Journal of Theoretical and Applied Information Technology
15th March 2023. Vol.101. No 5

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

SPATIAL TRANSFORMER NETWORKS FOR ADAPTIVE VIDEO ENCODING AND DECODING IN VIDEO TRANSMISSION

AJITHA G[1], SANTHIPRABHA I [2], AJITHA G[3]

[1] Research Scholar, Dept. of Electronics and Communications, University College of Engineering, JNTUK, Kakinada, India
[2] Professor, Dept. of Electronics and Communications, University College of Engineering, JNTUK, Kakinada, India
[3]Assistant Professor, Department of ECE, Institute of Aeronautical Engineering, Hyderabad

Email: ajithagphd@gmail.com

ABSTRACT

Video encoding is one of the methods for improving video transmission quality. Recently adaptive and scalable video encoding technology is emerging which compresses the video files while maintaining their quality. However, existing methods of video encoding and decoding are used to extract the contextual information, which does not suited for all kinds of devices, applications and services in wireless networks. Hence, the new video encoding and decoding technique is presented for video quality improvement using deep learning. For that purpose, Spatial Transformer Networks (STN) is presented in this paper. Firstly, coefficients are extracted in the set of frames of a given input video. This is done by means of fast curvelet transform method that splits the frame into multiple levels. Secondly, approximate coefficients generation is implemented by means of spatial transformer networks. Then the detailed coefficients are optimized by means of the Human Mental Search Optimization technique. Later, the side information is extracted using Inverse Curvelt Transform. Finally the, video quality is evaluated using Fuzzy Logic Inference System (FLIS) using video quality assessment metrics. This feedback is forwarded to the source node. The comparison is between the proposed approach with the existing approaches in terms of power consumption, time, PSNR, MSE and RMSE for different datasets.

Keywords: *Video Encoding And Decoding, Spatial Transformer Networks, Fuzzy Logic, Side Information Extraction, Video Quality Assessment*

1. INTRODUCTION

Multiple video applications and services, such as video telephony, HD and Ultrahigh-Definition (UHD) broadcasting, Internet protocol television (IPTV), and emerging mobile streaming, have advanced rapidly in the last decade. As a result, assessed the effectiveness of video that are broadcast and watched on youtube becomes an important research area. Video streaming has become extremely popular on the Internet, accounting for more than 55 percent of all traffic. Numerous scholars have put in a lot of effort to examine the efficiency of streamed communication systems. Similarly, organisations such as the International Telecommunications Union (itu (ITU) have many frameworks and standardisation processes in place for evaluating observed image quality in a range of application scenarios. Numerous media will not accept video signals recorded with contemporary devices for transmission or storage. As a result, and during encoding stage, it is important to reduce its size. It is possible to achieve this by lowering spatially and/or temporal resolution. From the other hand, multimedia displaying technologies (mobile phones, tablet devices, laptop, etc.) are rapidly evolving and can exhibit high definition video signals regardless of screen resolution. As a result, in during decoding, it really is essential to boost the spatial and/or temporal resolution of the video file (prior to presentation). As a result, interest in spatial/temporal video upscaling techniques has exploded in the previous decade. First, frames from an uncompressed video feed are eliminated in this study. With both the increasing demands for images and multimedia applications, the need for uniform images or video standard evaluation metrics has grown. Different ways to estimating the detection

Journal of Theoretical and Applied Information Technology
15th March 2023. Vol.101. No 5

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

performance of images and videos have been offered in the research. DVC's encoding procedure is intentionally simple. DVC's encoding procedure is intentionally simple. First, the incoming video sequence is divided into groups based on the cumulative movements of all the images in each block over a predetermined threshold. A GOP is the sampling frequency in each of these groups (Group of Pictures). The first frame in a GOP is considered as the core frames, and the other frames are known as Wyner-Ziv (WZ) frames. The H.264 main profile intra coder will be used to extract the so-called key frames. A larger GOP decreases the data rate by increasing the number of WZ frames between crucial images. The so-called WZ frame will be transformed block by block, with DCT performed to each 4x4 block. DCT coefficient bands will be formed by grouping the DCT coefficients of the full WZ frame. Every coefficients bands will be homogeneous vector quantization with pre-defined levels that after transformation coding. On the quantization bins, bitplane ordering will be accomplished. The Low Density Parity Check Accumulator (LDPCA) encoding will be used to compress each ordered bit-plane independently. The cumulative disorder of the encoded bit planes is represented by a series of binary representation computed by the LDPCA encoder. To guarantee adequate decoding, an 8-bit Cyclic Redundancy Check (CRC) sum will be given to the decode for every image pixel. These partial products will be held in a cache in the encoder and sent out in stages to the decoding, which will continuously demand additional bits through into the feedback connection in decode. The DVC decode method seems to be more difficult. H.264 main profile intra decoder will be used to decoding the key frames. The decoded key frames will be used to recreate side information (SI) at the decoding, that is a guess at the WZ frame that is only accessible at the encoding. For SI generating, a movement adjusted approximation in between two nearest reference frames is conducted. Connection distortion in the virtual channel can cause a discrepancy in between WZ frame and the associated SI. To produce a good estimation of the remainder, an online Laplacian model is applied (WZ-SI). The SI will be subjected to the DCT transform, yielding an approximation of the WZ frame's coefficients.

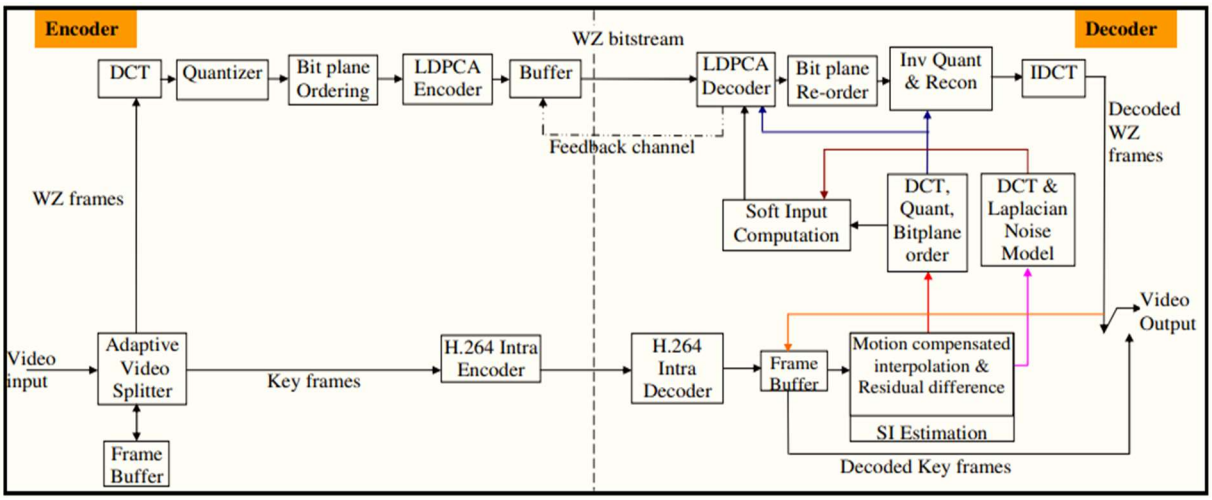

Fig.1. DVC based Encoder and Decoder Architecture

Soft input data for the information bits will be produced to use these DCT coefficients, taking into consideration the mathematical techniques of the artificial noises. By taking into account the earlier encoded bit planes and the value of input data, the probabilistic model produced for every Frequency component will be translated into unconditional bit probabilities. The presented method's goals are as follows:

- Perform encoding, which is the process of compressing content so that it requires less time;
- Compressing moving frames through video processing is yet another goal
- Ensure the quality of the videos at the destination side

This paper's contributions can be summarized as follows:

- Wavelet coefficient approximations are utilized to construct training (input, target) pattern data using a three-level decomposition using fast curvelet transform on both even and odd pre-defined counts of frames.

- Spatial transformer network (STNs) for the utilization of approximate coefficients extraction for the input frames
- Further the frames are optimized by means of inverse curvelet transform. This provides the SI frame information from previously computed estimated and detailed coefficients for each sub-band.
- Finally, the fuzzy logic inference system is used for video quality estimation at the destination node. Depends upon the information of destination node, the source node improves the quality of video by changing of bit rate adaptively by means of human mental search optimization algorithm.

2. LITERATURE REVIEW

With simultaneous and true decode in DVC, estimate of rapid and consistent side information is a significant issue. For elevated videos, the problem gets considerably more acute. To decrease the total cost of side information prediction, a computationally simple DVC codec is suggested in this paper, which uses a basic phase interpolation (Phase-I) technique. It is fast for all acquired images, and it produces considerable results for high-resolution recordings with such a large number of images (GOP). With just an increase in precision, the computing time for the proposed methods dramatically lowers. For high-resolution films and large GOP, it is 221 percent to 280 percent faster than standard frame interpolation, with little reduction in the visual clarity of predicted side information [21]. Side information (SI) is the predicted frames of the actual Wyner-Ziv (WZ) frame inside DVC) systems, and it has a huge effect on the overall system operation. The decoder uses high-quality SI to reconstruct Wyner-Ziv frames. As a result, we suggest a novel SI generation approach based on a study of existing SI generation systems. Motion compensation frame interpolation (MCFI) and optical flow approximation provide two kinds of SI, which would then be combined using a probabilistic fusing technique to gather highly precise SI. The experimental findings show that the Hybrid SI generating algorithm suggested in this paper may successfully increase SI quality [22]. In DVC, high-quality side information is essential for trustworthy soft-input information in the decoding to decode every DCT band of a Wyner-Ziv (WZ) frame, resulting in much more effective decoding. As a result, fewer error-correcting bits must be transferred from the encoding to the decoding to decoding each DCT band's bitplanes, resulting in improved compression effectiveness and rate-distortion performance. In this research, researchers look at how to enhance the rate capability of distributed video coding by gradually enhance the effectiveness of side information frames. A novel paradigm for consecutive side information refinement is described, which uses the specific data received after decode the preceding DCT bands of a WZ frame to improve the greater initial information frame. The matching side information frames is improved when additional information about the WZ frame becomes known after each DCT band of the WZ frame is decoded, and then used to decoding the next DCT band of the WZ frame. Simulations show that the suggested approach for side information frame refining improves the RD efficiency of distributed video coding significantly [23]. As comparison to progressive video coding, the DVC system is shown to have no performance degradation in theory. Unfortunately, there is a significant difference between mathematically maximum efficiency and real execution. At the decoding, the side information (SI) - a noise version of the original Wyner-Ziv frame (WZF) - is created, and the correlations noise - the difference between both the actual WZF and matching SI - is modelled. If the SI & correlations noise are computed as correctly as feasible, the DVC method will perform more effectively. So, utilizing information within decode WZFs as during decoding stage, this research presents a strategy to improve the quality of SI and also the correlations noise models. With each bitplane is decode, the original SI created by Movement Temporal Interpolation (MCTI) and previously decoded keyframes (KF) will be used as reference frames to improve the information sequentially. The experimental results showed that utilizing this strategy enhances the effectiveness of the distributed video coder considerably [24]. DVC is being looked at as a possible option for uplink applications including wireless video surveillance and multimodal sensor technologies. The accuracy of side information (SI) has a substantial impact on the codec's rate-distortion (RD) effectiveness in distributed video coding.

Nevertheless, the quality of the side information fluctuates throughout the sequence within each frame. As a result, in terms of improving side information, this research proposes a motion activity-adapted side information creation technique. The experiments demonstrate that utilising this strategy enhances the effectiveness of the distributed video coder substantially [25]. DVC enables adjustability in complex distribution seen between encoder and decoder. The decoder driven side information (SI) specifically for creating, like every building piece, is an important part of a DVC codec. The reliability of the SI produced at the decoder is crucial to the effectiveness of a DVC codec. The SI frame is a replica of the original Wyner-Ziv (WZ) frame. As a result, the higher the SI quality, the higher the codec's efficiency. The fundamental goal of this research is to increase the SI frame's quality in order to improve the DVC's current effectiveness. This paper mentions a hybrid SI generation technique that uses the principles of the discrete wavelet transform (DWT) and the extreme learning machine (ELM) algorithms in a transform domain-based DVC framework to achieve this goal. Results: Extensive simulations were conducted with the proposed and benchmark techniques for only certain standard video sequences. Multiple performance indicators, including such rate-distortion (RD), SI peak-signal-to-noise-ratio (PSNR) versus frame number, number of parity requests per SI frame, and so on, are used to assess the system system [26]. The presented technique's efficiency outperforms the benchmarking schemes, according to the empirical analysis and findings [27]. Many scholars have been working on SI generation algorithms throughout the last decade. The authors of this article offer a general model that can accomplish SI creation in the WZ video coding framework using an algorithm mixed with naive Bayesian theory. The suggested technique uses samples to construct the very first model, following which the algorithm filters samples and models based on the T 1 threshold. The algorithm then builds a generic model using filtering samples and models as parameters. Finally, using the motion vectors derived from of the created model, the suggested technique completes the development of SI. In comparison to state-of-the-art approaches, experimental results reveal that the suggested algorithm produces better rate-distortion performance and increase peak signal-to-noise ratio by up to 0.5 and 2 dB [28]. Since this SI is an approximation of each to frame, the production of Side Information (SI) is a crucial part of the decoder. Depth maps allow for the computation of an object's distance from the camera. Because the motion of depth frames and their associated texture frames (luminance and chrominance components) is highly connected, the additional depth information could be used to provide all the precise SI for the texture stream, boosting the process performance. Researchers compare multiple strategies for precise texturing SI generation with other state-of-the-art solutions in this research. On the reference decoder, the suggested system achieves improvements of up to 1.49 dB [29]. DVC's encoding efficiency and decoded speed are still facing major hurdles after decades of growth. For instance, DVC's rate-distortion (RD) performance lags considerably below conventional prediction-based video codecs (e.g., H.264), and the necessary decoding complexity limits its practical utility. To address these issues, we offer a real-time DVC methodology with continuous side information regeneration that takes advantage of GPGPU-based parallel processing. Given excellent RD efficiency and considerably enhanced decoding speed, the suggested method can eliminate one of DVC codec's Achilles' heels, namely the huge grouping of image issue [30].

3. PROPOSED WORK

In the proposed work, the SI method used in this study seeks to improve the reconstructed frame quality of videos. For video encoding and decoding, the proposed SI generation employs spatial transformer networks and a fast curvelet transform algorithm. Figure 1 depicts the flow of the suggested system model.

Journal of Theoretical and Applied Information Technology
15th March 2023. Vol.101. No 5

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

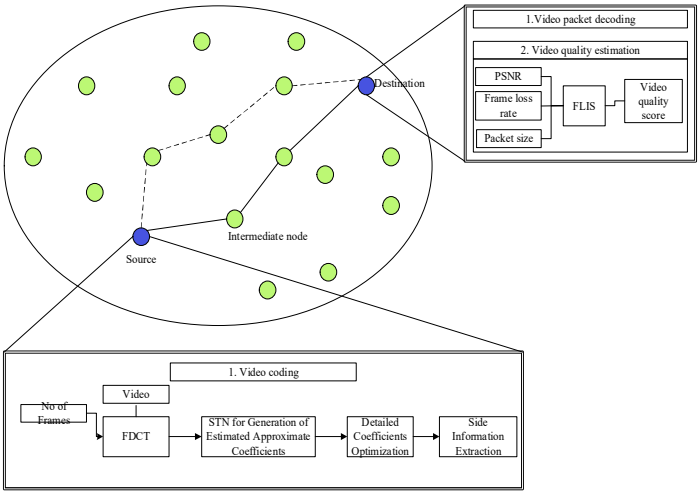

Fig.2. System Model

- ***Video encoding***

  Initially source node encodes the video before transmitting it to destination. In video encoding, **fast discrete curvelet transform** scheme is employed. This process is performed in application layer since our work is based on cross-layer approach; application layer is able to communicate with physical layer and network layer. In our work, FCT can adjust the parameters (such as number of enhancement layers, and rate of group of pictures (GOP)) accordance to congestion and video quality.

- ***Video quality estimation***

  Finally, the encoded video is received in the application layer of destination. Then the video is decoded and the quality of video is estimated by **Fuzzy Logic Inference System** scheme. In FLIS, peak to signal to noise ratio (PSNR), frame loss rate, and packet size difference are considered to estimate the quality of the video. After, estimation, the *Video Quality Score (VQS)* is sent to source through *Real-time Transmission Control Protocol (RTCP)*.

A. FDCT

Then coefficients (coefficients) are extracted by FDCT. This transformation method is used to mine the coefficients in the image. In FDCT, frequency coefficients are extracted that purpose is to extract the coefficients. However, curvelet transform is denoted by multi-scale object representation. It implementation is applied over the resized image and choose the coefficients from that. With the use of FDCT, the input image is decomposed into 4 levels, and the coefficients are calculated. The curvelet coefficients are coefficient vectors, which considered improving the object determination in the image, which executes based on the Fourier samples wrapping. The curvelet transform consists of multi-scale pyramids with different directions and positions. The FDCT is defined as the following:

$$c(j,l,k) = [f, \emptyset\_(j,l,k)] \quad (1)$$

Where j,l,k are defined as the parameters of scale, direction, and position. In FDCT, the image is given by the ROI segmented image, and the output will be the co-efficient extraction. Firstly, we initialize the data structure with shift condition. the input image is decomposed into the set of bands, in which Pyramid Scale Decomposition is determined using M_{pyr1} and M_{pyr2}. The pyramid is utilized to provide the various sub-bands. Each sub-band is windowed smoothly into "squares". Secondly, the smooth periodic extension of the high frequencies is mentioned in the Eqn.(4.1). Then we calculate the window scale and wrapping window to acquire the efficient curvelet coefficients. In wrapping, two processes are involved such as periodize and re-index the windowed frequency data. Thirdly, we calculate the normalization for x-coefficient, and y-coefficient by wrapping conditions. Later that, low-pass filter is applied. When the scale decomposition is completed, we determine the angular decomposition for horizontal, vertical

bands, left and right bands. At last, we identify that the curvelet level by following function.

$$Curvelet_{wav} = FFT_{SHIFT}\left(IFFT\ 2(L_{wed})\right) * \sqrt{sixe(roi_{ex})} \quad (2)$$

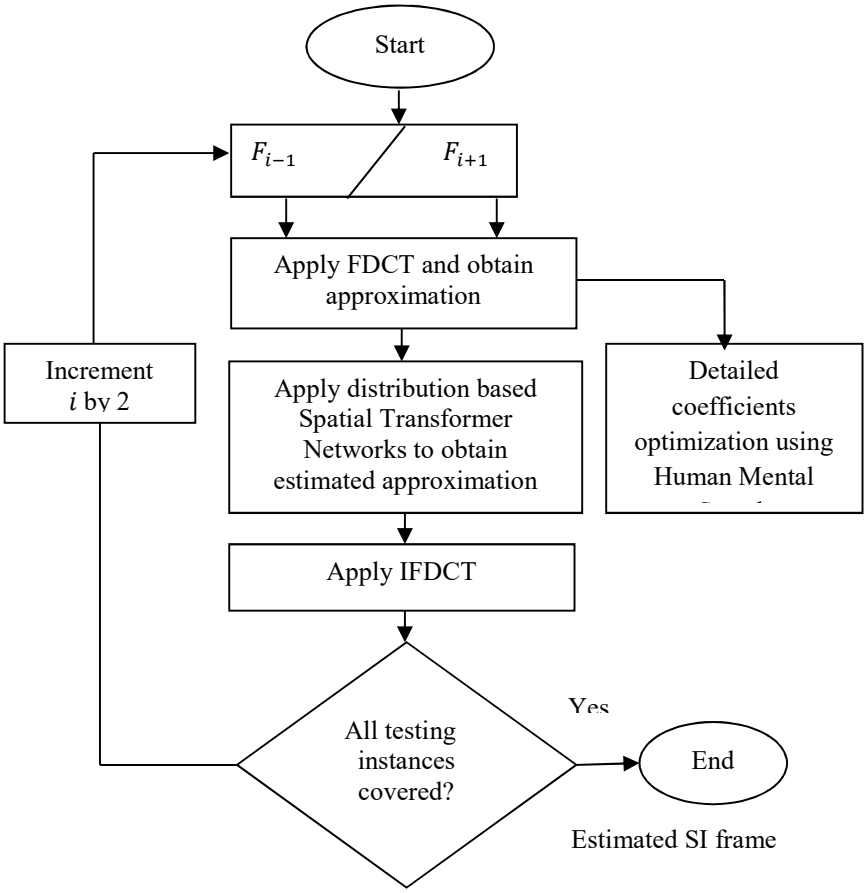

Fig.3.Spatial Transformer Networks Flowchart

Algorithm 1: Fast Discrete Curvelet Transform

Input: Coefficient Extracted image (roi_{ex})
Let N_s be the Number of scales
To initialize the data structure,
Finest, F_e=3;
Let shift condition,
Img_{sh}=fftshift(ifftshift(roi_{ex}))/$\sqrt{size(roi_{ex})}$
[N_e N_v]=size(Img_{sh})
Where,
N_e be the Number of edges
N_v be the Number of vertex
To find the pyramidal scale decomposition
M_{pyr1}=$\frac{N_e}{3}$; M_{pyr2}=$\frac{N_v}{3}$;
To compute smooth periodic extension of high frequencies,
If F_e ==1

SN_e=2*floor(2*M_{pyr1})+1
SN_v=2*floor(2*M_{pyr2})+1
Img_{sh}=$Img_{sh}(equ_{ind1}, equ_{ind2})$;
To compute windows length,
w_{len1}=floor(2* M_{pyr})-floor(M_{pyr1})-1-mod(($N_{e,3}$)==0);
w_{len2}=floor(2* M_{pyr2})-floor(M_{pyr2})-1-mod(($N_{v,3}$)==0);
co_x=0:(1/w_{len1}): 1;
co_y=0:(1/w_{len2}): 1;
To compute wrapping window,
w_{r1}=zeros(size(co_x))
w_{l1}=zeros(size(co_x))
$w_{r1}(co_x \leq 0)$=1
$w_{r1}(co_x > 0)$&($co_x < 1$)=exp(1-exp(1-1/co_x)($co_x > 0$)&($co_x < 1$));
$w_{l1}(co_x \geq 0)$=1
$w_{l1}(co_x > 0)$&($co_x < 1$)=exp(1-exp(1-1/co_x)($co_x > 0$)&($co_x < 1$));
Let find the low pass filter,
L_{pf} =[$w_{r1}, ones(1,2 * floor(M_{pyr1} + 1))$]
To compute angular decomposition,
n_{qua}=2+2*(nor_r);
for q_{re}=1: n_{qua}
M_{hor} =M_{pyr2}*(mod(q_{re},2)==1)+ M_{pyr1}*(mod(q_{re},2))==0;
M_{ver} =M_{pyr1}*(mod(q_{re},2)==1)+ M_{pyr2}*(mod(q_{re},2))==0;
If mod(n_{qua},2)
W_t=[W_{t_left} W_{t_rite}]
End
End
To compute left corner wedge,
l=l+1
Let regular wedge,
To compute right corner wedge,
$slow_{re}$=round(2*floor(4*M_{ver})2*n_{qua} + 1)- lcw_{re}(end)/floor(4*M_{ver}))
$midcw_{re}$=floor(4*M_{ver})-floor(M_{ver})+ $slow_{re}$
Let regular wedge,
L_{wed}=floor(4*M_{ver})-floor(M_{ver})
Wr_{data}=$\sqrt{2 * img(roi_{ex})}$
To compute wavelet level,
$curvelet_{wav}$=fftshift(ifft2(L_{wed}))* $\sqrt{sixe(roi_{ex})}$
End
End
Output: Coefficient Extraction F_{cur}

B. Spatial Transformer Networks

After the coefficient extraction, the optimum set of coefficients are extracted by the spatial transformer networks, and also it removes the redundant coefficient vectors. There are three reasons to use the spatial transformer networks, which are following:

(1). It considers the prior knowledge of the image (deep coefficients) for end-to-end training
(2). It resolves the two issues of CNN: translation invariance and max-pooling layers
(3). It considers only relevant coefficients and reduces the redundant coefficient vectors that significantly reduces the false positive rate

The spatial transformer networks address the translation invariance issues by performing two significant operations: Affine Transformation, and Interpolation. In traditional

CNN, this block is added to produces the effective outputs and finds the deep image coefficients for next step. It is well-suitable for implementing for large and complex image cases. In addition to scale invariant issues, CNN have another important issues such as scale, background cluster and viewpoints variation. Let's assume that the transformation matrix for the spatial transformer networks is AT_θ and θ is the linear transformation which is given as follows:

$$AT_\theta = \begin{bmatrix}\theta_1 & \theta_2 & \theta_3\\ \theta_4 & \theta_5 & \theta_6\end{bmatrix} = \begin{bmatrix}s_x & \varphi_x & t_x\\ \varphi_y & s_y & t_y\end{bmatrix} \quad (3)$$

Where s_x, s_y are the scale parameters, φ_x, φ_y are the skew parameters, t_x and t_y are the translation parameters. With the three sets of parameters, affine transformation matrix was calculated by the localization network. In this the grid generator computes the grid sampling$G \in K^{h\times w\times 3}$. It determines the arguments in the input data. This step finds the transformed output. We implements the grid warping on augmented input by sample kernel $G_i = (x_i^t, y_i^t)$, which provides the coefficient map V. Therefore the sample kernel which means bilinear sampling kernel is defined as follow:

$$V_i^c = \sum_n^h \sum_m^w u_{n\times m}^c \max(0,1 - (x_i^s - m)max(0,1 - |y_i^s - n|)) \quad (4)$$

Where x_i^s, and y_i^s are the source coordinates of the augmented input U that represents the sample points. The final predicted result from the spatial transformer networks are transformed into CNN for final optimum coefficients set prediction

C. Human Mental Search Optimization Algorithm

The clustering of extracted coefficients is performed in order to reduce the complexity in classification of modulation. For this purpose, a novel twin functioned-human mental search (TF-HMS) is proposed in this approach which was found to provide better performance than the Neutrosophic C-Means clustering algorithm. The HMS falls under population based optimization algorithm which utilizes bid space search. Initially, the populations of candidate solutions are generated randomly in the form of bid which are expressed as,

$$Bid = \{x_1, x_2 \dots x_{\mathrm{d}}\} \quad (5)$$

The evaluation of each bid is based on the specific objective function $OF(bid)$ such as the quality of candidate solution which is represented as,

$$OF(bid) = OF(x_1, x_2 \dots x_{\mathrm{d}}) \quad (6)$$

The mental search operation is performed to group the bids in ḍ dimensional space. The proposed TF-HMS considers the coefficient vectors $\mathfrak{F}_{\mathbb{V}}$ as the form of bids which encode the cluster centre. The twin functions considered for clustering are cluster purity and cluster entropy function. The number of clusters is represented by array length which is denoted as K. The upper and lower bound for each bid are computed which can be formulated as,

$$L = \min(\mathfrak{F}_{\mathbb{V}}) \quad (7)$$

$$U = \max(\mathfrak{F}_{\mathbb{V}}) \quad (8)$$

The formation of clusters is carried out based on three important coefficients namely inter-cluster distance, intra-cluster distance, and expression error respectively. The computation of weight of these parameters is represented as,

$$OF(Bid, \mathfrak{F}_{\mathbb{V}}) = \mathrm{C}_1 d_{max}(Z, Bid) + \mathrm{C}_2(Z_{max} - d_{min}(\mathfrak{F}_{\mathbb{V}}, Bid) + \mathrm{C}_3 I_e) \quad (9)$$

Where, $\mathrm{C}_1, \mathrm{C}_2, \mathrm{C}_3$ represents the weight values of parameters and Z_{max} denotes the maximum data value. The procedure of the proposed TF-HMS algorithm involves several stages which are described as follows,

Stage 1: Initially the parameters involved in the clustering process such as number of bids N, number of clusters K, number of clusters for grouping of bids K_{BG}, and mental search parameters such as M_{min} and M_{max} are initialized.

Stage 2: Based on the number of bids N, the population of the candidates P is formulated as,

$$P = \begin{bmatrix}Bid_1\\ \dots.\\ Bid_N\end{bmatrix} \quad (10)$$

Stage 3: The objective function OF is computed and the evaluation of bids based on OF is performed

Stage 4: The selection of best bid is executed based on the evaluation result from previous stage.

Stage 5: The selection of random value is performed, in which a random value between M_{min} and M_{max} is selected for each bid.

Stage 6: The creation of new bids is carried out in the vicinity of prevailing bids by using levy function which can be formulated as,

$$= Bid_i + S^{N_{POS}} \quad (11)$$

Where, S denotes the number of steps which is determined based on random values and maximum number of iterations (Max_I) respectively. The computation of S is formulated as follows,

$$= \left(2 - I * \left(\frac{2}{Max_I}\right)\right)^S * \Lambda \otimes Levy \quad (12)$$

Where, I denotes the current iteration, Λ denotes the random variables, and $\otimes$ denotes the element-wise multiplication respectively. The step size of S can be derived as follows,

$$S = \left(2 - I * \left(\frac{2}{Max_I}\right)\right) * 0.01 * \frac{U}{V^{\frac{1}{\beta}}} * (x^i - x_*) \quad (13)$$

Where U, V denotes two random variables, and x_* denotes the best position respectively.

Stage 7: The replacement of previous bid is carried out by the new bid when OF of new bid is better than the previous bid.

Stage 8: the grouping of bids into number of clusters is carried out in this stage based on the parameters.

Stage 9: In this step, the objective function of each cluster is computed and the evaluation of clusters is performed. The cluster possessing lower bound of mean objective function is considered as final cluster. The other clusters pass towards the best bid in the collection.

Stage 10: Go back to stage 2 and continue the other stages until the stop criterion is satisfied.

D. Fuzzy Logic Inference System

Fuzzy inference system is the key unit of a fuzzy logic approach, which is proposed for decision making as its main work. It uses the "IF…THEN" fuzzy rules to employ best decision making. The major characteristics of fuzzy inference system are following:

- The FIS output is always a fuzzy set irrespective of its input which can be crisp or fuzzy
- When it is considered as a controller, it must have defuzzified outputs
- In defuzzification, fuzzy variables are transformed into crisp variables.

Fundamental Elements of FIS are follows:

- **Database** – It defines the membership functions for fuzzy sets which is considered in rule base
- **Rule base** – It comprised of set of fuzzified rules (IF-THEN)
- **Decision Making Unit** – It allows to perform operations on rules
- **Fuzzification interface unit** – It transfers the crisp quantities into fuzzy quantities
- **Defuzzification interface unit** – It transfers fuzzified quantities in to crisp quantities

There are two methods are used in FIS that are: Mamdani Fuzzy Inference System and Takagi-Sugeno Fuzzy System (TS method). In Mamdani fuzzy inference system, output membership functions expect fuzzy sets. After the aggregation process, there is a fuzzy set for each output variable that requires defuzzification. The merits of Mamdani model are: (1). It is intuitive, (2). It has widespread acceptance and (3). It is fit for human input. In TS method, rule format is given as follows

$$IF\ x\ is\ A\ and\ Y\ is\ B\ THEN\ Z = f(x,y) \quad (14)$$

where AB are fuzzy sets in antecedents and $Z = f(x,y)$ is a crisp function in the consequent. Such type of IF-Then rule for a Sugeno model has the following form:

$$IF\ input\ 1 = x\ and\ input\ 2 = Y\ then\ Output\ is\ Z = ax + by + c \quad (15)$$

For a zero order sugeno model, the output level Z is a constant value $(a = b = 0)$. The merits of Sugeno method are follows: (1). It can be well-suited in linear models, (2). It can be efficient and flexible for adaptive and optimization technique, (3). It ensures continuity of the output surface, and (4). It is computationally efficient. However, fuzzy inference system is a precise and effective problem-solving methodology. It is able to handle linguistic knowledge and numerical data.

Journal of Theoretical and Applied Information Technology
15th March 2023. Vol.101. No 5

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

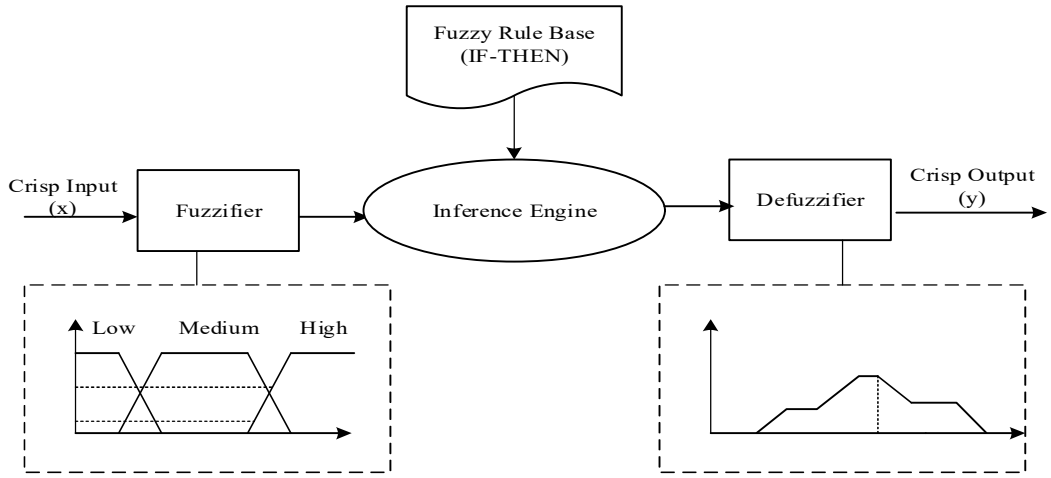

Fig.4. Fuzzy Inference System

In first FIS, we considered two input variables including bit rate and resolution. Totally there are 8 (2^3) fuzzy rules are used in first fuzzy inference system. In second FIS, two input variables are used including expected execution time and video frame rate. Here also we have 8 fuzzy rules. Finally, the sum of two FISs is added to compute the quality.

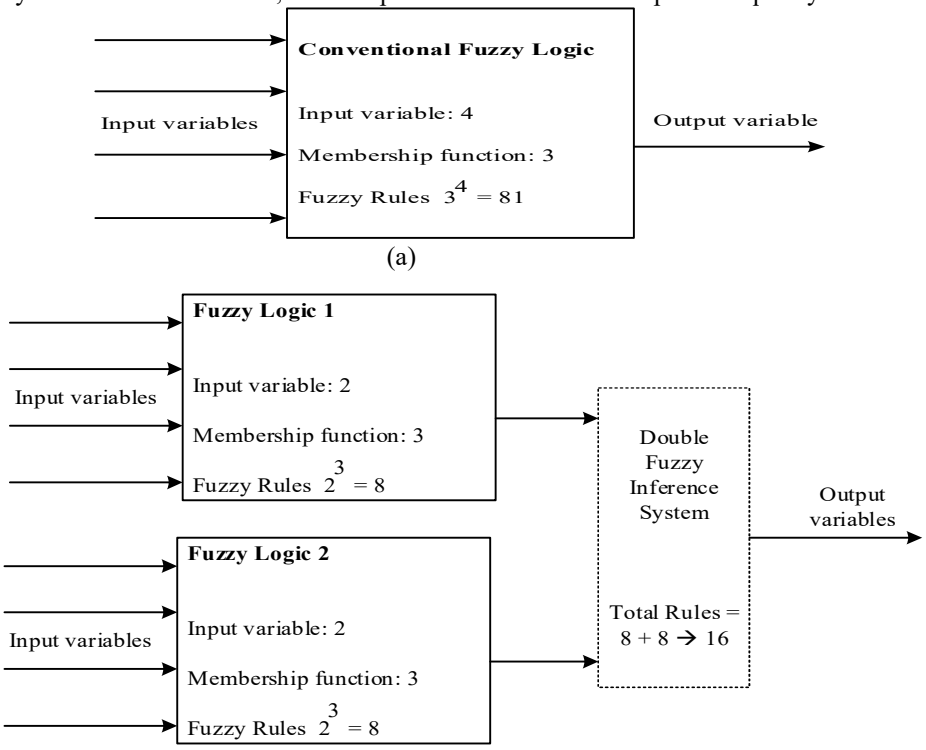

Fig.5. (a) Conventional Fuzzy Logic Unit and (b) Double Fuzzy Inference System

There are three linguistic states and term set of input variables for the double fuzzy inference system are shown low, high and medium. The five input variables are fuzzified using a trapezoidal fuzzifier that determines membership grade for each of the fuzzy sets. In inference module (the fuzzy rule base and inference), the fuzzified input measurements are then utilized by inference engine to measure control rules stored in the rule base based on the rule strength. Fuzzy

inference is the process of formulation of a mapping between two variables (input and output) using fuzzy logic algorithm. Various video quality enhancement parameters are used for the video decoding quality improvement.

4. EXPERIMENTAL EVALUATION

For evaluations, the proposed system was examined using MATLAB and conventional and accessible video sequences such as foreman and tennis sport. The input images were discovered to include a wide range of motion trajectories and speeds (from slow to fast), as well as multiple texture compositions, and were therefore used it to conduct a comprehensive evaluation of DVCs. Figure 5 shows the suggested spatial transformer networks-based SI estimation. The foreman video is used as an input in this presented design, including a prototype figure 6. Figure 7 shows the DCT frames that have been restored.

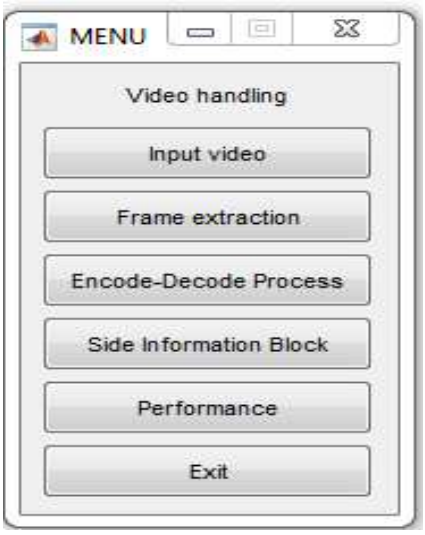

Fig.6.GUIfor side information estimation

Fig.7.Input (foreman)

Fig.8.Restored image from FDCT

In terms of PSNR, MSE, and RMSE, the proposed FDCT with STN-based SI scheme is compared against DWT, DCT, and Adaptive Video Coding (VC) methods (Root Mean Square Error). Tables 2, 3, and 4 show the power consumption and time values.

A. PSNR Comparison

The PSNR is computed using the formula to evaluate the actual and decoded frames.

$$PSNR = 10 \log \frac{255^2}{MSE} \quad (16)$$

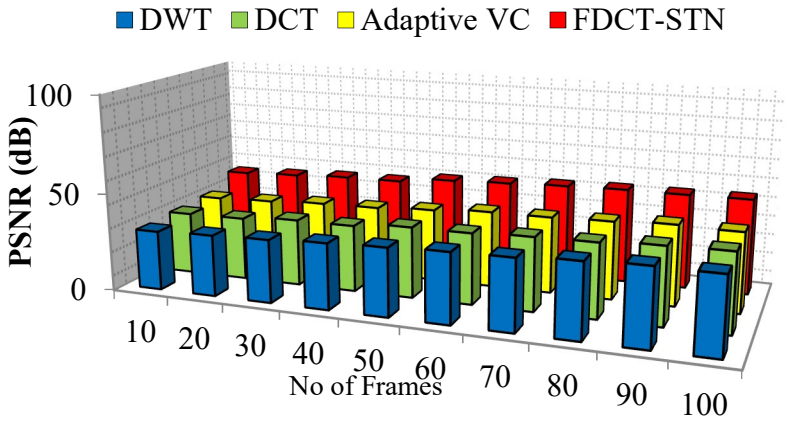

Fig.9.PSNR Performance with Respect No of Frames (Dataset 1)

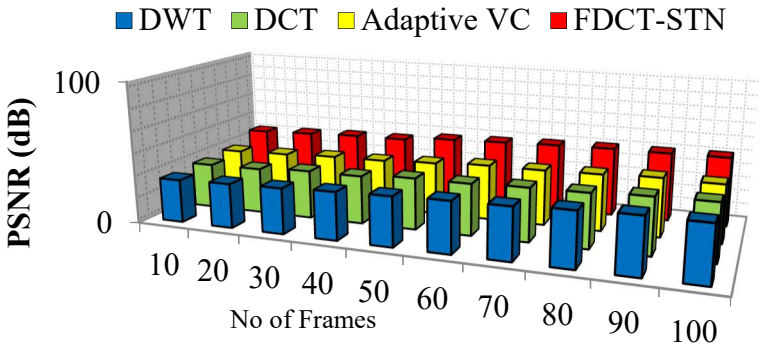

Fig.10.PSNR Performance with Respect No of Frames (Dataset 2)

The PSNR of the presented FDCT with STN approach is comparable to prior methods in Figure 11. The number of frames is measured on the x-axis, while PSNR is measured on the y-axis. PSNR (in dB) with the suggested strategy is much greater than that of DWT, DCT, Adaptive VC schemes in the majority of frames.

The suggested FDCT with STN approach obtains 42.16 dB for the tennis spot dataset, whereas previous techniques including such as DWT, DCT, and Adaptive VC achieve 32.15dB, 35.12dB, and 42.15dB for 100 frames, respectively. The results show that the suggested approach is superior because it produces superior quality SIs for the video surveillance videos.

Table 1 . PSNR Comparison

Dataset 1 (foreman)					Dataset 2 (tennis spot)				
No of frames	**DWT**	**DCT**	**Adaptive VC**	**FDCT-STN**	**DWT**	**DCT**	**Adaptive VC**	**DWT**	**FDCT-STN**
20	30.25	31.65	32.76	40.12	29.15	29.15	30.15	32.16	43.56
40	32.12	33.98	35.11	42.56	30.92	30.92	31.90	34.11	46.69
60	34.58	36.22	37.16	45.69	32.58	32.58	33.22	36.56	47.89
80	36.21	37.45	39.11	47.89	35.21	35.21	36.45	37.91	49.56
100	38.79	39.99	40.67	48.56	29.15	29.15	30.15	32.16	50.63

B. MSE comparison

The MSE is the difference between both the compression and original image's accumulated squares of the errors.

$$MSE=\frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2 \quad (17)$$

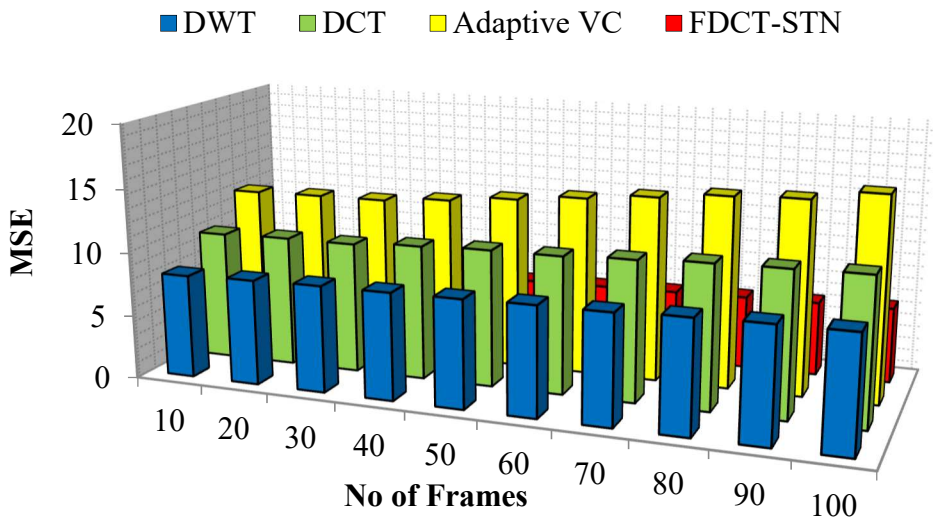

Fig.11.MSE for dataset 1

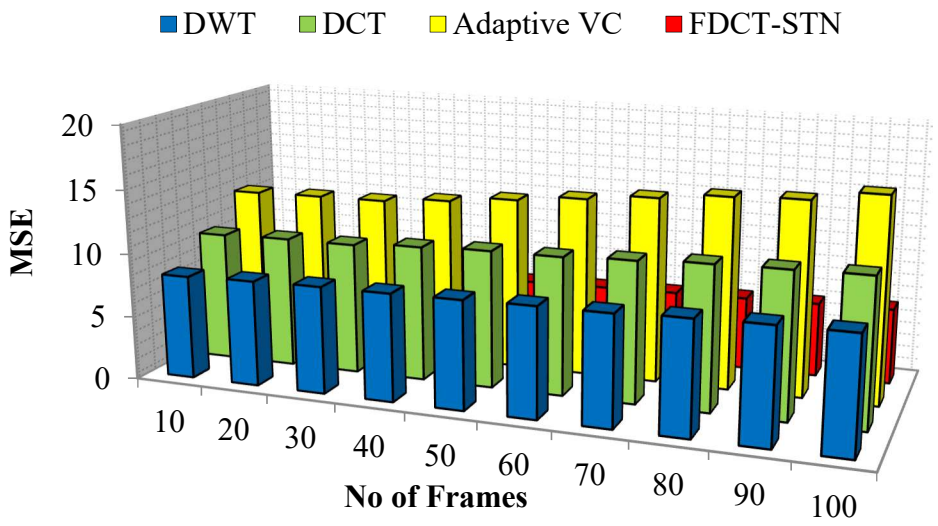

Fig.12. MSE for dataset 2

The MSE evaluation of the proposed FDCT-STN with conventional DWT, DCT, and Adaptive VC approaches is shown in Figures 11 and 12. The x-axis reflects the amount of frames, while the y-axis represents the MSE. The calculated approximation coefficients for the present frames are generated using FDCT-STN in this study.

For comprehensive coefficient optimization, the Human Mental Search optimization technique is used. It lowers the rate of mistake. The experimental findings show that the suggested system achieves 5.02 percent MSE for the tennis spot dataset, whereas other approaches are 30% higher than the proposed methods.

Table 2. MSE comparison

Dataset 1 (foreman)					Dataset 2 (tennis spot)			
No of frames	**DWT**	**DCT**	**Adaptive VC**	**FDCT-STN**	**DWT**	**DCT**	**Adaptive VC**	**FDCT-STN**
20	8.1	10.12	12.45	5	8.1	10.12	12.45	5
40	8.25	10.25	12.54	5.1	8.25	10.25	12.54	5.1
60	8.36	10.26	12.59	5.2	8.36	10.26	12.59	5.2
80	8.39	10.56	13	5.3	8.39	10.56	13	5.3
100	8.45	10.75	13.5	5.4	8.45	10.75	13.5	5.4

4.3 RMSE comparison

The proposed FDCT-STN method's RMSE is compared to the existing DWT, DCT, and Adaptive VC methods in Figure 13. The efficiency is enhanced by using an optimization strategy to choose the constant. The x-axis represents the number of frames, while the y-axis represents the RMSE. The suggested system obtains 1.5 RMSE for the tennis sport dataset, while other approaches including such DWT, DCT, and Adaptive VC achieve 2.5, 3.5, and 3.9 for 100 frames, correspondingly, according to experimental tests. The RMSE for dataset 1 and dataset 2 are shown in Figures 13 and figure 14. Table 4 compares the proposed HMO technique to previous techniques such as CSO and GA algorithms in terms of optimisation effectiveness. The optimisation techniques' reliability is increased for the amount of iteration, which implies that every type of input can be handled at any modality

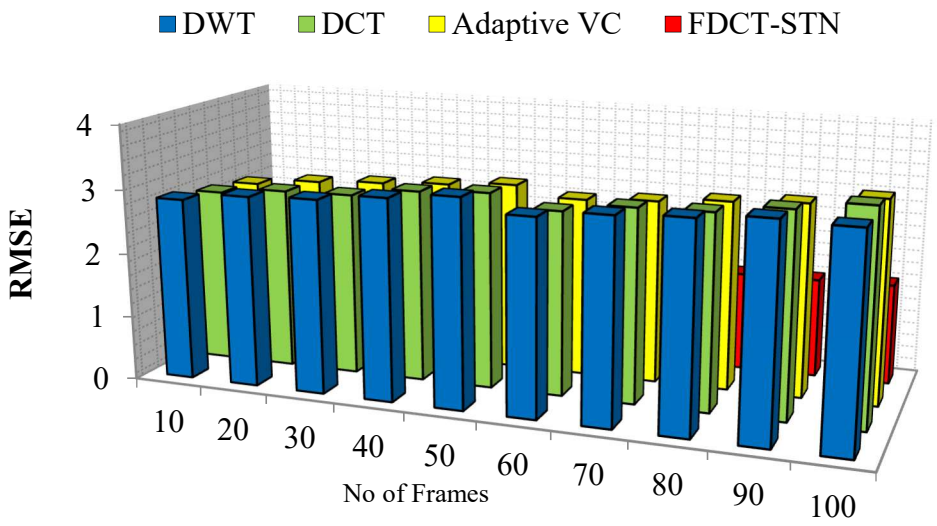

Fig.13.RMSE for dataset 1

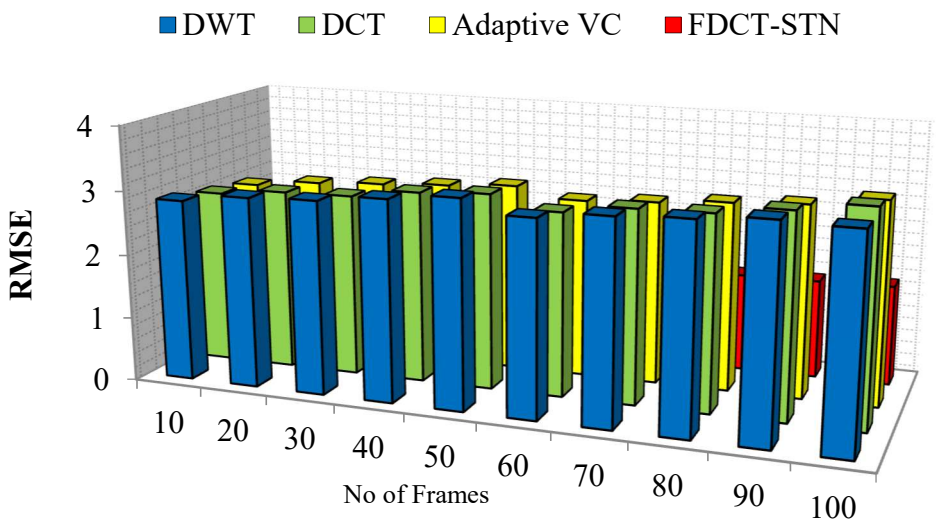

Fig.14.RMSE for dataset 2

Table 3. RMSE comparison

Dataset 1 (foreman)					Dataset 2 (tennis spot)			
No of frames	**DWT**	**DCT**	**Adaptive VC**	**FDCT-STN**	**DWT**	**DCT**	**Adaptive VC**	**FDCT-STN**
20	2.83	2.72	2.64	1.5	3.01	2.85	2.79	1.45
40	2.96	2.82	2.75	1.52	3.12	2.99	2.85	1.52
60	3.00	2.84	2.81	1.56	3.17	3.01	2.93	1.53
80	3.11	2.98	2.87	1.58	3.25	3.14	2.99	1.54
100	3.21	3.04	2.94	1.59	3.22	3.28	3.14	1.56

Table.4. Optimization Performance

Scenarios	Optimization Time (sec)		Total Generations (iterations)	
Dataset 1	CSO	112	CSO	75
	GA	95	GA	65
	HMO	85	HMO	70
Dataset 2	CSO	154	CSO	75
	GA	145	GA	70
	HMO	140	HMO	68

D. Power Consumption Comparison

Power Consumption is the of amount of energy consumed in a particular time period, which can obtained as a numerical value

$$P = \frac{E}{t}$$

Where P denotes the power consumption, E denotes energy consumed by the device, t time taken for an action.

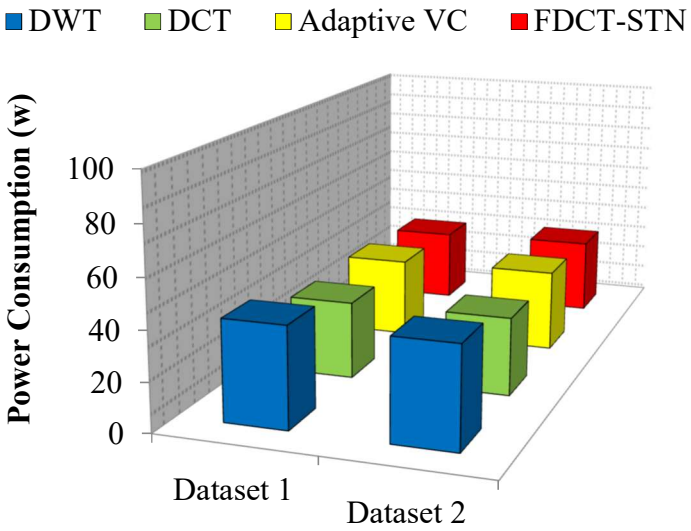

Fig.15. Power Consumption

Fig 15 describes the power consumption rate for the proposed as well as the existing approaches such as DWT, DCT, Adaptive VC, approaches. As compared to the proposed approach with the existing approaches, the performance of the existing approaches are poor in terms of obtained results for dataset 1 and dataset 2.

E. Time Comparison

It represents how long the device takes to perform a particular work. The simulation time is the time taken by the network to complete the process of encoding and decoding of videos. Fig 15 illustrates the performance of the proposed approach with the existing approaches.

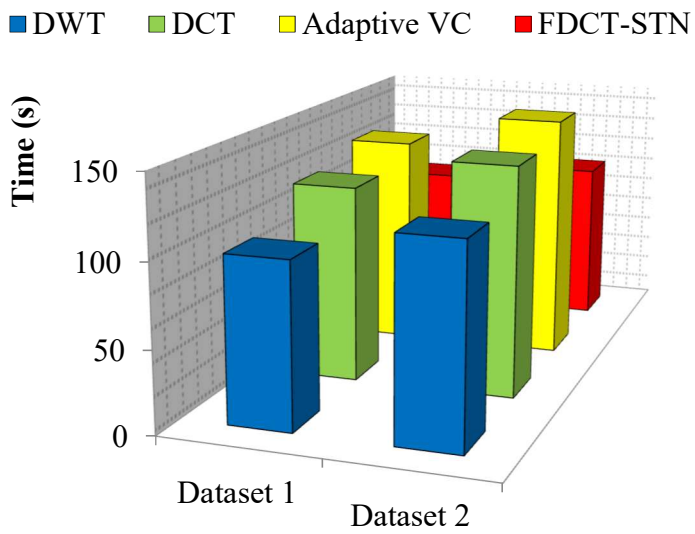

Fig.16. Time

Fig describes the time (s) for the proposed as well as the existing approaches such as DWT, DCT, Adaptive VC, approaches. As compared to the proposed approach with the existing approaches, the performance of the existing approaches are poor in terms of obtained results for dataset 1 and dataset 2.

5.CONCLUSION

To calculate SIs in DVCs, the suggested method used a Fast Discrete Curvelet Transform technique. The level-3 FDCT approach is used to mine approximation coefficients in this paper. STN is used to obtain approximation coefficients that are estimated. In this paper, comprehensive coefficients optimization is done using the HMO technique. IFDCT is utilised in the last stage to get approximated SIs in the wavelet coefficients. The SI-quality of this proposed work has indeed been developed to increase codec efficiency. Furthermore, as comparing to certain other systems, the experimental data of the suggested scheme's simulations achieved higher performance measure results in MSE, PSNR, Power Consumption, Time, and RMSE. As comparing against current solutions, the proposed approach has outperformed them. In future, this work extends into the lightweight artificial intelligence approach for the resource constrained based video transmission devices and services. Furthermore, security is added to improve the sensitivity of the video while transmitting from the source node.

DECLARATION:

Ethics Approval and Consent to Participate:
No participation of humans takes place in this implementation process
Human and Animal Rights:
No violation of Human and Animal Rights is involved.
Funding:
No funding is involved in this work.
Conflict of Interest:
Conflict of Interest is not applicable in this work.
Authorship contributions:
There is no authorship contribution
Acknowledgement:
There is no acknowledgement involved in this work.

REFERENCES

[1] W.Zhang, Q.Liu, H. Li, and C.W. Chen, "Wyner-Ziv video coding using progressive encoding and decoding," *2011 Visual Communications and Image Processing (VCIP)*,2011,pp.1-4. DOI:10.1109/VCIP.2011.6116019.
[2] V.Kumar, and S. Sengupta, " Decoder driven multi resolution side information refinements and mode decisions for improved rate-distortion performance in distributed video coding," *2011 IEEE International Conference on Multimedia*

and Expo, 2011, pp.1-6. DOI:10.1504/IJCVR.2014.065570.

[3] D.Zhang, Y. Yang and L. Xie, " Distributed Compressive Video Sensing with Adaptive Reconstruction Based on Temporal Correlation," *2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC)*, 2018, pp.546-550. https://doi.org/10.1155/2021/5539697 ,

[4] S.Shimizu, Y. Tonomura, H. Kimata and Y. Ohtani, "Improved view interpolation for side information in multiview distributed video coding,"*2009 Third ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC)*,2009,pp.1-8. DOI:10.1109/NORSIG.2006.275235.

[5] S.Wang, L. Cui, L. Stanković, V. Stanković, and S. Cheng, " Adaptive Correlation Estimation With Particle Filtering for Distributed Video Coding,"*IEEE Transactions on Circuits and Systems for Video Technology, Vol.22*, 2012,pp.649-658. DOI:10.1109/TCSVT.2011.2171263.

[6] L.Liu, A. Wang, Z. Li, and K.Zhu, " An Improved Distributed Compressive Video Sensing Based on Adaptive Sparse Basis,"*2011 First International Conference on Robot, Vision and Signal Processing*, 2011,pp.137-140. DOI:10.1587/transinf.2016PCP0009.

[7] Y.Tonomura, T. Nakachi, T. Fujii, and H.Kiya, "Parallel Processing of Distributed Video Coding to Reduce Decoding Time," *IEICE Trans. Fundam. Electron. Commun. Comput. Sci., Vol.92-A*,2020, pp.2463-2470. **https://doi.org/10.3390/info13070342**

[8] W.Wang, and J.Chen, "Side Information Generation Scheme Based on Coefficient Matrix Improvement Model in Transform Domain Distributed Video Coding," *Entropy*, Vol.22.2020. DOI: 10.3390/e22121427

[9] N.T.Thao, V.H.Tiến, H.V. Xiem, L.T. Ha and D.T. Duong, "Side information creation using adaptive block size for distributed video coding,"*2016 International Conference on Advanced Technologies for Communications (ATC),* 2016, pp.339-343.

[10] V.Anton, and G.Marat, "A rate-distortion adaptive order of bitplanes decoding for distributed video coding," *2016 XV International Symposium Problems of Redundancy in Information and Control Systems (REDUNDANCY)*, 2016,pp.166-171. DOI:10.1109/RED.2016.7779355.

[11] Y.Zhu, L. Song, R. Xie, and W.Zhang, " SJTU 4K video subjective quality dataset for content adaptive bit rate estimation without encoding,"*2016 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB),2016,pp.* 1-4.

[12] D.Park, H.J. Shim, and B.Jeon, " Fast Side Information Generation Method using Adaptive Search Range,"2012,DOI:10.5909/JEB.2012.17.1.179.

[13] M.Akinola, " Intelligent side information generation in distributed video coding"2015.

[14] H.V.Luong, L.L. Rakêt, X. Huang, and S. Forchhammer, " Side Information and Noise Learning for Distributed Video Coding Using Optical Flow and Clustering," *IEEE Transactions on Image Processing*, Vol.21, 2012, pp.4782-4796.

[15] M.Zheng, " Side information exploitation, quality control and low complexity implementation for distributed video coding," 2013.DOI:10.1109/ICSPCC.2012.6335688.

[16] N.Cen, Z. Guan, and T. Melodia, "Interview Motion Compensated Joint Decoding for Compressively Sampled Multiview Video Streams,"*IEEE Transactions on Multimedia*, Vol.19, 2017, pp.1117-1126. **DOI:** 10.1109/TIP.2012.2215621

[17] L.Meng, Y. Zhao, J. Pan, H. Bai, and A.Wang, " GOP-Flexible Distributed Multiview Video Coding with Adaptive Side Information," *ICCCI*,2010.

[18] J.Kim, J. Kim, and K.Seo, " A Side Information Generation Using Adaptive Estimation and Its Performance Comparison in PDWZ CODEC,"*The Journal of the Korean Institute of Information and Communication Engineering*,Vol. 14, pp.383-393.

[19] R.Oh, H.J. Shim, and B.Jeon, "Adaptive Hard Decision Aided Fast Decoding Method in Distributed Video Coding,"*Journal of the Institute of*

Electronics Engineers of Korea, Vol.47,2013,pp. 66-74.

[20] R.Parseh, and F.Lahouti, " Multi-Mode Nested Quantization in Presence of Uncertain Side Information and Feedback," *IEEE Transactions on Communications*, Vol.61,2013,pp.743-752. DOI:10.1109/IWCIT.2015.7140211

[21] S.Khursheed, N. Badruddin, V. Jeoti, and M.Ahmed, "Fast Side Information Generation for High-Resolution Videos in Distributed Video Coding Applications," *International Journal of Advanced Computer Science and Applications*, Vol.11,2020.

[22] W.Wang, and J.Chen, "Hybrid Side Information Generation Algorithm Based on Probability Fusion for Distributed Video Coding,"*2020 5th International Conference on Computer and Communication Systems (ICCCS)*, 2020.pp.291-295. DOI:10.3390/e22121427.

[23] Y.Mohammad Taheri, M.O. Ahmad, and M.N.Swamy, "Successive refinement of side information frames in distributed video coding," *Multimedia Tools and Applications*,2019,pp. 1-26.

[24] T.H.Vu, T.N. Huong, M.N. Ngoc, and X.HoangVan, " Improving performance of distributed video coding by consecutively refining of side information and correlation noise model," *2019 19th International Symposium on Communications and Information Technologies (ISCIT)*, 2019,pp.502-506. DOI:10.1109/ISCIT.2019.8905187

[25] T.Nguyen, T.B. Huong, T.V. Huu, and S.VuVan, " Content based side information creation for distributed video coding,"*2019 6th NAFOSTED Conference on Information and Computer Science (NICS)*, 2019,pp.223-227.

[26] Q.H.Van, L.D. Hue, V.D. Du, V.N. Hong, and X.HoangVan, "Complexity Controlled Side Information Creation for Distributed Scalable Video Coding,"*2019 3rd International Conference on Recent Advances in Signal Processing, Telecommunications & Computing (SigTelCom)*, 2019,pp.104-108. DOI:10.1109/SIGTELCOM.2019.8696264

[27] B.Dash, S. Rup, A. Mohapatra, B. Majhi, and M.N. Swamy, " Multi-resolution extreme learning machine-based side information estimation in distributed video coding," *Multimedia Tools and Applications*,Vol. 77,2018,pp. 27301-27335.

[28] Y.Cao, L. Sun, C. Han, and J.Guo, " Improved side information generation algorithm based on naive Bayesian theory for distributed video coding," *IET Image Process.,*Vol. *12*, 2018,pp.354-360.DOI:10.3390/e22121427

[29] M.Zamarin, A. Ukhanova, and S.Forchhammer, " Texture side information generation for distributed video coding of video-plus-depth," 2018,DOI:10.1109/ICIP.2013.6738350

[30] Y.Shen, H. Cheng, J. Luo, Y. Lin, and J.Wu, " Efficient Real-Time Distributed Video Coding by Parallel Progressive Side Information Regeneration,"*IEEE Sensors Journal,* Vol.*17*, 2017,pp.1872-1883. DOI:10.1109/JSEN.2017.2653100