 61402268, 61373081, 61401260, 61572298, 61601269, 61602285, 61601268), the Technology and Development Project of Shandong (No. 2013GGX10125), the Natural Science Foundation of Shandong China (No. BS2014DX006, ZR2014FM012, ZR2015PF006, ZR2016FB12) and the Taishan Scholar Project of Shandong, China.

REFERENCES

[1] J. Slepian and J. Wolf, Noiseless Coding of Correlated Information Sources, *IEEE Trans. on Information Theory*, vol. 19, no. 4, July 1973.

[2] A. Wyner and J. Ziv, The Rate-Distortion Function for Source Coding with Side Information at the Decoder, *IEEE Trans. on Information Theory*, vol. 22, no. 1, January 1976.

[3] R. Puri and K. Ramchandran, PRISM: A New Robust Video Coding Architecture Based on Distributed Compression Principles, *Proc. Allerton Conf.*, October 2002

[4] A. Aaron, R. Zhang and B. Girod, Wyner-Ziv Coding of Motion Video, *Proc. Asilomar Conference on Signals and Systems* , Pacific Grove, CA, Nov. 2002. Invited Paper.

[5] Z. Y. Wei and King N. Ngan, Spatio-Temporal Just Noticeable Distortion Profile For Grey Scale Image/Video in DCT Domain *IEEE Trans. Circ. Syst. Video Tech.*, vol. 19, no. 3, March 2009.

[6] A. Aaron, S. Rane, B Girod, Wyner-Ziv vidoe coding with Hash-based motion compensation at the receiver, *Proceeding of the IEEE International Conference on Image Processing* : (ICIP 04) Vol.5, Oct 24-27, 2004, Boston, MA, USA. Piscataway, NJ, USA: IEEE, 2004: 3097-3100.

[7] X. Fan, O. C. Au, N. M. Cheung. Transform-domain adaptive correlation estimation (TRACE) for Video Technology, 2010, 20(11): 1423-1436.

[8] R. Hansel, E. Muller, Improved adaptive temporal inter-/extrapolation schemes for distributed video coding, *Proceeding of the Picture Coding Symposium*,(PCS12), May 7-9, 2012, Krakow, Poland.Piscataway, NJ, USA: IEEE, 2012:213-216.

[9] D. Kubasov, J. Nayak and C. Guillemot, Optimal reconstruction in Wyner-Ziv video coding with multiple side information, *EEE Multimedia Signal Processing Workshop*, Oct. 2007.

[10] K. Misra, S. Karande and H. Radha, Multi-hypothesis distributed video coding using LDPC codes, *Proc, Allerton Conference on conmmunication, control and computing*, Sep. 2005.

[11] L. L. Meng, J. X. Zong, Bayesian Multi-Hypothesis Wyner-Ziv Video Coding, *Journal of Information Hiding and Multimedia Signal Processing*, vol. 8, no. 2, March 2017.

[12] C. H. Chou, C. W. Chou, A perceptually optimized 3-d subband codec for video communication over wireless, *IEEE Trans. Circ. Syst. Video Tech*, 6(2) (1996) 143-156.