DNN Based Moth Search Optimization for Video Forgery Detection

Malle Raveendra, K Nagireddy

Abstract: ***Recently, video tampering process becomes easier due to the rapid advancements in user-friendly editing software and multimedia technology (e.g., Mokey by Imagineer Systems, and Photoshop and Premiere by Adobe). This technologies may highly tamper the original videos, so that the audience gets mislead. Nowadays, MPEG-4 codec is included in a large proportions of video cameras and surveillance systems. Therefore the double compression detection process included as an initial step in the video forensic is receiving a high significance. In this paper, the double compression artifacts is detected by adopting the Markov based features, which identifies the interpolated original videos. The double compressed frames are then segmented by introducing an SLIC super pixel segmentation technique. Here, the feature extraction is performed by applying the scale information that is obtained from the multi-scale Gabor filters. The features of this Gabor scale accurately extract the structural features and also reduce too much of redundancy. This extracted features are then provided to DNN (deep neural network) for forgery detection. In this video forensic process, DNN classifier is included for forgery detection. The CNN classifier is included in various existing forgery detection techniques. But, in our work we include DNN because it contains number of hidden layers which provide accurate results for this forgery detection process. To improve the DNN performance, Moth Search Optimization (MSO) algorithm is introduced in this forgery detection technique. Every nook and corner of this world we can able to find the surveillance cameras for security purpose. But, some fraudsters perform forgeries in this recorded videos for their own benefits. To identify this, a lot of forgery detection techniques are coming into existence. So in this work, we introduce the DNN based MSO to perform the forgery detection in videos.This implementation is processed in python simulation platform. The parametric evaluations are taken in terms of F1-Score, average accuracy, Precision, Recall and. Experimental results will provide improved performance in video forgery detection.***

Keywords: Video forgery, Markov statistics, Multi-scale Gabor filter, DNN, Moth Search Optimization (MSO), SLIC super pixel segmentation.

I. INTRODUCTION

The digital video technology's rapid growth makes tampering or editing a video sequence verysimpler than before. An object can be removed in a video sequence using dominantsoftware for video editing like Apple Final Cut Pro, Adobe Premiere, and Adobe after Effects. Recently, to a greater extentthe research of video tampering detection was focussed by many researchers.

Detection of object forgery is a novel theme in the field of passive digital video forensic research [1-2].A video comprises of a series of images called frames.

Revised Manuscript Received on October 15, 2019

Malle Raveendra, Ph.D Scholar in Electronics and Communication Engineering, Jawaharlal Nehru Technological University, Anantapuramu.

Dr.K Nagireddy, Professor in Electronics and Communication Engineering NBKR Institute of Science and Technology.

In the temporal, spatio-temporal, and spatial fields the video forgery attacks are performed. Only in the both spatio-temporal and spatial domains the copy-paste and Region splicing tampering arises and in the temporal field the Frame insertion, removal, copying and shuffling will occur. It is instinctivethat also on videos the detection process of image tampering can also be used[3].

Howeversuitable results as expected will not be producedowing to complexphasesin videos such asfrequent moving objects or noise sustainedas a result of compression. Moreover, excluding the information of the temporal domain causes heavy computation cost. Into two types the Video tampering recognitionmethods can be classed i.e. active and passive (correspondinglytermed as blind). In active methods digital signature and watermarking comes, to validate the video certainimportantinformation's will be purposely embedded into it. Tampering takes place if any change is made to the embedded information [4-7].

Only a few devices has the facility to embed a digital signature or a watermark to the captured video. Once tampering is done, in such situations these methods possibly fails before inserting digital signature or watermark. Mostly in naturalconditionscertain information of the videos such as metadata will not be available. Because of the intelligent models development the making of undoubted fake video content has enlarged. For some years the selective modification of the image contend was done, but the use of same approaches to video demands excessively labour regarding its mass use. In a video if every frame is considered as aseparate image, there could be possiblya lot of images to process effectively. These issues could be handled with better computing power, also with the DNNs evolution.

Recently in many applications the Deep learning methods have seen a huge success.So in various new fields these approach has been used, such as identification of camera model [8,9], steganalysis [10], image recapture forensics [11], detection of image manipulation [12], copy-move forgery detection (CMFD) of image [13], and so on. Particularly the Generative Adversarial Networks (GANs) are appliedfor alteringthe original video to recreate facial expressions of human [14], change the meteorological conditions [15] and for face-swapping applications [16]. The re-enactment of Human face is certainly a new howeveris a common research area where simply, the speaking head is altered visually to imitate the a second actors facial expression [14, 17, 18] or to pair with another audio track [19, 20].

These are having some of the simple applications for instancemovie re-dubbing in a several language or producing new movie scenes with an iconic actor's old video tapes, although fake

contents can also be produced . In certain situations, to reliably fool human eyes the fake content is definitely enough. From [18] the authors realised that apart from random guessing the performance of human viewers is better when aiming to determine whether the footages of facial re-enactment was genuine or integrated. However, the DNN possibly canextricateeasily the authentic and forged footage.

The main contribution of this work is to identify the video forgery. Stronger forensic evidences are provided by video sequences than the motionless images. Thus, surveillance video, as important evidence, is often used in the case investigation. The development of forgery software's increase the risk of forgery in videos. This forgery is performed in an accurate manner, so it takes time to detect this forgery and also found expensive. To minimize this cost and time of forgery detection, various research process are continuing in this field. In this SLIC superpixel segmentation process is included to segment the frames into number of regions. The features from this segmented portions are extracted by multi-scale Gabor filters. This extracted features are very much valuable for DNN to perform the forgery detection process.

The organization for this entire paper is: In Section 2. Some existing methods that are implemented for video forgery detection is discussed. The complete framework of this proposed method is discussed in Section. 3, here some details regarding Markov statistics based double compression, SLIC segmentation, multi-scale Gabor filters, DNN, and MSO are provided. The experimental analysis and outcome of this proposed method is discussed in Section. 4. Lastly, the conclusion for the proposed method is provided in Section. 5.

II. RELATED WORK

In the digital multimedia forensicsdomain, a major research advances have been made and some of them are based on image forgery and video forgery and detection [21, 22, 23, 24, 25, and 26]. This works performs frame duplication, deletion and insertion types of forgeries in video frames. Ashort review of related researches is presented in this sectiontowards the detection of fore-mentionedcategories of forgeries in videos.The tampered video detection faces lot of challenges. However the tampering detectors are accessible they are concerned with some of the particular tampering techniques.

In [18] the authors generatedadataset to date, entirely on the digital re-enactment strategy of [17]. For detecting the manipulated video with minimum errors a DNN was employed,although in [27] it was shown that this approach could not be transmitted to new video manipulation approaches. Singh and Aggarwal [28] after reviewing the video content authentication methods, noticed that for the genuine doctored videos there is no reliable database available. For image rebroadcast detection ahuge dataset was produced in [29]. Demonstration was conducted by them on this dataset that outperformed some of the previous techniques, CNN based diverse dataset, obtained an accuracy of about 97% while detecting the forged and real videos.Compared with the detection techniques [30] the manipulated techniques are more effective, there are many approaches for altering an image or video in digitalised manner but only a few methods are available to identifythese types of manipulations. Because of this, it is essential to develop a detection methods that are not based on the nature of video manipulation.

To detect tampering and to find the reliabilities and patterns within data [18, 31] the Machine learning techniques are proved to be good, but regarding their large data requirements a new techniques isrequired. For detecting inter-frame tampering the authorsin [22] used the compression elements of macro-block compression type. In the video sequence that are encoded within MPEG-2, the deleted frames were determined with 95% accuracy using machine learning approaches. To identifythe video forgery based on chroma–key composition the authors in [31] proposed the Auto-encoder with recurrent neural network. The authors for detecting this forgery segregated the patches of 128×128 size from frames. By applying a single highpass third-order derivative filter some handcrafted features were mined from every patch. The auto-encoders were employed for producing an anomaly score. The auto encoder's parameters were learned from the authentic frames of the handcrafted features. The generated feature vector in the testing phasethat fails to fit with the intrinsic network model therefore large error is produced.

To determine the quality of image, the authors in [32] developed a process of estimating the HEVC frames QP directly from pixels using DNN. Accurate outcomes were achieved by means of a patch-wise method and also by dataset obtained from UCID[33]. In this approach QP estimation is examined fromsequences of H.264/AVC video. This H.264/AVC is a prevalent video compression standards which is utilised on YouTube, public datasets and broadcast videos. The frame patches of a video sequences is classified by CNN utilizing its quantisation parameters as labels. Disparate [32], further investigation can be done to detectthe forgeries in videos using these features.

III. VIDEO FORGERY DETECTION BY DEEP LEARNING BASED MSO

A. Overview

Video forgery detection plays a major role in various fields. This forensic is widely applied to identify the criminal activities that are performed by offenders in video. Recently, number of research process are performed in this forensic field for forgery detection. An advancement in technologies develop various software to accomplish the forgery in video. Particularly, Adobe Photoshop and Video Editor are some of the multimedia tools and software that are developed recently to tamper or edit the medial files.

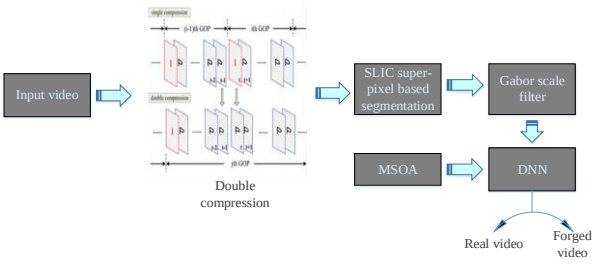

Figure. 1 Overview of proposed method

In this work, initially the video is subjected for double compression process. Here, the Markov statistics based double compression process is applied. The frames obtained from double compression is then provided as an input for SLIC super-pixel. This SLIC approach is employed to perform the segmentation process in double compressed frames. This frames are segmented into a number of regions by SLIC super-pixel segmentation process. The segmented frames are then given to Gabor scale filter, which utilizes the scale information for feature extraction. The extracted features are then passed into DNN for forgery detection. Based on this extracted features, the forged frames are identified by this DNN classifier. The performance of this classifier is further improved by introducing an optimization algorithm (i.e. MSO).

B. Double compression by Markov statistics

In JPEG images, Markov statistics achieves high performance for both single and double compression. The blocks that are found in MPEG-4 is encoded by the JPEG-like scheme, this shows that Markov statistics also found to be effective for this double MPEG-4 compression detection [34]. The procedure for feature extraction from MPEG-4 by Markov statistics process is shown in Figure.1 [34].

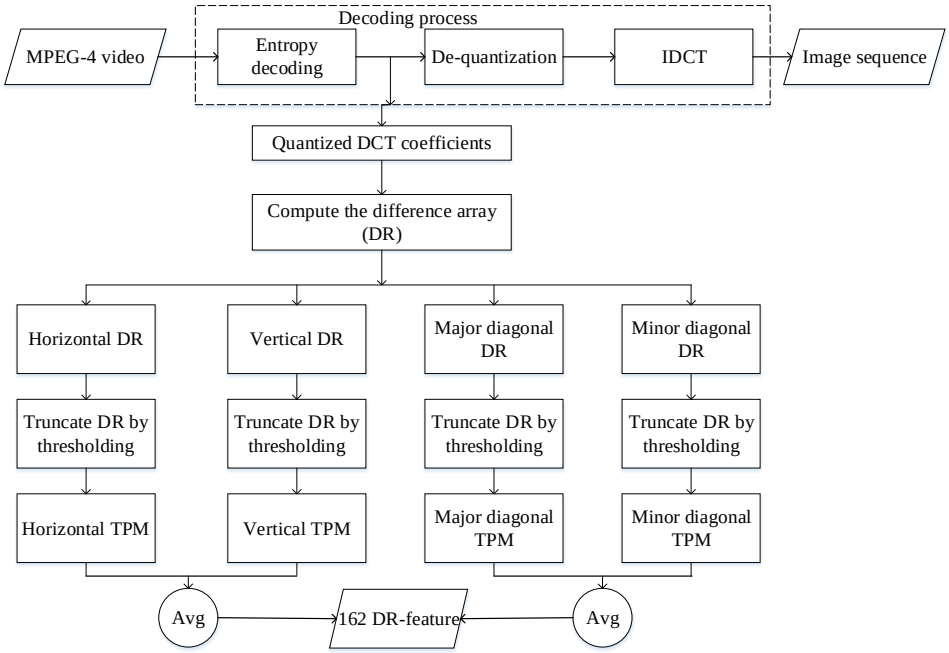

Figure. 2 Markov statistics based double compression

In this process, initially, the quantized DCT coefficients $q(i, j)$ are obtained while executing the decoding process. Let us assume, only the magnitudes of coefficients that are obtained in y component. Then, the difference ray (DR) along vertical (v), horizontal (h), major (M) diagonal and minor (m) diagonal directions are computed. For (h) direction, the artifacts of double compression is extruded by the differential operation,

$$DR^{h}_{i,j} = q_{i,j} - q_{i,j+1} \quad (1)$$

In this third step, (DR), is truncated by applying the threshold operation. If the obtained value is found to be $> t$ or $< -t$, then the corresponding value is normally referred as t or $-t$. Where this t is set as 4, because the statistical analysis that is performed on MPEG-4 video shows that almost all the elements of the (DR) may fall in the interval [-4, 4].

In the fourth step, model the first-order Markov random technique for each (DR), along the same direction. The Markov transition probability matrix (TPM) (P^h) for horizontal direction is calculated by using following equation,

$$P^{h}_{a,b} = \Pr\left(DR^{h}_{i,j+1} = b / DR^{h}_{i,j} = a\right), \quad (2)$$

Where, $a, b \in [-T, T]$. By using this process, the 9×9 TPM on several direction is obtained.

Finally, the method that is discussed in [39], is applied to minimize the feature dimensionality. Then, the vertical, horizontal matrices and major, minor diagonal matrices are averaged separately to attain the feature sets f and F as

$$f = \frac{1}{2}\left(P^{h} + P^{v}\right) \quad (3)$$

$$F = \frac{1}{2}\left(P^{M} + P^{m}\right) \quad (4)$$

The concatenation of both f and F provide the final feature F' which is 162-DR.

The statistical artifacts is produced by rounding errors among the DR elements, whereas this rounding errors is produced by double quantization technique. Based on the random process theory, the DRs are characterized by the one-step Markov TPM. So, gradually the machine learning framework can be employed to identify the double MPEG-4 compression.

C. SLIC super pixel segmentation

The decompressed frames are then subjected to segmentation process. Here, the most popular SLIC (Simple Linear Iterative Clustering) super-pixel segmentation is applied for frame segmentation. It segments the frames into a number of regions. This segmented regions may exhibit some certain properties like spectral homogeneity or compacity. Frequently, this approach is included to extract the objects from the image. A less ambitious version that is developed for the segmentation issues is superpixel segmentation. The main purpose of this superpixel segmentation is to split the frame into minute, compact, same size, and homogeneous segments [35]. The most popular superpixel approach is SLIC approach, it is found in the class of gradient-ascent-based techniques and also affordshigh segment adherence for boundaries than the other segmentation approaches. An example image for this SLIC superpixel segmentation is shown in Figure. 3,

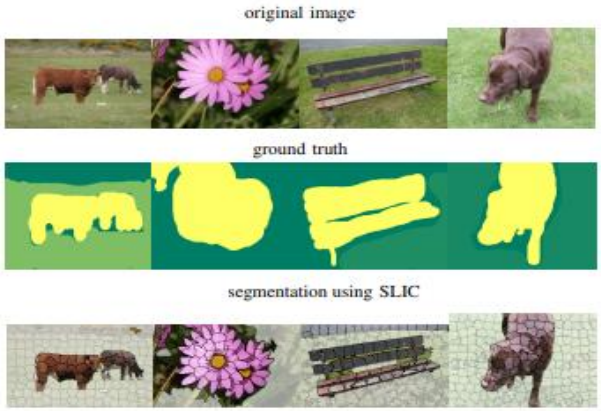

Figure. 3 Example image for SLIC superpixel segmentation

Here, the frame is represented as F having n pixels of $X \times Y$ dimensions, whereas the each pixels P_N at positions (x_N, y_N) contains R features called $f_N^r, N = [1,..,n], r = [1,..,R]$. This SLIC algorithm is same as that of K-means clustering algorithm. Every segment is denoted by the centroid $C_t = (X_t, Y_t, F_t^r), t = [1,..,T]$. The spatial mean (gravity's geometric center) is represented by the couple (X_t, Y_t). The mean features on the t segment is represented as F_t^r.

During initialization, the frame is divided into T square segments succeeding the regular grid. The algorithm parameter normally referred as spatial width (SpW) parameter is applied as an original grid size (square grid elements side length). This parameter may provide the initial size of the super-pixels and it also defines the total number of super-pixels. Then, the subsequent steps of this algorithm is discussed below.

In iterative phase, the entire image pixels are associated along with its adjacent centroid, which is found to be same as that of the K-means principle. The distance that exist between this iterative steps is estimated as weighted sum of spatial and Euclidean distances of feature. The parameter of DW (distance weight) permits the manipulator to provide high significance to either segment homogeneity or segment compacity.

$$d(P_N, C_t) = d_{feature} + DW \times d_{spatial} \quad (5)$$

$$= \sqrt{\sum_{r=1}^{R} (F_t^r - f_N^r)^2} + DW\sqrt{((X_t - x_N)^2 + (Y_t - y_N)^2)} \quad (6)$$

Here, the candidate centroids are selected within the limited search radius, so that the unnecessary distances are avoided. After updating the entire pixels, recalculate the centroids, and then began the subsequent iterative step. The iteration process will stop after reaching the convergence criterion, i.e., either the total residual (which contains maximum square distance that is obtained among the centroids of two successive steps) is found minimum than an available value or else a maximal number of iterations.

D. Multi-scale Gabor filters

The segmented images are then given to this multi-scale Gabor filter for feature extraction. The Gabor filter is introducedin 1946 by Gabor. The properties of orientation and spatial frequency selectivity, spatial localization, are represented by this Gabor wavelet [36]. Gabor related researches are widely performed in image understanding areas and also in pattern recognition. The Gabor wavelet is defined as follows:

$$\psi_{b,c}(a) = \frac{\|K_{b,c}\|^2}{\delta^2} e^{\left(-\|K_{b,c}\|^2 \|a\|^2 / 2\delta^2\right)} \left[e^{iK_{b,c}a} - e^{-\delta^2/2}\right] \quad (7)$$

Where, $a = (p, q)$ represents the image pixels, the Gabor filters orientation is represented as b, c represents the scales of Gabor filters, whereas the symbol $\|\cdot\|$ represents the norm operation.

$$K_{b,c} = K_c e^{p\varphi_b}, \quad K_c = K_{\max} / f^c, \varphi_b = \pi b/4 \quad (8)$$

Where, $K_{\max}$ represents the maximum frequency. In the frequency domain, the spacing between kernels is represented as f. For an image $I(a) = I(p, q)$, the Gabor representation normally referred as Gabor image, is defined as the convolution of image with the Gabor kernels.

$$G(a, b, c) = I(a) * \psi_{b,c}(a) \quad (9)$$

There are two Gabor parts for each image pixels, they are the imaginary and real part. The coefficient of Gabor filter $G(a, b, c)$ is a complex number which is rewritten as:

$$G(a, b, c) = M(a, b, c) \cdot \exp(p\theta(a, b, c)) \quad (10)$$

Here, the magnitude is represented as $M(a, b, c)$

and $\theta(a,b,c)$ represents the phase value. It is clear, that the local energy dissimilarities is included in the magnitude information. Most of feature extraction approach based on Gabor transform may concatenates the magnitude coefficients to define the augmented Gabor feature. The algorithm for Gabor scale feature extraction is shown below,

Table. 1 Gabor scale feature extraction algorithm

Input: an image $I(a)=I(p,q)$, Gabor kernel functions $\psi_{b,c}(a)$ for $\forall b,c$

Convolution of both Gabor kernel function and image $\psi_{b,c}(a), G(a,b,c)=I(a)*\psi_{b,c}(a)$,which is rewritten as:

$$G(a,b,c)=M(a,b,c)\cdot\exp(p\theta(a,b,c))$$

At the same scale, the orientation information are cumulated to attain the Gabor scale features.

$$Gaborscale(a,c)=\sum_{b}M(a,b,c)$$

This Gabor filter generates redundant features but with high dimension, the dimension gets increased due to the introduction of Gabor filter with multiple scales and orientation. So, down sampling is performed in this redundant features to minimize the dimensionality of Gabor features.

This Gabor scale feature minimize redundancy and also accurately attain the structural information. The dimensionality reduced features extracted by this Multi-scale Gabor filters are then provided to the DNN classifier for forgery detection. In this, DNN the training and testing phases are performed to clearly classify the forgery part from the entire video. DNN has a number of hidden layers, the weight parameter of this NN is further optimized within this hidden layers.

E.Deep Neural Network (DNN) with MSO

DNN

For classification the DNN technique is described in this section. This method is suitable for learning the discriminative and active features. For a large amount of unlabelled data the DNN can be applied. Since we aim at classification in which data acquisition is consider as a challenging task as it produces complications in pattern learning. In order to deal with this concern, with the sparse auto encoders an improved DNN is presented. To learn the feature pattern according to proposed approach, an auto-encoders that are adaptive is combined with the ideal de-noising method. Moreover to the neural network classifier these learned features are provided as input. The entire procedure of this proposed approach is explained in the resulting subsections. Initially, the designing of adaptive auto-encoders are discussed. Though, a substantial performance is shown by the neural networks in terms of classification on the other hand by using the neural network, it fails to tackle the multi-objective function and the data that having high-dimension [37].

Adaptive Auto-Encoder

An auto-encoder is symmetrical in nature and is a type of neural network. The auto-encoders with huge data is applied for feature learning. The reduced training and minimized loss function provides significant performance in auto-encoders. The auto encoder's main aim is to attain faultless input data restoration at the output layer. A fundamental issue is been faced by the conventional encoders i.e. the input layer has been copied by them to the hidden layer and this duplicated layer fails to provide substantial details about the input data. The conventional auto-encoders extension improves the systems performance.

Assume an input dataset S with $k\times l$ dimension and it is expressed as $Y=\{y(1),y(2),...,y(i),...y(N)\}\in R_l$. In this method, 'l' represents the dataset length and 'k' represent the sample number considered for estimation. In this adaptive auto-encoder, the corresponding data matrix is reflected as an input matrix. Selecting an appropriate activation function is identified as a prominent procedure in this DNN, as it may affect the overall performance [37].

In this method, sigmoid activation function is selected due to its performance enhancement ability. The next aim in this method is to learn the features and to understand the feature appearance. Finally, this justified features are provided as an input data Y and the expression for this is provided as,

$$h(y(m),w,b)=\sigma(wy(m)+b), m=1,2,....k) \quad (11)$$

Next, the output layer function is expressed by

$$outputlayer=\sigma(w^I h(y(m)+b), m=1,2,...,k) \quad (12)$$

Moreover, a term is introduced to alleviate the repeated input features through the involvement of an objective function. After that, this variable is provided to the hidden layer, afterthat the operating neurons are controlled during present cycle. Here, $a_n(m)$ represents the activation function for n^{th} hidden layer. The back propagation approach is considered to predict the software deficiency. The back propagation method propagates the learned weights from outer nodes to inner nodes. This propagation is performed for reduction of learning error which is obtained during the computation of network weights gradient. This method is applied for the input feature set Y, and the expression for the hidden layer can is given in Equation. (13).

$$a=sigmoid(wy+b) \quad (13)$$

Where, b indicates bias and 'w' represents the weight for input feature set. This hypothesis is used to give expression for the average weight of activation function as:

$$P_n=\frac{1}{k}\sum_{m=1}^{k}[a_n(m)] \quad (14)$$

During the initial stage of DNN training, we have to consider the average activation function value nearer to zero due to neuron idleness. For this, we are applying penalty for the average activation function. This penalty is applied only for the cases that diverges from the substantial average activation function value. The expression for penalty is given in Equation. (15)

$$P_{penalty}=\sum_{n=1}^{s_2}\rho // \rho_n \quad (15)$$

Where, s_2 indicates the total number of hidden layer

neurons, KL (.) represents the Kullback–Leibler divergence (KL divergence) and it is given as:

$$KL(\rho // \rho_n) = \rho_{\log} \frac{\rho}{\rho_n} + (1-\rho)\log \frac{1-\rho}{1-\rho_n} \quad (16)$$

$KL(\rho // \rho_n) = 0$ for $\rho_n = \rho$ otherwise it tolerates an increment that produce divergence, commonly referred as adaptive constant. The adaptive constant is obtained by applying the cost function and its expression is shown below,

$$C_{adaptive}(w,b) = C(w,b) + \beta \sum_{n=1}^{s2} KL(\rho // \rho_n) \quad (17)$$

β represents the penalty weight applied by KL divergence method. The identification of weight 'w' and biases 'b' becomes a prominent task due to its cost function, therefore, these two constraints are found to be directly relative to oneanother so that it can disturb the performance of entire system. It is solved by formulating an additional optimization problem and the output of this corresponding issue is consider to reduce the, $C_{adaptive}(w,b)$. This emerged optimization issue is solved by introducing the novel MSO algorithm. This MSO performs the optimization process to iteratively update the bias and weight. This can be written as:

$$P_n = {2(k+1-n)}/{k(k+1)}, n = 1,2,.....,k \quad (18)$$

And

$$b_m(i) = b_m(i) - \varepsilon \frac{\partial}{\partial b_m(i)} C_{adaptive}(w,b) \quad (19)$$

In this ε, represents the learning rate of DNN.

Moth search optimization (MSO) algorithm

Moth is a type of insect, normally belongs to the butterfly family of order Lepidoptera. In this entire world, total of about 160,000 moth species are found, among them most of them are nocturnal. While comparing with the other moth characteristics, the phototaxis and Levy flights are identified as the most typical features which are discussed in following subsections [38]. The weight parameter of neural network is provided as an input for this MSO algorithm. This algorithm identifies the optimized weights by performing the searching operation.

A. Phototaxis

The process in which the moths fly encircling the source light is referred as phototaxis. But still, the accurate procedure of phototaxis is unknown, so various hypotheses are developed to explain this phototaxis process. Among them, one of the most important hypotheses is celestial navigation as it is applied in transverse orientation during flying. In order to maintain a fixed angle towards the celestial light (i.e., the moon), the moths will travel in straight line. Frequently, the angle that exists between both light source and moth may get changed, but we can't able observe that change, because the celestial object is found in a far-away distance. To move towards the light source, the moth will naturally adjust its flight orientation to the best, thereby it allows the airborne moths to fall downwards. They form a spiral-path to move much closer towards the source light.

B. Levy flights

Heavy-tailed, non-Gaussian statistics is identified as the commonly employed techniques in various applications like the behavior of enormous insects and animals. Levy flight is a type of random movement, so in natural environments it is considered as one of the major flight pattern. Other than this moth fly, the Drosophila also exhibit this Levy flight, whereas this flight can be approximated over a scales range as a power law distribution having the feature exponent nearer to 3/2. Normally, this Levy distribution is expressed in the form of power-law and it is described in subsequent equation,

$$L(s) \sim |s|^{-\beta} \quad (20)$$

Where, $1 < \beta \leq 3$ is identified as an index.

The moths those having the distance as close to the best one, will fly in the Levy flight manneraround the fittest one. Or else they will update their positions by executing the Levy flight with respect to equation. (21), whereas the moth j, is updated by using subsequent equation:

$$X_j^{a+1} = X_j^a + \delta L(s) \quad (21)$$

Where, X_j^a and X_j^{a+1} represents the original and updated position at generation a, whereas the current generation is represented as, a. The step that obtained from the Levy flight is represented as, $L(s)$. For the problem of interest, the scale factor is represented by the parameter, δ. The expression for this δ is shown in below equation:

$$\delta = W_{\max} / a^2 \quad (22)$$

Here, $W_{\max}$ represents maximum walk step and the value for this $W_{\max}$ is set based on the available problem.

Levy distribution $L(s)$ in the above equation can be rewritten as

$$L(s) = \frac{(\alpha-1)\Gamma(\alpha-1)\sin\left(\frac{\pi(\alpha-1)}{2}\right)}{\pi s^{\alpha}} \quad (23)$$

In this, s is found higher than zero. $\Gamma(x)$, represents the gamma function. As mentioned earlier, from the Levy distribution having $\alpha = 1.5$ the moths Levy flight can be drawn.

C. Fly straight

The moths present apart from the source light will flutterin straight line towards the light source. The flight expression for moth j is formulated as

$$X_j^{a+1} = \lambda \times \left(X_j^a + \varphi\left(X_{best}^a - X_j^a\right)\right) \quad (24)$$

Where X_{best}^a represents the fittest moth at a generation, λ, represents the scale factor. φ, is an acceleration factor.

Or else the moth will fly beyond the light source that is towards the final location. In this case, the formula applied to calculate the final position of moth j is given in equation. (25),

$$X_j^{a+1} = \lambda \times \left(X_j^a + \frac{1}{\varphi}\left(X_{best}^a - X_j^a\right)\right) \quad (25)$$

For simplicity, the position of moth j is updated by applying the equations. (24 & 25), with half percentage possibility. The original, updated, and best position of moth is represented as X_j, $X_{j,new}$ and, X_{best}. λ, control the algorithm's convergence speed and also enhance the population diversity.The optimized weight furthermore improves the performance of the DNN classifier. Due to this, an optimized results is obtained by this method. The DNN is applied for classification purposes. In this method, the real and forged videos are identified and classified by this DNN. The DNN contains a number of hidden layers, so number of iterations are performed within this classifier. Finally, an accurate classification is provided by this DNN classifier. An application of MSO algorithm along with DNN, furthermore enhances the classification performance.The algorithm for entire process is shown in below table

Table 2. Algorithm for entire process

Input: the MPEG-4 videos is given as an input.
Output: Forged frame
Step 1: Input the video
Step 2: Double compression artifacts are detected by adopting Markov statistics
Step 3: Segment double compressed frames by SLIC super-pixel segmentation.
Step 4: Extract features from segmented images by multi-scale gabor filter
Step 5: Based on these extracted features, the forged frames are identified by hybrid DNN-MSO.
Step 6: Update weight (W) parameter of DNN with MSO fitness function
Step 7: DNN-MSO extracted the forged frame from whole video.
Step 8: Evaluate the performance by means of F-measure, classification accuracy, precision, recall.

IV. EXPERIMENTAL ANALYSIS

In this part, initially the details that are required for implementation process is discussed, after that performance metrics are evaluated and then this metrics are compared with some prevailing techniques. The parametric evaluations are taken in terms of Recall, Precision, and accuracy. For this experiment process, totally of 4000 frames are collected from 20 different videos, among that 2100 are forged, whereas 1900 are authentic. Each frames resolution is found to be 640× 480, similarly the frame rate for this video is found to be 25fps (frames per second). Frame duplication is the process in which the single frame from the entire video is selected and duplicated, after that this selected frame is then placed in another location in the similar video sequence.

An example for this video forgery process is shown in Figure. (4)[43]. In this, the frames of real video is placed in top row, whereas the frames of forged video is place in bottom row. From this, it is clearly seen, that in the forged video, the frames 1 & 2 are again duplicated and place again instead of frames 4 & 5.

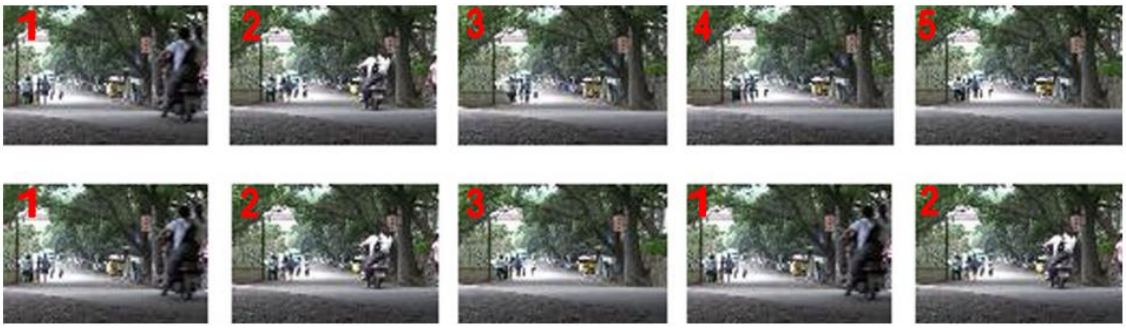

Figure. 4 Original video and Forged video (1 & 2 are again placed instead of 4 & 5 frames)

A. Evaluation standards

The DNN output is mapped to identify the forged and actual video. Here the accuracy value for our proposed approach is determined by using following equation. (26),

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \times 100\% \quad (26)$$

Here,the number of forged videos that are normally designated as forged is represented as TP, FP indicates that the number of real videos that are incorrectly recognised as forged, the number of real videos that are correctly identified as real is represented as TN, and the number of videos that are mistakenly identified as authentic is indicated as FN [42].

Table. 3 Performance evaluation for proposed method

	Positive (%)	Negative (%)
True	96.05	92.85
False	6.57	5.95

$$P = \frac{TP}{TP + FP} \quad (27)$$

$$R = \frac{TP}{TP + FN} \quad (28)$$

TPR (TP rate) and FPR (FP rate) are normally applied for parameter tuning. The FPR needs to be minimum, whereas the TPR value is to be high. The equation for F1-score is shown in equation. (29), which is obtained from both recall, and precision.

$$F1 - score = \frac{2 \times P \times R}{P + R} \quad (29)$$

The accuracy, P, R, and $F1-score$ values that obtained by evaluating our proposed method is shown in below table. This method is also compared with two prevailing techniques. The evaluation outputs of this prevailing techniques is also provided in table. 4. The graphical representation for this evaluation metrics are shown in below figures.

Table. 4 Comparison of performance metrics of proposed method with two prevailing methods

Methods	Accuracy	P	R	F1-score
Proposed	94.375	93.58	94.805	94.19
Su [40]	89.6	92.2	90.5	91.34
Zhang	83.7	89.9	77	82.95
[41]				

The evaluation metrics like accuracy, P, R, and $F1-score$ are evaluated to analyse the effectiveness of our proposed forgery detection technique. This forgery detection process achieves high accuracy value than other prevailing forgery detection techniques (Su et al [40] & Zhang et al, [41]). The accuracy outcome of proposed and other prevailing techniques are shown in Figure. (5). The accuracy of this proposed forgery detection process is found to be 94.37, which is higher than Su et al, (89.6) and Zhang et al, (83.7).

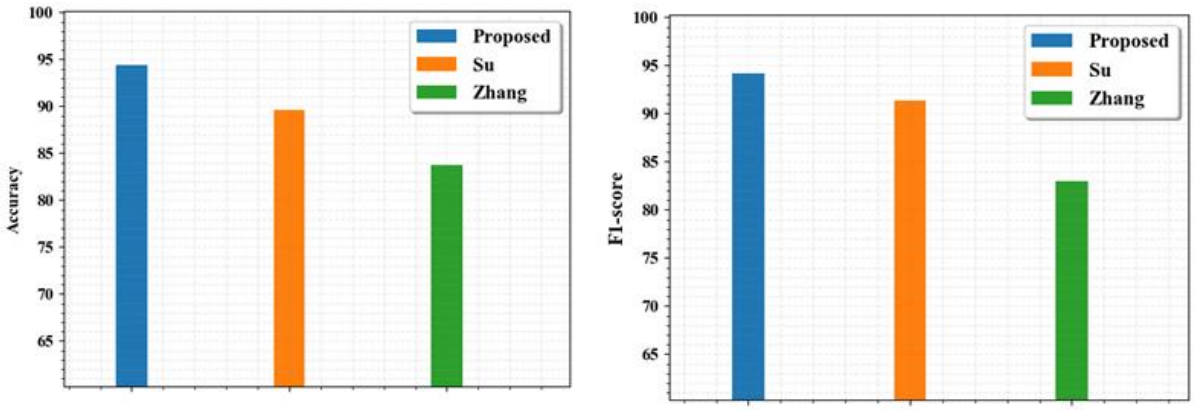

Figure. 5 F1-score and Accuracy results of proposed and prevailing forgery detection techniques

The, $F1-score$ is obtained from both recall and precision values. The equation to evaluate the accuracy, P, R, and $F1-score$ are defined in Equations. (26, 27, 28, & 29). The graphical representation for P and R of proposed and two prevailing forgery detection methods is depicted in Figure. (6).In the existing techniques, the neural network is not included for forgery detection, but in this proposed work the DNN along with MSO is included for detecting the forgery. The major advantage of this neural network is that it enhances the detection accuracy.

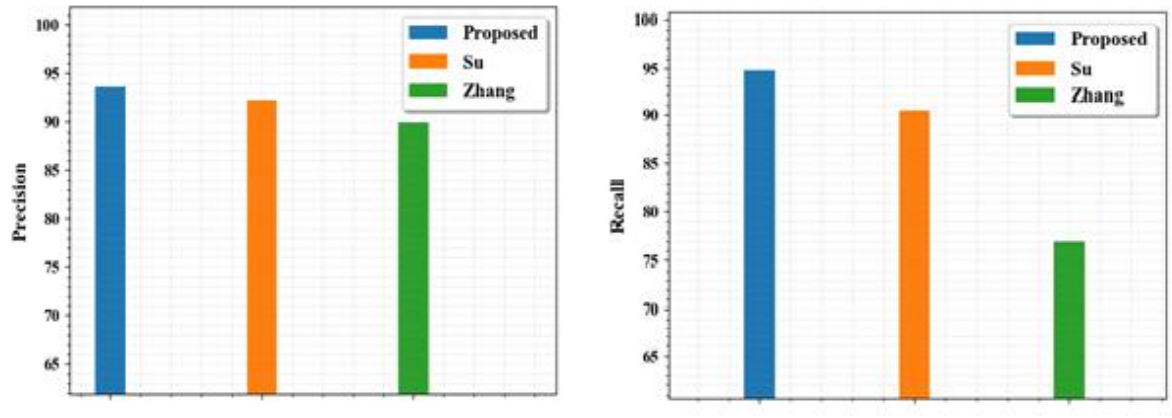

Figure. 6 Graphical representation for precision and recall of prevailing and proposed forgery detection methods

This obtained result shows the performance of this approach. It is compared with two existing [40, 41] techniques to prove its effectiveness. The precision (93.58) and recall (94.8) value is attained by this proposed method. In 40 & 41, the forgery detection process is performed without including the machine learning process. But here, DNN based on MSO is included which improves the system performance than the other prevailing techniques. Various methods depicts better performance in video forgery detection, but here, the DNN introduction along optimization algorithm enhances this forgery detection process. The comparison results provided above indicates that the proposed

method attains high accuracy, P, R, and F1-score rate than the other forgery detection algorithms.

V. CONCLUSION

Video forgery detection occupies a huge gain in digital world. Number of detection techniques are developed for this detection purpose. But they failed to detect the forgery more accurately. Various approaches include neural networks for forgery detection, however they failed to bring the optimal result. By taking all these into consideration, we include DNN based MSO for forgery detection, which accurately and optimally identify the forged video. Here, the DNN based forgery detection process is experimented by including various videos that are collected from internet or YouTube. This method shows very high accuracy than other existing techniques. The main advantage of this that, here the MSO algorithm is included to further enhance the performance of DNN classifier. The experimental analysis indicates the effectiveness of this proposed approach and it is compared with two prevailing techniques. The evaluation metrics like accuracy, precision, recall and F1-score are evaluated to analyse the effectiveness of our proposed forgery detection technique. The results of this evaluation metrics depicts that this DNN-MSO based forgery detection process provide robust and more effective forgery detection results than the other algorithms.

REFERENCES

1. P. Johnstonand E. Elyan, "A review of digital video tampering: from simple editing to full synthesis." *Digital Investigation* (2019).
2. Y. Yao, Y. Shi, S. Weng and B. Guan. "Deep learning for detection of object-based forgery in advanced video." *Symmetry* 10, no. 1, 2017, pp. 3.
3. P. Johnston, E. Elyan and C. Jayne, "Video tampering localisation using features learned from authentic content." *Neural Computing and Applications*, 2019,pp.1-15.
4. L. Bondi, S. Lameri, D. Güera, P. Bestagini, E. Delp and S. Tubaro, "Tampering detection and localization through clustering of camera-based CNN features." In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2017, pp. 1855-1864. IEEE.
5. V. Joshi and S. Jain, "Tampering detection in digital video-a review of temporal fingerprints based techniques." In *2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom)*, 2015, pp. 1121-1124. IEEE.
6. T.M. Mohammed, J. Bunk, L. Nataraj, J.H. Bappy, A.Flenner, B. S. Manjunath, S.Chandrasekaran, A.K. Roy-Chowdhury and L.A. Peterson. "Boosting Image Forgery Detection using Resampling Features and Copy-move Analysis." *Electronic Imaging* 2018, no. 7 (2018),pp.1-7.
7. L. Chen, X.Peng, J.Tian and J. Liu,"A learning-based approach for leaf detection in traffic surveillance video". *Multidimensional Systems and Signal Processing*, 2018, 29(4), pp.1895-1904.
8. L.Bondi, L.Baroffio, D.Güera, P.Bestagini, E.J.Delp and S.Tubaro,"First steps toward camera model identification with convolutional neural networks". *IEEE Signal Processing Letters*, 2016, 24(3), pp.259-263.
9. A.Tuama, F.Comby and M. Chaumont,"Camera model identification with the use of deep convolutional neural networks". In 2016 *IEEE International workshop on information forensics and security (WIFS)*2016, December, pp. 1-6. IEEE.
10. G.Xu, H.Z. Wu and Y.Q. Shi,"Structural design of convolutional neural networks for steganalysis". *IEEE Signal Processing Letters*, 2016, 23(5), pp.708-712.
11. P. Yang, R. Ni and Y.Zhao, "Recapture image forensics based on Laplacian convolutional neural networks". In International Workshop on *Digital Watermarking*,2016September, pp. 119-128. Springer, Cham.
12. B. Bayar and M.C.Stamm, "A deep learning approach to universal image manipulation detection using a new convolutional layer". In Proceedings of the 4th ACM Workshop on *Information Hiding and Multimedia Security*, 2016, June,pp. 5-10, ACM.
13. Y. Rao and J. Ni,"A deep learning approach to detection of splicing and copy-move forgeries in images". In 2016 IEEE International Workshop on *Information Forensics and Security* (WIFS) (2016, December, pp. 1-6). IEEE.
14. S. Suwajanakorn, S. M. Seitz, I. Kemelmacher-Shlizerman, "Synthesizing obama: learning lip sync from audio", *ACM Transactions on Graphics* (TOG) 2017, 36 (4), 95.
15. M.-Y. Liu, T. Breuel, J. Kautz, "Unsupervised image-to-image translation networks", in: *Advances in Neural Information Processing Systems*, 2017, pp. 700–708.
16. H. Dong, P. Neekhara, C. Wu, Y. Guo, "Unsupervised image-toimage translation with generative adversarial networks", arXiv preprint arXiv:1701.02676.
17. J. Thies, M. Zollh¨ofer, M. Stamminger, C. Theobalt, M. Nießner, "Face2face: Real-time face capture and reenactment of rgb videos", in: *Computer Vision and Pattern Recognition* (CVPR), 2016 IEEE Conference on, IEEE, 2016, pp. 2387–2395.
18. A. R¨ossler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, M. Nießner, "Faceforensics: A large-scale video dataset for forgery detection in human faces", arXiv preprint arXiv:1803.09179.
19. T. Karras, T. Aila, S. Laine, A. Herva, J. Lehtinen, "Audio-driven facial animation by joint end-to-end learning of pose and emotion", *ACM Trans actions on Graphics (TOG)*, 2017, 36 (4), 94.
20. L. Chen, Z. Li, R. K Maddox, Z. Duan, C. Xu, "Lip movements generation at a glance", in: *The European Conference on Computer Vision (ECCV)*, 2018.
21. L.Yu, et al.,"Exposing frame deletion by detecting abrupt changes in video streams". *Neurocomputing*,2016, 205, pp.84–91
22. T.Shanableh, "Detection of frame deletion for digital video forensics". *Digit. Investig*, 2013, 10(4), pp.350–360.
23. J.A.Aghamaleki, A.Behrad, "Inter-frame video forgery detection and localization using intrinsic effects of double compression on quantization errors of video coding",*Sig. Process.: Image Commun*, 2016, 47, pp.289–302.
24. Y. Liu, T.Huang, "Exposing video inter-frame forgery by Zernike opponent chromaticity moments and coarseness analysis",*Multimed. Syst.* 2017, 23(2), pp.223–238.
25. C. Long, E. Smith, A.Basharat, A.Hoogs, "A C3D-based convolutional neural network for frame dropping detection in a single video shot". In: *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, July 2017, pp. 1898–1906.
26. Y. Su, W.Nie, C.Zhang, "A frame tampering detection algorithm for MPEG videos". In: *6th IEEE Joint International Information Technology and Artificial Intelligence Conference*, 2011, vol. 2, pp. 461–464. https://doi.org/10.1109/ ITAIC.2011.6030373
27. A.Khodabakhsh, R.Ramachandra, K. Raja, P.Wasnik, C. Busch,"Fake face detection methods: can they be generalized?" In: 2018 *international conference of the biometrics special interest group (BIOSIG)*. IEEE,2018, pp. 1–6
28. R.D. Singh, N. Aggarwal,"Video content authentication techniques: a comprehensive survey". *Multimedia Syst*, 2017, 24, pp.1–30
29. S. Agarwal, W. Fan, H.Farid,"A diverse large-scale dataset for evaluating rebroadcast attacks". In: *IEEE international conference on acoustics, speech, and signal processing*.2018.
30. M. Huh, A. Liu, A. Owens, A. A.Efros,"Fighting fake news: image splice detection via learned self-consistency",In: *The European conference on computer vision (ECCV)*,2018.
31. D.D'Avino, D.Cozzolino, G.Poggi, L.Verdoliva, "Autoencoder with recurrent neural networks for video forgery detection". *Electron. Imaging*, 2017, (7), pp.92–99.
32. S.Bosse, D.Maniry, T.Wiegand, W.Samek,"A deep neural network for image quality assessment". In: *IEEE international conference on image processing (ICIP)*. IEEE,2016, pp. 3773–3777
33. G. Schaefer, M. Stich,"UCID: an uncompressed color image database". In: *Storage and retrieval methods and applications for multimedia* 2004, vol 5307. International Society for Optics and Photonics, 2003, pp. 472–481
34. X. Jiang, W. Wang, T. Sun, Y.Q. Shi and S. Wang,"Detection of double compression in MPEG-4 videos based on Markov statistics". *IEEE Signal processing letters*, 2013, *20*(5), pp.447-450.
35. D.Derksen, J.Inglada and J. Michel,"Scaling Up SLIC Superpixels Using a Tile-Based Approach". *IEEE Transactions on Geoscience and Remote Sensing*, 2019, 57(5), pp.3073-3085.
36. B.Q. Zhang, Z.C. Mu, H. Zeng and H.B. Huang,"Ear recognition based on Gabor

scale information". In *2013 International Conference on Wavelet Analysis and Pattern Recognition*, 2013, July,pp. 153-157. IEEE.
37. C.Manjula and L. Florence,"Deep neural network based hybrid approach for software defect prediction using software metrics". *Cluster Computing*, 2018, pp.1-17.
38. G.G. Wang,"Moth search algorithm: a bio-inspired metaheuristic algorithm for global optimization problems". *Memetic Computing*, 2018, 10(2), pp.151-164.
39. T.Pevny, P. Bas and J.Fridrich,"Steganalysis by subtractive pixel adjacency matrix". *IEEE Transactions on information Forensics and Security*, 2010, 5(2), pp.215-224.
40. L. Su, T. Huang and J.Yang, "A video forgery detection algorithm based on compressive sensing". *Multimedia Tools and Applications*, 2015, 74(17), pp.6641-6656.
41. J. Zhang, Y. Su, M. Zhang,"Exposing digital video forgery by ghost shadow artefact". In: *Proceedings of the First ACM workshop on Multimedia in forensics*. ACM, 2009, pp. 49–54
42. Li, Q., Wang, R. and Xu, D., 2018. An Inter-Frame Forgery Detection Algorithm for Surveillance Video. Information, 9(12), p.301.
43. Yang, J., Huang, T. and Su, L., 2016. Using similarity analysis to detect frame duplication forgery in videos. Multimedia Tools and Applications, 75(4), pp.1793-1811.

AUTHORS PROFILE

MalleRaveendra born in 1983 in a remote village in Andhra Pradesh, INDIA and completed B.Tech from S.V University in the year 2006 and obtained M.Tech from Bharath University in the year 2009. Presently he is Pursuing Ph.D. in Jawaharlal Nehru Technological University, Anantapuramu, Andhra Pradesh, INDIA. His areas of interest include Image and video processing.

Dr.K.Nagi Reddy born in 1974 in a remote village in Andhra Pradesh, INDIA and completed AMIETE in the year 1996, obtained M.Tech from JNT University in the year 2001 and obtained a Ph.D. from S.VUniversity. Presently he is working as Professor in NBKR. Institute of Science & Technology, Vidyanagar, Nellore (dt), Andhra Pradesh, INDIA. He is a life member of ISTE, IETE. His areas of interest include Image and Video Processing.