I. J. Computer Network and Information Security, 2013, 9, 58-67
Published Online July 2013 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijcnis.2013.09.08

Video Forensics in Temporal Domain using Machine Learning Techniques

Sunil Jaiswal, Sunita Dhavale
Defence Institute of Advanced Technology, Pune- 411025, India
sunj04@gmail.com, sunita.dhavale@diat.ac.in

Abstract **—** In defence and military scenarios, Unmanned Aerial Vehicle (UAV) is used for surveillance missions. UAV's transmit live video to the base station. Temporal attacks may be carried out by the intruder during video transmission. These temporal attacks can be used to add/delete objects, individuals, etc. in the live transmission feed. This can cause the video information to misrepresent facts of the UAV transmission. Hence, it is needed to identify the fake video from the real ones. Compression techniques like MPEG, H.263, etc. are popularly used to compress videos. Attacker can either add/delete frames from videos to introduce/remove objects, individuals etc. from video. In order to perform attack on the video, the attacker has to uncompress the video and perform addition/deletion of frames. Once the attack is done, the attacker needs to recompress the frames to a video. Wang and Farid et. al. [1] proposed a method based on double compression technique to detect temporal fingerprints left in the video caused due to frame addition/deletion. Based on double MPEG compression, here we propose a video forensic technique using machine learning techniques to detect video forgery. In order to generate a unique feature vector to identify forged video, we analysed the effect of attacks on Prediction Error Sequence (PES) in various domains like Discrete Cosine Transform (DCT), Discrete Fourier Transform (DFT), Discrete Wavelet Transform (DWT) domain etc. A new PES feature γ is defined and extracted from DWT domain, which is proven robust training parameter for both Support Vector Machine (SVM) and ensemble based classifier. The trained SVM was tested for unknown videos to find video forgery. Experimental results show that our proposed video forensic is robust and efficient in detecting video forgery without any human intervention. Further the proposed system is simpler in design and implementation and also scalable for testing large number of videos.

Index Terms **—** Digital forensics, Temporal forensics, Discrete Cosine Transform, Discrete Fourier Transform, Discrete Wavelet Transform, Support Vector Machine, Ensemble based classifier

I. INTRODUCTION

Recent trends have seen an extensive use of video for communication, education, medical, entertainment, security, surveillance, defence makes it a rich source for exploitation in the cyber domain. With the advent in video technology, numerous tools are available to easily edit/tamper video even by a novice user. Video sequences can be maliciously tampered to misrepresent audio & visual information being conveyed. Numerous video editing tools are available in the market that allows easy editing of video information. Malicious individuals can alter video by deleting/adding frames in a video. Frames can be deleted to remove objects from a video such as, individuals, weapon etc. Similarly, frames can be added to introduce objects in the video. Deletion/addition of frames is done to misrepresent the video information or for the purpose of covert communication i.e. video steganography. Frames can be deleted from the video to remove objects/persons from CCTV footage to avoid charges in the court of law. Especially in case of defence scenario, where UAV's transmit live video to the base station, these temporal attacks may cause to misrepresent critical UAV data. Detecting the traces of video tampering is a challenging and cumbersome task for video forensic analyst while analyzing large number of frames in videos.

Digital forensic is a significant tool that can be used to find digital evidences of alteration in the video. Video forensics deals with the acquisition and analysis of traces left due to the source device or the compression history. Very few works has been reported on video forensics based on the compression history. Wang et. al. [1] proposed a method based on double compression technique to detect temporal fingerprints left in the video caused due to frame addition/deletion. Stamm, Lin and Liu [2] extended this work for temporal forensics and anti-forensics techniques along with a game theoretic framework. The technique proposed by Wang relies on visual detection of temporal fingerprints in *DFT* domain. Detecting the peak location by visual inspection is a tedious task in case of analyzing large number of videos and also prone to human error.

Section II briefly explains the *MPEG* compression, Section III describes the temporal forensic of video, Section IV analyses the feature extraction for machine learning, Section V describes machine learning techniques like *SVM* [3-4] and Ensemble based classifier. Section VI proposes a novel technique to classify forged videos. Section VII analyses the experimental setup and results of the proposed technique. Section VIII concludes the paper followed by Appendix in last section.

II. *MPEG* COMPRESSION

MPEG [5-7] is a widely used and popular video compression standard. It is extensively used for video streaming, digital video broadcasting, security and surveillance related applications. In a *MPEG* encoded video sequence, any video is considered as a sequence of images. Group of Pictures (*GOP*) is then defined as a set of video frames that follow a predefined frame pattern. Frame pattern consists of Intra frames (*I*), Bi-directional frame (*B*) and Prediction frame (*P*). Frame pattern can be of the form *IBPIBP*, *IPPPIPP*, *IBBPBBPBB* etc. It is applied to the whole video sequence. I-frames are used as reference/anchor frame to predict *P/B*-frames in a *GOP*. *P*-frames are predicted from preceding I-frame/P-frame. *B*-frames are predicted using preceding and following I/P- frames within a *GOP*. *B*-frames can be optional in a video sequence. Fig. 1 demonstrates an inter-frame coding of P-frames and B-frames within a *GOP*.

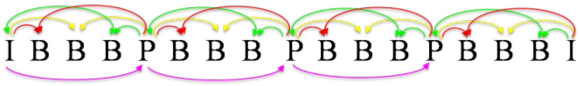

Figure 1: Inter-frame coding in a GOP

Spatial redundancy [5],[6],[7] is eliminated in the similar manner as for a *JPEG* standard using Discrete Cosine Transform (*DCT*) and entropy encoding techniques. Motion compensation techniques [5],[6],[7] are applied to eliminate temporal redundancies caused due to the motion of objects between frames. Motion estimation computes motion vectors between two frames, to find displacement of objects moving between the frames. Motion compensation uses these motion vectors to predict future *P*-frame/*B*-frame. The difference between the anchor frame and predicted P-frame forms *PES* for the video. If *L* is the total number of prediction error frames in the video and e_i be the prediction error of the i^{th} *P*-frame then *PES* is given as $P_{es}(n)=[e(1), e(2),...e(L)]$, where $e(L)$[2].

Prediction error is computed between I-frame / P-frame and the predicted frame. I-frames, motion vectors and *PES* are transmitted as a bit stream to the decoder, instead of the whole video frames. Various video compression standards such as *MPEG-2*, *MPEG-4*, *H.263*, *H.264* etc. use these motion compensation techniques.

III. TEMPORAL VIDEO FORENSIC

A. *Temporal Attack on Video*

In order to delete frames from a video, it has to be uncompressed so as to access frames in the video. The frames are then deleted from the uncompressed sequence of images. After deleting the frames the sequence of images are re-compressed to form the MPEG compressed video. Fig. 2 shows the flowchart for frame addition/deletion for an MPEG video. Fig. 3 shows a sample "Carphone" video used in the conducted experiments.

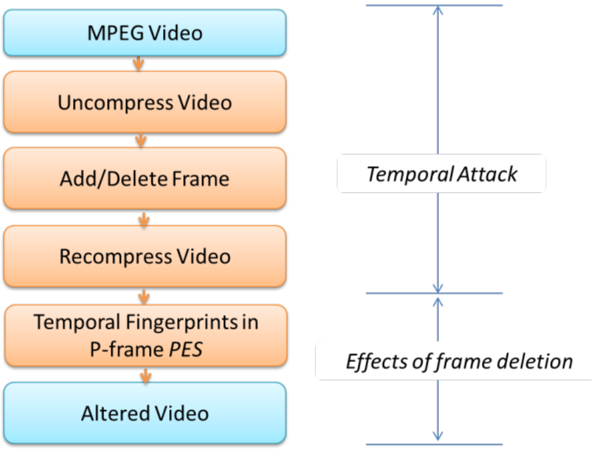

Figure 2: Frame addition/deletion

Figure 3: "Carphone" video sequence

B. *Prediction Error Sequence (PES)*

Prediction frame is computed in the MPEG encoder during motion estimation. The prediction frame is the difference of the reference frame (I/P- frame) and the predicted P-frame. The prediction frame is averaged to obtain the prediction error. The set of prediction errors computed in a video sequence forms the Prediction Error Sequence (*PES*)[2].This PES is extracted from the temporal properties of the motion compensated video. Fig. 4 shows the *PES* of the original "Carphone" video sequence and Fig. 5 shows the *PES* of the 6-frames deleted "Carphone" video sequence

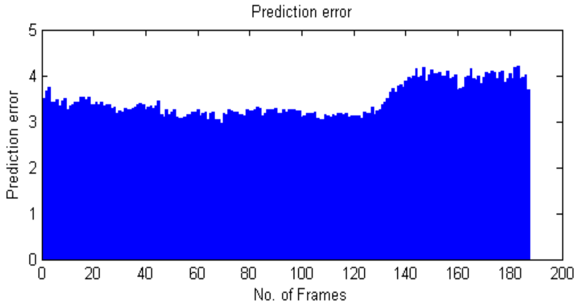

Figure 4: *PES* of the original "Carphone" video

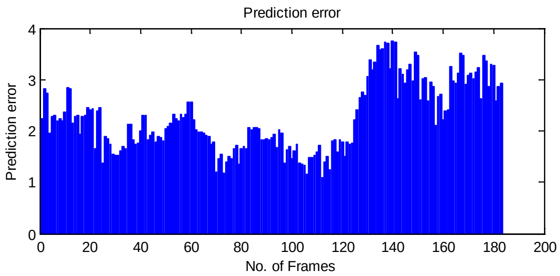

Figure 5: *PES* of the 6–frames deleted “Carphone” video

IV. *PES* Feature Extraction

Transformations like *DCT, DFT, DWT* etc. are proven as a suitable candidate for pattern recognition and classification problems [8]. The machine learning approach is proposed to classify the tampered video from the original videos. Machine learning [8] techniques require feature vectors for classification of data. We propose to use various transformations such as Discrete Cosine Fourier Transform (*DFT*) [1],[2],[8], Discrete Cosine Transform (*DCT*) [8] and Discrete Wavelet Transform (*DWT*) [8-9] to generate statistical feature . As compared to the original input feature vectors, transform domain features show high information packing properties [8]. The basic approach followed is to transform a given set of measurements to a new set of statistical feature. This reduces the necessary feature space dimension. This task is referred as dimensionality reduction techniques. Using transform-based features will help remove information redundancies from the dataset. The length of *PES* depends on the number of frames in a video. For video forensic analysis, the number of frames in the video under test may vary from the original video. This affects the length of *PES* in a video. Hence, analysis in frequency domain is better strategy compared to the time domain analysis. The transform coefficients are a suitable candidate to obtain feature vectors. This section explains the use of *DCT, DFT, DWT* domain coefficients to generate unique feature for temporal video forensic.

A. Discrete Fourier Transform (DFT)

Discrete Fourier Transform (*DFT*) [1],[2],[8] is an important tool in numerous applications of digital signal processing, image processing. Wang and Farid [1] illustrated that *DFT* can be used to detect video tampering in the video compression history. *DFT* for *N* input samples *x(0), x(1), . . . , x(N -1)* is defined as,

$$y(k) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x(n) \quad exp\left(-j\frac{2\pi}{N}kn\right).$$

$$, k = 0, 1, 2, \dots N-1 \qquad (1)$$

Fig. 6 shows the *DFT* plot of the *PES* for the original “*Carphone*” video sequence with 250 frames following *IPPP* frame pattern. Fig. 7 shows the *DFT* plot of the *PES* for the 6-frames deleted at the starting position in the “*Carphone*” video sequence. The *DFT* of the frame deleted video shows spikes in the *DFT* plot. These spikes are the fingerprint of the video frame deletion and caused due to the process of *double MPEG compression* [1]. The technique proposed by Wang relies on visual detection of temporal fingerprints in *DFT* domain. Detecting the peak location by visual inspection is a tedious task in case of analyzing large number of videos and also prone to human error.

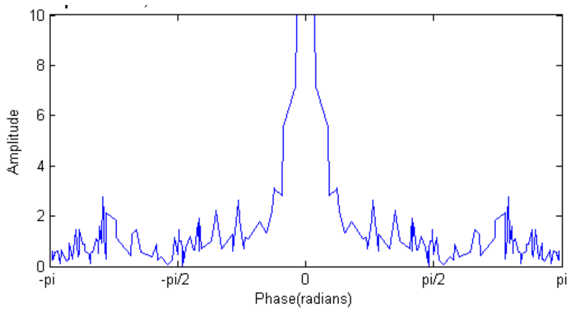

Figure 6: (|*DFT{PES}*| of the original “*Carphone*” video

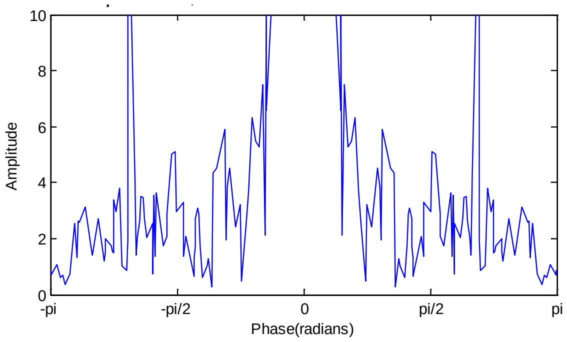

Figure 7: (|*DFT{PES}*| of the6–frames deleted “*Carphone*” video

B. Discrete Cosine Transform (DCT)

DCT is used in data compression, frequency spectrum analysis and information hiding. *DCT* has better energy compaction properties i.e few transform coefficients can represent the majority of the energy in a sequence. Mathematically a *DCT* of an input sequence x is given as ,

$$y(k) = \alpha(k) + \sum_{n=1}^{N} x(n) \cos\frac{\pi(2n-1)(k-1)}{2N}$$

$$k = 1, 2, \dots N \qquad (2)$$

where, $\alpha(k) = \begin{cases} 1/\sqrt{N} & k=1 \\ \sqrt{2/N} & 2 \leq k \leq N \end{cases}$

DCT have very good information-packing properties for images. i.e. they concentrate most of the energy in a few coefficients. The DCT transform of an image shows that the high-intensity coefficients are concentrated in a small area. This size changes with various transforms as it depends on the property of energy compaction properties of the transform. Fig. 8 shows the *DCT* plot of the *PES* for the original "*Carphone*" video sequence with 250 frames following *IPPP* frame pattern. Fig. 9 shows the *DCT* plot of the *PES* for the 6-frames deleted at the starting position in the "*Carphone*" video sequence. No prominent difference found in DCT domain compare to DFT domain.

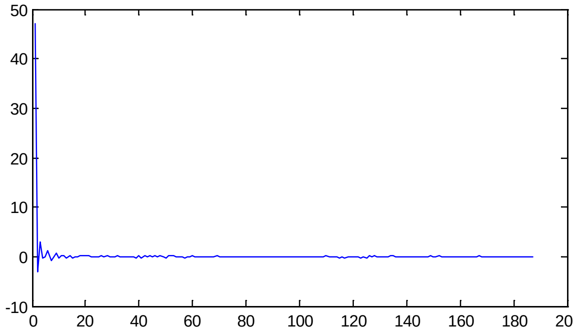

Figure 8: DCT{*PES*} of the Original "Carphone" video

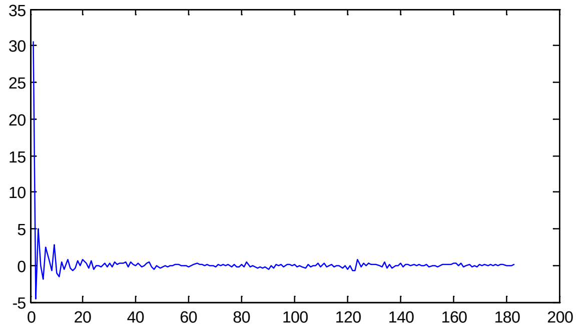

Figure 9: DCT{*PES*} of the 6–frames deleted "Carphone" video

C. Discrete Wavelet Transform

Discrete Wavelet Transform (DWT) [8],[9] represents a signal in time-frequency form.DWT is based on sub-band coding. Digital filtering techniques are utilised to represent a digital signal in time-scale form. The signal to be analysed is passed through filters with different cutoff frequencies at different scales. In *DWT*, the most prominent information in the signal appears in high amplitudes and the low amplitude signals are less prominent information.

To carry out the forensic analysis of the video in question, we need to find or extract unique feature describing overall characteristic of *PES* vector. *DWT* is used to generate feature vector for a classifier. Hence, first level *DWT* is applied on PES to obtain coarse and detail sub-bands. *DWT* applies low pass filter to form low frequency component, approximate coefficient (*cA*) and high pass filter to form high frequency component, detail coefficient (*cD*). The length of *cA* is *L/2* and that of *cD* is *L/2*. Then the difference vector signal *Diff(n)* = *cA(n)-cD(n)* , where $1 \leq n \leq L/2$ .It is calculated in order to extract unique feature γ related to original or tampered video. Fig. 10 shows the steps for γ extraction.

$$\gamma = \frac{1}{(L/2)}\sum_{i=1}^{L/2}\left|Diff(i)\right| \qquad (3)$$

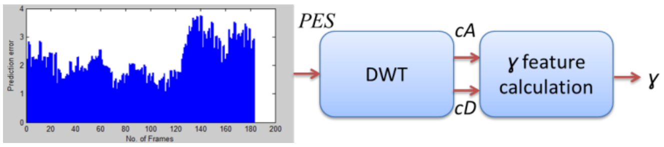

Figure 10: γ feature extraction using *PES*

Fig. 11 shows the *cA* coefficients of the *DWT {PES}* for the original "*Carphone*" video sequence with 250 frames following *IPPP* frame pattern. Fig. 12 shows the *cA* coefficients of the *DWT{PES}* for the 6-frames deleted at the starting position in the "*Carphone*" video sequence. Fig. 13 shows the *cD* coefficients of the *DWT {PES}* for the original "*Carphone*" video sequence with 250 frames following IPPP frame pattern. Fig. 14 shows the *cD* coefficients of the *DWT {PES}* for the 6-frames deleted at the starting position in the "*Carphone*" video sequence.

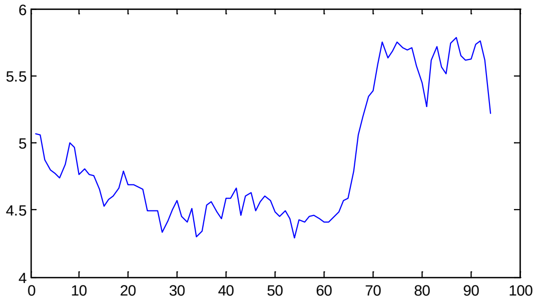

Figure 11: cA of the (DWT {PES} of the original "Carphone" video

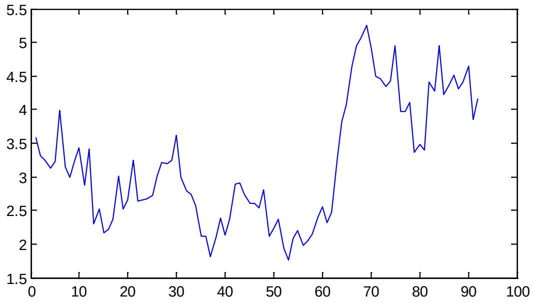

Figure 12: cA of the (DWT{PES} of the 6–frames deleted "Carphone" video

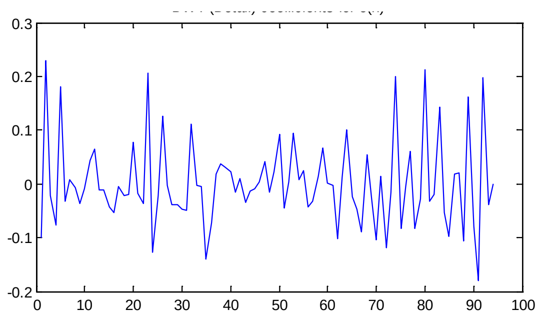

Figure 13: cD of the (DWT{PES} of the original "Carphone" video

Copyright © 2013 MECS

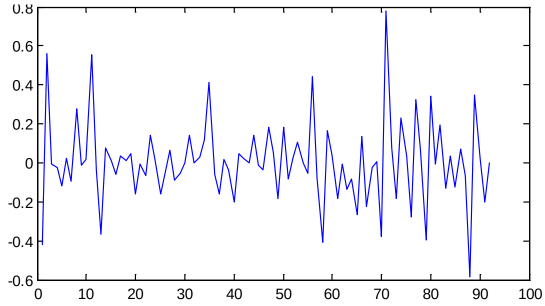

Figure 14: cD of the (DWT{PES} of the 6–frames deleted "Carphone" video

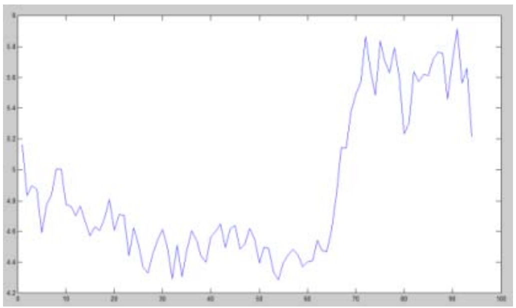

Figure 15: |Diff(n)| values for original "Carphone" video with 250 frames

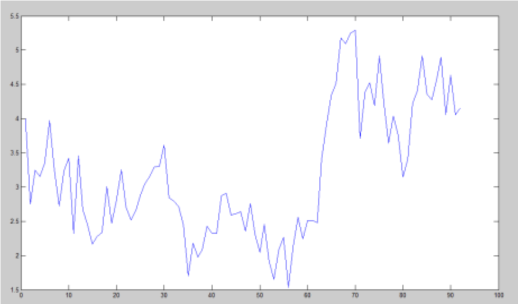

Figure 16: |Diff(n)| values for "Carphone" video with 6-frames deleted at the start

γ is computed on *PES* of original videos as well as on *PES* of altered videos after deleting different number of frames e.g. 3, 6, 9 etc. Also experimentally, the γ-values of original and framed deleted videos are found to differ a lot, hence γ can be used as a robust feature for training *SVM* to classify forged videos.

V. Machine Learning Techniques[8]

The real world applications deal with classifying data and recognising patterns from the data. There exists numerous supervised, semi supervised and unsupervised data classification algorithms[8] such as Artificial Neural Networks (ANN), decision trees, Support Vector Machine (SVM)[3],[4],[8] Ensemble based classifiers[8],[13],[14],[15] etc. These classification algorithms can be used in digital forensic applications to detect tampering and gather evidences from digital content. The proposed technique utilises SVM and Ensemble based classifiers to classify the tampered video from the original video, using the PES.

A. Support Vector Machine

Support Vector Machine is a set of supervised learning methods introduced by Vapnik (1995) and Cortes. It is a learning system that distinguishes between two classes by an optimal separating hyper plane. If the distance between closest input vectors to the hyperplane is maximum and the separation is without any error, the set of vectors are optimally separated.

Kernel methods map data from input space to feature space and perform learning in feature space. The kernel defines the classifier. *SVM* employs numerous kernel functions such as linear, polynomial, quadratic, radial basis function (*RBF*) and Multilayer perceptron (*MLP*). These kernel functions are used as approximating functions. These functions solve the problem of curse of dimensionality.

The goal of a classification problem is to discriminate between two classes, without loss of generality and *SVM* helps achieve it. This will help the classifiers to work well on unseen problems. *SVM* is a robust classification algorithm used for text classification, face recognition and several pattern recognition problems. The robustness makes *SVM* a suitable choice for carrying out classification of video forgery using *PES*.

B. Ensemble Based Classifier

In machine learning, ensembles [13],[14],[15] are used as they produce better results in term of accuracy and stability. The idea of ensemble based classifiers is to combine a set of classifiers instead of a single classifier. Weak learners will produce varying results making the classifier unstable and less accurate. The use of multiple classifiers, instead of a single classifier will reduce bias, because multiple classifiers will learn better. Hence, ensembles are used to improve the performance of these unstable learners. Ensemble based methods are used to detect video tampering with respect to original videos using the temporal property of the video (*PES*).

Ensemble based classification algorithm are proven to be suitable for multimedia detection scenarios. In our proposed technique, it is used to classify the videos as either original or forged. Ensemble methods are used to map training data to kernel space. γ feature is used to train the ensemble classifier. Various learning algorithms are employed by ensembles to improve the classifiers performance. *Boosting* & *bagging* are the popular ensemble methods.

Boosting [13-15] is a technique to improve the performance of any learning algorithm. *AdaBoost* (Adaptive Boosting) is such boosting algorithm formulated by Yoav Freund and Robert Schapire. It is an adaptive algorithm, as it tweaks the classifiers for data misclassified by previous classifiers. Even though the classifiers used may be weak, the algorithm will finally improve the model. The weak classifiers are considered, because they will have negative values in the final aggregation of the classifiers and will work as their inverses.. The technique repeatedly runs a weak learner on various distributed training data. These classifiers can be merged to form a strong classifier, hence reducing the

Copyright © 2013 MECS

error rate to increase the accuracy. Boosting algorithms have certain limitations. The algorithm fails to perform for insufficient data. It also fails to perform if the data is noisy.

Bagging [13-15] ensemble is a technique where each classifier is trained on a set of training data that are drawn randomly. The training is again carried out by replacing the training data with a dataset from the original training set. This training set is called a bootstrap replicate of the original training dataset. In bootstrapping, the replicate consists of average, 63.2% of the original training set, with multiple problems being trained many times. The predictions are made by considering the majority of votes in the ensemble. The aim of bagging is to reduce the error due to variance of the classifier.

VI. The Proposed Video Forensic Technique

In order to overcome the drawback of the method proposed by Wang & Farid, here we propose a novel method to automatically detect video forgery. Also analysing videos is a stressful and cumbersome job for human being. Hence, we propose to apply transforms to *PES* and extract feature vectors to classify video forgery using machine learning techniques like *SVM*, *Ensemble based classifier*. Fig. 17 shows the block diagram of our proposed temporal forensic method.

Here, we propose a novel technique to detect frame deletion using *DWT* and *SVM*, after analyzing the effect on *PES* in *DWT* domain (refer section IV). *DWT* is applied to *P*-frame *PES* to compute a new statistical parameter 'γ' as given in equation (3). This feature is related to the difference vector obtained from first level *DWT* coarse and detail sub-bands. *SVM* [3-4] is already proven as a robust technique to classify original and forged images/videos compared to other classification techniques such as neural networks, decision trees etc. Hence here we use *SVM* to classify the forged videos from the original videos, based on calculated 'γ' values. Experimental results show the robustness of using 'γ' for video forgery detection and also making our scheme scalable over analyzing large number of videos automatically, without human intervention.

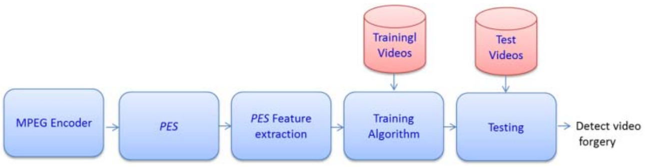

Figure 17: Proposed approach for video forensics using Machine Learning

VII. Simulation and experiment

A. Experimental Setup

The proposed approach is simulated in *MATLAB 7.1*. *MPEG-2* like encoder [10] by Steve Hoelzer is customized for detecting video alteration. The codec was modified and used for automatic frame deletion and detecting fingerprints in a tampered video using temporal video forensic.The following video sequences [11] were used in the experiments: *Carphone, Akiyo, Bowing, Foreman, Grandma, Intros, Mad900, Mother_daughter, News,Pamphlet, Paris,Sign_irene, Silent, Vtc1nw_422* and *mthr_dotr*. These are uncompressed QCIF videos of .y4m format, with dimensions of 176x144. All the selected videos are of similar nature, comprised of person carrying out verbal conversation. These videos have lip movement of the person speaking, while the rest of the background is almost static. Each video is tested with a fixed video length of 250 frames. In our simulations, these raw uncompressed videos are first encoded using a fixed length *GOP* of frame pattern *IPPP*. The scaling factor used in *MPEG* compression module is set to the fixed value of 31.

The first experiment is carried out on total 16 different .y4m videos. Out of which, 10 different videos are used in the training phase of *SVM* and remaining 6 different videos are used in the testing phase of *SVM*. The videos selected in *SVM* training phase are altered by deleting the first 3-frames/6-frames at the beginning of the video sequence and compressed using the *MPEG* encoder. *PES* is extracted from each video and processed using the *DWT* to calculate γ from the difference vector $|Diff(n)|$. Table I represents the γ -values for selected 10 videos with no deletions, 3-frame deleted and 6-frames deleted at the start position. These γ-values of selected videos are used to train the *SVM* in the training phase. Fig. 18 shows the difference in γ-values for original, 3-frame, 6-frame deleted videos in the beginning position. Table I is also testing using Ensemble based classifier.

Further, another experiment was carried out to test the performance of the proposed forensic technique. The trained *SVM* was tested for a dataset of 18 test videos with frames deleted at positions from 1 to 30 in the start position of the video sequence. These videos form a dataset of size 30 x 18 = 540 γ-values. Table III shows the extracted γ values of these 18 videos. The trained SVM is used to test these test videos. Also, Ensemble based classifier are used to test these 18 videos.

B. Result and Analysis

Table I represent the results of the various kernel functions used by the *SVM* and Ensemble based classifier for detecting video forgery. Table II show that all the five *SVM* kernel functions: linear, polynomial, quadratic, *MLP* and *RBF* detected the original and forged videos correctly.

TABLE I. Training Videos And Their Statistical Feature value Used For Training *SVM* And Ensemble Based Classifier

Sr. no.		γ - values		
	Video	*No deletions*	*3-frames deleted @ start position*	*6-frames deleted @ start position*
1.	Carphone_qcif	0.3514	2.669	3.198
2.	akiyo_qcif	0.2261	1.653	1.708

3.	bowing_qcif	0.2722	2.236	2.273
4.	foreman_qcif	0.5565	3.511	4.248
5.	grandma_qcif	0.0784	1.360	1.376
6.	intros_422_qcif	0.3068	2.249	2.341
7.	mad900_qcif	0.1646	1.85	1.873
8.	mother_daughter_qcif	0.2114	1.744	1.818
9.	news_qcif	0.1882	2.221	2.178
10.	pamphlet_qcif	0.3213	2.388	2.366

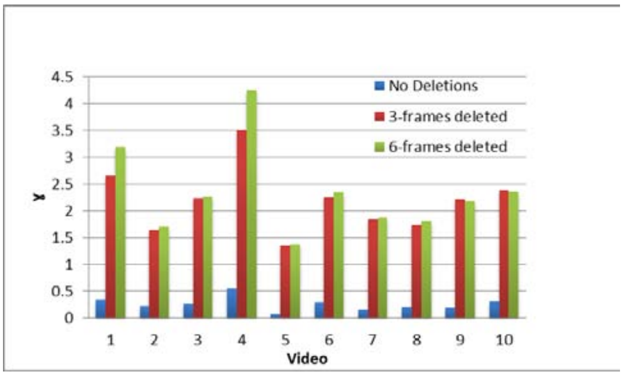

Figure 18: Difference in γ-values for original, 3-frame, 6-frame deleted videos

TABLE II. RESULTS FOR VARIOUS SVM KERNEL FUNCTIONS TO DETECT VIDEO FORGERY USING THE PROPOSED APPROACH

Sr. no.		SVM Detection Result			
	Video	*SVM* Kernel function	No deletions	3-frames deleted @ start position	*6-* frames deleted @ start position
1.	paris_qcif	Linear	Original	Forged	Forged
		Quadratic	Original	Forged	Forged
		Polynomial	Original	Forged	Forged
		RBF	Original	Forged	Forged
		MLP	Original	Forged	Forged
2.	sign_irene_qcif	Linear	Original	Forged	Forged
		Quadratic	Original	Forged	Forged
		Polynomial	Original	Forged	Forged
		RBF	Original	Forged	Forged
		MLP	Original	Forged	Forged
3.	silent_qcf	Linear	Original	Forged	Forged
		Quadratic	Original	Forged	Forged
		Polynomial	Original	Forged	Forged
		RBF	Original	Forged	Forged
		MLP	Original	Forged	Forged
4.	students_qcif	Linear	Original	Forged	Forged
		Quadratic	Original	Forged	Forged
		Polynomial	Original	Forged	Forged
		RBF	Original	Forged	Forged
		MLP	Original	Forged	Forged
5.	vtc1nw_422_qcif	Linear	Original	Forged	Forged
		Quadratic	Original	Forged	Forged
		Polynomial	Original	Forged	Forged
		RBF	Original	Forged	Forged
		MLP	Original	Forged	Forged
6.	mthr_dotr_qcif	Linear	Original	Forged	Forged
		Quadratic	Original	Forged	Forged
		Polynomial	Original	Forged	Forged
		RBF	Original	Forged	Forged
		MLP	Original	Forged	Forged

C. Receiver Operating Characteristics (ROC)

ROC curves [16], [17] are a measure to evaluate the performance of a detector. ROC is a plot of the true positive rate against the false positive rate for the different possible cutpoints of a test. Sensitivity is the proportion of tampered videos that are tested positive. Specificity is the proportion of original videos that are tested negative. Sensitivity and specificity describe how well the test discriminates between tampered and original videos. ROC curve shows the tradeoff between sensitivity and specificity. The test is supposed to be more accurate if the curve is closer to the left axis and top axis of the ROC plot. If the curve is closer to the 45-degree line of the ROC space, the test is less accurate.

Fig. 19 shows the ROC curve for the videos mentioned in the Table I. Various kernel functions successfully detected video forgery using the proposed approach. Experimental result shows a robust and efficient classification of original and tampered videos using SVM.

TABLE III. TEST VIDEOS AND THEIR STATISTICAL FEATURE VALUE ɣ USED FOR TRAINING SVM AND ENSEMBLE BASED CLASSIFIER

Frames Deleted	*PES*	γ - values			
		carphone	*akiyo*	*bowing*	*deadline_qcif*
1	186	2.7774	1.651	2.0254	2.7451
2	186	3.212	1.7234	2.2626	2.8851
3	185	2.6698	1.6529	2.236	3.1267
4	184	0.3506	0.2236	0.2787	0.1636
5	183	2.7981	1.6606	2.0433	2.7514
6	183	3.1983	1.7087	2.2737	2.8808
7	182	2.6772	1.6787	2.2581	3.137
8	181	0.3542	0.2265	0.2772	0.1724
9	180	2.778	1.6533	2.0564	2.7434
10	180	3.1983	1.7261	2.2987	2.8833
11	179	2.6586	1.6559	2.2679	3.1259
12	178	0.3566	0.2235	0.2849	0.164
13	177	2.8031	1.6629	2.0753	2.7481
14	177	3.1937	1.7106	2.3106	2.8724
15	176	2.6804	1.6815	2.2886	3.1344
16	175	0.3607	0.2278	0.2832	0.1709
17	174	2.7771	1.6563	2.0888	2.7454
18	174	3.1999	1.7288	2.3369	2.8869
19	173	2.6748	1.6576	2.2986	3.1312
20	172	0.3573	0.2218	0.291	0.1626
21	171	2.7982	1.6655	2.1088	2.7495
22	171	3.1833	1.713	2.3476	2.8769
23	170	2.6884	1.6826	2.3236	3.136
24	169	0.3616	0.2241	0.2896	0.1723
25	168	2.7781	1.6589	2.1227	2.7473
26	168	3.191	1.7303	2.3758	2.8854
27	167	2.6772	1.6597	2.3333	3.1328
28	166	0.3639	0.2213	0.2984	0.1657
29	165	2.8036	1.663	2.144	2.7548
30	165	3.1874	1.7122	2.3876	2.8871

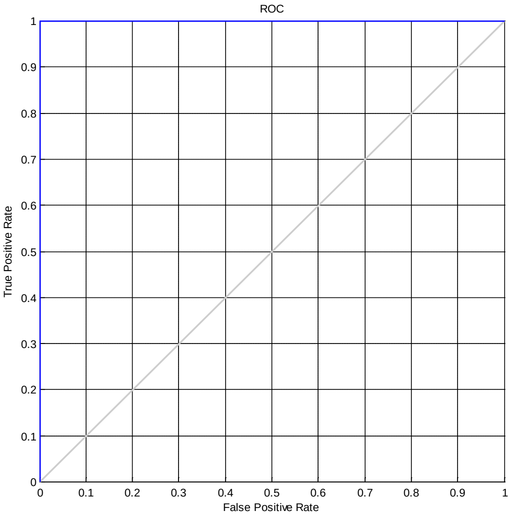

Figure 19: ROC curve for SVM classification

Fig. 20 shows the ROC curve for few of the videos mentioned in the Table III. This ROC shows that SVM is robust and efficient in classifying original and tampered videos.

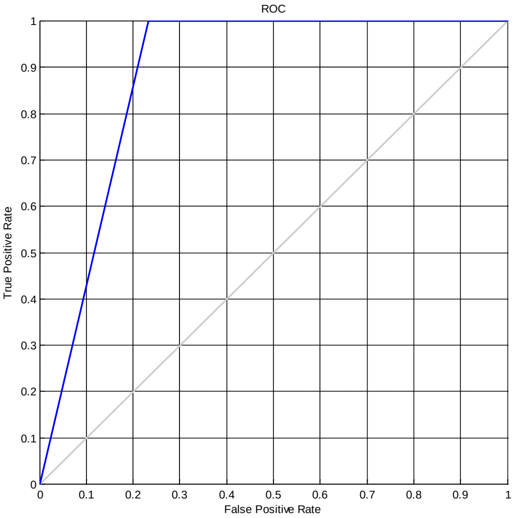

Figure 20: ROC curve for SVM classification

Fig. 21 shows the *ROC* curve for Ensemble based video forensic classification mentioned in the Table I. Experimental results show that the performance of ensemble based classifier was similar to *SVM*. Ensemble based classifier also successfully detected video forgery using the proposed approach.

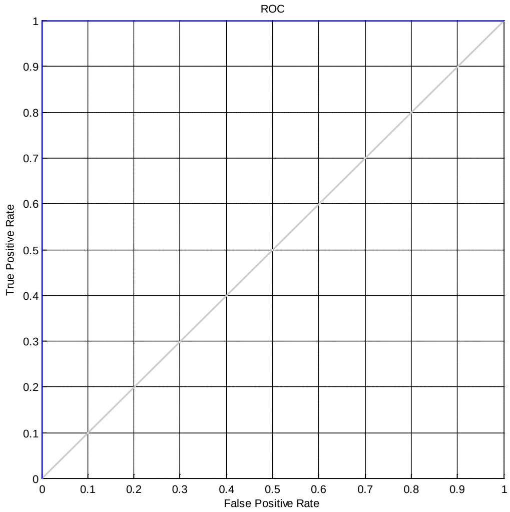

Figure 21: ROC curve for Ensemble based classification

Fig. 22 shows the ROC curve for few of the videos using Ensemble based video forensic classification mentioned in the Table III. Ensemble based method successfully detected most of the videos with good efficiency. The results of ensemble based classifier matched with the SVM classifier.

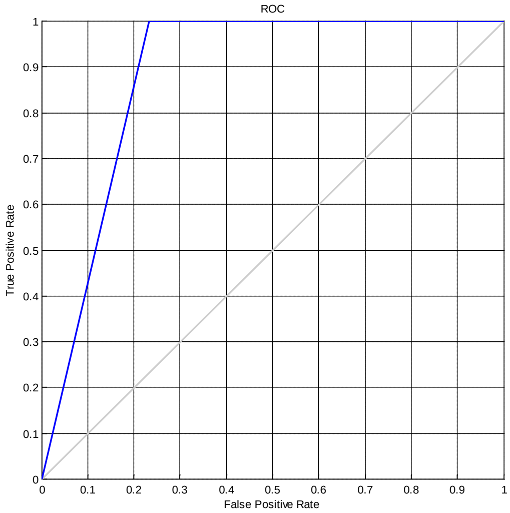

Figure 22: ROC curve for Ensemble based classification

VIII. Conclusions

In this thesis, we proposed a novel automatic video forgery detection technique based on double *MPEG* compression using *SVM* and Ensemble based classifier in the *DWT* domain. The *PES* feature 'γ' related to the difference vector of first level *DWT* coarse and detail sub bands is proven as a robust training feature for automatic detection of temporal attacks. Experimental results shows all the test video samples were correctly detected by both *SVM* as well as Ensemble based classifier. The proposed scheme exhibits simpler design and implementation. The experimental results have validated the efficiency of our video forgery detection technique. Also the scheme automates the detection process without the need of human intervention.

Appendix

The .y4m videos [11] used in the experiments as shown in the appendix. Fig. 23(a)-(p) are used in the *SVM* and Ensemble training phase and testing phases.

(a) carphone (b) akiyo (c) bowing (d) foreman

(e) grandma (f) intros_422 (g) mad900 (h)mother_daughter

(i) news (j) pamphlet (k) paris (l) sign_irene

(m) silent (n) students (o)vtc1nw_42 2 (p) mthr_dotr

Figure 23: Videos used in the training and testing phases of the proposed

References

[1] Weihong Wang, Hany Farid, "Exposing Digital foregeries in Video by Detecting Double MPEG Compression," in Proc. ACM Multimedia and Security Workshop, Geneva, Switzerland, 2006, pp. 37–47.

[2] Matthew C. Stamm, W. Sabrina Lin and K. J. Ray Liu, "Temporal Forensics and Anti-Forensics for Motion Compensated Video," in IEEE Transactions On Information Forensics and Security, Vol. 7, No. 4, August 2012, pp. 1315-1329.

[3] Asma Rabaoui, Manuel Davy, Stéphane Rossignol, and Noureddine Ellouze , "Using One-Class SVMs and Wavelets for Audio Surveillance" in IEEE Transactions On Information Forensics and Security, VOL. 3, NO. 4, December 2008,pp. 763-775

[4] Cheng-Liang Lai,Yi-Shiang Chen "The Application of Intelligent System to Digital Image Forensics", Proceedings of the Eighth International Conference on Machine Learning and Cybernetics, Baoding, 12-15 July 2009, pp. 2991-2998

[5] *The MPEG Handbook*, 1st ed., Elsevier Group, John Watkinson,2001, pp. 140-208.

[6] A beginners guide for MPEG-2 standard.Available: http://www.fh-friedberg.de/fachbereiche/e2/telekom-labor/zinke/mk/mpeg2beg/beginnzi.htm

[7] Basic Video Coding and MPEG.Available: http://www-ee.uta.edu/dip/courses/ee5356/631pub04_sec12videoMPEG.ppt

[8] "Pattern Recognition in Matlab", K.Koutroumbas ,Elsevier,2009

[9] *Digital Image Processing*, 3rd ed., Pearson Education, Rafael C. Gonzalez,Richard E. Woods,2009, pp. 510-512.

[10] MPEG-2 overview and MATLAB codec project.Available: http://www.cs.cf.ac.uk/Dave/Multimedia/Lecture_Examples/Compression/mpegproj/

[11] Xiph.org Video Test Media.Available : http://media.xiph.org/video/derf/

[12] "Image processing and Pattern Recognition", Frank Y. Shih, Wiley, 2010

[13] Jan Kodovský, Jessica Fridrich and Vojtěch Holub, "Ensemble Classifiers for Steganalysis of Digital Media", IEEE Transactions on Information Forensics and Security, Vol. 7, No. 2, April 2012, pp. 432-444

[14] Gabriele Zenobi, Pádraig Cunningham " Using Diversity in Preparing Ensembles of Classifiers Based on Different Feature Subsets to Minimize Generalization Error", Department of Computer Science, Trinity College Dublin , pp 1-15

[15] Lior Rokach, "Ensemble Methods for classifiers" Chapter 45, Department of Industrial Engineering, Tel-Aviv University, pp. 957-962

[16] Lena Kallin Westin, "Receiver operating characteristic (ROC) analysis" ,Department of Computing Science, Umeå University, Available : www8.cs.umu.se/research/reports/2001/018/part1.pdf

[17] "Introduction to ROC Curves" , Available : http://gim.unmc.edu/dxtests/ROC1.htm

Sunil Jaiswal, born in 1980, currently pursuing M.Tech. in Computer Science and Engineering (Cyber Security) at Defence Institute of Advanced Technology. His main research interests include Digital Forensics and Cyber Security.

Sunita Dhavale, working as an Assistant Professor in Computer Science and Engineering Department, Defence Institute of Advanced Technology. Her main research interests include Digital Forensics, Steganography and Digital Watermarking.

Copyright © 2013 MECS