

Decoder driven side information generation using ensemble of MLP networks for distributed video coding

Bodhisattva Dash¹ · Suvendu Rup¹ ·
Anjali Mohapatra¹ · Banshidhar Majhi² ·
M. N. S. Swamy³

Received: 31 October 2016 / Revised: 9 June 2017 / Accepted: 11 August 2017 /
Published online: 26 August 2017
© Springer Science+Business Media, LLC 2017

Abstract This paper proposes an ensemble of multi-layer perceptron (MLP) networks for side information (SI) generation in distributed video coding (DVC). In the proposed scheme, both three-layer and four-layer MLP structures are used to form the ensemble model. The proposed model includes four sub-modules. The first sub-module involves the training of the individual networks. The second sub-module selects ‘M’ number of trained MLPs based on the mean square error (MSE) performance metric. Next, the third sub-module involves the testing phase of each of the selected MLPs. Finally, in the last sub-module, the overall ensemble SI is generated using a dynamically averaging (DA) method. The primary goal of this work is to minimize the estimation error between the SI and the corresponding Wyner-Ziv (WZ) frame so that the overall efficiency of DVC codec can be increased. The proposed scheme is evaluated with respect to different parameters such as Rate-Distortion (RD), Peak Signal to Noise Ratio (PSNR), Structural Similarity Index (SSIM), and number of parity requests made per estimated frame. The evaluation indicates that the proposed ensemble model shows better generalization capabilities with improved PSNR (in dB) as compared to each of the individual selected networks. Additionally, the comparative analysis also exhibits that the proposed SI generation scheme generates better SI frames in comparison with the contemporary techniques. Further, using a statistical test, namely, ANOVA with significance level of 5%, it has been validated that the proposed technique yields a significant enhancement in the performance as compared to that of the benchmark schemes.

✉ Bodhisattva Dash
bdash.fac@gmail.com

¹ Department of Computer Science and Engineering, International Institute of Information Technology, Bhubaneswar, 751003 India

² Pattern Recognition Research Laboratory, Department of Computer Science and Engineering, National Institute of Technology, Rourkela, 769004 India

³ Department of Electrical and Computer Engineering, Concordia University, Montreal, QC, Canada

Keywords Distributed video coding (DVC) · Transform domain Wyner-Ziv video coding (TDWZ) · Ensemble neural network · Side information (SI) · Multi-layer perceptron (MLP) · Structural similarity index (SSIM) · Rate-Distortion (RD)

1 Introduction

Possible customary video coding solutions, particularly those accepted by different standards, follow either discrete cosine transform (DCT) or differential pulse code modulation (DPCM). An extensive computation is involved at the encoder to decide the best modes so as to predict and encode each of the macroblocks. The decoder performs the reverse operation and reconstructs the video sequence. This kind of framework is also alluded to as the predictive video coding. Recent digital video coding standards have complex encoder and a simple decoder. These standards follow a broadcast-oriented model which is mainly suitable for applications like video broadcasting, video-on-demand, Blu-Ray discs, and so on. Here, the video information is coded once but decoded many times.

On the contrary, the increase in the availability of the affordable consumer hand-held devices like mobile camera phones, low-power surveillance systems, sensor networks, multi-view image acquisition, and inter-connected camcorders, demands a low complexity encoder and a smart decoder. Again, in a realistic video communication, video encoder works in a reverse complexity mode as there exists a scarcity in the available battery power and the computational resources. Since early 21st century, a video coding archetype, usually termed as distributed video coding (DVC), derived from two important theoretical results known as Slepian-Wolf (SW) [38] and Wyner-Ziv (WZ) [41] theorems, has drawn a lot of attention. DVC shows a reverse complexity aiming towards a reduction of computational complexity from the encoder to the decoder. The SW theorem proclaims that a lossless compression can be obtained in-between two correlated information A and B , by implementing separate encoding and joint decoding. It also suggests that the minimal rate required by this scheme is equivalent to that of the technique used in the conventional approach. WZ theorem extends the idea of the lossless case to the lossy case. Here, the compression of source A is done when B is known to the decoder, with the condition that both are jointly Gaussian sequences. The source B is called as the side information (SI).

The first feasible DVC solution was developed at Stanford University [1] and is one of the most endorsed architecture in the literature. Its primary aim is to estimate frames (also called as SI) from the previous traditionally transmitted frames (also called as odd or key frames). The estimates are the replica of the remaining frames (also called as even or WZ frames). The encoding and decoding of the WZ frames can be done either in the pixel domain (PD) or the transform domain (TD) [17]. In the case of PD, the pixels are directly quantized and encoded, whereas in the case of TD, a block-based DCT is employed. The resulting DCT coefficients are then quantized and encoded.

Initially, Aaron et al. proposed the Stanford-based TD framework [3]. This framework explores the intra-frame statistical reliance and shows better coding efficiency as compared to the Stanford-based PD architecture. Further, Puri et al. proposed a suitable TD video codec entitled Power-efficient, Robust, hIgh compression Syndrome based Multimedia coding (PRISM) [34, 35]. A spatial resolution based reduction technique for the current coding standards is presented by Mukherjee and group in [30]. Currently, the most accepted DVC architecture is certainly the Stanford-based framework and its extensions. From the development of DVC until to date, the DISCOVER video codec [6] provides the most promising results in terms of the coding efficiency. It is a TD based DVC framework, which

is an extension of the Stanford-based architecture [1]. A complete assessment of the codec performance is presented in the DISCOVER project [16].

Regardless of the modern developments in DVC, the Rate-Distortion (RD) performance is yet to reach the performance offered by the customary video codecs, predominantly for the videos with acute and non-uniform motion characteristics. It is mainly due to the inferior estimation of the original even (WZ) frame at the decoder. It is also evident that the estimation quality dominates the coding efficiency and a superior estimation quality signifies a higher correlation and lower bit-rate requirement. So, to enhance the RD behavior, an ensemble of multi-layer perceptron (MLP) networks for SI creation in a DVC framework is proposed. The original SI generation module of the Stanford-based transform domain DVC architecture (see Fig. 1) is replaced by the proposed SI generation module, and the overall codec performance is then assessed. The ensemble of the neural network is a technique where the outcomes of a set of individually learned neural networks are integrated to produce a combined output [45]. The primary advantage of using the ensemble method is that the individual components tend to make errors, whereas their combined output tends to reduce the effect of the individual errors.

The article is organized as follows. Section 2 briefly outlines the operational workflow of the Stanford-based transform domain DVC architecture. Section 3 elaborates the relevant literature on SI generation in DVC. The proposed SI generation scheme using an ensemble of MLP networks is formulated and critically discussed in Section 4. The comprehensive simulation along with the results are deliberated in Section 5. Finally, Section 6 outlines the concluding remarks and scope for future work.

2 Stanford based transform domain DVC architecture

The transform domain DVC paradigm (see Fig. 1) followed in this paper is based on the original Stanford architecture [1], started in the Image group of the Instituto Superior

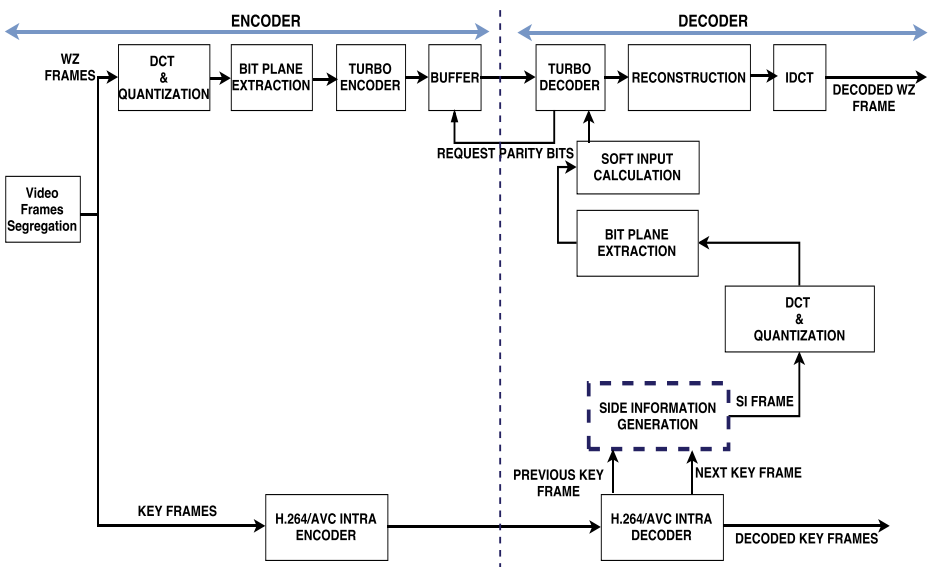


Fig. 1 Stanford-based transform domain DVC architecture

Técnico (IST). It is also termed as the IST-TDWZ framework. The operational workflow is briefly discussed below.

Usually, at the encoder side, the frames in each of the video sequences are represented either as the odd (Key) frames or the even (WZ) frames. Although the key frames are encoded using the conventional video coding standards such as H.264/AVC Intra, the even frames are encoded using the WZ video coding solutions. Initially, upon each WZ frame, a 4×4 integer block-based DCT is employed. Then, the resulting coefficients from the corresponding locations of each 4×4 block are clustered to constitute the DCT coefficient band. Further, depending on the desired output quality, each of the coefficient bands is uniformly quantized using the quantization matrices defined in [12]. After obtaining the symbol stream, a binarization technique is applied to extract the bit planes. An array of bit planes is formed starting with the most significant bit (MSB) to the least significant bit (LSB). These bit planes are then individually SW encoded using an SW encoder consisting of a turbo encoder (TE) and a parity buffer (PB).

The TE includes two similar recursive systematic convolutional (RSC) encoders and a pseudo-random interleaver. The pseudo-random interleaver is used to minimize the correlation between the inputs applied to the RSC encoders. Both the RSC encoders are considered to be of rate half and can be represented by a matrix given in (1). The encoder produces two outputs: 1) the same input sequence, and 2) a parity bit corresponding to each of the input bits in the sequence. The resulting input bit stream is rejected and the parity bit stream is kept in a buffer. Upon request by the decoder, the stored parity bits are transmitted to the decoder, chunk-by-chunk, based on a pseudo-random puncturing pattern with a span of 48.

$$\begin{bmatrix} 1 & \frac{1 + D + D^3 + D^4}{1 + D^3 + D^4} \end{bmatrix} \quad (1)$$

At the decoder, an estimation of the original WZ frame (also termed as SI) is made from the adjacent decoded key frames depending on the group of picture (GOP) size. If GOP is two, the two neighboring key frames will be the immediate past and the immediate next of the WZ frame. For larger GOP's, the already estimated and decoded WZ frames will act as the reference frames for creating and decoding the next frames. Upon the estimated frame, same steps are followed as done on the encoder side; namely, the PD to TD conversion, followed by quantization and bit-plane extraction technique. Further, the process of turbo decoding is employed. The input to the turbo decoder is the conditional bit probabilities of the obtained SI bit plane (also termed as the soft-input), starting with the MSB to the LSB. In many DVC-based solutions, the soft-input is calculated using a Laplacian distribution model that explores the interdependence between the original WZ and its corresponding estimated frame (SI) by calculating the estimation error (also called as the noise). The parameters required by the model can be calculated either in offline or online mode at different levels [11].

Starting with the MSB to LSB, each of the soft-input associated with the current bit plane is iteratively decoded using the turbo decoder. The decoder uses a logarithmic maximum a posteriori (Log-MAP) algorithm and produces the decoded quantized symbol stream with the help of the received parity bits. Further, the decoder calculates an error probability P_e for the current bit plane. If $P_e > 10^{-3}$, requests for the additional amount of party bits is made to reduce the error, or else, the current bit plane is considered to be successfully decoded. After obtaining all the decoded streams, the DCT coefficients are regenerated by applying the method presented by Kubasov et al. [26]. Upon the regenerated coefficients, an inverse discrete cosine transform (IDCT) is carried out to get the pixel values. These pixel values

are reorganized to generate the final decoded WZ frame. Lastly, all the decoded (key and WZ) frames are ordered to build the video sequence.

The overall DVC framework depends on different modules, namely, the SI generation, intra-frame coding, noise modeling, and so on. Significant efforts have been made to these modules to improve the performance of the codec. In general, from most of the observations, it is seen that the overall codec performance mainly relies on the quality of the generated SI frame. Better the equivalence between the generated SI and the corresponding WZ frame, the minimum will be the requests for parity bit by the decoder.

3 Related work on SI generation for DVC

Side information generation is considered to be one of the decisive factors that affect the overall efficiency of DVC codec. In Stanford-based DVC solutions, the estimation of the original WZ frames is done at the decoder side of the codec using the conventionally decoded key frames.

Since 2002, in DVC, a significant amount of research has been conducted so as to improve the accuracy of the estimated SI frame. Among these, Girod et al. at Stanford University, and Ramchandran et al. at the University of California, Berkeley, have formulated some of the promising SI generation frameworks. In this article, the Stanford-based DVC architecture is adopted, and some of the relevant literature in the context of SI generation based on Stanford-based DVC framework is highlighted below.

Aaron et al. presented two approaches which are based on hierarchical frame inter-dependency [2]. In the first approach, an extrapolation technique is used to generate the SI either from an odd (key) frame or an even (WZ) frame. In the second approach, an increased temporal resolution along with a bi-directional interpolation technique is employed. Both of these techniques fail to create a good quality SI when the video sequences with a longer group of pictures (GOPs) and intensified motion characteristics are used. Two other techniques, namely, the motion compensated interpolation (MC-I) and motion compensated extrapolation (MC-E) [3], have also been proposed by the same authors. In MC-I, a bi-directional block matching algorithm is used to create the SI using two previously decoded key frames at temporal index $(t - 1)$ and $(t + 1)$. The major limitation of this scheme is that it involves a high computational burden for the motion estimation task. In MC-E, the SI generation depends on the motion vectors obtained using the reconstructed WZ frame at $(t - 2)^{th}$ time instant and the decoded key frame at $(t - 1)^{th}$ time instant. The limitation of this scheme is that the error due to reconstruction process leads to a degradation in the quality of SI.

Meanwhile, Aaron et al. proposed two other schemes, namely, previous extrapolation (Prev-E) and average interpolation (AV-I) [3], to address low complexity video codecs. Brites et al. presented a frame interpolation scheme which is based on motion vectors (MVs) estimation. The MVs are estimated in two steps, 1) the forward, and 2) the bi-directional (forward and backward). This scheme also introduces the smoothing of MVs in spatial domain [7]. Further, a useful SI refinement methodology has also been proposed by the same authors in [12]. It is also referred to as Instituto Superior Técnico Transform Domain Wyner - Ziv (IST-TDWZ) codec and is adopted and extended by many DVC researchers. The same group also presented an SI generation framework in a European Project named as DIStributed CODing for Video sERvices (DISCOVER) codec [16]. Both IST-TDWZ and DISCOVER codec shows similar performance, but in DISCOVER, it employs a lower density parity check (LDPC) code, whereas the IST-TDWZ implements a turbo code.

Adikari et al. proposed a multiple SI stream method for DVC [5]. Here, two SI streams are used which are generated using the motion extrapolation and compensation (ME-C) technique. Ye et al. [43] proposed an SI generation technique that exploits the spatio-temporal dependencies between the video frames. Choi et al. [15] presented an advanced motion compensated interpolation scheme which results in an enhancement of the temporal resolution. This method overcomes the issues of traditionally overlapped block-based motion compensation. Cheng et al. proposed a classification based motion estimation scheme [14], which attempts to compute the accurate motion vectors in each block of two key frames. Further, a multiple block motion interpolation technique is also proposed which uses various motion vectors for efficient generation of SI. Ascenso et al. presented a two-mode block level based flexible structure for generating SI [8].

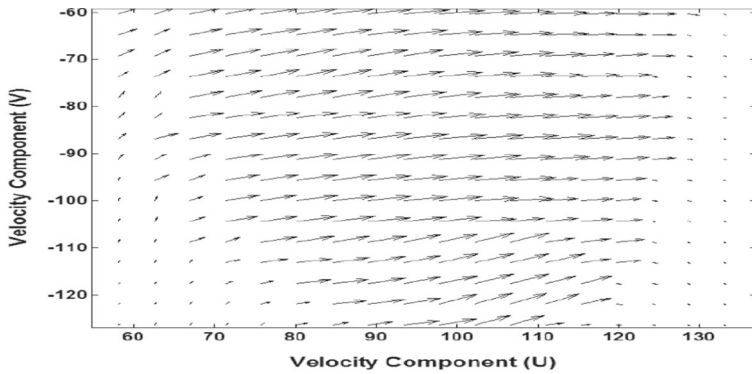
Ko et al. [24] proposed a WZ coding technique where the quality of SI can be enhanced using a side matching feature in frame interpolation. This matching feature helps in reducing the errors present in the SI, and hence, increases the codec efficiency. Hänsel et al. proposed two global motion guided adaptive interpolation or extrapolation techniques [19] which include fast camera motion characteristics with the help of a global motion estimation and refinement method. Tagliasacchi et al. focused on the effect of motion modeling for generating SI. In the proposed scheme, a Kalman filter [40] is used to improve the SI quality at the decoder. Rup et al. proposed an SI generation method [37] using multi-layer perceptron (MLP) where the WZ frames are predicted from two decoded key frames adjacent to it.

From all the above discussions, it can be noticed that SI generation is one of the most crucial tasks in DVC framework. It is due to the fact that the compression efficiency of DVC strongly depends on the correlation between SI frame and the corresponding original WZ frame. It may also be noted that the video sequences constitute non-linear motion patterns and very limited focus has been paid towards the use of soft computing based techniques like artificial neural networks (ANNs) and its variant for SI generation in DVC. With this in mind, this article proposes an ensemble of multi-layer perceptron (MLP) networks for SI frame generation in DVC.

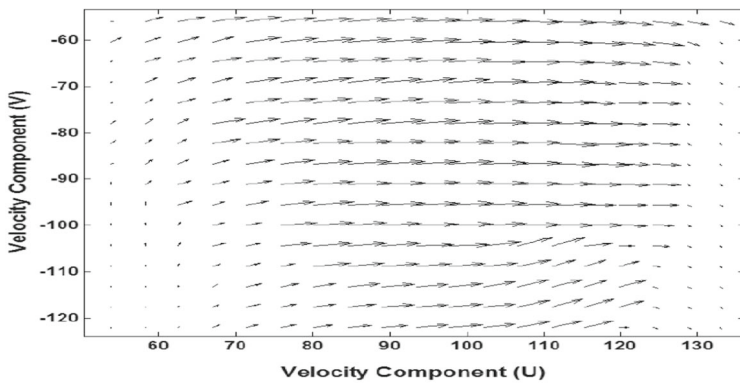
4 Proposed SI generation methodology

As discussed in Section 3, several methods have been proposed to estimate the WZ frame at the decoder of the DVC framework. However, artificial neural network (ANN) based techniques have been poorly explored for the problem under consideration. Using ANNs, different kinds of non-linearity issues that are tough using conventional methods can be solved. Most of the video sequences exhibit intra-/inter- frame variations and are subject to the effects of non-linearity [37]. Furthermore, for better clarification, the motion behavior analysis between the backward (f_{i-1}) and the present (f_i) frame, and between the present (f_i) and the forward (f_{i+1}) frame for different video sequences are studied. Figures 2 and 3 show the inter-frame pixel movements in terms of the motion vectors for *Foreman* and *Carphone* video sequences (See Table 2), respectively.

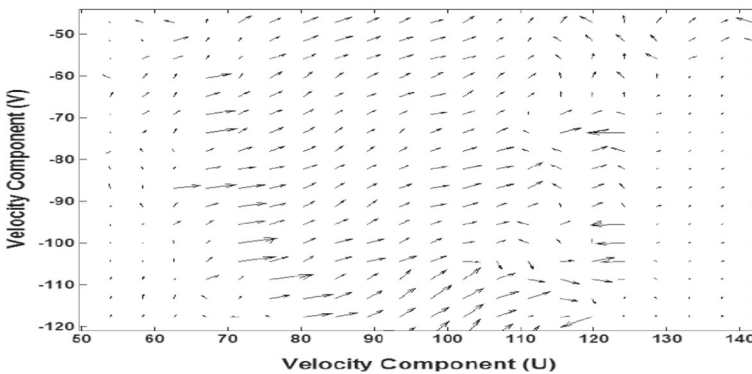
Additionally, the variation in the motion vector components for *Foreman* sequence is exhibited in Table 1. It may be noted that the difference between the motion vectors across the frames of a video sequence manifests non-linearity. Similar findings are observed with *Carphone* and other video sequences as well. So, it becomes apparent that the problem of estimating feasible WZ frames (so-called SI) can be efficiently handled using ANNs. Further, the two primary benefits of an ANN are its autodidacticism capability, and the ability to



(a)



(b)



(c)

Fig. 2 Motion vector plot for *Foreman*: (a) between 103th and 104th frame; (b) between 104th and 105th frame; and (c) difference between (a) and (b)

estimate the non-linear relationship between the input and the output of a complex network. There exist several configurations of ANN, out of which, the most popular feed-forward neural network (FFNN) [31] configuration is the MLP network. The MLP network is trained

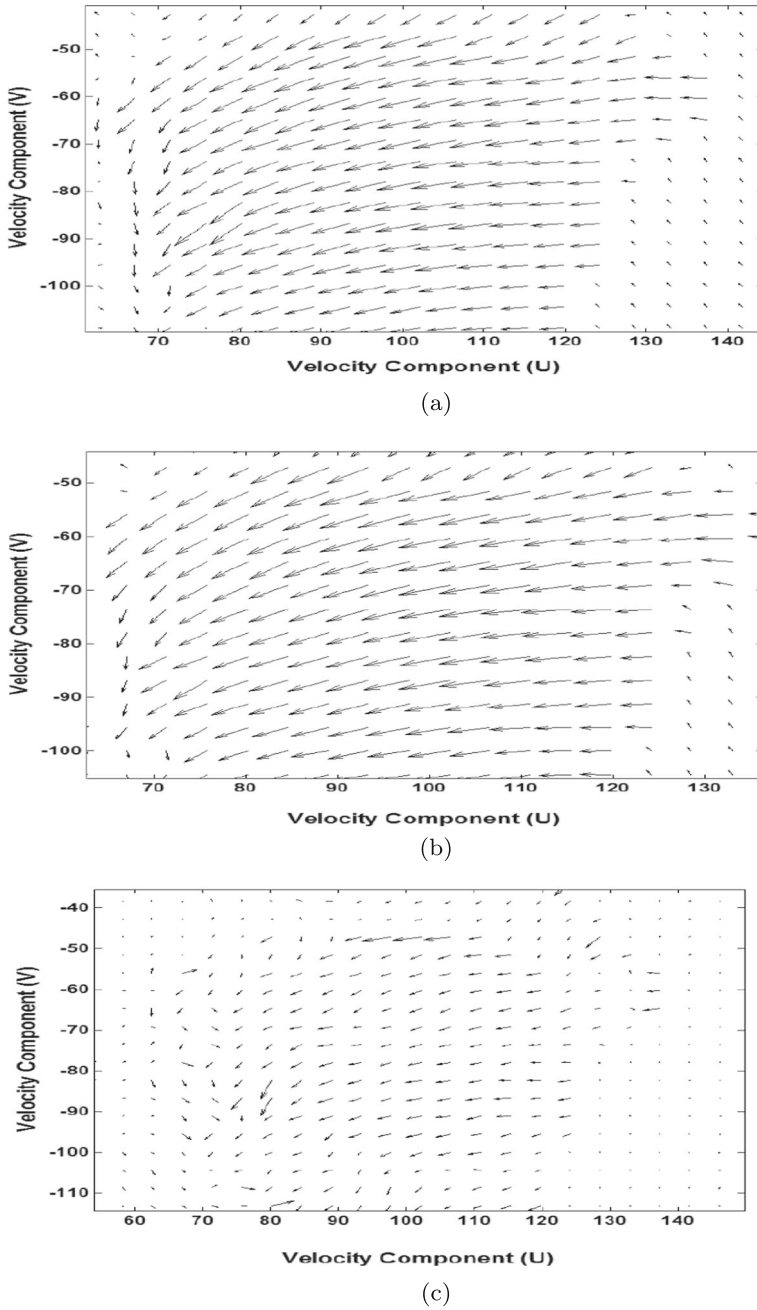


Fig. 3 Motion vector plot for *Carphone*: (a) between 117th and 118th frame; (b) between 118th and 119th frame; and (c) difference between (a) and (b)

Table 1 Variation in motion vectors in three consecutive frames of *Foreman*

Location of Motion	Motion Vector Component (U,V)	
Vector (X,Y)	Previous Frame ($(f_{i-1})^{th}$) to Current Frame ($(f_i)^{th}$)	Current Frame ($(f_i)^{th}$) to Next Frame ($(f_{i+1})^{th}$)
(52.8, -86.8)	(0.748, 1)	(0.707, 0.734)
(62.6, -78)	(0.77, 1.02)	(0.902, 0.76)
(97.8, -100)	(4.81, 0.773)	(4.24, 0.29)
(111, -118)	(3.34, 1.37)	(3.05, 1.27)
(129, -60.4)	(0.845, -0.226)	(1.25, -0.557)

for a particular task with the help of a learning algorithm, namely, the back-propagation (BP) algorithm [20]. Moreover, once the network is trained, it can be used to produce the desired outputs. For better understanding, details of an MLP network is discussed in [9, 21].

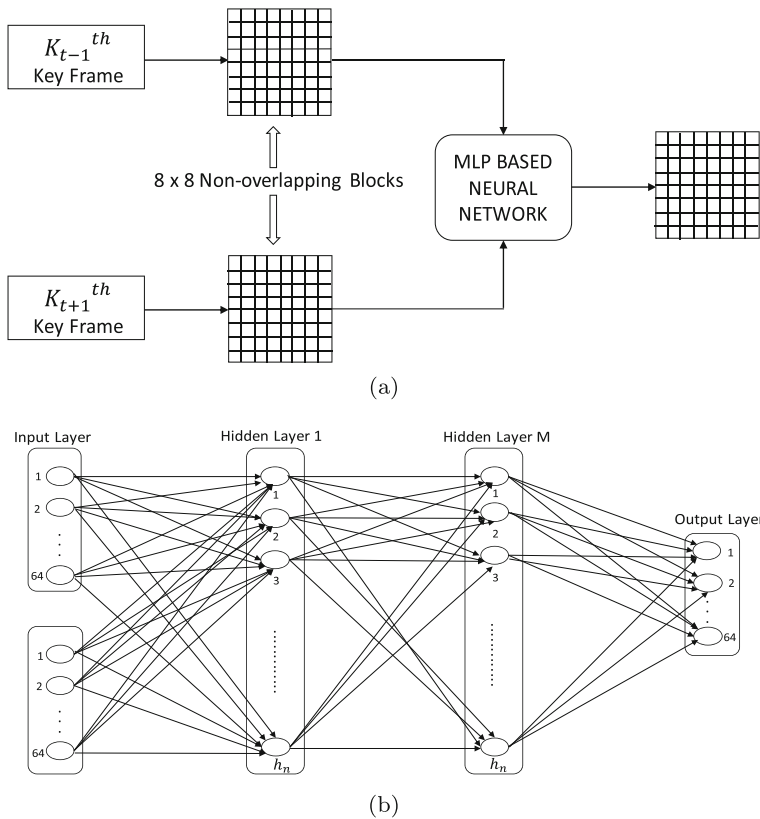


Fig. 4 Architecture of MLP network as SI frame estimator: (a) Generalized; (b) Internal structure

The basic block diagram of a generalized MLP network as SI frame estimator is shown in Fig. 4a. Figure 4b depicts the complete internal structure of the MLP network for the problem under consideration. It consists of 1) an input layer with 128 neurons representing the inputs from 8×8 block of both the key frames; 2) ‘ M ’ number of hidden layers with $h_{n_1}, h_{n_2}, \dots, h_{n_M}$ be the number of hidden neurons in each layer, respectively; and 3) an output layer with 64 neurons which represents the predicted output for the original WZ frame under consideration. Here, in this article, an ensemble of MLPs for SI generation in a DVC framework is proposed. An ensemble is a training model wherein various neural networks (NNs) trained for a particular task are grouped together, and the output of each NN is combined to generate the final ensemble output [28]. The works in [44, 45] have shown that ensemble of NN can produce a substantial enhancement in the generalization capability. Additionally, different weight initialization techniques [23, 27, 45] can be exploited to produce better accuracy as compared to the individual networks. With all the properties mentioned above, the ensemble technique becomes a new and emerging topic of interest for researchers working in many research domains like traffic flow prediction [29], pattern recognition [42], time series prediction [4], and so on.

Both theoretical and experimental analyses [18, 25, 32, 33] have exhibited that the efficacy of an ensemble firmly relies on both heterogeneity and accuracy of the distinct networks. Hansen et al. [18] showed that ensembling of a limited number of NNs could remarkably improve the generalization capability of the overall NN structure. Islam et al. [22] proposed that a group of diversified networks with lesser accuracy can be merged into an ENN with better accuracy. Hence, these properties are explored for the SI estimation problem under consideration. Figure 5 represents the proposed ensemble of MLP networks for SI generation. The proposed model is partitioned into four sub-modules. The first sub-module involves the learning phase of each of the ‘ N ’ learners. In second sub-module, based on a mean square error (MSE) metric, ‘ M ’ number of trained NNs are selected ($M \leq N$). Here, in this work, $M < N$ is considered. Next, the third sub-module involves the testing phase of each of the selected learners. Finally, in the last sub-module, the SI generated by the individual NNs are merged to produce the eventual SI of the overall ensemble model. The merging process is based on the principle of Dynamically Averaging (DA) [28] as shown in (2).

$$(SI)_f = \sum_{m=1}^N [\beta_m (SI)_m] \quad (2)$$

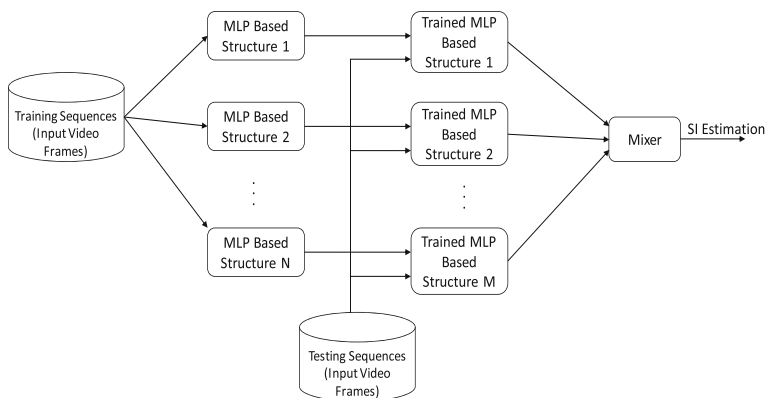


Fig. 5 Proposed ensemble model using MLP networks

where β_m 's represents the weight assigned to each of the selected networks, and are computed using (3).

$$\beta_m = \frac{\left(\frac{1}{(RMSE)_{out(m)}}\right)}{\sum_{n=1}^N \left(\frac{1}{(RMSE)_{out(n)}}\right)} \quad (3)$$

The detailed procedure involved in the training and testing phase of the proposed SI generation scheme is presented in Algorithm 1 and Algorithm 2, respectively.

Algorithm 1 Training Phase of the Proposed Scheme

Input : *Epochs* : Total number of epochs; μ : Learning Parameter; N_B : Total number of training blocks;
 N_M : Total number of MLP networks; *PB* : Previous Block; *NB*: Next Block; *TB*: Target Block;

Output: *Net* : Individual Trained Networks; $RMSE_{out}$: Root Mean Square Error (1: N_M);

```
// Function GenNet() generates the MLP network;
// Function WgtIni() initializes random weights for the
// created network;
// Function GenTrainPat() generates the training pattern;
// Function FwdPass() calculates the output of the created
// network;
// Function BwdPassErr() calculates the error of the
// created network;
// Function WgtUpdt() updates the weights of the network
// based on the calculated error;
// Function CalRMSE() computes the Root Mean Square
// Error.
// Function FuncSelectNet() returns the final selected
// networks to form the ensemble network.
```

```
1 for i ← 1 to  $N_M$  do
2    $Net(i) \leftarrow GenNet(i)$ ;
3    $(Net)_{wgt}(i) \leftarrow WgtIni(Net)$ ;
4   for j ← 1 to Epochs do
5     for k ← 1 to  $N_B$  do
6        $B_P \leftarrow Read(PB)_k$ ;  $B_N \leftarrow Read(NB)_k$ ;  $B_T \leftarrow Read(TB)_k$ ;
7        $TrainPatterns \leftarrow GenTrainPat(B_P, B_N, B_T)$ ;
8        $NetOut \leftarrow FwdPass(Net, TrainPatterns)$ ;
9        $RMSE_1 \leftarrow CalRMSE(NetOut, B_T)$ ;
10       $BwdPassErr(TrainPatterns, NetOut, Net)$ ;
11       $WgtUpdt(TrainPatterns, NetOut, Net, \mu, (Net)_{wgt}(i))$ ;
12     $RMSE_2 \leftarrow mean(RMSE_1)$ ;
13   $RMSE_{out} \leftarrow mean(RMSE_2)$ ;
14  $FuncSelectNet(N_M, MSE)$ ;
```

Algorithm 2 Testing Phase of the Proposed Scheme

```

input :  $BS$  : Block Size;  $f_i$  : Starting frame index;  $NB$  : Total number of blocks;
         $N$  : Total number of testing frames;  $T_{PB}$  : Total number of pixels in a block;
         $Net$  : Individual Trained Networks;  $RMSE_{out}$  : Obtained Root Mean
Square Error from Algorithm 1;  $N_M$  : Total number of MLP networks;

output:  $(EstSI)_f$  : Ensembled estimated SI frame;
         $(EstSI)_1$  : Network output for each block;

// Function CalcNetOut() calculates the output from the
// trained network;
// Function FuncMerge() performs the merging operation of
// each network output to obtain the ensembled output.

1 Initialization:  $T_{PB} \leftarrow BS$ ;
2 for  $j \leftarrow f_i$  to  $N$  step 2 do
3    $IP \leftarrow Read(frame)_{j-1}$ ;  $IN \leftarrow Read(frame)_{j+1}$ ;
4    $[r, c] \leftarrow size(IP)$ ;  $NB \leftarrow (r \times c) / T_{PB}$ ;
5   for  $k \leftarrow 1$  to  $N_M$  do
6     for  $l \leftarrow 1$  to  $NB$  do
7        $S_{IP} \leftarrow Read(IP)_l$ ;  $S_{IN} \leftarrow Read(IN)_l$ ;
8        $(EstSI)_1(l) \leftarrow CalcNetOut(Net(k), S_{IP}, S_{IN})$ ;
9        $(EstSI)_2(k) \leftarrow concatenate((EstSI)_1)$ ;
10       $(EstSI)_f \leftarrow FuncMerge((EstSI)_2, N_M, RMSE_{out})$ ;

```

5 Discussion and analysis of the results

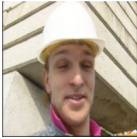
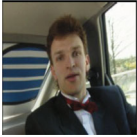




To exhibit the competency of the proposed SI generation scheme, simulations are carried out in MATLAB on different standard video sequences [10, 39] as listed in Table 2, with GOP size as 2, in all cases. During the extensive simulations, the video sequences with diversified motion characteristics and texture features are considered. To compare the overall performance of the proposed scheme, various other benchmark video codecs such as H.263 (Intra), H.264/AVC (Intra), H.264/AVC (NO MOTION), IST-TDWZ [12], and MLP-SI [37], are also simulated on the same video sequences. Comparative analyses are done with different parameters such as convergence characteristic, rate-distortion (RD), peak signal to noise ratio (PSNR), structural similarity index (SSIM), the number of parity requests made per estimated frame, and decoding time.

Further, to obtain a better clarity in the performance analysis of the proposed SI generation scheme, the overall simulation is grouped into eight different experiments. Each of the experiments is described below in detail.

Experiment 1: Convergence analysis and selection of network components for ensemble model

The primary objective of this article is to analyze the effectiveness of the proposed ensemble model in terms of the accuracy of SI estimation. An ensemble is a highly robust technique wherein the results of the independently trained networks are merged to obtain a

Table 2 Standard test video sequences

Video Sequence	Snapshot	Resolution	Frames per second	Format
<i>Foreman</i>		176 × 144	15, 30	QCIF, CIF
<i>Carphone</i>		176 × 144	15	QCIF
<i>Miss America</i>		176 × 144	15	QCIF
<i>Coastguard</i>		176 × 144	15	QCIF
<i>Kristen-Sara</i>		1280 × 720	24	SD
<i>Kimono</i>		1920 × 1080	24	HD

joint prediction. In general, a desirable performance of an ensemble model can be attained with diversified NNs. Therefore, during the training phase, it becomes crucial to maintain the heterogeneity among the individuals [13]. The heterogeneity can be attained by employing any one of the following steps: a) changing the set of initialized network weights; b) diversifying the network's internal structure by using different input and/or hidden units; c) facilitating various NN configurations; d) employing several training algorithms; e) consideration of trained networks based on randomized input sample space. In this work, twelve different MLP topologies are used by diversifying the internal structure of the MLPs, i.e., the number of hidden layers and the number of hidden units in each of the individual layers are altered, while keeping both the input units and the sample size fixed.

As far as different application scenarios are concerned, the structure of NN is an important criterion. So, an experimental procedure is adopted to select the best structure which fits the particular application. In the proposed model, both three-layer (I, H, O) and four-layer (I, H_1, H_2, O) structures are used. The tangent-sigmoid and pureline functions are

used as the activation function for the hidden and the output layer, respectively. To train the network, scaled conjugate gradient learning algorithm is used. Further, the number of hidden units (h_n) used in the single hidden layer structure are 10, 14, 15, 20, 25, and 30. Similarly, (10, 5), (14, 5), (15, 5), (20, 5), (25, 5), and (30, 5), are used in the two hidden layer structure.

For the proposed model, four different topologies, namely, (128, 10, 64), (128, 15, 64), (128, 14, 5, 64), and (128, 25, 5, 64) are selected. The selection of the topologies is made based on the MSE values. A comparative MSE-based training convergence behavior for three-layer and four-layer structures is shown in Fig. 6a–b, respectively. For each of the training samples, the pixel values from two 8×8 blocks of $(i - 1)^{th}$ and $(i + 1)^{th}$ frames (key frames) acts as the input pattern. Similarly, the pixel values from the corresponding 8×8 block of the $(i)^{th}$ frame (WZ frame) acts as the target.

To affirm that the training samples represent all possible motion and texture features, a total of 30,000 training samples are collected at a random from 40 frames of *Foreman*, 40 frames of *Container*, 30 frames of *Coastguard*, 30 frames of *News*, and 30 frames of *Kimono* video sequences. Further, in the testing phase, the remaining frames that are not included in the training phase along with other video sequences are considered to validate the performance of the proposed scheme.

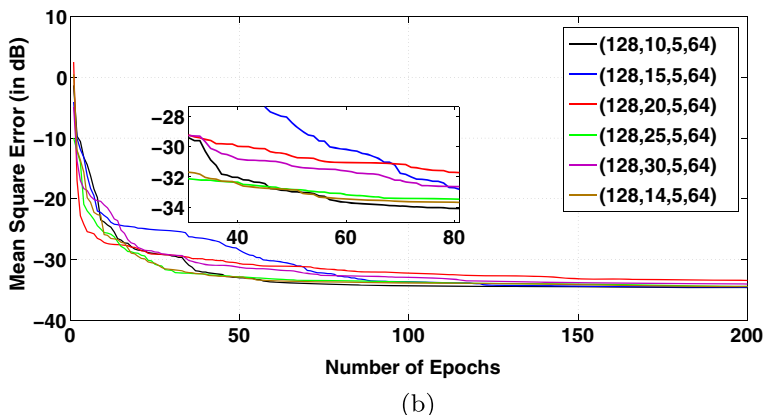
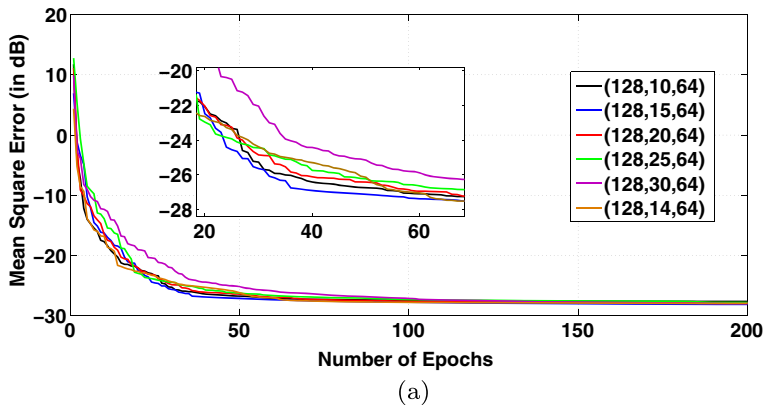


Fig. 6 Convergence characteristics for: (a) Three Layer; and (b) Four Layer

Experiment 2: Performance Analysis of SI estimation with respect to PSNR (in dB)

As discussed in the previous experiment, the primary goal of this work is to generate a better quality of SI using an ensemble of MLPs. The selected NN components are integrated to form the ensemble model which is then used to create the SI for the corresponding WZ frame. In this experiment, the quality of SI frames is measured in terms of PSNR (in dB) between the estimated SI and the original WZ frame, for different video sequences. Similar results are computed for IST-TDWZ [12] and MLP-SI [37] schemes as well.

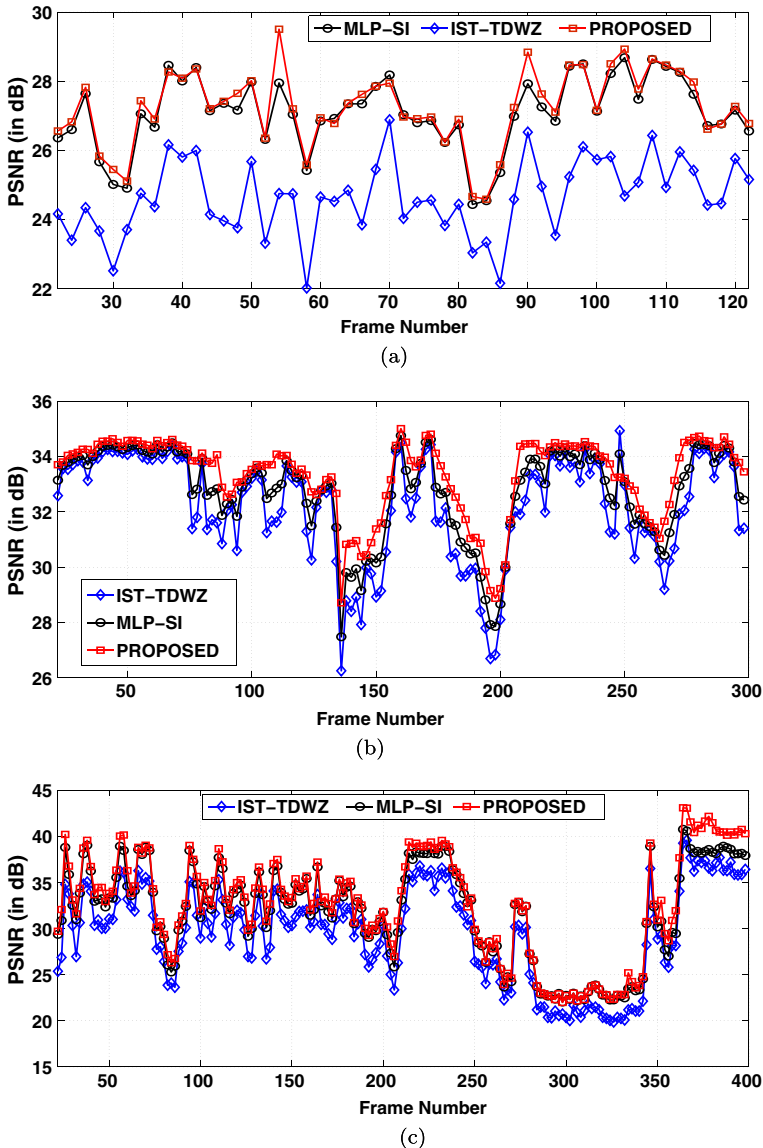


Fig. 7 PSNR (in dB) plot per estimated SI frame of: (a) *Carphone*; (b) *Kimono*; and (c) *Foreman*

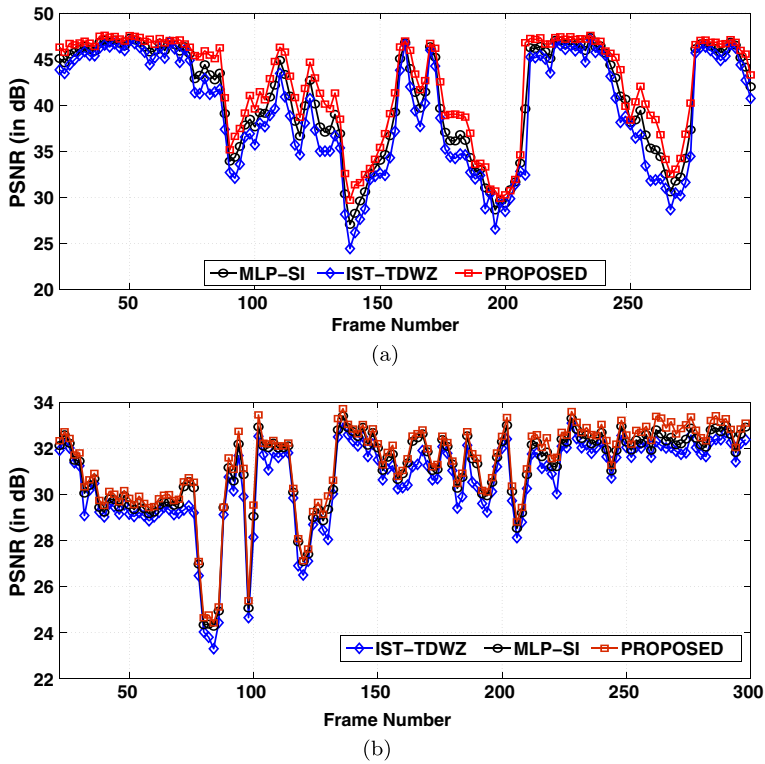


Fig. 8 PSNR (in dB) plot per estimated SI frame of: (a) *Kristen-Sara*; and (b) *Coastguard*

Table 3 Comparison of PSNR (in dB) between the ensemble model and individual MLPs

Video sequence	Network structure				Ensemble
	128:10:64	128:15:64	128:14:5:64	128:25:5:64	
Carphone	29.9	29.97	30.13	30.01	31.52
Coastguard	30.41	30.2	30.75	30.77	31.32
Foreman	29.95	30.56	30.81	31.02	32.58
Kristen-Sara	38.93	39.03	40.48	41.61	42.25
MissAmerica	41.02	40.86	40.21	41.33	41.85

Table 4 Average SSIM Values for different video sequences

Video sequence	SSIM Values		
	IST-TDWZ	MLP-SI	PROPOSED
Coastguard	0.8744	0.9004	0.9314
Foreman	0.8361	0.8522	0.9165
Mother-Daughter	0.9209	0.9759	0.9837
Carphone	0.8856	0.9083	0.9277
Kristen-Sara	0.8151	0.9740	0.9945

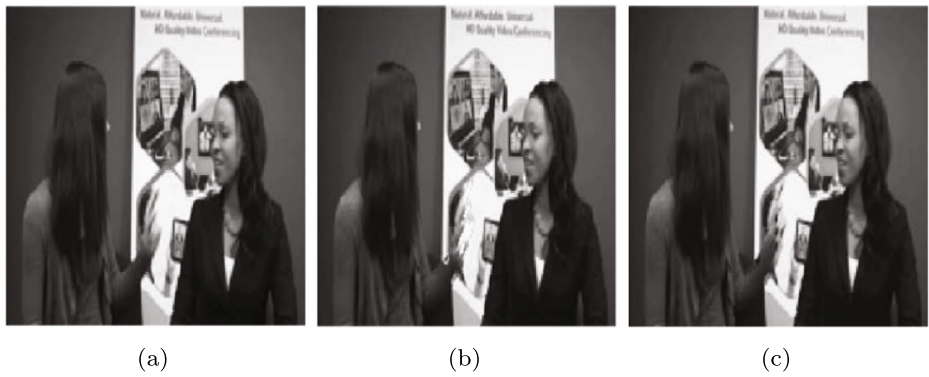


Fig. 9 140th frame of *Kristen-Sara*: (a) Original, (b) MLP-SI, and (c) Proposed

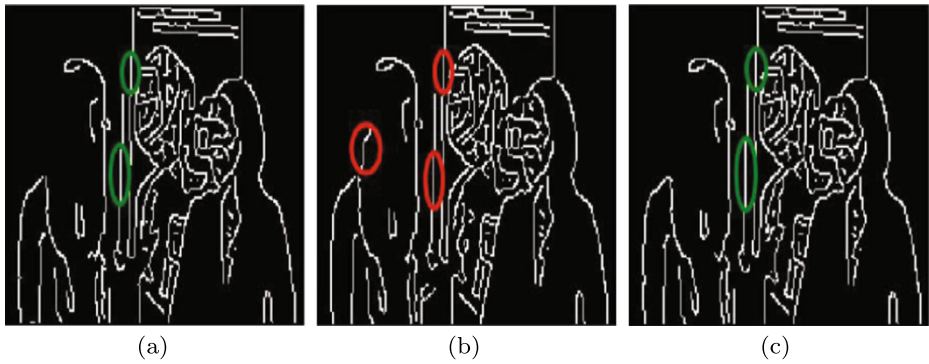


Fig. 10 Edge detected output of *Kristen-Sara*: (a) Original, (b) MLP-SI, and (c) Proposed

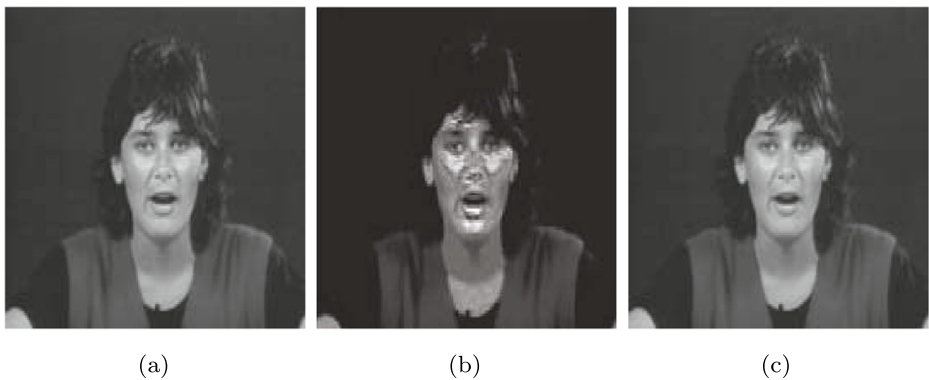


Fig. 11 22nd frame of *Miss America*: (a) Original, (b) IST-TDWZ and (c) Proposed

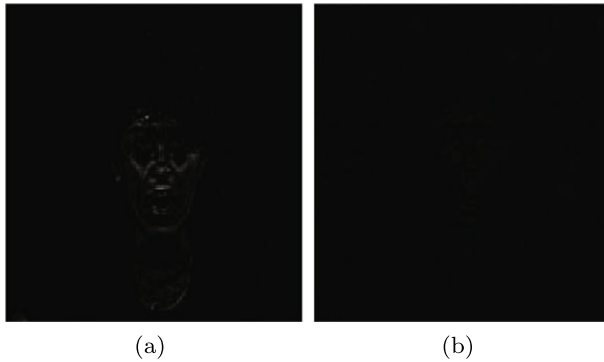
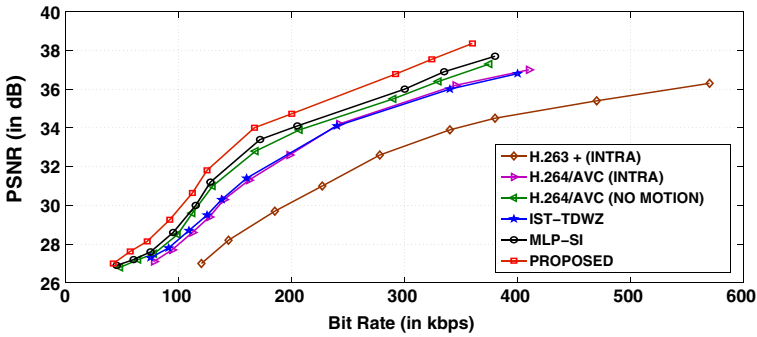
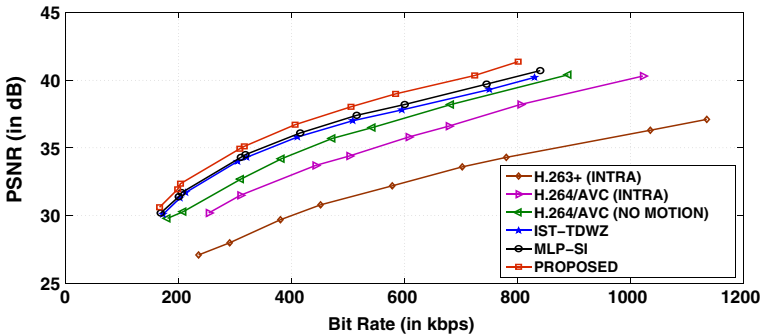


Fig. 12 Difference of *Miss America* between: (a) Original & IST-TDWZ, and (b) Original & Proposed

Figure 7a–c show the PSNR (in dB) comparison among the schemes for *Carphone*, *Kimono*, and *Foreman* video sequences, respectively. Similarly, Fig. 8a–b represent the plot for *Kristen-Sara*, and *Coastguard* video sequences, respectively. It is observed that in majority number of frames, the PSNR (in dB) with the proposed scheme is notably higher than that of MLP-SI and IST-TDWZ schemes. The results obtained reflect that the proposed



(a)



(b)

Fig. 13 Rate Distortion for *Foreman* for: (a) *QCIF* (at 15 fps); (b) *CIF* (at 30 fps)

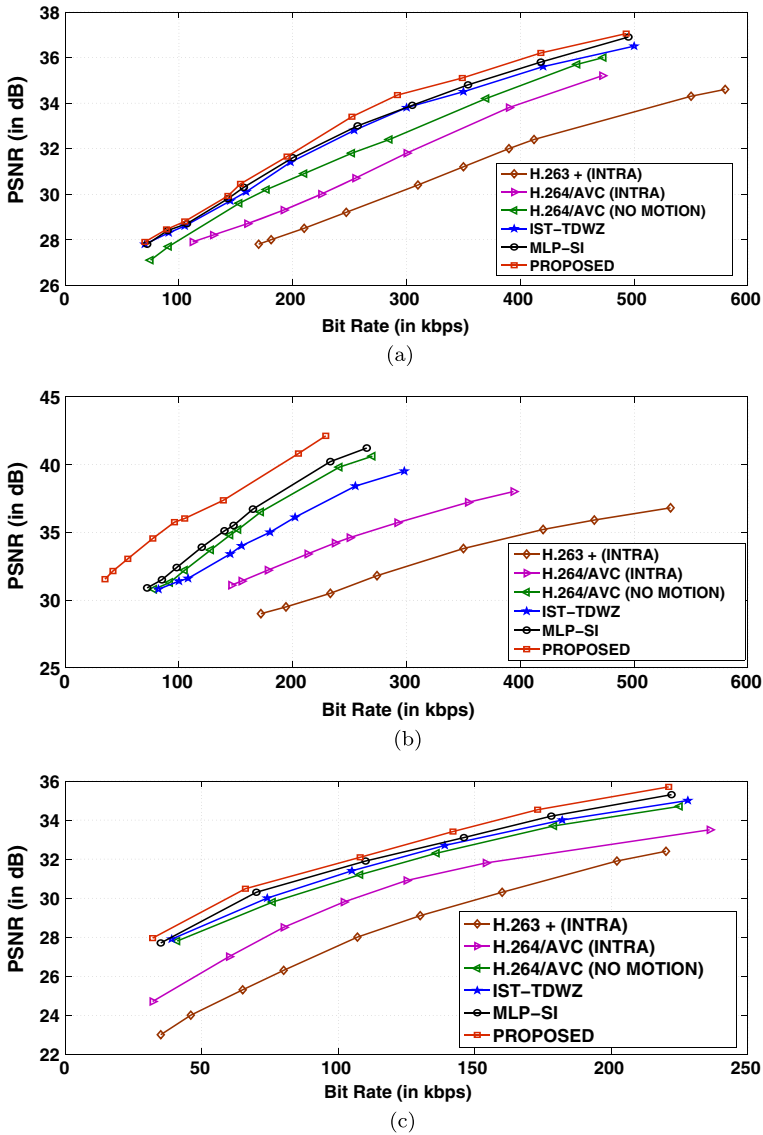


Fig. 14 Rate Distortion for *QCIF* (at 15 fps) for: (a) *Coastguard*; (b) *MissAmerica*; and (c) *Carphone*

Table 5 Average PSNR Gain (in dB) of the proposed scheme over MLP-SI scheme

Video Sequence	Average PSNR Gain (in dB) (at 15 fps)	
	Lower Bit Rate	Higher Bit Rate
Foreman	0.47	0.66
Coastguard	0.1	0.26
MissAmerica	2.24	2.98
Carphone	0.2	0.35

Table 6 Average PSNR Gain (in dB) of the proposed scheme over IST-TDWZ scheme

Video sequence	Average PSNR Gain (in dB) (at 15 fps)	
	Lower Bit Rate	Higher Bit Rate
Foreman	0.6	2.52
Coastguard	0.15	0.47
MissAmerica	2.57	4.01
Carphone	0.41	0.65

method can generate qualitative SI for video sequences with different resolution. Furthermore, it is also noticed that the ensemble model has a better generalization capability as compared to each of the individual MLP networks selected to form the ensemble model. For better understanding, a PSNR (in dB) comparison between the ensemble model and individual MLPs is shown in Table 3.

Experiment 3: Study of Perceptive Measure of Side Information

This experiment evaluates the perceptive measure in terms of SSIM for the proposed and the benchmark schemes. SSIM measures the structural similarity between two images and determines the degradation in the picture quality caused by some processing techniques like data compression or transmission. If the images are similar, then the SSIM value lies close to 1. The average SSIM obtained with the proposed and the benchmark schemes for different video sequences are listed in Table 4. From the table, it can be observed that the proposed method shows superior performance as compared to its counterparts.

For visual (subjective) analysis, the original 140th frame of *Kristen-Sara* and the corresponding estimated SI frames with MLP-SI and the proposed technique are shown in Fig. 9a–c, respectively. Figure 10a–c represent the edge detected output of the respective frames. It may be noticed that the portion marked with red color indicates the incorrect estimations with MLP-SI scheme (See Fig. 10b), and green color shows the correct estimations, as in the original frame, with the proposed method (See Fig. 10c). Similarly, Fig. 11a–c represent the original, estimated SI frame with MLP-SI and proposed. The difference between the original & IST-TDWZ, and original & proposed scheme, for 22nd frame of *Miss America* sequence, is shown in Fig. 12a–b, respectively. It may be noticed that the proposed method generates better approximation than IST-TDWZ technique does. Similar findings are also observed with other video sequences as well.

Table 7 Average PSNR Gain (in dB) of the proposed scheme over H.264/AVC (NO MOTION) scheme

Video Sequence	Average PSNR Gain (in dB) (at 15 fps)	
	Lower Bit Rate	Higher Bit Rate
Foreman	0.61	1.06
Coastguard	0.92	0.98
MissAmerica	2.41	3.31
Carphone	0.57	0.98

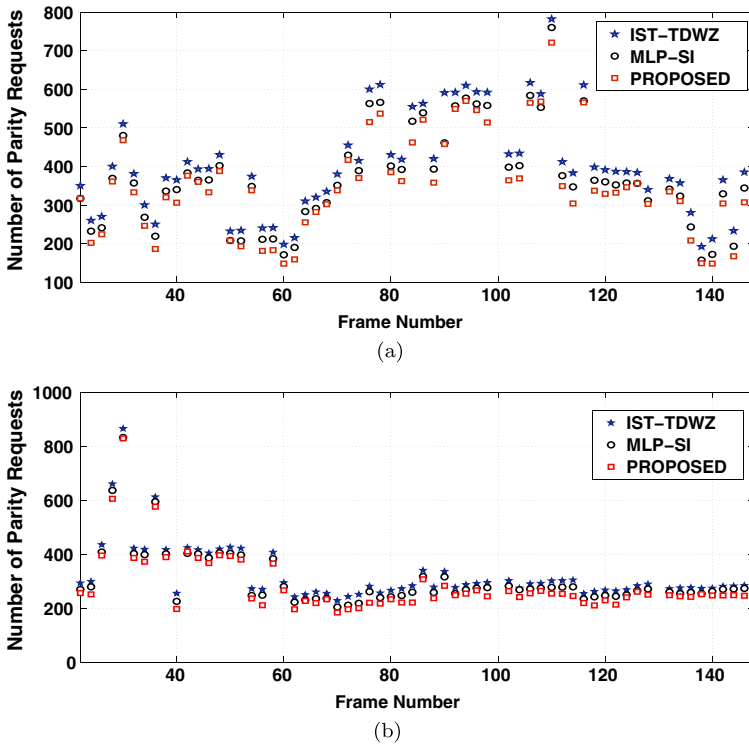


Fig. 15 Number of Parity Requests (at 15 fps) for: (a) *Foreman*; (b) *Coastguard*

Experiment 4: Study of comprehensive RD characteristics

This experiment measures the RD performance of the proposed DVC codec. For the computation of RD performance, only the *Y* (Luminance) component of the video frames is used. Figure 13 shows the RD plot for *Foreman* (at 15, 30 fps). Similarly, the RD plot for *Coastguard*, *MissAmerica*, and *Carphone* at 15 fps are exhibited in Fig. 14a–c, respectively. Additionally, the average PSNR gain (in dB) achieved with the proposed technique over MLP-SI, IST-TDWZ, and H.264/AVC (NO MOTION) schemes, for lower and higher bit rates, are shown in Tables 5, 6 and 7, respectively.

Table 8 Average PSNR (in dB) of all decoded frames for different quantization matrices (Q_i 's) of *Foreman* Sequence

Q_i	PSNR (in dB)		
	IST-TDWZ	MLP-SI	Proposed
4	32.04	32.38	32.99
5	32.13	32.41	33.04
6	33.38	33.62	34.40
7	35.56	35.84	36.49
8	37.42	37.88	38.54

Table 9 Average PSNR (in dB) of all decoded frames for different quantization matrices (Q_i 's) of *Carphone* Sequence

Q_i	PSNR (in dB)		
	IST-TDWZ	MLP-SI	Proposed
4	27.44	27.73	28.01
5	28.72	29.01	29.30
6	30.61	30.92	31.39
7	32.74	33.23	33.80
8	34.71	35.05	35.75

Experiment 5: Assessment of number of Parity Requests per SI frame

As the efficiency of the decoder is critically dependent on the number of parity requests, it becomes essential to assess the number of additional parity bits required by the decoder to correct the error that exists between the original WZ and the generated SI frame. Figure 15a–b represent the number of requests initiated per SI frame with the proposed, IST-TDWZ, and MLP-SI schemes, for *Foreman*, and *Coastguard* sequences, respectively. During the experimental process, a noiseless channel is considered for transmission of parity bits.

From the experimental result, it is noticed that a maximum of 721 requests is made with the proposed scheme for the 110th frame of *Foreman* sequence, whereas a maximum number of 760, and 782 requests are made with MLP-SI, and IST-TDWZ schemes, respectively. Similarly, for *Coastguard* sequence, a maximum of 829 requests is made, in contrast to 832, and 865 for MLP-SI and IST-TDWZ schemes, respectively. In general, similar improvements are observed with the proposed SI generation method for other video sequences as well.

Experiment 6: Analysis of temporal evaluation

In DVC, to get the eventual decoded WZ frame, the error between the original WZ and the estimated SI frame is further corrected using additional number of parity bits. Higher PSNR (in dB) value reflects superior quality of the decoded WZ frame. So, to investigate the quality of the decoded WZ frames, PSNR (in dB) is considered as the performance metric.

The average PSNR (in dB) values obtained with the proposed and benchmark techniques for *Foreman*, *Carphone*, *Coastguard*, and *MissAmerica* sequences are reported in Tables 8, 9, 10 and 11, respectively. For temporal evaluation, the quantization matrices Q_4 to Q_8 are considered [12]. Moreover, for $Q_i = 8$, it is observed that the proposed SI generation scheme attains an average PSNR gain of 0.66 dB, 0.70 dB, 0.55 dB, and 0.90 dB, over MLP-SI scheme for *Foreman*, *Carphone*, *Coastguard*, and *MissAmerica* video sequences,

Table 10 Average PSNR (in dB) of all decoded frames for different quantization matrices (Q_i 's) of *Coastguard* Sequence

Q_i	PSNR (in dB)		
	IST-TDWZ	MLP-SI	Proposed
4	29.28	29.57	30.10
5	30.72	31.02	31.31
6	32.24	32.63	32.90
7	34.21	34.68	35.24
8	36.27	36.65	37.20

Table 11 Average PSNR (in dB) of all decoded frames for different quantization matrices (Q_i 's) of *MissAmerica* Sequence

Q_i	PSNR (in dB)		
	IST-TDWZ	MLP-SI	Proposed
4	29.44	29.57	30.37
5	29.72	30.01	30.65
6	32.61	32.92	33.57
7	34.74	35.23	35.82
8	36.71	37.05	37.95

respectively. Similarly, it is also to be noted that the proposed scheme exhibits an average PSNR gain of 1.12 dB, 1.04 dB, 0.93 dB, and 1.24 dB, over IST-TDWZ scheme for *Foreman*, *Carphone*, *Coastguard*, and *MissAmerica* video sequences, respectively.

Experiment 7: Assessment of Decoding Time

In DVC, the decoder complexity is considerably higher than that of the encoder. So, to assess the decoder complexity, the average decoding time requirement (in seconds) with the proposed and other benchmark schemes for the quantization matrices (Q_4 , and Q_8) [12] is reported in Table 12. It is observed that the proposed scheme requires considerably less decoding time as compared to that for other competent schemes. Similar findings are also observed for other quantization matrices as well.

Experiment 8: Statistical Analysis

Statistical analysis is a scientific approach employed to make judgments with a measurable confidence. Analysis of variance (ANOVA) is a statistical method used to verify whether the means of several groups are all equal. Initially, in ANOVA, a null and alternative hypothesis is defined. The null hypothesis states that there is no significant difference among the groups against the alternative hypothesis that there is a significant difference. The rejection or acceptance of the null hypothesis critically depends on the resulting p-value of the ANOVA test. If $p \leq 0.05$ (considered significance level of 5%), the null hypothesis fails to be accepted. Further, for better understanding, the detailed analysis of ANOVA is presented in [36].

Table 12 Comparison of Total Decoding Time (in seconds)

Sequence	Total Decoding Time (in secs)		
	IST-TDWZ	MLP-SI	Proposed
<i>Foreman</i> (Q_4)	1124	1088	837
<i>Miss America</i> (Q_4)	422	337	283
<i>Coastguard</i> (Q_4)	717	621	564
<i>Carphone</i> (Q_4)	821	788	739
<i>Foreman</i> (Q_8)	3698	3565	3498
<i>Miss America</i> (Q_8)	1207	1127	882
<i>Coastguard</i> (Q_8)	3013	2953	2908
<i>Carphone</i> (Q_8)	2805	2709	2623

Table 13 ANOVA test with respect to PSNR (in dB) for *Foreman*

Data Summary				
	Techniques			
	IST-TDWZ	MLP-SI	Proposed	Total
Sample Size	189	189	189	567
ΣX	5587.9823	6023.4823	6157.3476	17768.8122
Mean	29.566	31.8703	32.5786	31.3383
ΣX^2	170426.943	197343.3709	206797.3185	574567.6323
Variance	27.7256	28.5815	32.9777	31.3134
Std. Deviation	5.2655	5.3462	5.7426	5.5958
Std. Error	0.383	0.3889	0.4177	0.235

Standard Weighted-Means Analysis				
ANOVA Summary		Independent Techniques (k=3)		
Source	SS	df	MS	
Treatments (Between-Groups)	937.8431	2	468.9215	F = 15.76
Within-Groups	16785.5434	564	29.7616	p-value = 0.0001
Total	17723.3865	566		

In the present work, ANOVA is used to validate that the proposed method produces statistically significant enhancement as compared to the benchmark schemes with respect to different parameters like PSNR (in dB), SSIM, and so on. The detailed analysis of the ANOVA test with respect to PSNR (in dB) for *Foreman*, and *Kristen-Sara* sequence is reported in Tables 13, and 14, respectively. It is noticed that the p-values obtained (.0001 for *Foreman*, and 0.000331 for *Kristen-Sara*) is considerably less than the set significance

Table 14 ANOVA test with respect to PSNR (in dB) for *Kristen-Sara*

Data Summary				
	Techniques			
	IST-TDWZ	MLP-SI	Proposed	Total
Sample Size	139	139	139	417
ΣX	5480.3953	5671.1505	5872.2646	17023.8103
Mean	39.4273	40.7996	42.2465	40.8245
ΣX^2	221472.007	235971.3473	252090.0339	709533.3883
Variance	39.0927	33.264	29.0389	34.9642
Std. Deviation	6.2524	5.7675	5.3888	5.9131
Std. Error	0.5303	0.4892	0.4571	0.2896

Standard Weighted-Means Analysis				
ANOVA Summary		Independent Techniques (k=3)		
Source	SS	df	MS	
Treatments (Between-Groups)	552.5084	2	276.2542	F = 8.17
Within-Groups	13992.5879	414	33.7985	p-value = 0.000331
Total	14545.0963	416		

Table 15 ANOVA test with respect to SSIM for *Carphone*

Data Summary				
	Techniques			
	IST-TDWZ	MLP-SI	Proposed	Total
Sample Size	180	180	180	540
ΣX	163.500716	166.086246	159.415972	489.002934
Mean	0.908337	0.922701	0.885644	0.905561
ΣX^2	149.203226	153.733416	141.941608	444.87825
Variance	0.003852	0.002712	0.004222	0.003815
Std. Deviation	0.062061	0.052075	0.064978	0.061765
Std. Error	0.004626	0.003881	0.004843	0.002658
Standard Weighted-Means Analysis				
ANOVA Summary		Independent Techniques (k=3)		
Source	SS	df	MS	
Treatments (Between-Groups)	0.125672	2	0.062836	F = 17.48
Within-Groups	1.930597	537	0.003595	p-value = 0.00001
Total	2.056268	539		

level of 5%. Similarly, the analysis with respect to SSIM for *Carphone* sequence is shown in Table 15. The obtained p-value of 0.00001 is less than 5% significance level. Moreover, similar findings have been observed with other parameters as well. Hence, in general, it can be validated that the proposed technique produces statistically significant improvement as compared to the benchmark schemes.

6 Conclusion

In this study, an ensemble of MLP networks for SI generation in a DVC framework has been proposed and assessed employing the Stanford-based TDWZ video codec. It has been demonstrated that the proposed model is capable of generating efficient SI for DVC. The proposed scheme estimates SI for the current WZ frame using two adjacently decoded key frames as input. In the proposed methodology, the training of the individual MLPs is done in an offline mode using the training (input, target) patterns which are collected across different video sequences with diversified motion and texture features. The proposed model selects ' M ' number of trained networks based on MSE metric to form the ensemble model.

Determining an appropriate number of hidden layers and number of hidden units in each of the individual layers is one of the essential assignment in ANN architecture. Therefore, in this study, both (I, H, O) and (I, H_1, H_2, O) structures have been analyzed. The number of hidden units used in the single hidden layer structure are 10, 14, 15, 20, 25, and 30. Similarly, for the two hidden layer structure, the number of hidden units considered are (10, 5), (14, 5), (15, 5), (20, 5), (25, 5), and (30, 5). Out of these, four best topologies, namely, (128, 10, 64), (128, 15, 64), (128, 14, 5, 64) and (128, 25, 5, 64) are selected. A dynamically averaging (DA) approach is employed to integrate the SI frames generated from each of the selected networks. The proposed ensemble model shows better generalization capabilities as compared to the individual MLPs.

Comparisons have been made with respect to the existing contemporary video codecs, and from the exhaustive simulations, it has been observed that the proposed ensemble scheme exhibits a better performance in terms of both qualitative and quantitative measures. Additionally, with the help of the statistical test like analysis of variance (ANOVA), it has been further validated that the proposed methodology produces significant enhancement (with 5% significance level) over the benchmark techniques. It has also been exhibited that the proposed scheme is capable of minimizing the estimation error between the generated SI and the corresponding WZ frame.

In future, some advanced machine learning algorithms, namely, convolutional neural network, extreme learning machine, and so on, along with ensemble of these non-linear predictors could be explored to generate a better quality of SI in a DVC framework. Moreover, creating SI frames simultaneously from each of the individual NNs using a MapReduce framework could be another possible extension of the proposed work. Further, a hardware implementation of the proposed framework is also possible.

References

1. Aaron A, Zhang R, Girod B (2002) Wyner-Ziv coding of motion video. In: 2002 IEEE Conference record of the thirty-sixth asilomar conference on signals, systems and computers, vol 1, pp 240–244
2. Aaron A, Setton E, Girod B (2003) Towards practical Wyner-Ziv coding of video. In: IEEE International conference on image processing, ICIP 2003, Proceedings, vol 3, pp III–869
3. Aaron A, Rane SD, Setton E, Girod B (2004) Transform-domain Wyner-Ziv codec for video. In: Electronic imaging 2004. International Society for Optics and Photonics, pp 520–528
4. Adhikari R (2015) A neural network based linear ensemble framework for time series forecasting. *Neurocomputing* 157:231–242
5. Adikari ABB, Fernando WAC, Arachchi HK, Weerakkody WARJ (2006) Multiple side information streams for distributed video coding. *Electron Lett* 42(25):1447–1449
6. Artigas X, Ascenso J, Dalai M, Klomp S, Kubasov D, Ouaret M (2007) The DISCOVER codec: architecture, techniques and evaluation. In: Picture coding symposium (PCS'07) (No. MMSPL-CONF-2009-014). Lisbon
7. Ascenso J, Brites C, Pereira F (2005) Improving frame interpolation with spatial motion smoothing for pixel domain distributed video coding. In: 5th EURASIP conference on speech and image processing, multimedia communications and services. Smolenice, pp 1–6
8. Ascenso J, Brites C, Pereira F (2010) A flexible side information generation framework for distributed video coding. *Multimed Tools Appl* 48(3):381–409
9. Bhandari S, Patel N (2017) Nonlinear adaptive control of a fixed-wing UAV using multilayer perceptrons. In: AIAA Guidance, navigation, and control conference, pp 1524
10. Brites C (2005) Advances on distributed video coding. Instituto Superior Técnico, MS Thesis
11. Brites C, Pereira F (2008) Correlation noise modeling for efficient pixel and transform domain Wyner-Ziv video coding. *IEEE Trans Circ Syst Vid Technol* 18(9):1177–1190
12. Brites C, Ascenso J, Pedro JQ, Pereira F (2008) Evaluating a feedback channel based transform domain Wyner-Ziv video codec. *Signal Process Image Commun* 23(4):269–297
13. Cao MS, Pan LX, Gao YF, Novák D, Ding ZC, Lehký D, Li XL (2015) Neural network ensemble-based parameter sensitivity analysis in civil engineering systems. *Neural Comput Applic* 1–8
14. Cheng MH, Leou JJ (2008) A new side information generation scheme for distributed video coding. *Adv Multimed Inf Process-PCM* 2008:782–785
15. Choi BD, Han JW, Kim CS, Ko SJ (2007) Motion-compensated frame interpolation using bilateral motion estimation and adaptive overlapped block motion compensation. *IEEE Trans Circ Syst Vid Technol* 17(4):407–416
16. DISCOVER-Distributed Coding for Video Services (2005). [Online]. Available: <http://www.discoverdvc.org/>. Accessed 29 Jul (2009)
17. Girod B, Aaron AM, Rane S, Rebollo-Monedero D (2005) Distributed video coding. *Proc IEEE Special Issue Adv Vid Cod Deliv* 93(1):71–83

18. Hansen LK, Salamon P (1990) Neural network ensembles. *IEEE Trans Pattern Anal Mach Intell* 12(10):993–1001
19. Hänsel R, Müller E (2011) Global motion guided adaptive temporal inter-/extrapolation for side information generation in distributed video coding. In: 18th IEEE International conference on image processing (ICIP), pp 2629–2632
20. Hameed AA, Karlik B, Salman MS (2016) Back-propagation algorithm with variable adaptive momentum. *Knowl-Based Syst* 114:79–87
21. Hossain MS, Ong ZC, Ng SC, Ismail Z, Khoo SY (2017) Inverse identification of impact locations using multilayer perceptron with effective time-domain feature. *Invers Probl Sci Eng* 1–19
22. Islam MF, Kamruzzaman J (2006) ANN ensemble and output encoding scheme for improved transformer tap-changer operation. In: Power systems conference and exposition, PSCE'06. IEEE PES, pp 1063–1068
23. Jiménez D (1998) Dynamically weighted ensemble neural networks for classification. In: 1998 IEEE World Congress on computational intelligence. The 1998 IEEE international joint conference on neural networks proceedings vol 1, pp 753–756
24. Ko B, Shim H, Jeon B (2007) Wyner-Ziv video coding with side matching for improved side information. *Adv Image Vid Technol* 816–825
25. Krogh A, Vedelsby J (1995) Neural network ensembles, cross validation, and active learning. *Adv Neural Inf Process Syst* 7:231–238
26. Kubasov D, Nayak J, Guillemot C (2007) Optimal reconstruction in Wyner-Ziv video coding with multiple side information. In: IEEE 9th Workshop on multimedia signal processing, MMSP 2007, pp 183–186
27. Liu Y, Yao X (1999) Ensemble learning via negative correlation. *Neural Netw* 12(10):1399–1404
28. Maqsood I, Khan MR, Abraham A (2004) An ensemble of neural networks for weather forecasting. *Neural Comput Appl* 13(2):112–122
29. Moretti F, Pizzuti S, Panziera M, Annunziato M (2015) Urban traffic flow forecasting through statistical and neural network bagging ensemble hybrid modeling. *Neurocomputing* 167:3–7
30. Mukherjee D, Macchiavello B, de Queiroz RL (2007) A simple reversed-complexity Wyner-Ziv video coding mode based on a spatial reduction framework. In: Electronic imaging 2007. International Society for Optics and Photonics, pp 65081Y–65081Y
31. Neelakanta PS, DeGross D (1994) Neural network modeling: statistical mechanics and cybernetic perspectives. CRC Press, Boca Raton
32. Optiz DW, Shavlik JW (1996) Actively searching for an effective neural network ensemble. *Connect Sci* 8(3-4):337–354
33. Optiz DW, Shavlik JW (1996) Generating accurate and diverse members of a neural-network ensemble. *Adv Neural Inf Process Syst*. 535–541
34. Puri R, Ramchandran K (2002) PRISM: A new robust video coding architecture based on distributed compression principles. In: Proceedings of the annual allerton conference on communication control and computing, vol 40, No 1. The University, pp 586–595, 1998
35. Puri R, Majumdar A, Ramchandran K (2007) PRISM: a video coding paradigm with motion estimation at the decoder. *IEEE Trans Image Process* 16(10):2436–2448
36. Rencher AC, Trenkler G (1996) Methods of multivariate analysis. *Comput Stat Data Anal* 22(3):334
37. Rup S, Majhi B, Padhy S (2014) An improved side information generation for distributed video coding. *AEU-Int J Electron Commun* 68(3):201–209
38. Slepian D, Wolf J (1973) Noiseless coding of correlated information sources. *IEEE Trans Inf Theory* 19(4):471–480
39. Standard Video Sequences (2017). <https://media.xiph.org/video/derf>
40. Tagliasacchi M, Tubaro S, Sarti A (2006) On the modeling of motion in Wyner-Ziv video coding. In: 2006 IEEE International conference on image processing, pp 593–596
41. Wyner A, Ziv J (1976) The rate-distortion function for source coding with side information at the decoder. *IEEE Trans Inf Theory* 22(1):1–10
42. Yang WA, Zhou W (2015) Autoregressive coefficient-invariant control chart pattern recognition in autocorrelated manufacturing processes using neural network ensemble. *J Intell Manuf* 26(6):1161–1180
43. Ye S, Ouaret M, Dufaux F, Ebrahimi T (2009) Improved side information generation for distributed video coding by exploiting spatial and temporal correlations. *EURASIP J Image Vid Process* 2009(1):683510
44. Zhou ZH, Chen SF (2002) Neural network ensemble. *Chin J Comput-Chin Edn* 25(1):1–8
45. Zhou ZH, Wu J, Tang W (2002) Ensembling neural networks: many could be better than all. *Artif Intell* 137(1–2):239–263



Bodhisattva Dash received his M.Tech degree in Electronics and Telecommunication Engineering from Silicon Institute of Technology (SIT), Bhubaneswar, Odisha, India. He is continuing his Ph.D. degree in Computer Science and Engineering from International Institute of Information Technology, Bhubaneswar (IIIT-Bh), Odisha, India. His research interest includes Image and Video Processing, Distributed video coding, Scalable Video Coding, Image Compression, and Image Restoration.



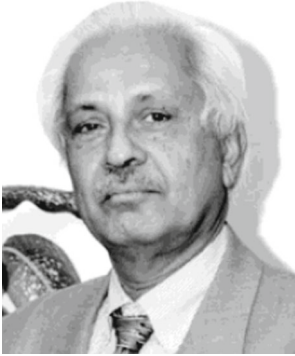
Suwendu Rup received his M.Tech degree in Computer Science and Engineering, from Jadavpur University, Kolkata, India. He received his Ph.D. degree in Computer Science and Engineering from National Institute of Technology (NIT), Rourkela, Odisha, India. Since 2010, he is with the Department of Computer Science and Engineering, International Institute of Information Technology Bhubaneswar (IIIT-Bh), India, and currently serving as an Assistant Professor. His research interest includes Image and Video Processing, Distributed video coding, Image Compression, Digital Image Watermarking, and Video Object Detection and Tracking.



Anjali Mohapatra received her M.Tech and Ph.D. degree in Computer Science from Utkal University, Bhubaneswar, Odisha, India. In 2008, she joined the Department of Computer Science and Engineering, International Institute of Information Technology Bhubaneswar (IIIT-Bh), India, and currently serving as an Assistant Professor. Her research areas include novel application areas of computer science such as Molecular Biology, Soft Computing, Image Processing and Algorithms.



Banshidhar Majhi received his M.Tech degree and Ph.D. in Computer Science and Engineering in the year 1998 and 2003, respectively, from National Institute of Technology (NIT), Rourkela, Odisha, India. Since 1991, he is with the Department of Computer Science and Engineering, NIT, Rourkela, and currently serving as the Professor and Dean Academics. His research interest includes Image and Video Processing, Data Compression, Soft Computing, Bio-metrics and Network Security.



M. N. S. Swamy received the B. Sc. (Hons.) degree in mathematics from Mysore University, India, in 1954, the Diploma in electrical communication engineering from the Indian Institute of Science, Bangalore, India, in 1957, the M. Sc. and Ph. D. degrees in electrical engineering from the University of Saskatchewan, SK, Canada, in 1960 and 1963, respectively, and the Doctor of Science degree in Engineering (Honoris Causa) from Ansted University, Penang, Malaysia, in 2001. He is presently a Research Professor and the Director of the Center for Signal Processing and Communications, Department of Electrical and Computer Engineering, Concordia University, Montreal, QC, Canada, where he served as the Chair of the Department of Electrical Engineering from 1970 to 1977 and Dean of Engineering and Computer Science from 1977 to 1993. Since July 2001, he holds the Concordia Chair (Tier I) in Signal Processing. He has also taught in the Electrical Engineering Department, Technical University of Nova Scotia, Halifax, NS, Canada, and the University of Calgary, Calgary, AB, Canada, as well as in the Department of Mathematics, University of Saskatchewan. He has published extensively in the areas of number theory, circuits, systems, and signal processing and holds four patents. He is the coauthor of two book chapters and three books: *Graphs, Networks and Algorithms* (Wiley, 1981), *Graphs: Theory and Algorithms* (Wiley, 1992), and *Switched Capacitor Filters: Theory, Analysis and Design* (Prentice-Hall, 1995). A Russian Translation of the first book was published by Mir Publishers, Moscow, in 1984, while a Chinese version was published by the Education Press, Beijing, in 1987.