Motion Vector Estimation based Uncompressed Video Watermarking Using Extreme Learning Machine

Anurag Mishra
Department of Electronics
Deendayal Upadhyay College
University of Delhi, New Delhi
anurag_cse2003yahoo.com

Preeti Chuttani
GGSIP University
New Delhi

preeti03.ch@gmail.com

Abstract— Multimedia processing with real time constraints requires fast execution of processing algorithms. This is one of the most challenging tasks of the present day research. Moreover, copyright protection of video content is another important issue to ponder with. This paper presents a novel uncompressed video watermarking scheme using a newly developed Single Layer Feed-forward Network (SLFN) popularly known as Extreme Learning Machine (ELM) by computing its motion vectors in uncompressed domain. To achieve this, firstly RGB uncompressed video is divided into frames and extract the frames of blue component. From these frames which are having maximum motion are determined by using block matching method. Secondly, every selected video frame is transformed using DCT and these coefficients are used to train ELM. The trained ELM produces a normalized output vector used as watermarks to be embedded in the low frequency DCT coefficients of the frame. The embedded frames are further examined for its visual quality by average PSNR of all watermarked frames put together. The resultant video exhibits good visual quality. The average PSNR is high which shows that visual quality of frames post embedding is good. The extraction of the watermarks from these frames yield high normalized correlation (NC) and low bit error-rate (BER) values which indicate successful watermark recovery. The result indicate that the watermarking scheme successfully carry out embedding and extraction processes. Further, it shows that the proposed scheme is robust against common video processing attack. Three different video processing attacks are carried out over signed frames to examine robustness of the embedding scheme. Attacked frames also yield high NC and low BER values which clearly indicate that watermark recovery post attack is good. The proposed scheme encompassing ELM training, embedding and extraction completes its execution in seconds which is indicative of the fact that the scheme is quite suitable for real time video watermarking applications.

Index Terms— Single layer feed forward network, Uncompressed RGB video, AVI format, Motion Estimation, Video Watermarking

1. INTRODUCTION

The digital revolution especially in the domain of internet technologies has given a new dimension to the multimedia distribution. Advancement in technology has given multimedia users the ability to tamper with, produce copies of and illegally redistribute digital content. Digital watermarking of multimedia content is one such mechanism which has acquired much importance during last more than one decade. In this case, the location and presence of the embedded information is unknown to unauthorized parties who have illegal access to copyright data. A complete digital watermarking system must include three essential requirements: watermark generation, watermark embedding, watermark detection followed by extraction [1]. Many researchers have given various image/video based watermarking schemes in compressed/uncompressed domain. These researchers had also used different transform domain technique along with various artificial neural networks (ANN) [2-7].

Hartung et al. [2] have revealed one of the pioneering works for watermarking of compressed and uncompressed video. The authors have given a watermarking scheme in which encrypted noise signal is used as a watermark. In their research work they have used DCT as the working domain and claims that the scheme is suitable for hybrid coding also. The author also proves that the scheme is robust against various video processing attacks.

Biswas et al. [3] had developed a compressed video watermarking scheme in which a number of binary images have been obtained from a single watermark and then is used for watermark processing. The authors assert that the scheme is highly robust against spatial attacks such as scaling, rotation, frame averaging and filtering besides temporal attacks like frame dropping and temporal shifting.

Rajab et al. [4] and Faragallah [5] presents an efficient, robust and imperceptible video watermarking technique based on SVD decomposition performed in DWT domain.

El' Arbi et al. [6] have delved upon a novel video watermarking algorithm based on multi resolution motion estimation and artificial neural network (ANN). In this case a multi resolution motion estimation algorithm is adopted to preferentially allocate the watermark to coefficients containing the motion. They have employed a back propagation neural network (BPNN) to memorize the relationship between coefficients in 3×3 block of the image. The authors claim that the proposed scheme is robust against common video processing attacks.

Recently, a novel video watermarking algorithm is developed using a newly developed Single Layer Feed forward Network (SLFN) popularly known as Extreme Learning Machine (ELM) [7]. This machine is very fast and completes its training in few milliseconds. The authors have shown that this machine can be successfully utilized for developing real time watermarking applications.

In this paper, we propose ELM based approach to implement watermarking of frames. The description of ELM is described in section 2. In this scheme we have selected two standard uncompressed videos. In order to avoid watermarking of complete video we select few frames out of total number of frames, which is carried out with the help of motion vector estimation. We are selecting only those frames which are having maximum motion and that to only 10% of the total number of frames. The selection criterion is given in the listing 1 of section 3. After selecting frames, embed the watermark into low frequency component of uncompressed AVI video frames by using DCT- ELM watermarking scheme using the embedding procedure given in listing 2 of section 3. The scheme utilized here reduces the issue of time complexity.

The signed video sequences are found to have good visual quality which is indicated by high PSNR values. The extraction of the watermark from these frames yield high normalized correlation (NC) and low bit error-rate (BER) which indicate that the extracted watermark is highly correlated with the original watermark. The signed video frames are also examined for robustness by executing three different video processing attacks. The attacks used in the present work are: (1) Scaling % (20, 40, 60, 80 and 100), (2) Gaussian Noise (with mean = 0 and variance 0.001, 0.02, 0.04, 0.08, 0.016), (3) JPEG (QF = 75, 80, 85, 90 and 95). This paper is organized into 4 sections. Section 2 presents a brief theoretical description of ELM algorithm and mathematical formulations. Section 3 describes the proposed embedding and extraction procedure. Section 4 delves upon the results obtained in this simulation and its discussion. Finally, Section 5 presents the conclusion followed by list of references.

2. EXTREME LEARNING MACHINE

The Extreme Learning Machine [8-9] is based on a Single hidden Layer Feed forward Neural Network (SLFN) architecture. This technique differs from the conventional training algorithms such as Back Propagation (BP) algorithms which may face difficulties in manual tuning control parameters and local minima. On the other side, training of ELM is very fast; it has a good accuracy and offers a solution in the form of system of linear equations. For a given network architecture, ELM does not have any control parameters like stopping criteria, learning rate, learning epochs etc., and thus, the implementation of this network is very simple. In this algorithm, the input weights and hidden layer biases are randomly chosen which are based on some continuous probability distribution function. The output weights are then analytically calculated using a simple generalized inverse method known as Moore-Penrose generalized pseudo inverse [10]. By using this method the ELM training is carried out in millisecond time. It is concluded that the proposed ELM based fast embedding and extraction scheme is suitable for real time applications which is one of the most important consideration for multimedia processing.

2.1 Mathematics of ELM Model

Given a series of training samples $(x_i, y_i)_{i=1,2,...,N}$ and $\hat{N}$ the number of hidden neurons where $x_i = (x_{i1},..., x_{in}) \in \Re^n$ and $y_i = (y_{i1},..., y_{im}) \in \Re^m$, the actual outputs of the single-hidden-layer feed forward neural network (SLFN) with activation function $g(x)$ for these N training data is mathematically modelled as

$$\sum_{k=1}^{\hat{N}} \beta_k g(\langle w_k, x_i \rangle + b_k) = o_i, \quad \forall i = 1,..., N \tag{1}$$

where $w_k = (w_{k1},..., w_{kn})$ is a weight vector connecting the k^{th} hidden neuron, $\beta_k = (\beta_{k1},..., \beta_{km})$ is the weight vector connecting the k^{th} hidden neuron and output neurons and b_k is the threshold bias of the k^{th} hidden neuron. The weight vectors w_k are randomly chosen. The term $\langle w_k, x_i \rangle$ denotes the inner product of the vectors w_k and x_i and g is the activation function.

The above N equations can be written as

$$H\beta = O \tag{2}$$

and in practical applications $\hat{N}$ is usually much less than the number N of training samples and $H\beta \neq Y$,where

$$H = \begin{bmatrix} g(\langle w_1, x_1\rangle + b_1) & \cdots & g(\langle w_{\hat{N}}, x_1\rangle + b_{\hat{N}}) \\ \vdots & & \vdots \\ g(\langle w_1, x_N\rangle + b_1) & \cdots & g(\langle w_{\hat{N}}, x_N\rangle + b_{\hat{N}}) \end{bmatrix}_{N\times\hat{N}}$$

$$\beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_{\hat{N}} \end{bmatrix}_{\hat{N}\times m}, \qquad O = \begin{bmatrix} o_1 \\ \vdots \\ o_N \end{bmatrix}_{N\times m}$$

And

$$Y = \begin{bmatrix} y_i \\ \vdots \\ y_N \end{bmatrix}_{N\times m} \quad (3)$$

The matrix H is called the hidden layer output matrix. For fixed input weights $w_k = (w_{k1},...,w_{kn})$ and hidden layer biases b_k, we get the least-squares solution $\hat{\beta}$ of the linear system of equation $H\beta = Y$ with minimum norm of output weights β, which gives a good generalization performance. The resulting $\hat{\beta}$ is given by $\hat{\beta} = H^{+}Y$ where matrix H^{+} is the Moore-Penrose generalized inverse of matrix H [10]. The above algorithm may be summarized as follows:

2.2 The ELM Algorithm

Given a training set

$S = \{(x_i, y_i) \in \Re^{m+n}, y_i \in \Re^m\}_{i=1}^{N}\sum$, for activation function $g(x)$ and the number of hidden neurons $\hat{N}$;

Step1: For $k = 1,...,\hat{N}$ randomly assign the input weight vector $w_k \in \Re^n$ and bias $b_k \in \Re$.

Step2: Determine the hidden layer output matrix H .

Step3: Calculate H^{+} .

Step4: Calculate the output weights matrix $\hat{\beta}$ by $\hat{\beta} = H^{+}T$.

Many activation functions can be used for ELM computation. In the present case, sigmoid activation function is used to train the ELM.

2.3 Computing the Moore-Penrose Generalized Inverse of a matrix

A matrix G of order $\hat{N}\times N$ is the Moore - Penrose generalized inverse of real matrix A of order $N\times\hat{N}$ $AGA = A, GAG = G$ and AG, GA are symmetric matrices.

Several methods, for example orthogonal projection, orthogonalization method, iterative methods and singular value decomposition (SVD) methods exist to calculate the Moore-Penrose generalized inverse of a real matrix. In ELM algorithm, the SVD method is used to calculate the Moore-Penrose generalized inverse of H. Unlike other learning methods, ELM is very well suited for both differential and non – differential activation functions. As stated above, in the present work, computations are done using "Sigmoid" activation function.

3. Proposed DCT - ELM Based Uncompressed Video Watermarking Scheme

First, the video is divided into non-overlapping frames of size M × N. While embedding the watermark into blue channels of frames of the uncompressed video, it is observed that to obtain the complete signed video is a bit costly in time. Therefore, a selection criterion is applied to select the frames fit to be watermarked. This way, not more than 10% of the total frames of the video are watermarked. The selection criterion is based on motion vector estimation. Listing 1 stimulates the steps used for these selection criteria.

Listing 1: Criterion to select required number of frames

- The motion vectors between two neighboring frames are computed by using block matching approach. In this process full search method is used
- As a result of step 1, each frame is divided into blocks of size 8 x 8, the motion range is found to exist between 0-98. To determine the locations of maximum motion, set a threshold of 75. The locations where the motion is found to be greater than 75 are selected
- Thereafter, we set another threshold to select only those frame combinations where the maximum motion occurs in more than 50 locations in accordance with the criterion stipulated in step no 2 depending upon the type and nature of video
- A total of maximum 10% of the frames are finally selected for watermark embedding after computing step 3

Watermark Generation

Second, the blue channel of the selected frame is divided into 8

x 8 pixel blocks and DCT of all such blocks is computed to transform the blocks into frequency domain. Zigzag scanning of each block is done to select first 21 AC coefficients barring the DC coefficient. Thus, a dataset of size 1024*21 is created which holds 21 selected coefficients from each of 1024 blocks in all. From this dataset, the mean of each row is computed and placed at the first column position. This results in creation of another dataset of size 1024 x 22 wherein the mean values (labels) are fixed in column 1. This is used as the training dataset for the ELM to be used in regression mode. Once the ELM algorithm is trained, it gives as its output the predicted values for the mean calculated in the form of a vector of size 1024 * 1. The predicted output values corresponding to each block are kept in a separate data file. This output vector is used as watermark to be embedded within the original frame using Cox's formula given in Eqn. (4) [11].

$$v_i^{'} = v_i(1.0 + \alpha \times f(x)) \tag{4}$$

Where $f(x)$ is output of ELM after training, v_i are DCT coefficients and $v_i^{'}$ are coefficients of the signed frame. The parameter α is known as scaling factor and is optimized to be 0.5 for all our practical calculations. The computed watermark is embedded into each block and the inverse DCT of each block is taken to retrieve the signed image. Listing 2 gives watermark embedding algorithm.

Listing 2: Watermark Embedding Algorithm

- Block code blue channel of each frame to size 8 x 8.
- Transform the blue channel of the frame in frequency domain by computing block wise DCT.
- Apply zigzag scanning to all AC coefficients of each block and select first 21 coefficients from each block barring the DC coefficient. Thus develop a dataset of size 1024 * 21 using these coefficients.
- Compute the mean of all 21 coefficients for each row and place it in first column as label. Thus, recreate a dataset of size 1024 * 22.
- Train the ELM in regression mode by supplying this dataset to the machine. As a result, the ELM produces an output vector of size 1024 * 1 which is used as watermark to be embedded within the host image using Cox's formula.
- Take Inverse DCT (IDCT) to obtain blue channel of the signed frame.
- Then combine the modified blue channel with original R and G to obtain the colored signed frame. and then placing that signed frame at their position to get the signed video.

The embedded frames are further examined for its visual quality by computing PSNR individually and taking an average PSNR of all frames put together. Eqn. (5) and (6) respectively give mathematical formulae for PSNR and AVG_PSNR. The computed results are presented in detail in Sect. 4.

$$\mathrm{PSNR} = 10\log_{10}\left(\frac{255^2}{\mathrm{MSE}}\right) \tag{5}$$

Where MSE =mean square error

$$\mathrm{AVG\_PSNR} = \sum_{i=1}^{T}\frac{\mathrm{PSNR}}{T} \tag{6}$$

Where T is total number of frames

Eqn. (5) is also making use of another parameter known as the Mean Square Error (MSE) which is given in the Eqn. (7).

$$\mathrm{MSE} = \frac{1}{N^2}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}(C_{ij} - R_{ij})^2 \tag{7}$$

Where C_{ij} = Sample of current block
R_{ij} =Sample of the reference area
N = size of block.

Watermark Extraction

During watermark extraction process, initially 8 x 8 block wise DCT of blue channels of both original and signed frames are computed and the coefficients of the original frame which are used in embedding process are subtracted from the respective coefficients of the signed frame. In this manner, both the original and recovered watermark sequence is known. Listing 3 depicts the watermark extraction algorithm.

Listing 3: Watermark Extraction Algorithm

- Divide blue channels of original and watermarked frames into 8 x 8 size blocks
- Compute DCT of all blocks of both these frames
- Subtract only those computed coefficients of the original frame from the respective coefficients of signed frame which are used in embedding process and thus recover the watermark
- Compute the normalized cross-correlation coefficient or NC (W, W') and Bit Error Rate or BER (W, W') using original and recovered watermark sequences W and W' respectively. These parameters are given by Eqns. (8) and (9) respectively

$$\mathrm{NC}(W, W') = \sum_{i=1}^{x}\sum_{j=1}^{y}\frac{[W(i,j)*W'(i,j)]}{[W(i,j)]^2} \tag{8}$$

$$\mathrm{BER}(W, W') = \frac{1}{xy}\sum_{j=1}^{xy}[W'(j) - W(j)] \tag{9}$$

4. Results and Discussion

The performance of the proposed watermarking scheme is evaluated on two standard video in RGB uncompressed AVI format. These video sequences are Suzie and Car respectively. As described in Section 3, the watermark is generated in the

form of real numbers produced by the ELM and is embedded into the low frequency coefficients of the blue channel of the selected frames. The entire simulation is carried out using MATLAB R2013a. The selected frames for two video sequences which are used for watermark embedding are given in Table 1.

Table1: Selected Frames of the given video

Name of Video	Selected Frames
Suzie	47,50,53,56,59,62,65,71
Car	4,12,19,27,34,42,49,59

Figs. 1-2 depict two original frames of Suzie and Car video sequences. Their respective signed frames are depicted in Figs. 3 and 4. Their respective PSNR values are mentioned below these frames. High computed PSNR values indicate that the visual quality of these frames is very good.

Fig. 1: 59th original video frame of Suzie video

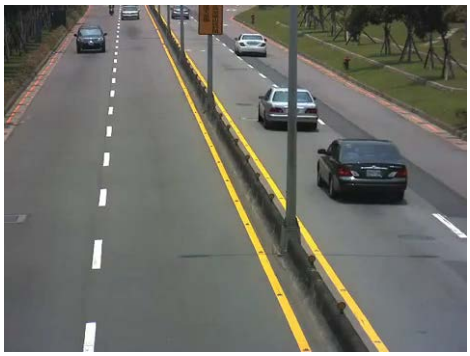


Fig. 2: 34th original video frame of Car video

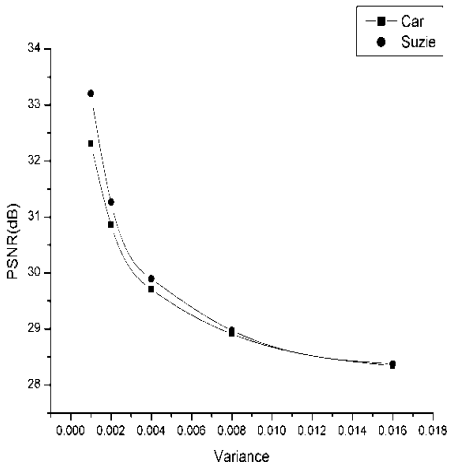
Fig. 3: 59th signed video frame of Suzie video (45.459 dB)

Fig. 4: 34th signed video frame of Car video (37.241 dB)

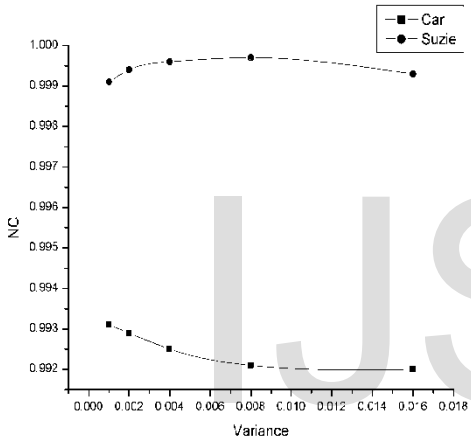
The performance of the proposed watermarking scheme is evaluated by computing parameters PSNR for visual quality, NC and BER for similarity of the recovered watermark. However, as the video comprises of number of frames, the visual quality is examined by calculating Average PSNR for the watermarked frames. Similarly, average values of NC and BER are also computed. The Average PSNR in our simulation is 43.84 dB and 36.68 dB, respectively for Suzie and Car sequences by using a scaling factor $\alpha = 0.5$.

Average NC and Average BER values for Suzie are NC = 0.9996 and BER = 0.1937 respectively. These values for Car are NC = 0.9987 and BER = 0.3781 respectively. Significantly high values of NC and low values of BER clearly indicate that watermark recovery is completely successful for the proposed scheme.

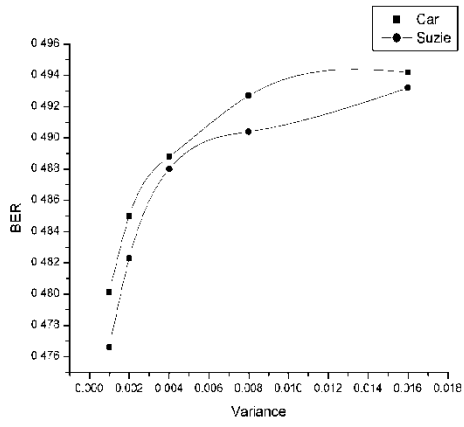
To examine the issue of robustness of the proposed watermarking scheme, three different video processing attacks are implemented over the signed video frames. These attacks are Scaling (Scale % = 20, 40, 60, 80 and 100), Gaussian noise (Noise variance = 0.001, 0.002, 0.004, 0.008 and 0.016) and JPEG compression (QF = 75, 80, 85, 90 and 95). PSNR, NC and BER are calculated with respect to variation in respective attack parameter and plots are shown in Figs. 5 - 7.



(a)



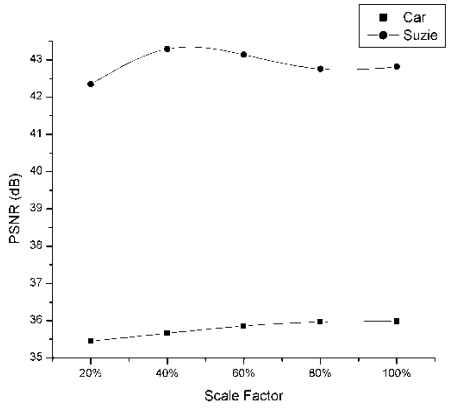
(b)



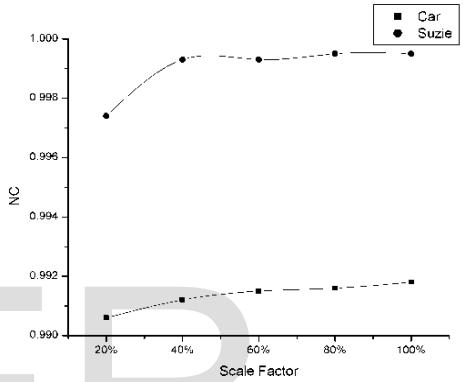
(c)

Fig 5(a–c): Plots of PSNR, NC and BER w. r. t. Gaussian noise variance

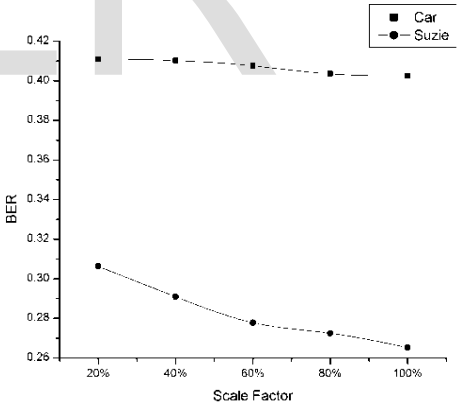
The plots shown in Fig. 5 clearly indicate that both PSNR and NC substantially decrease as a result of increase in noise variance. As usual, the BER increases with the increase in the noise variance which indicates successful watermark recovery after carrying out Gaussian noise attack. Note that the set behavior is observed for all two video sequences used in the present work.



(a)



(b)



(c)

Fig 6(a–c): Plots of PSNR, NC and BER w. r. t. Scaling %

It is clear from Fig. 6 that both PSNR and NC are found to increase with the increase in % of scaling coefficients. Agarwal et al. [11] have also reported a similar behavior in case of scaling attack. On the contrary, the BER is found to decrease substantially with the increase in scaling percentage. Large values of PSNR and NC clearly indicate that the proposed watermarking algorithm resists the scaling attack carried out over signed frames. It is found to exhibit good visual quality as well as successful watermark recovery even after implementing large scaling ratios.

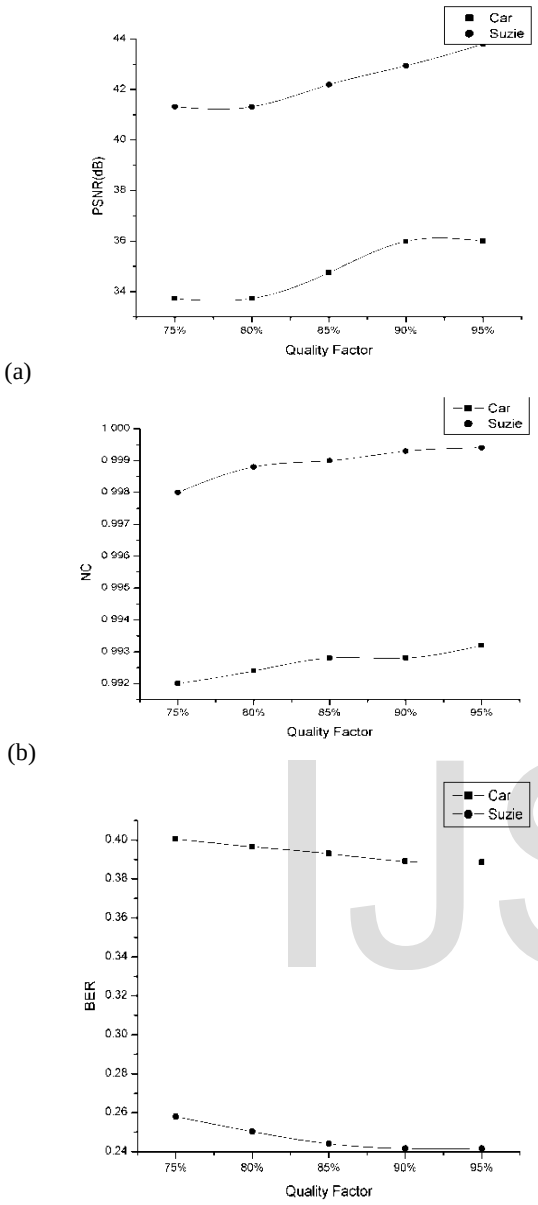


Fig 7(a–c): Plots of PSNR, NC and BER (%) w. r. t. Quality factor

Fig. 7 shows the plots of PSNR, NC and BER w. r. t JPEG quality factor parameter. The PSNR and NC are found to increase with the increase in quality factor and BER decreases with the increase in quality factor. This is in accordance with the expected behavior to indicate that the proposed embedding scheme resists the JPEG attack carried over the signed frames. To analyze the issue of time complexity of the proposed watermarking scheme, we take into account ELM training time, embedding time, extraction time for those frames which are embedded with the watermark. The training of ELM is an integral part of both embedding and extraction procedure. Table 1 compiles the average training time, embedding and extraction time spans for results for selected frames in two video sequences.

Table1: Time (seconds) consumed by different processes of proposed scheme

Video / Time (Sec)	Suzie	Car
ELM Training	0.0370	0.0760
Embedding	0.2320	0.3939
Extraction	0.01567	0.0156
Total Time	0.28467	0.4855

This indicates that the selected frames of the given video are successfully watermarked within a time span ranging between milliseconds to seconds for the entire video. This makes the proposed Motion vector computation and ELM machine based watermarking scheme a suitable candidate for developing real time video watermarking applications.

In this work, the results clearly indicate good optimization of visual quality and robustness obtained by using ELM algorithm with minimum time complexity. The training of ELM is an integral part of both embedding and extraction procedures. We therefore conclude that the proposed watermark embedding and extraction scheme developed using DCT-ELM is capable to implement real time video watermarking applications.

5. CONCLUSION

In this paper, we successfully demonstrate a video watermarking scheme for two given videos in RGB uncompressed AVI format. This scheme is implemented in DCT domain using a fast executing neural network commonly known as Extreme Learning Machine (ELM). Selection of frames for watermark embedding is determined by computing its motion vectors between the ongoing frames. The fast training of this machine is suitable for optimized video watermarking on a real time scale. Successful watermark recovery is indicated by high normalized cross correlation values and low bit error rate values between embedded and extracted watermarks. The robustness of the proposed scheme is examined by carrying out three different video processing attacks. This scheme is found to be robust against selected attacks. It is concluded that the proposed scheme produces best results due to optimized embedding facilitated by training of ELM in minimum time and overall this scheme is suitable for developing real time video watermarking applications.

REFERENCES

[1] K. R. Chetan and K. Raghavendra, "DWT Based Blind Digital Video Watermarking Scheme for Video Authentication," International Journal of Computer Applications, vol. 4, no. 10, pp. 19-26, 2010.

[2] F. Hartung and B. Girod, "Watermarking of Uncompressed and Compressed Video," Signal Processing, vol. 66, no. 3, pp. 283-301, 1998

[3] S. Biswas, S. R. Daz and E. M. Petriu, "An Adaptive Compressed MPEG-2 Video Watermarking Scheme," Instrumentation and Measurement, IEEE Transactions on, vol. 54, no.5, pp. 1853-1861, 2005

[4] L. Rajab, T. Al-Khatib and A. Al-Haj, "Video Watermarking Algorithms Using the SVD Transform," European Journal of Scientific Research, vol. 30, no. 3, pp. 389-401, 2009

[5] O. S. Faragallah, "Efficient Video Watermarking Based on Singular Value Decomposition in the Discrete Wavelet Transform Domain," AEU-International Journal of Electronics and Communications, vol. 67, no. 3, pp. 189-196, 2013

[6] M. El'Arbi, C. B. Amar and H. Nicolas, "Video Watermarking Based on Neural Networks," In Multimedia and Expo, IEEE International Conference, pp. 1577-1580, July2006

[7] C. Agarwal, A. Mishra, A. Sharma, and G. Chetty, "A Novel Scene Based Robust Video Watermarking Scheme in DWT Domain Using Extreme Learning Machine," In Extreme Learning Machines 2013: Algorithms and Applications, pp. 209-225, Springer International Publishing 2014

[8] M. B. Li, G. B. Huang, P. Saratchandran and N. Sundararajan, "Fully Complex Extreme Learning Machine," Neurocomputing, vol. 68, pp. 306-314, 2005

[9] G. B. Huang, Q. Y. Zhu and C. K. Siew, "Extreme Learning Machine: Theory and Applications," Neurocomputing, vol. 70, no.1 pp. 489-501, 2006

[10] G. B. Huang, The Matlab code for ELM is available on http://www.ntu.edu.sg

[11] A. Mishra, A. Goel, R. Singh, G. Chetty, and L. Singh, "A Novel Image Watermarking Scheme Using Extreme Learning Machine," In Neural Networks (IJCNN), the 2012 International Joint Conference, pp. 1-6, 2012