Regular Article

Improving TDWZ Correlation Noise Estimation: A Deep Learning based Approach

Tien Vu Huu[1], Thao Nguyen Thi Huong[1], Xiem Hoang Van[2], San Vu Van[1]

[1] Posts and Telecommunications Institute of Technology, Hanoi, Vietnam
[2] University of Engineering and Technology, Vietnam National University, Hanoi, Vietnam

Correspondence: Tien Vu Huu, tienvh@ptit.edu.vn
Communication: received 3 May 2020, revised 19 May 2020, accepted 21 May 2020
Online publication: 10 June 2020, Digital Object Identifier: 10.21553/rev-jec.254
The associate editor coordinating the review of this article and recommending it for publication was Prof. Vo Nguyen Quoc Bao.

***Abstract*– Transform domain Wyner-Ziv video coding (TDWZ) has shown its benefits in compressing video applications with limited resources such as visual surveillance systems, remote sensing and wireless sensor networks. In TDWZ, the correlation noise model (CNM) plays a vital role since it directly affects to the number of bits needed to send from the encoder and thus the overall TDWZ compression performance. To achieve CNM with high accurate for TDWZ, we propose in this paper a novel CNM estimation approach in which the CNM with Laplacian distribution is adaptively estimated based on a deep learning (DL) mechanism. The proposed DL based CNM includes two hidden layers and a linear activation function to adaptively update the Laplacian parameter. Experimental results showed that the proposed TDWZ codec significantly outperforms the relevant benchmarks, notably by around 35% bitrate saving when compared to the DISCOVER codec and around 22% bitrate saving when compared to the HEVC Intra benchmark while providing a similar perceptual quality.**

***Keywords*– Transform domain Wyner-Ziv video coding (TDWZ); correlation noise model (CNM); deep learning (DL); DISCOVER CODEC, High Efficiency Video Coding (HEVC).**

1 Introduction

In conventional video coding standards, such as H.264/AVC [1] and HEVC [2] the compression performance is obtained by exploiting spatial and temporal redundancies. However, due to the complicated motion estimation process, the encoder usually has high complexity. On the contrary, the decoder is very light because the original video is simply reconstructed by following the instructions of the received information. This architecture is naturally designed for downlink applications in which the video sequence is encoded once and decoded many times. Clearly, this becomes disadvantageous for uplink applications such as wireless sensor networks and surveillance systems (see Figure 1) in which many encoders deliver data to a central decoder and devices only have constrained resources in terms of battery and processing capability. To overcome this problem, some researches focus on low complexity video algorithms for predictive video coding [3, 4].

Another approach to meet this scenario is distributed video coding (DVC) which has been introduced in the last decade [5–8]. DVC is developed based on the Slepian-Wolf [5] and Wyner-Ziv [6] theorems. The Slepian-Wolf theorem states that when two statistically dependent signals are independently encoded but jointly decoded, the same rate is achieved when compared to jointly encoded and decoded systems. The Wyner-Ziv theorem, an extension of Slepian-Wolf for the lossy compression, becomes the theoretical basis for distributed video coding. Based on this concept, the high burden motion estimation part can be shifted from the encoder to the decoder.

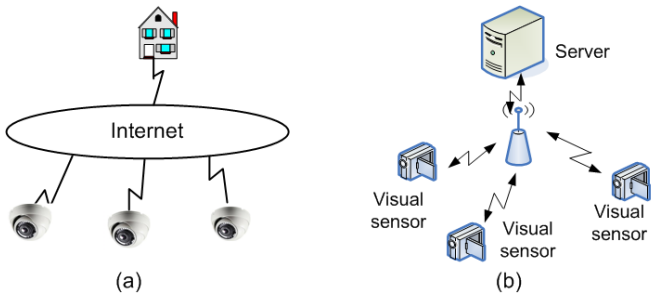

Figure 1. Examples of uplink video applications: (a) surveillance system; (b) wireless sensor networks.

Following the theoretical developments, some practical Wyner-Ziv (WZ) video codecs have been introduced [7, 8]. One of the most popular DVC approaches is the Stanford DVC codec [8], proposed by Stanford University using a feedback channel to support the decoding process. Later, hundreds of advances have been proposed by many researchers in order to improve DVC rate-distortion (RD) performance. So far, DISCOVER project [9] has been commonly used as the performance benchmark in DVC research community. In this architecture, video frames are split into key frames and WZ frames. While the key frames are encoded by using predictive coding solutions, the WZ frames are encoded by using distributed coding principles. Parity bits for WZ frames are generated by using channel

encoder and only these parity bits are transmitted to the decoder while the systematic bits are eliminated. In order to reconstruct the original WZ frames at the decoder, its estimation called side information (SI) is created. In this case, SI is considered as a noisy version of the WZ frames and its errors can be corrected by using parity bits received. Therefore, the compression performance of DVC codec is improved if the difference or noise between the created SI and the original WZ frame is estimated more accurately. However, noise correlation modeling is very complicated because the SI is only available at the decoder while the original WZ frame only exists at the encoder. In addition, SI quality changes frame by frame and even within each frame. In other words, estimating the distribution of noise needs take into account non-stationary characteristics, both in the temporal and spatial direction.

In the literature, correlation noise modeling parameters can be estimated based offline or online processing. Offline CNM estimation [8–10] refers to the case CNM parameters are estimated at the encoder using the original WZ frame and online CNM estimation [11–13] means that CNM parameters are estimated at the decoder without using the original WZ frame. Although offline approaches give better RD performance than online approaches but it receives little attention because it is an undesirable scenario. The encoder must perform the complex motion estimation to create the SI as the decoder and consequently encoder complexity is increased. Another approach direction on estimating correlation noise is proposed in [14–16]. In these works, the correlation noise model determines the number of least significant bit (n_{LSB}) bitplanes which is encoded and transmitted to the decoder and n_{LSB} is computed at both the encoder and the decoder. While in [14], an asymmetric CNM solution in which n_{LSB} is seperately computed at encoder and decoder by different SI generation solutions has been proposed, the solution in [15] uses the same way in determining correlation information at both encoder and decoder. In order to avoid the correlation information mismatch between the encoder and decoder but keeping the low complexity encoder, adaptive CNM is proposed in [16] using rate distortion optimization approach. This helps the codec maintain the low complexity while providing the better RD performance.

In DVC codecs, CNM is usually modeled by the Laplacian distribution [17] because it provides the balance between complexity and model accuracy. In order to further explore the noise correlation, several distributions have been examined in the past. In [18], an exponential power model, sometimes named "Generalized Gaussian", is used. Another approach [19] proposes a combination of two distributions to model the noise distribution adaptively upon the content of video sequence. In this work, Laplacian distribution is still used for AC coefficients but DC coefficients use alternatively Gaussian distribution and Laplacian distribution for low motion frames and high motion frames, correspondingly. To estimate CNM more precisely, the parameter of CNM is continuously updated after decoding each bitplane or band [20, 21]. That is because the more information obtained from previously decoded bitplanes/bands is exploited for decoding next bitplanes/bands. The authors in [22] propose a clustering method for DCT blocks to estimate the Laplacian parameter of CNM. Results showed that although the proposed method is performed on cluster level, it can outperform the noise model at coefficient level.

Recently, neural networks have been applied and obtained significant success in many areas including video compression. For traditional video compression algorithms, a lot of neural network based methods have been proposed for particular modules such as intra prediction and residual coding [23], entropy coding [24] in order to improve the performance of system. For distributed video coding, several deep learning based SI generation methods [25, 26] have been proposed. Authors in [25] use a deep belief network with four 16×16 key frame blocks as the input blocks to predict the side information. In [26], extreme learning machine neural network is used to estimate transformed coefficients of the WZ frame. These proposed SI generation schemes have obtained improvements in terms of both qualitative and quantitative measures.

With remarkable results of using neural networks for video compression, this paper aims to exploit strong abilities of neural networks for further performance enhancement of the TDWZ codec. In this paper, deep-learning based correlation noise modeling (DL-CNM) technique that estimates CNM parameters at band level is proposed. The learning process is carried out on the residual frame which is created based on the decoded key frames at the decoder. Experimental results shown that the advanced TDWZ with DL-CNM significantly outperforms the relevant benchmarks, notably by around 35% bitrate saving when compared to the DISCOVER codec and around 22% bitrate saving when compared to the HEVC Intra benchmark while providing a similar perceptual quality.

The rest of this paper is structured as follows. In Section 2, the proposed transform domain Wyner-Ziv architecture is presented. The proposed deep learning based correlation noise modeling method is introduced on Section 3. The experimental results and analyses are described in Section 4. Finally, the conclusions are presented in Section 5.

2 Architecture of the Proposed Transform Domain Wyner-Ziv Codec

The transform domain Wyner-Ziv codec proposed and used to evaluate in this paper is depicted in Figure 2 with the novel modules highlighted. Basically, it follows the structure of DISCOVER DVC codec [9] with the exception of DL- CNM block, SI generation block proposed in [27] and using HEVC Intra [28] instead of H.264/AVC Intra for key frame coding. Therefore, we will present in this Section the walkthrough of the proposed TDWZ encoder and decoder.

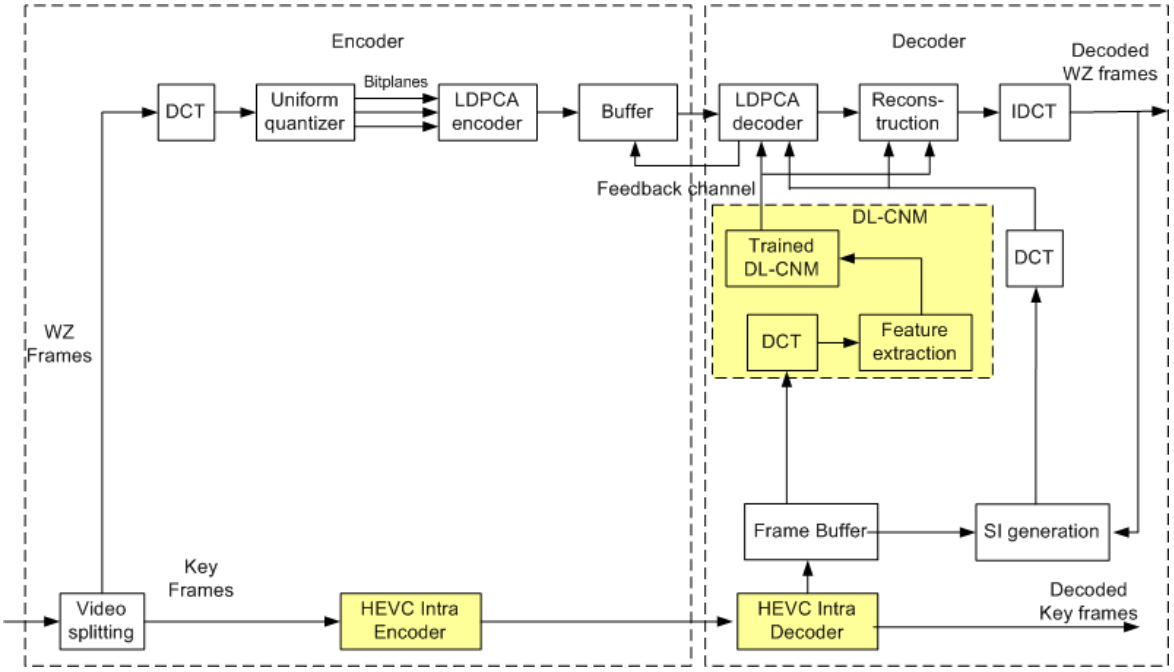

Figure 2. Architecture of proposed TDWZ codec.

2.1 TDWZ Encoder

The video sequence is divided into two kinds of frames: key frames and Wyner-Ziv frames. In this paper, the size of Group of Pictures (GOP) is equal 2, this means there is one WZ frame between two key frames.

The key frames are intra encoded where the temporal redundancy is not exploited by using a conventional video coding standard which is adopted as commonly used in DVC literatures [7, 9, 17]. Different from DISCOVER [9], key frames in this codec are coded by HEVC Intra instead of H.264/AVC Intra. Using the same principle as H.264/AVC Intra coding, HEVC Intra coding extends it to allow representing a larger range of textural and structural information in images [28]. HEVC Intra coding saves 22.3% bitrate compared to H.264/AVC Intra coding with the same objective quality [28]. For distributed video coding, HEVC Intra coding is also utilized and assessed in [29]. Experiments performed allow to conclude that when key frames are coded by HEVC Intra instead of H.264/AVC Intra, compression performance of the system is significantly improved. It is the reason why HEVC Intra coding is chosen for coding key frames in this paper.

The WZ frames are encoded based on distributed video coding principles. Each WZ frame is divided into block size of 4×4 and each block is transformed into the DCT coefficients by using a blockwise 4×4 DCT transform. These transformed coefficients are arranged into bands in which coefficients with the same positions from different blocks belong the same band. In this case, 16 bands are generated and uniformly scalar quantized. Quantization matrices corresponding to different rates are chosen as in [30]. Quantized DCT bands are binarized and bits with same significance are grouped into bitplanes. Bitplanes are given into low density parity check accumulate encoder (LDPCA) to generate parity bits. The parity bits, together a Cyclic Redundancy Check (CRC) computed for each encoded bitplane, are stored in the buffer. Depending on the request from the decoder through the feedback channel, parity bits are transmitted in chunks to the decoder and CRC will be used to aid the decoder in detecting errors.

2.2 TDWZ Decoder

First, HEVC Intra decoder is used to decode the key frames and decoded key frames are stored in the buffer. After that, the side information is created based on decoded key frames by using a SI generation technique. In this paper, the advanced SI generation method proposed in [27] is used. The previously decoded key frames are also used to create the residual frame that expresses the difference between the original WZ frame and corresponding SI. The detail of proposed correlation noise modeling used in this paper is introduced in Section 3. The LDPCA decoder corrects the errors in the side information by using correlation noise information and parity bits sent from the encoder. After that, the original DCT coefficients are reconstructed and then inversely DCT transformed to get the original WZ frame.

3 Proposed Deep Learning based Correlation Noise Model for TDWZ Codec

To understand the proposed DL-CNM method, this Section will start by introducing the TDWZ noise modeling. After that, it introduces the architecture of the DL-CNM and the training process. Finally, the Section will conclude by describing how to use the DL-CNM in TDWZ.

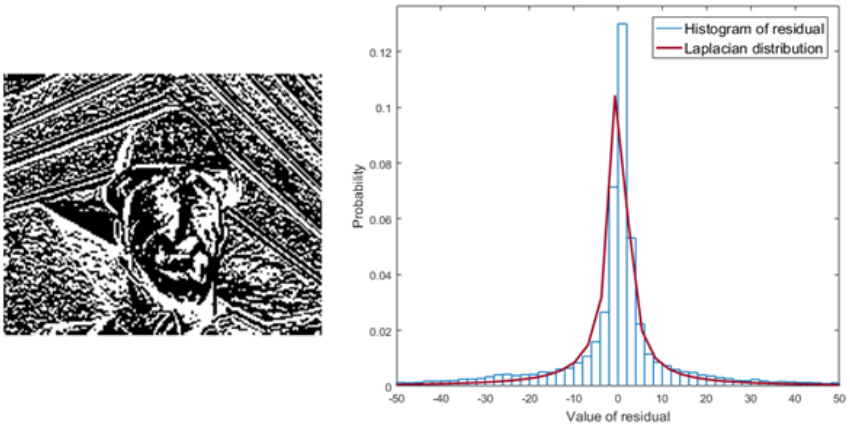

Figure 3. Example of noise distribution.

3.1 Correlation Noise Model for TDWZ Codec

In TDWZ codec, SI frame is considered as a corrupted version of corresponding WZ frame and it provides the input information for LDPCA decoder. If the estimated SI is more similar to the WZ frame, the number of errors that need to be corrected by the decoder is fewer. So, estimating the correlation noise between SI frame and the original frame is very important for the RD performance of the codec.

In the literature, this statistical noise can be modeled by various distributions such as Laplacian distribution [17], Generalized Gaussian [16] or Gaussian distribution [19]. However, Laplacian distribution is often chosen because it provides the good compromise between the model accuracy and the complexity. The residual frame $R = WZ(x,y) - SI(x,y)$ is modeled as Laplacian distribution in Equation (1) below:

$$f_R(r) = \frac{\alpha}{2} e^{-\alpha|r|}, \tag{1}$$

where $f_R(\cdot)$ is the probability density function and the Laplacian distribution parameter, α, is computed by:

$$\alpha = \sqrt{\frac{2}{\sigma^2}}, \tag{2}$$

where σ^2 is the variance of the residual frame.

For TDWZ codec, Laplacian distribution parameter α can be estimated at different granularity levels: frame level, DCT band level and coefficient level [17]. Figure 3 illustrates the real histogram of a residual frame in pixel domain for *Foreman* sequence.

3.2 Proposed Deep Learning based Correlation Noise Model

As mentioned above, the Laplacian distribution is usually chosen to estimate the correlation noise in DVC codecs. Therefore, in this work, the Laplacian distribution is also selected for modeling the correlation noise of the DCT coefficients of the residual frame. Normally, α is deduced based on the residual frame which is computed from two motion compensated key frames. However, α estimated in this manner is different from the actual value computed by WZ frame at the encoder and the SI frame at the decoder. Therefore, to improve further the correctness of α, a deep learning based correlation noise model (DL-CNM) in which inputs are features of DCT coefficients is proposed. The detail of this method is explained in the following sub-section.

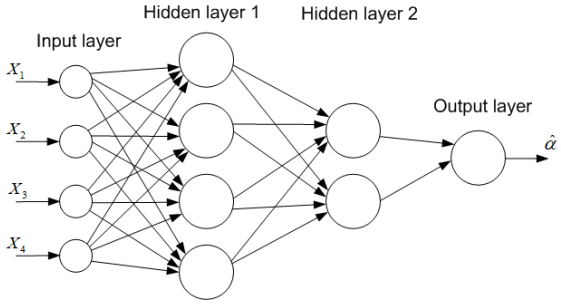

Figure 4. Architecture of DL-CNM.

3.2.1 Architecture of DL-CNM: In this study, we implement a neural network including one input layer, two hidden layers and an output layer as shown in Figure 4. The input layer with four values X_1, X_2, X_3, X_4 appropriately are four features *Min*, *Max*, *Mean*, *Variance* of DCT coefficients in a band of the residual frame. All layers in network are fully connected. In the hidden layers 1 and 2, the ReLU activation function is used. Assuming that X is the set of four inputs, $\widehat{Y}_{i,k}$ is the output at the k^{th} neuron of the i^{th} layer. The output at a neuron is computed as follows:

$$\begin{aligned} \widehat{Y}_{1,k} &= g\left(W_{1,k} * X + B_1\right), \; k = \overline{1,4} \\ \widehat{Y}_{2,k} &= g\left(W_{2,k} * \widehat{Y}_1 + B_2\right), \; k = \overline{1,2} \end{aligned} \tag{3}$$

where $W_{i,k}$ and B_i are weight matrix and bias parameters of layer i, $\widehat{Y}_1$ is a matrix of outputs at hidden layer 1. $g(\cdot)$ is activation function ReLU and is defined as follows:

$$g(x) = \begin{cases} 0 & \text{if } x < 0, \\ x & \text{if } x \geq 0. \end{cases} \tag{4}$$

At the output layer, the linear activation function is used to compute the predicted value $\widehat{\alpha}$ for a DCT coefficient band in the correlation noise frame as the follows:

$$\widehat{\alpha} = W_3 * \widehat{Y}_2 + B_3, \tag{5}$$

where W_3 is a weight matrix and $\widehat{Y}_2$ is matrix of outputs at hidden layer 2.

3.2.2 Training: To obtain the optimal Laplacian parameters, we use ten video sequences *Coastguard, Hall-Monitor, News, Container, Flower Garden, Mobile, Mother, Claire, Grandma, Harbour* having resolution of 176×144 (QCIF) and number of frames per sequence is 300 and frame rate is 15 fps. The reason to select these video sequences for training is that content of the sequences includes both low and high motion characteristics. To extract features for inputs of DL-CNM, ten sequences are HEVC Intra encoded and decoded with four quantization parameter (QP) values. The features of k^{th} frame in a video sequence are extracted at decoder as the following steps:

- *Step 1: Computing the residual frame*

$$R_k(x,y) = F'_{k-1}(x,y) - F'_{k+1}(x,y), \tag{6}$$

where (x,y) is coordinate of pixel in a frame, R_k is the residual frame, F'_{k-1}, F'_{k+1} are two motion compensated decoded key frames.

- *Step 2: Transforming the residual frame*
Residual frame R_k is divided into 4×4 blocks then DCT transform is applied block by block to obtain DCT coefficients.

$$T_k(u,v) = DCT\left[R_k(x,y)\right], \tag{7}$$

where (u,v) is coordinate of blocks 4×4 in a frame.

- *Step 3: Extracting features of a DCT transformed frame*
DCT coefficients of frame T_k are grouped into sixteen bands in which $T_{k,0}$ includes DC coefficients and $T_{k,b}$ $(b = \overline{1,15})$ refers to AC coefficients from AC_1 to AC_{15}. In each band, four features are computed as the followings:

$$\begin{aligned} X_1 &= Min\left\{T_{k,b}(i)\right\}, \\ X_2 &= Max\left\{T_{k,b}(i)\right\}, \\ X_3 &= \frac{1}{N}\sum_{i=1}^{N} T_{k,b}(i), \\ X_4 &= \frac{1}{N}\sum_{i=1}^{N} T^2_{k,b}(i) - \left(\frac{1}{N}\sum_{i=1}^{N} T_{k,b}(i)\right)^2, \end{aligned} \tag{8}$$

where b is index of bands, N is number of 4×4 blocks in a frame, i is index of coefficients in a band.

In order to extract the target values of α parameter, the WZ frame at the encoder is used together with the SI frame at the decoder to compute oracle correlation noise parameter. The target value $\overline{\alpha}_{k,b}$ for b band of k^{th} frame is extracted as follows:

- *Step 1: Computing the actual residual frame* $\overline{R}_k(x,y)$

$$\overline{R}_k(x,y) = WZ(x,y) - SI(x,y) \tag{9}$$

- *Step 2: Transforming the residual frame* $\overline{R}_k(x,y)$
The frame $\overline{T}_k(u,v)$ is created by using the Equation (7) with $R_k(x,y)$ is replaced by $\overline{R}_k(x,y)$.
- *Step 3: Computing the average variance* $\overline{\sigma}_{k,b}$
The average variance $\overline{\sigma}_{k,b}$ for b band of k^{th} frame is computed as Equation (10):

$$\overline{\sigma}^2_{k,b} = \frac{1}{N}\sum_{i=1}^{N} \overline{T}^2_{k,b}(i) - \left(\frac{1}{N}\sum_{i=1}^{N} \overline{T}_{k,b}(i)\right)^2, \tag{10}$$

where $\overline{T}_{k,b}$ is the b band of the residual frame $\overline{R}_k(x,y)$

- *Step 4: Computing the oracle value*
The oracle value is computed as followings:

$$\overline{\alpha}_{k,b} = \sqrt{\frac{2}{\overline{\sigma}^2_{k,b}}}. \tag{11}$$

After encoding and decoding ten video sequences, the deduced dataset including input and target values is fed into DL-CNM network to train. In this work, DL-CNM is implemented and trained by using Google Colaboratory [31] with 500 epochs and batch-size equals to 4. The result of training process is a set of weights.

3.2.3 Using DL-CNM in Transform Domain DVC Codec: At the decoder, the residual frame and features of DCT coefficients are computed as Equation (6), Equation (7) and Equation (8). Using the set of weights in DL-CNM learned from the above training process, the predicted parameter $\widehat{\alpha}$ corresponding to the residual frame is obtained.

Figure 5. First frames of test sequences: (a) Akiyo; (b) Foreman; (c) Carphone; (d) Soccer.

16	8	0	0
8	0	0	0
0	0	0	0
0	0	0	0

(a)

32	16	8	4
16	8	4	0
8	4	0	0
4	0	0	0

(b)

64	32	16	8
32	16	8	4
16	8	4	4
8	4	4	0

(c)

128	64	32	16
64	32	16	8
32	16	8	4
16	8	4	0

(d)

Figure 6. Four quantization matrices.

4 Experimental Results and Analyses

4.1 Experimental Setup

As mentioned above, with a low complexity encoder, DVC codec is well suited for applications such as visual surveillance networks with low resolution due to small data volumn. Therefore, in this experiment, four video sequences *Foreman*, *Akiyo*, *Carphone* and *Coastguard* with the size of 176×144 are adopted to test. These sequences are chosen because of diverse characteristics and variety of texture contents. Figure 5 illustrates the first frames of these video sequences. WZ frames of these sequences are encoded with four 4×4 quantization matrices (QM) describe in Figure 6. To achieve the similar quality of WZ frames, key frames are HEVC Intra encoded using suitable quantization

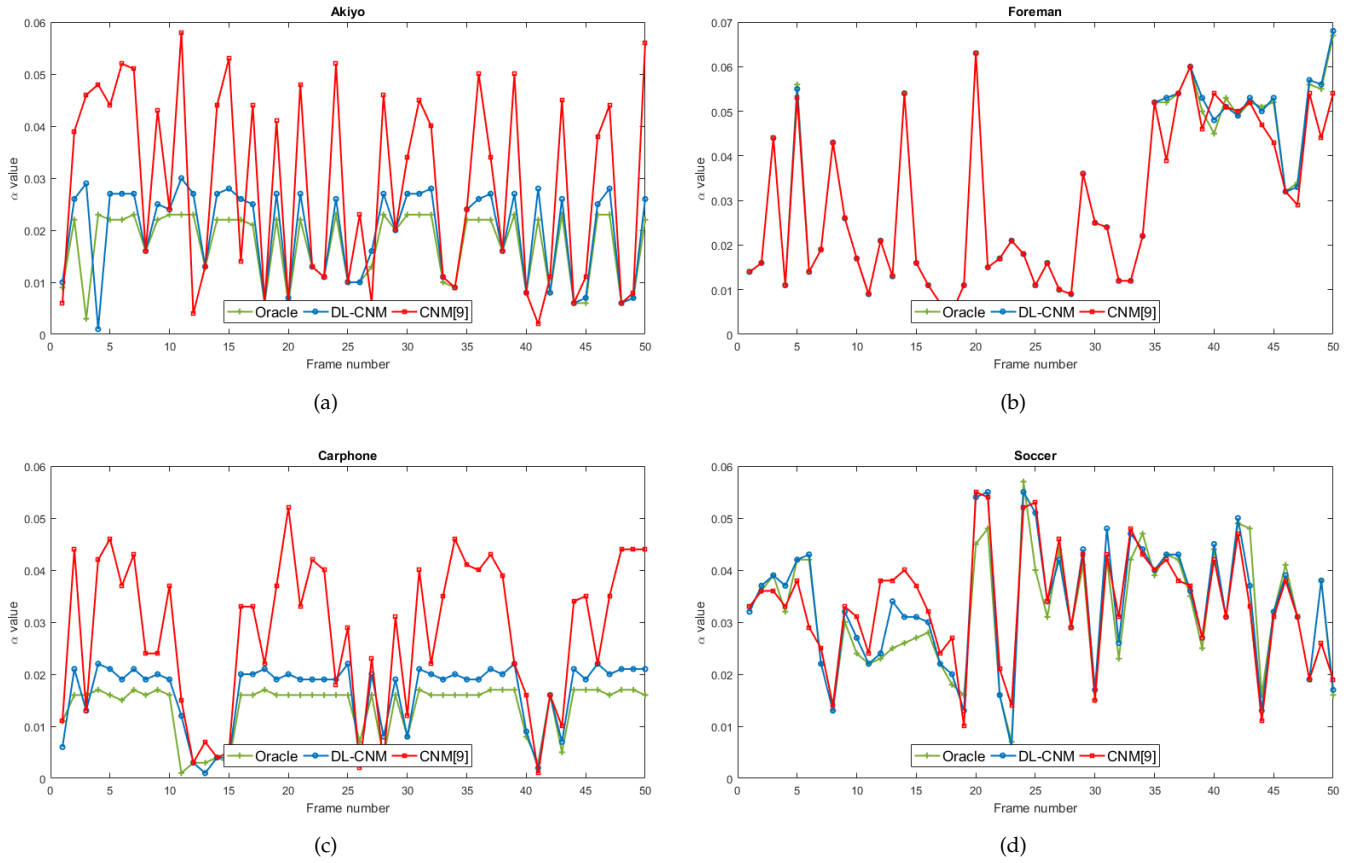

Figure 7. Comparison of estimated α values.

parameters (QP). For *Akiyo*, *Foreman* and *Carphone* sequences, $QPs = 40, 34, 29, 25$ and for *Soccer* sequence, $QPs = 44, 36, 31, 25$.

The performance of our proposed video codec named DL-CNM codec is assessed and compared with the following benchmark schemes:

HEVC Intra: This benchmark codec uses the HEVC reference software HM [32] with Intra coding mode.

DISCOVER-HEVC codec: This codec is transform domain DVC architecture DISCOVER [9] with key frames are coded by HEVC Intra instead of H.264/AVC Intra.

TDWZ codec [27]: This codec refers to TDWZ codec described in [27] in which SI generation is performed by refining progressively in decoding process.

4.2 DL-CNM accuracy assessment

In this sub-section, α parameter, which is estimated by proposed DL-CNM method and denoted by $\widehat{\alpha}$, is compared with the α parameter computed in CNM of DISCOVER codec [9]. If the estimated parameter is closer to the oracle parameter, the estimation is considered more accurately. In this assessment, four video sequences *Akiyo*, *Foreman*, *Carphone*, *Soccer* are used. Figure 7 illustrates the comparison of α parameters which are computed by CNM [9] and proposed DL-CNM method with the oracle parameter. As shown in the figures, $\widehat{\alpha}$ value estimated by DL-CNM method is closer to the target value $\overline{\alpha}_{k,b}$ than the parameter α computed by CNM [9], especially with the low motion video sequences such as *Akiyo* and *Carphone*. This shows that the proposed neural network has improved the accuracy of CNM.

4.3 Decoded Frame Quality Assessment

In this sub-section, the decoded frame qualities of the proposed TDWZ codec, measured in terms of PSNR, are compared with relevant benchmarks. A comparison of decoded frame qualities achieved with different video codecs named HEVC Intra, DISCOVER-HEVC, TDWZ [27] and DL-CNM TDWZ is presented in Table I and is illustrated in Figure 8.

- **DL-CNM TDWZ codec versus HEVC Intra:** HEVC Intra is used as a benchmark for comparison because it represents low complexity conventional video codec. As demonstrated in Table I, the proposed codec achieves higher PSNR value than HEVC Intra codec for almost sequences with exception of *Carphone* sequence. The improvements for low motion sequences and high motion sequences are different. For low motion sequences, such as *Akiyo*, the PSNR gains up to 1.37 dB but the result is not good for the high motion sequence *Carphone*. The reason is that the *Carphone* sequence is considered high motion with abrupt changes in content. In particular, in this sequence, scene changes occur at the 89^{th} and 115^{th} WZ frames. This leads to an decrease in SI quality and CNM accuracy. Consequently, the PSNR is dramatically dropped at these frames.

Table I
AVERAGE PSNR (*dB*) VALUES OF THE DECODED FRAMES

Sequence	Codec	QP1	QP2	QP3	QP4	Average
Akiyo	HEVC Intra	30.92	35.21	38.98	41.97	36.77
	DISCOVER-HEVC	28.34	32.79	36.68	40.55	34.59
	TDWZ [27]	30.97	35.53	39.98	43.74	37.56
	DL-CNM TDWZ	**31.80**	**36.39**	**40.46**	**43.91**	**38.14**
Foreman	HEVC Intra	29.18	33.08	36.66	39.71	34.66
	DISCOVER-HEVC	29.69	33.71	37.42	40.92	35.44
	TDWZ [27]	29.77	33.79	37.49	40.98	35.51
	DL-CNM TDWZ	**29.97**	**33.97**	**37.74**	**40.92**	**35.65**
Carphone	HEVC Intra	**29.94**	**34.04**	**37.73**	**40.80**	**35.63**
	DISCOVER-HEVC	26.69	31.54	34.98	38.39	32.90
	TDWZ [27]	29.31	33.01	36.34	39.68	34.59
	DL-CNM TDWZ	29.79	33.22	36.39	39.64	34.76
Soccer	HEVC Intra	28.22	32.45	35.32	39.47	33.86
	DISCOVER-HEVC	28.83	32.60	35.83	39.81	34.27
	TDWZ [27]	28.87	32.66	35.90	39.88	34.33
	DL-CNM TDWZ	**28.87**	**32.67**	**35.93**	**39.91**	**34.35**

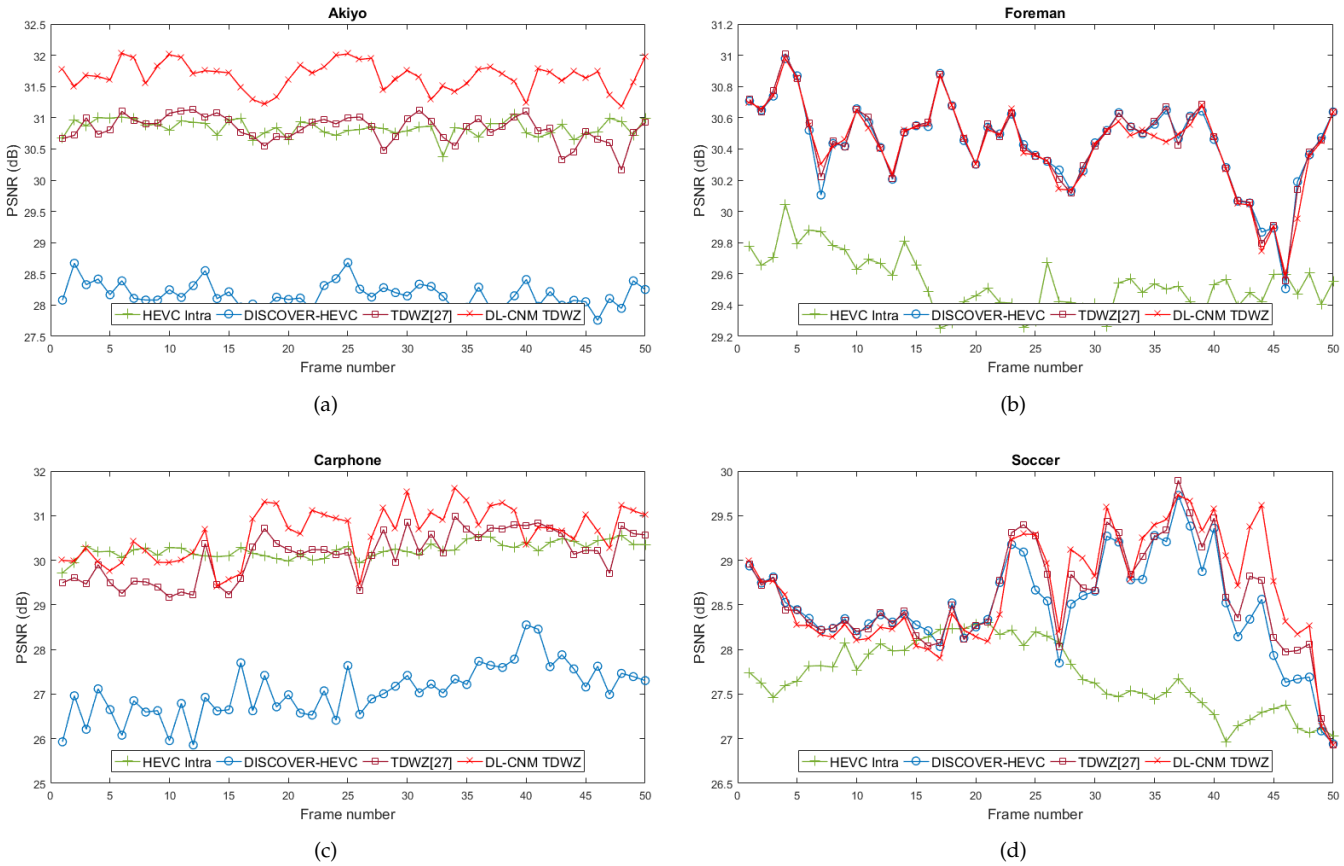

Figure 8. PSNR values of decoded frames with QP1.

- **DL-CNM TDWZ codec versus other DVC codecs:** The other DVC codecs refers to DISCOVER-HEVC, TDWZ [27]. Our proposed codec achieves better results than the others for all video test sequences. In comparison with DISCOVER-HEVC codec, the PSNR of proposed DL-CNM TDWZ codec has been improved up to 3.55 dB e.g *Akiyo* sequence. Compared with TDWZ [27] codec, similar improvements are obtained.

4.4 TDWZ Compression Performance Assessment

In this assessment, the proposed method is compared with relevant benchmarks in terms of bitrate and PSNR of each luminance frame. In addition, the

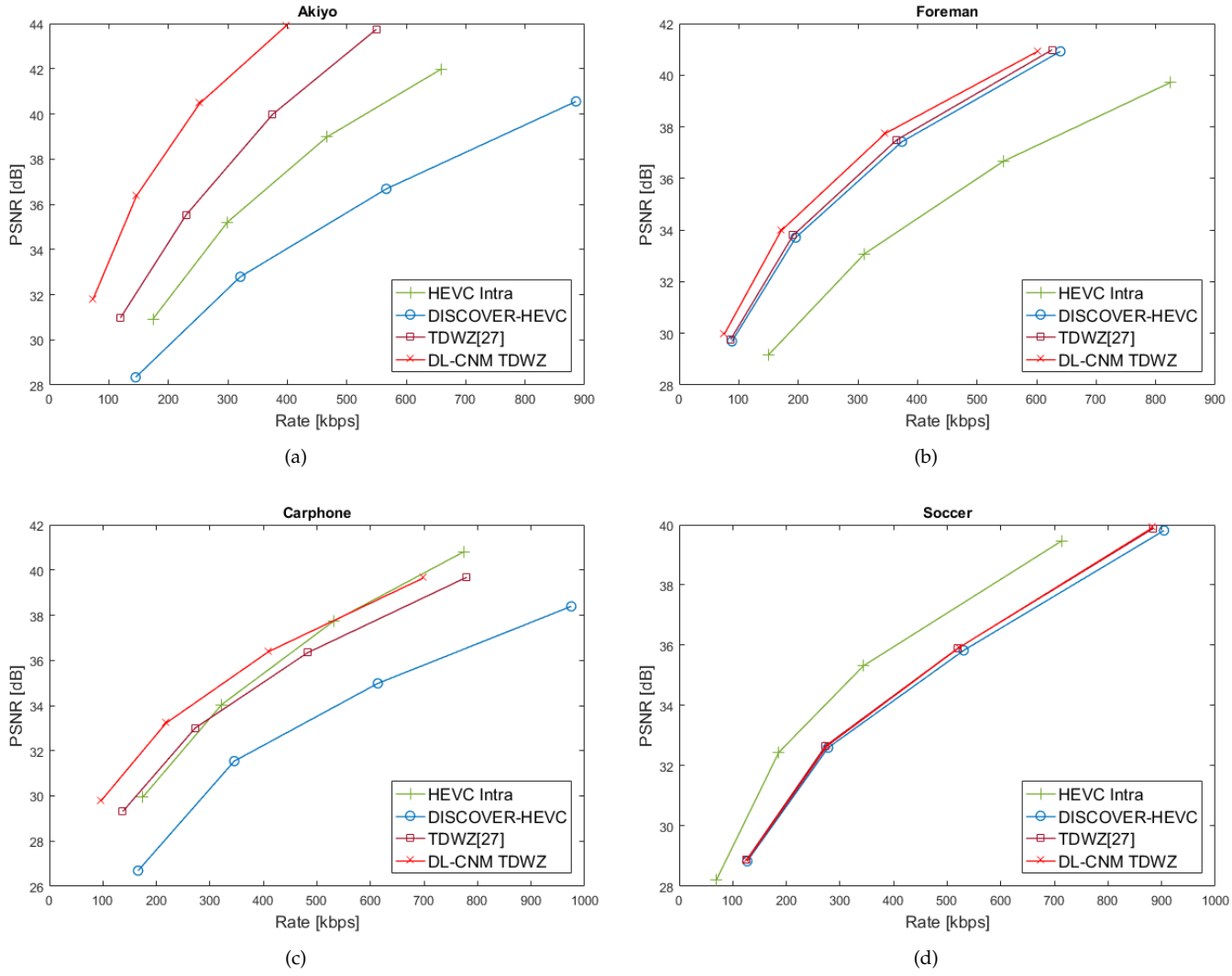

Figure 9. RD performance for the video sequences: Akiyo, Foreman, Carphone and Soccer.

Table II
A Comparison of BD Rate and BD PSNR between DL-CNM TDWZ and HEVC Intra

Sequence	DL-CNM TDWZ vs. HEVC Intra	
	BD Rate	BD PSNR
Akiyo	-57.34	6.58
Foreman	-50.59	4.00
Carphone	-17.99	0.94
Soccer	37.88	-1.62
Average	**-22.01**	**2.47**

Table III
A Comparison of BD Rate and BD PSNR between DL-CNM TDWZ and other DVC Codecs

Sequence	vs. DISCOVER-HEVC		vs. TDWZ [27]	
	BD Rate	BD PSNR	BD Rate	BD PSNR
Akiyo	-72.76	8.94	-52.62	5.37
Foreman	-14.46	0.86	-11.24	0.65
Carphone	-51.46	4.15	-20.79	1.25
Soccer	-2.43	0.14	0.52	-0.03
Average	**-35.27**	**3.52**	**-21.03**	**1.81**

Bjontegaard metrics [33] including bitrate saving (BD rate) and PSNR gain (BD PSNR) are used to compare two RD performance curves. The RD plots for *Akiyo*, *Foreman*, *Carphone* and *Soccer* sequences are shown in Figure 9. BD Rate, BD PSNR gains obtained with the proposed TDWZ codec over other benchmark schemes are presented in Table II and Table III. From the results achieved, the following observations are drawn:

- **DL-CNM TDWZ codec versus HEVC Intra:** The RD performance of the DL-CNM TDWZ codec is better than that of HEVC Intra for almost all test video sequences except the highly complex motion sequence *Soccer*. For low motion sequences, the proposed codec overcomes HEVC Intra because of good quality SI and accurate CNM. Measured by Bjontegaard bitrate metric, the proposed codec saves up to 57.34% for low motion sequences such as *Akiyo*. For four test sequences, an average 22.01% birate saving and 2.47 dB BD-PSNR gain are obtained.
- **DL-CNM TDWZ codec versus other DVC codecs:** The proposed DL-CNM TDWZ RD performance is

significantly better than the other DVC codecs for all test video sequences. RD improvements for low motion sequences are higher than for complex motion sequences. In comparison with DISCOVER-HEVC codec, BD-PSNR gain up to 8.94 dB and BD-rate reduces 72.76% for *Akiyo* sequence. For complex and high motion sequences, it is difficult in generating good quality SI and correct CNM. Therefore, it is hard to obtain such big improvement. However, our proposed codec achieved an average bitrate reduction of 35.27% when compared with DISCOVER-HEVC and 21.03% when compared with TDWZ [27].

5 Conclusion

In this work, a method to improve the accuracy of correlation noise model is proposed for transform domain Wyner-Ziv video coding. In this proposal, the α parameter is estimated by deep learning network with two hidden layers. Based on the trained model, the α parameter is predicted more accurately. The experimental results show that the proposed codec can significantly improve RD performance when compared with relevant benchmark schemes. In particular, compared with low complexity conventional video coding HEVC Intra, RD performance of our proposed codec is better for almost test video sequences, especially the low motion sequences. Compared with previous DVC codecs, such as DISCOVER-HEVC, our proposed codec can achieve significant improvements for all test sequences.

References

[1] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H. 264/AVC video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, 2003.

[2] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649–1668, 2012.

[3] N. H. Phat, V. Tran-Quang, and T. Miyoshi, "Low-complexity motion estimation algorithm using edge feature for video compression on wireless video sensor networks," in *Proceedings of the 13th Asia-Pacific Network Operations and Management Symposium*. IEEE, 2011, pp. 3–10.

[4] ——, "Video compression schemes using edge feature on wireless video sensor networks," *Journal of Electrical and Computer Engineering*, vol. 2012, pp. 1–20, 2012.

[5] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions on Information Theory*, vol. 19, no. 4, pp. 471–480, 1973.

[6] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Transactions on Information Theory*, vol. 22, no. 1, pp. 1–10, 1976.

[7] R. Puri, A. Majumdar, and K. Ramchandran, "PRISM: A video coding paradigm with motion estimation at the decoder," *IEEE Transactions on Image Processing*, vol. 16, no. 10, pp. 2436–2448, 2007.

[8] A. Aaron, S. D. Rane, E. Setton, and B. Girod, "Transform-domain Wyner-Ziv codec for video," in *Proceedings of the 2004 Visual Communications and Image Processing*, vol. 5308. International Society for Optics and Photonics, 2004, pp. 520–528.

[9] X. Artigas, J. Ascenso, M. Dalai, S. Klomp, D. Kubasov, and M. Ouaret, "The DISCOVER codec: architecture, techniques and evaluation," in *Proceedings of the Picture Coding Symposium (PCS'07)*, 2007.

[10] S. Milani and G. Calvagno, "A distributed video coder based on the H. 264/AVC standard," in *Proceedings of the 15th European Signal Processing Conference*. IEEE, 2007, pp. 673–677.

[11] C. Brites, J. Ascenso, and F. Pereira, "Improving transform domain Wyner-Ziv video coding performance," in *Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing*, vol. 2, 2006, pp. 525–528.

[12] R. Martins, C. Brites, J. Ascenso, and F. Pereira, "Statistical motion learning for improved transform domain Wyner–Ziv video coding," *IET Image Processing*, vol. 4, no. 1, pp. 28–41, 2010.

[13] C. Brites, J. Ascenso, and F. Pereira, "Studying temporal correlation noise modeling for pixel based Wyner-Ziv video coding," in *Proceedings of the International Conference on Image Processing*. IEEE, 2006, pp. 273–276.

[14] H. V. Xiem, J. Ascenso, and F. Pereira, "Correlation modeling for a distributed scalable video codec based on the HEVC standard," in *Proceedings of the 16th International Workshop on Multimedia Signal Processing (MMSP)*. IEEE, 2014, pp. 1–6.

[15] ——, "HEVC backward compatible scalability: A low encoding complexity distributed video coding based approach," *Signal Processing: Image Communication*, vol. 33, pp. 51–70, 2015.

[16] ——, "Adaptive scalable video coding: An HEVC-based framework combining the predictive and distributed paradigms," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 8, pp. 1761–1776, 2016.

[17] C. Brites and F. Pereira, "Correlation noise modeling for efficient pixel and transform domain wyner–ziv video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 9, pp. 1177–1190, 2008.

[18] C. Brites, J. Ascenso, and F. Pereira, "Learning based decoding approach for improved wyner-ziv video coding," in *Proceedings of the Picture Coding Symposium*. IEEE, 2012, pp. 165–168.

[19] T. Maugey, J. Gauthier, B. Pesquet-Popescu, and C. Guillemot, "Using an exponential power model for wyner-ziv video coding," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2010, pp. 2338–2341.

[20] H. Qin, B. Song, Y. Zhao, and H. Liu, "Adaptive Correlation Noise Model for DC Coefficients in Wyner-Ziv Video Coding," *ETRI Journal*, vol. 34, no. 2, pp. 190–198, 2012.

[21] J. Park, B. Jeon, D. Wang, and A. Vincent, "Wyner-Ziv video coding with region adaptive quantization and progressive channel noise modeling," in *Proceedings of the IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*, 2009, pp. 1–6.

[22] H. Van Luong, X. Huang, and S. Forchhammer, "Adaptive noise model for transform domain Wyner-Ziv video using clustering of DCT blocks," in *Proceedings of the IEEE 13th International Workshop on Multimedia Signal Processing*, 2011, pp. 1–6.

[23] T. Chen, H. Liu, Q. Shen, T. Yue, X. Cao, and Z. Ma, "Deepcoder: A deep neural network based video compression," in *Proceedings of the IEEE Visual Communications and Image Processing (VCIP)*. IEEE, 2017, pp. 1–4.

[24] R. Song, D. Liu, H. Li, and F. Wu, "Neural network-based arithmetic coding of intra prediction modes in HEVC," in *Proceedings of the IEEE Visual Communications and Image Processing (VCIP)*, 2017, pp. 1–4.

[25] B. Tian and W. Xiong, "A Side Information Generation method using Deep Learning for Distributed Video Cod-

ing," in *Journal of Physics: Conference Series*, no. 6, 2018, pp. 1–6.
[26] B. Dash, S. Rup, A. Mohapatra, B. Majhi, and M. Swamy, "Multi-resolution extreme learning machine-based side information estimation in distributed video coding," *Multimedia Tools and Applications*, vol. 77, no. 20, pp. 27 301–27 335, 2018.
[27] T. V. Huu, T. N. T. Huong, M. N. Ngoc, and H. V. Xiem, "Improving performance of distributed video coding by consecutively refining of side information and correlation noise model," in *Proceedings of the 19th International Symposium on Communications and Information Technologies (ISCIT)*. IEEE, 2019, pp. 502–506.
[28] J. Lainema, F. Bossen, W.-J. Han, J. Min, and K. Ugur, "Intra coding of the HEVC standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1792–1801, 2012.
[29] C. Brites and F. Pereira, "Distributed video coding: Assessing the HEVC upgrade," *Signal Processing: Image Communication*, vol. 32, pp. 81–105, 2015.
[30] R. Martins, C. Brites, J. Ascenso, and F. Pereira, "Refining side information for improved transform domain Wyner-Ziv video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19, no. 9, pp. 1327–1341, 2009.
[31] Google, colaboratory: frequently asked questions. [Online]. Available: http-s://research.google.com/ colaboratory/faq.html (accessed: 6-21-2018)
[32] HEVC reference software. [Online]. Available: https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/
[33] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves," *VCEG-M33, 13th ITU-T VCEG Meeting, Austin, TX, USA*, 2001.

Tien Vu Huu received the B. Eng. in Electrical Engineering from Hanoi University of Technology, Hanoi, Vietnam in 2002. He received the Ph.D. degree from Chulalongkorn, Thailand, in 2011. He is currently working at Multimedia Faculty, Posts and Telecommunications Institute of Technology. His research interests are digital image processing, video communications and virtual reality.

Thao Nguyen Thi Huong received the B. Eng. in Electrical Engineering from Posts and Telecommunications Institute of Technology in 2003. She is currently a lecturer and pursuing the Ph.D. degree in Posts and Telecommunications Institute of Technology. Her research interests are digital image processing, video communications.

Xiem Hoang Van is the Deputy Head, Manager of the Department of Robotics Engineering, Faculty of Electronics and Telecommunications, Vietnam National University - University of Engineering and Technology (VNU-UET). He received the Ph.D. degree (with Distinctions) from Lisbon University, Portugal, in 2015, the M.Sc. degree from Sungkyunkwan University, South Korea, in 2011, and the B.E degree from Hanoi University of Science and Technology, Vietnam, in 2009, all in Electrical and Computer Engineering. He is an executive committee member of VNU-UTS Joint Innovation and Technology research center. His research interests are machine learning, image, video communications and robot vision.

San Vu Van received the Ph.D. degree from Electronics and Telecommunications Research Institute, Republic of Korea, in 2000. He is currently an Associate Professor at Posts and Telecommunications Institute of Technology. His research interests are Transmission and Digital Signal Processing.