A Side Information Generation method using Deep Learning for Distributed Video Coding

TIAN Bo, XIONG Weizhi
School of Big data, Tongren University1, Tongren 554300, China

Abstract. In order to improve quality of reconstructed frame in distributed video coding, a Deep Learning for Side Information (DL-SI) method was proposed. According to the fact that non-linear of pixel movement in a video sequence, the deep learning network was achieved and trained online. Therefore, the four key frames as the input block to the trained to predict the side information. Experiment results reveal that compared with the typical method IST-TDWZ, the peak signal noise ratio is improved by 2.4dB-3.7dB. This indicated that the quality of reconstructed video is significantly improved with deep learning.

1. Introduction
With the rapid development of communications and computer technology, Video transmission is becoming an important application under the wireless networks. The emerging device like the mobile phone, video sensor, demand low complexity encoder and high cost decoder due to the encoder has finite computation ability and memorizer space. The conventional video coding, on the contrary, has highly complex encoding. Therefore, the Distributed video coding (DVC), which based on two theorems, Slepian-Wolf and Wyner-Ziv, as a solution for wireless video transmission application.

The Wyner-Ziv coding has widespread used in current DVC. In fact, 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. And WZ theorem extends the SW theorem to a packet lossy transmission environment [11]. The original signal, which call Side Information (SI), is crucial to reconstruct the video frame. In DVC, in order to further decrease the computational of encoder, the motion estimation is performed at the decoder to generate the SI. And the quality of SI frame at the decoder significantly influences the RD performance of the DVC. At facet, the DVC has shifts the complexity from the encoder to decoder at the same time ensure the video quality. So the current research work is focus on the problem of obtaining and transmission SI.

Many method for generating accurate SI have been proposed in recent literatures. In order to classify the side information with respect to its impact on the end-to-end compression performance, a set of metrics other than the PSNR and explicitly have been designed[2], and introduce a set of new metrics that are better adapted for side information effectiveness evaluation, and that are based on a suitable power of the absolute difference between side information and the original image, or on the Hamming distance between the respective transform coefficients [3]. A novel SI fusion algorithm is proposed which motion detection is performed to extract the moving object which can be predicted by utilizing both temporary correlations and spatial correlations [4]. Side information generation using adaptive block size for distributed video coding[5]. ESI is generated based on motion vector which is estimated using log search algorithm. The motion vector is calculated from initially reconstructed key frames and WZ frames[6]. Each frame is viewed as independent sources in the above literature, and
independent coding between each frame. This coding types is the conventional intra-frame coding. In [7] SI is computed according to the time temporal difference of motion vector at decoder.

In consideration of problems that in side information generation proceeding, in this paper, a Deep Learning for Side Information (DL-SI) method was proposed. According to the fact that non-linear of pixel movement in a video sequence, the deep learning network was achieved and trained online. Therefore, the four key frames as the input block to the trained to predict the side information. And we integrate the DL-SI with the DISCOVER architecture and performance measures are computed. As far as we know, there is no side information generation method which using deep learning network. The method proposed in this paper has some reference sense in resolving problems on side information generation.

2. Deep Learning for Side Information method

2.1. Deep Learning network

Deep Belief Network (DBN) is a typical deep learning method which has significant representational power for complex structures. It has obtained superior performance in the field of image and video processing [8]. DBN using unsupervised learning algorithm to initialize parameter. It can be trained quickly and avoid falling into local minimum value while using the traditional method. The architecture of DBN is shown in the following figure.

![DBN Architecture](image)

Fig. 1. DBN architecture

DBN consists of visual layer and hidden layer, which expressed as $v = [v_1, v_2, \ldots, v_n]$ and $h = [h_1, h_2, \ldots, h_m]$, respectively. According to references [9], the Gaussian-Bernoulli RBM and Bernoulli-Bernoulli RBM has exploited in this paper, the energy function are defined as follows:
\[ E_{vb}(v, h) = -\sum_{j=1}^{N_v} \sum_{i=1}^{N_h} v_i h_{ji} w_{ji} - \sum_{i=1}^{N_v} v_i b_i^v - \sum_{j=1}^{N_h} h_j b_j^h \]  
\[ E_{vh}(v, h) = -\sum_{j=1}^{N_v} \sum_{i=1}^{N_h} v_i h_{ji} \sigma_i + \frac{\sum_{i=1}^{N_v} (v_i - b_i^v)^2}{2 \sigma_i} + \sum_{j=1}^{N_h} h_j b_j^h \]  

Where \( w_{ji} \) is the link weight of \( i \)th visual node and \( j \)th hidden layer. \( b_i^v, b_j^h \) represent the offset. Therefore, the joint probability distribution of \( (v, h) \) is defined as follows:

\[ p (v, h; \theta) = \frac{1}{Z} e^{-E(v, h; \theta)} \]

Where \( Z = \sum_{v} \sum_{h} e^{-E(v, h; \theta)} \). The updating method of node value in visual layer and hidden layer could get accordance to reference [10][11].

### 2.2. DL-SI algorithm

The proposed scheme is motivated by the fact that in a video sequence, the pixel locations are shifted frame by frame due to motion. This inter-frame pixel movement is found out to be nonlinear due to 3-D motions of an object moving back and forth in horizontal, vertical, and diagonal direction. Method using neural network to generate side information have been studied in reference[12]. But it still has a lot of weak points in accuracy and the quality of reconstructed frame. Considering the predominance of describe nonlinear system, the block diagram of side information generation method based deep learning is given in Fig.2.

![Fig. 2 Information of side information calculation](image)

Four frames from key frames \( k_{i-1} \) and \( k_{i-2} \) were used to predict the WZ frame. It can improve the prediction accuracy quality of reconstruction frame. In order to improve the efficiency of training and computing process. The 16×16 blocks was the input of network. The initial weights and biases of the artificial network are obtained by offline raining method. Therefore the WZ frame could be predicted. The steps of the DL-SI algorithm is shown as following:

step 1 The key frames \( k_{i-1} \), \( k_{i-2} \) and \( k_{i+1} \), \( k_{i+2} \) are divided 16×16 blocks;

step 2 The blocks of size 16×16 in key frame as input and block of size 16×16 as target. The
network parameter is obtained by training the deep learning network while achieve the convergence.

step 3 The decoded key frame in receiver used as the input to network, then The WZ frame is obtained.

step 4 The obtained WZ frame is using as side information in receiver.

3. experiment and result analysis

In this section, we evaluate the performance and efficacy of our proposed DL-SI algorithm. It is developed based on Stanford DISCOVER. To demonstrate the RD performance, consider the frames from the News, Foreman and Carphone video sequences. And the WZ GOP size is 2, the turbo encoder is using in the experiment. The performance of our proposed algorithm is compared with IST-TDWZ.

The PSNR performance with different bit rate for each video sequences are shown in Fig 3.

From the result, it can be seen that the DL-SI shows graceful degradation and achieve an average PSNR gain of 2.4-4.1dB over IST-TDWZ. The reason of this phenomenon is that the proposed DL-SI can improve the quality of reconstructed WZ frame with deep learning network.
The PSNR of WZ frame with different sequences is shown in Fig 4. It is very obvious that the proposed DL-SI has a superior PSNR performance when compared to IST-TDWZ.

4. Conclusions
In order to improve quality of reconstructed frame in distributed video coding, a Deep Learning for Side Information (DL-SI) method was proposed. According to the fact that non-linear of pixel movement in a video sequence, the deep learning network was exploited to predicted side information. Therefore, the train set of network is consist of four key frames. The 16×16 blocks of side information can be predicted while 16×16 blocks of key frames as input. Experiment results reveal that compared with the typical method IST-TDWZ, the peak signal noise ratio is improved by 2.4dB-3.7dB. This indicated that the quality of reconstructed video is significantly improved with deep learning.

Acknowledgements
This study was supported by the Science and Technology Cooperation Project of Guizhou province under Grant No. Qiankehe LH [2015] 7251. Science Foundation of TongRen University under Grant No. trxyDH1503.

References
[1] Nikos Deligiannis, Adrian Munteanu, Shuang Wang. Peter Maximum likelihood laplacian correlation channel estimation in layered wyner-ziv coding. IEEE Transactions on Signal Processing, 2014, 62(4): 892—904.
[2] Jerome Gauthier, Marco Cagnazzo, Beatrice Pesquet Popescu. Evaluation of side information effectiveness in distributed video coding thomas maguey. IEEE Transactions on Circuits and Systems for Video Technology, 2013, 23(12): 2116—2126.
[3] R.Puri and K. Ramchandran, PRISM: A new robust video coding architecture based on distributed compression principles, in Proc. Allerton Conference on Communication, Control and Computing, 2012: 1—5.
[4] Hui Yin, Mengyao Sun, Yumei Wang, Yu Liu. Fusion side information based on feature and motion extraction for distributed multiview video coding. 2014 IEEE Visual Communications and Image Processing Conference, 2014, 414—417.
[5] NTH Thao, VH Tien, VV San.Side information creation using adaptive block size for distributed video coding[C].International Conference on Advanced Technologies for Communications.2016,Ho Chi Minh City, Viet Nam.
[6] V. Angayarkanni,V. Akshaya,S. Radha.Distributed compressive video coding using Enhanced side information for WSN[C]. International Conference on Wireless Communications,
Signal Processing and Networking (WiSPNET), 2016, Chennai, India.

[7] Yun-Chung Shen, Han-Ping Cheng, Ji-Ciao Luo. Efficient real-time distributed video coding by parallel progressive side information regeneration. IEEE Sensors Journal, 2017, 17(6): 1872—1883.

[8] Pratik Prabhanjan Brahma, Dapeng Wu, Yiyuan She. Why deep learning works: a Manifold Disentanglement Perspective. IEEE Transactions on Neural Networks and Learning Systems, 2016, 27(10): 1997—2008.

[9] Ian McLoughlin, Haomin Zhang, Zhipeng Xie. Robust sound event classification using deep neural networks. IEEE Transaction on Audio, Speech, and Language Processing, 2015, 23(3): 540—552.

[10] Li XianHui, Yu ZhengTao, Wei SiChao, et al. List ranking method based on listwise. Pattern Recognition and Artificial Intelligence, 2015, 28(11): 977—983.

[11] Yajie Miao, Hao Zhang, Florian Metze. Speaker Adaptive Training of Deep Neural Network Acoustic Models Using I-Vectors. IEEE Transactions on Audio, Speech, and Language Processing, 2015, 23(11): 1938-1949.

[12] Suvendu Rup, Banshidhar Majhi, Sudarshan Padhy. An improved side information generation for distributed video coding. International Journal of Electronics and Communications, 2014, 68: 202—209.