Spatial Video Forgery Detection and Localization using Texture Analysis of Consecutive Frames

Mubbashar SADDIQUE[1], Khurshid ASGHAR[2], Usama Ijaz BAJWA[3], Muhammad HUSSAIN[4], Zulfiqar HABIB[5]
[1,3,5]*Department of Computer Science, COMSATS University Islamabad, Lahore Campus, Pakistan*
[2]*Department of Computer Science, University of Okara, Pakistan*
[4]*Department of Computer Science, King Saud University, Riyadh, Saudi Arabia*
[5]*drzhabib@cuilahore.edu.pk*

***Abstract*—Now-a-days, videos can be easily recorded and forged with user-friendly editing tools. These videos can be shared on social networks to make false propaganda. During the process of spatial forgery, the texture and micro-patterns of the frames become inconsistent, which can be observed in the difference of two consecutive frames. Based on this observation, a method has been proposed for detection of forged video segments and localization of forged frames. Employing the Chrominance value of Consecutive frame Difference (CCD) and Discriminative Robust Local Binary Pattern (DRLBP), a new descriptor is introduced to model the inconsistency embedded in the frames due to forgery. Support Vector Machine (SVM) is used to detect whether the pair of consecutive frames is forged. If at least one pair of consecutive frames is detected as forged, the video segment is predicted as forged and the forged frames are localized. Intensive experiments are performed to validate the performance of the method on a combined dataset of videos, which were tampered by copy-move and splicing methods. The detection accuracy on large dataset is 96.68 percent and video accuracy is 98.32 percent. The comparison shows that it outperforms the state-of-the-art methods, even through cross dataset validation.**

***Index Terms*—forensics, image classification, machine learning, multimedia systems.**

I. INTRODUCTION

Digital videos contain rich information and evidence about an event [1]. These videos can easily be tampered with user-friendly video editing tools like Adobe Premier, After Effect, GNU Gimp, and Vegas etc., and uploaded on the social media for propaganda and malicious purposes. It also reduces the credibility of videos on the social media and in the court of law [2-3]. Such issues have made forensic analysis essential to ensure the authenticity and integrity of videos. Many fields of life such as media groups, insurance companies, marriage bureaus, investigation agencies, social media, and courts of law require the authentication of videos [4].

The existing video forgery detection methods can broadly be categorized into two groups: active and passive. Passive techniques do not require embedded information (watermark, digital signature) unlike active techniques, and hence can be applied to authenticate any video. Due to this reason, passive techniques have become a hot research area in the field of information security.

Videos can be forged by three different ways: (i) spatial or object-based tampering, (ii) temporal or frame based tampering, and (iii) spatio-temporal tampering [5]. The focus of this research is on spatial tampering, an example of which is shown in Fig. 1. The object in red rectangle is present in original video Fig. 1(a) but this object is deleted from the frames to tamper the video as shown in Fig. 1(b). In spatial forgery, the actual information is concealed by deleting an object from different frames and viewers are misguided by providing them false information. The purpose of this type of forgery is not merely retouching or changing format, but to hide the facts for propaganda or other criminal intentions. As such, this type of forgery is dangerous and has a negative impact on the society.

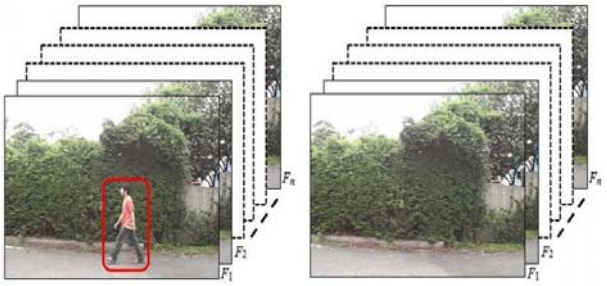

Figure 1. Tampering in spatial domain (object based)

Despite the severity of spatial forgery, the research in this area is at its early stage. A number of techniques have been proposed to detect image forgery [6-7]. These techniques are not applicable for the detection of spatial or object-based forgery in videos due to the following reasons. Firstly, videos are encoded and compressed before storage and transmission due to large amount of data in video frames. Secondly, computational complexity of reported techniques becomes very high when applied to video frames. Thirdly, forgery traces are available in consecutive frames of a video, which is not possible in case of an image [8]. Lastly, a model, which is trained on images, cannot be applied to video frames because it is not trained to consider the contextual forgery information embedded in consecutive

This research is supported by the University of Okara and PDE-GIR project which has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 778035.

Digital Object Identifier 10.4316/AECE.2019.03012

frames.

In spatial domain, forgery can be done in two different ways (i) copy move and, (ii) splicing. In copy move forgery, the object is copied and pasted in the frames of the same video, whereas in splice forgery, the object is taken from another video and pasted in the frames of a video. Spatial video forgery detection aims to find whether the video is forged or not? Whereas, the localization digs out which frames of the video are forged and the exact regions, where an object or some parts are tampered in these frames. In this research, the focus is on both the detection of forged video segments (VSs) and localization of forged frames.

During spatial domain video tampering, the texture of micro-patterns is changed in tampered frames, which is a very strong clue to detect this kind of forgery. Subramanyam and Emmanuel [9] applied texture descriptor Histogram of Oriented Gradients (HOG) to model the tampering traces in video frames. This descriptor and its variants employ gradient orientation, which cannot describe local texture micro-patterns and variations effectively. HOG gives only shape information due to occurrences of gradient orientation, hence not robust to noise and scale variations [12] Local Binary Pattern (LBP) is another popular texture descriptor, which is investigated for image forgery detection [6]. It has been used for many classification tasks [10-11]. This descriptor is robust against monotonic illumination changes and contrast variation; however, it is sensitive to noise and small gray-level fluctuations. Moreover, LBP also does not incorporate the edge strengths.

Keeping in view the above limitations, a descriptor is required to represent the changes occurred in video frames due to spatial forgery for developing a reliable video forgery detection model. For this purpose, a video is divided into video segments and then difference of consecutive frames (DOCFs) is calculated. Subsequently, proposed descriptor is used to extract feature vectors from DOCFs. Finally, the model is trained using SVM classifier. If at least one consecutive pair of frames is found forged, then video segment is declared forged, otherwise it is authentic.

This study has following contributions: (i) An architecture of video forgery detection and localization is presented. (ii) A new descriptor based on the difference of consecutive frames is developed to extract the discriminative features, which helps in video forgery detection and locating forged frames. (iii) The parameters of SVM are tuned to classify a given video as authentic or forged and localization of forged frames. (iv) Good accuracy is achieved through cross dataset validation as well. The results indicate that the proposed method outperforms state-of-the-art methods.

A. Related Work

Different techniques have been proposed for detection of video forgery in the spatial domain and can be divided into various categories based on their types of feature values. In first category, statistical features are used to detect the forgery. C. Hsu, T. Hung, and C. Lin [13] used wavelet co-efficient thresholding and Bayesian Classifier. C. Richao, Y. Gaobo, and Z. Ningbo [1] employed statistical moments to extract features and used SVM for classification. The method achieved good accuracy but tested on very limited dataset. A. Subramanyam and S. Emmanuel [9] exploited the HOG features. Singh et al., [14] exploited the correlation coefficient to find the duplicated regions in the videos. Su et al., [15] examined the k-singular value decomposition K-SVD algorithm and K-means. S. Chen, S. Tan, B. Li, and J. Huang [16] used motion residual and steganography features to detect video forgery. Though the feature vector of these techniques are small in length relative to other categories of algorithms but unable to detect forgery in the presence of different types of post-processing operations.

This category of techniques exploited the noise characteristics to detect the forgery. Kobayashin et al., [17] employed noise characteristics. D.K. Hyun, S.J. Ryu, H.Y. Lee, and H.K. Lee [18] detected forgery using sensor pattern noise (SPN). R. D. Singh and N. Aggarwal [19] used pixels correlation, noise inconsistency and discrete fractional Fourier transformation. Panday et al., [20] worked with scale invariant feature transform (SIFT), noise residual and correlation to detect copy-move forgery. Goodwin and Chetty [21] also used noise residual, quantization features and their transformation in cross-model subspace to detect the copy-move forgery. The techniques of this category although performed well but are dependent on the hardware.

The techniques in this category work based on optical flow and motion residual. A. Bidokhti and S. Ghaemmaghami [22] proposed a technique based on optical flow to detect a copy-move forgery from MPEG videos. In [23] moment feature of wavelet co-efficients and optical flow are combined with SVM to detect the facial expression re-enacted forgery (FERF). Al-sanjary et al., [24] proposed optical flow inconsistencies and dynamic time warping (DTW) matching algorithm to detect copy-move forgery in videos. L. Su and C. Li [25] detected copy-move forgery in the first frame of the video by employing MISIFT and used spatio-temporal context learning to detect the forged areas from remaining frames of the video. In [26] block-based motion estimation is used to extract motion from the adjacent frames and then the magnitude and orientation are employed to differentiate the authentic and forged video. The technique presented in [27] expose the forgery in videos that have ballistic motion.

This category of techniques utilized the deep convolution neural network. Zampoglou et al., [28] employed Q4 and Cobalt forensic filters with pre-trained GooleNet and ResNet networks to detect the video forgery. Y. Yao, Y. Shi, S. Weng, and B. Guan [29] utilized a CNN to extract high dimension features and used an absolute difference between successive frames to cut down the temporal redundancy, a max pooling layer to minimize the computational complexity and high pass filter layer to increase the residual left during the forgery process. The techniques of this category give high-features and produced good accuracy; however, the small size of tampering cannot be detected with the help of these algorithms and the computation is very high which are developed so far.

Most of the existing techniques detect the video forgery in two main steps. First part is the extraction of features from video frames by using some descriptors. In the second part, these features are passed to a classifier in vector form to train the model and then classify the video as forged or authentic. In this process, the extraction of features is a very important step because these features encode the traces left

in the tampered frames of video. So, we need such type of descriptor which can encode and extract the discriminant features from these traces easily.

B. Motivation and Objectives

It is an important and challenging task to extract discriminative information from frames, based on which a video can be identified as forged or authentic. The texture and pixel statistics of frames are changed during the tampering process. The traces left during the process of tampering, can be seen easily by decompressing the video into frames and calculating the difference of consecutive frames (DOCFs). When the difference is calculated in forged and authentic frames, the traces left in tampered part in forged video can be seen but in an authentic video these traces do not exist. This fact is also shown in Fig. 2.

The authentic and forged video frames (1, 2, 3 and 4) are presented in Fig. 2(a) and Fig. 2(b) respectively. In Fig. 2(b), red rectangle in the frames (1, 2, 3 and 4) illustrates the tampered part in the frames. The zoom in view (tampered area) of the difference between consecutive frames (1 and 2, 2 and 3, 3 and 4) of the authentic video is shown in Fig. 2(c), 2(d), and 2(e) respectively. Similarly, Fig. 2(f), 2(g), and 2(h) demonstrates the zoom in view of the difference between consecutive frames of the highlighted forged part in the forged video. In the authentic video, it can be easily observed from Fig. 2(c), 2(d) and 2(e) that the difference remains unchanged but in the forged video it is changed as shown in Fig. 2(f), 2(g) and 2(h). This change in the difference is due to traces left in the forged video. In this research, this change in texture that exists in the consecutive forged frames is investigated by applying a proposed descriptor to extract the most discriminant features.

Following are the objectives of this study: (i) to propose a descriptor which encodes tampering traces in a precise way, (ii) to localize the forged frames in the video segment (iii) to ensure the robustness of the proposed method.

C. Organization of the Paper

The rest of the paper is organized as follows. The proposed method is elaborated in Section II. The evaluation methodology is described in Section III. The results are discussed in Section IV. Finally, Section V concludes the paper with future work.

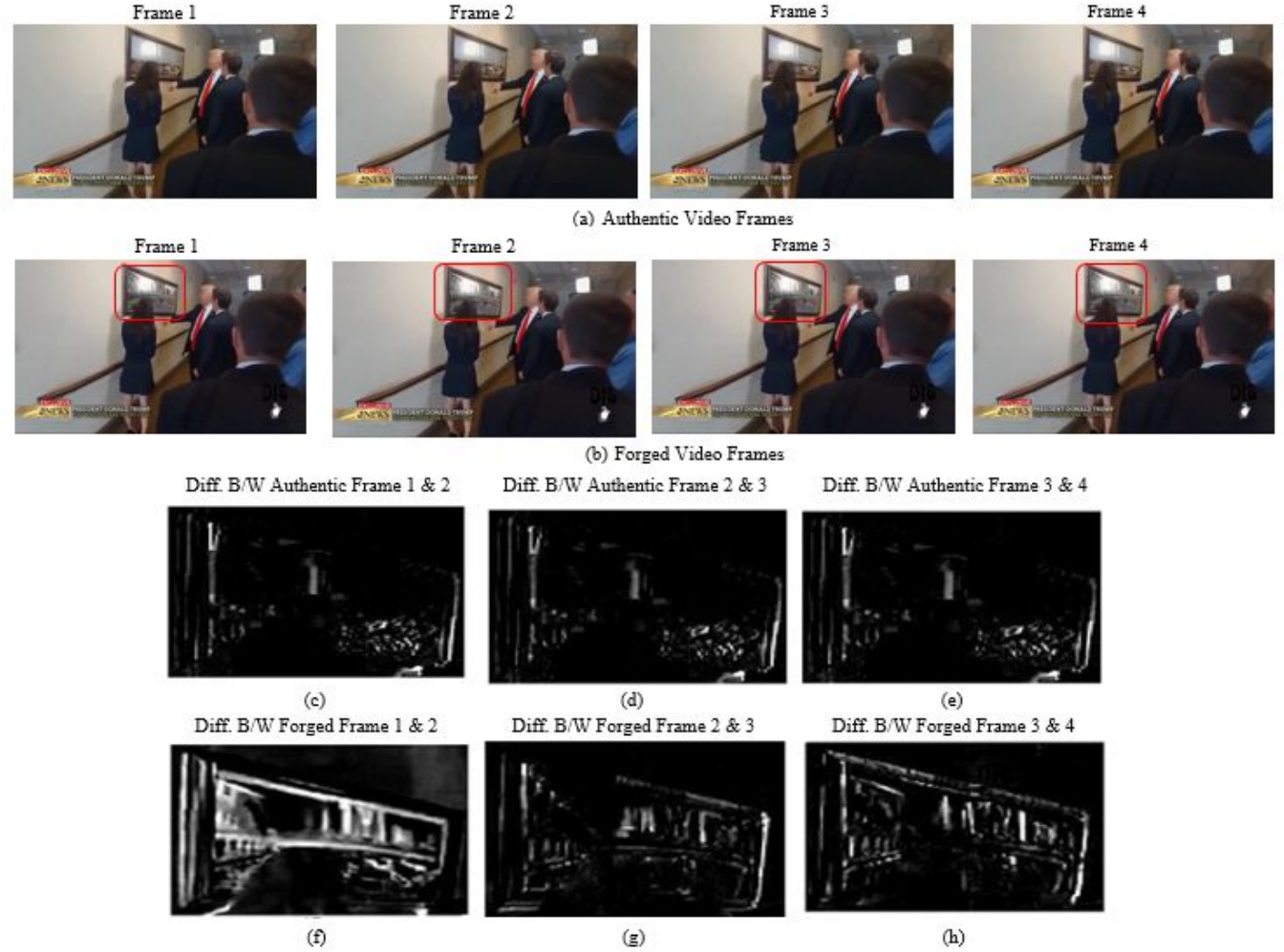

Figure 2. (a) Frames of the authentic video. (b) Frames of the forged video (forged part is highlighted by a red rectangle in frame 1, 2, 3 and 4 in a forged video). (c) Zoom in view of difference between forged part of frame 1 and 2 of the original video. (d) Zoom in view of difference between forged part of frame 2 and 3 of the original video. (e) Zoom in view of difference between forged part of frame 3 and 4 of the original video. (f) Zoom in view of difference between forged part of frame 1 and 2 of a forged video. (g) Zoom in view of difference between forged part of frame 2 and 3 of a forged video. (h) Zoom in view of difference between forged part of frame 3 and 4.

II. Proposed Video Forgery Detection System

The goal of this study is to develop a robust system, to classifiy whether a video is forged or authentic and to identify the forged frames. For this purpose, an encoding scheme is developed to model the embedded forgery traces, which are embedded in the forged part of the video during forgery. Firstly, a video is divided into video segments (VSs) of length 30 frames each and frames are extracted from VSs. The overlapping DOCFs is calculated and then features are extracted from DOCFs with the help of proposed descriptor (explained with detail in next sub section), which are then passed one-by-one to SVM based decision model. The decision model returns a decision d_i, where $i = 1,2,..., N-1$. For one video segment, if all the DOCFs are found authentic i.e., all $d_i = -1$, the video segment is authentic otherwise it is forged, and the consecutive frames are declared forged for which $d_i = 1$. The proposed video forgery detection and localization system is presented in Fig. 3. The detail of the system is given below.

A. Proposed Descriptor

Researchers used LBP and HOG texture descriptors for detection of forgery in videos and images. LBP is robust for illumination and contrast while it is sensitive to noise. HOG gives shape information, but not robust to noise and scale variations. Spatial (object-based) forgery has two important clues of tampering detection in video frames. First, the texture of the forged/tampered object and second, the shape formed by the edges of the forged/tampered object. LBP and HOG do not differentiate texture and shape properly, so there is a dire need for such descriptor that can differentiate between both properly. Recently, a robust texture descriptor DRLBP [30] has been proposed which has fused both texture and shape (edge) information into a single representation. This single representation helps to represent the sharp transitions such as discontinuities and inconsistencies occurred during forgery and present in the form of edges, lines, and corners in the DOCFs as shown in Fig. 2(f), 2(g) and 2(h). So, these inconsistencies and discontinuities are highlighted unequivocally by proposed descriptor. Due to this fact, in this research, a new descriptor named CCD-DRLBP (Chrominance value of Consecutive frame Difference and Discriminative Robust Local Binary Pattern) is proposed to extract discriminant features. The process to compute features from DOCFs through CCD-DRLBP is explained below.

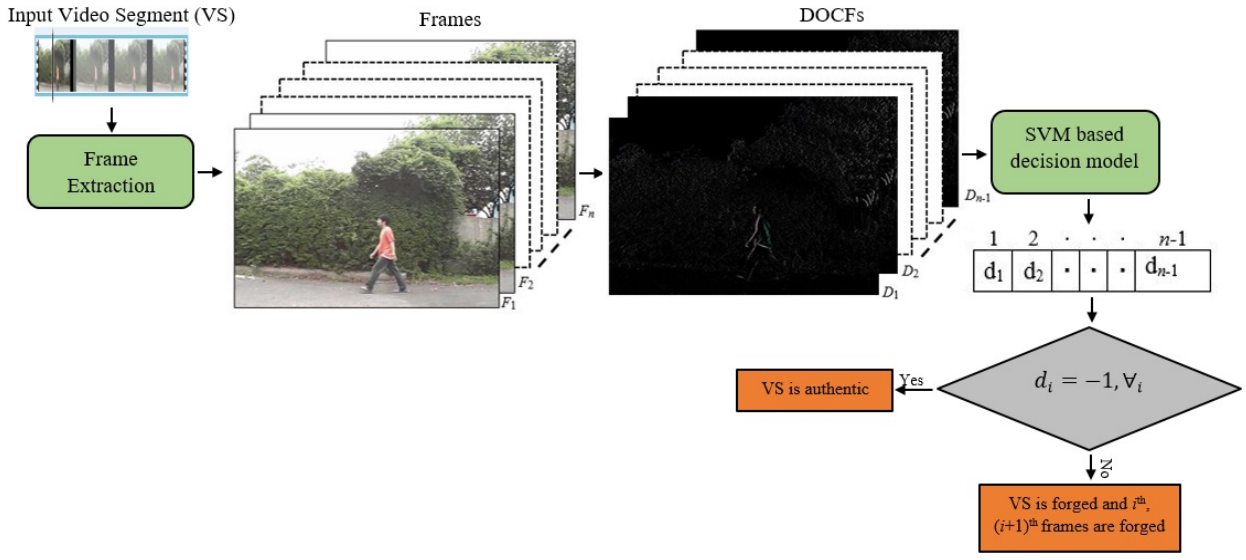

Figure 3. Video forgery detection and localization system

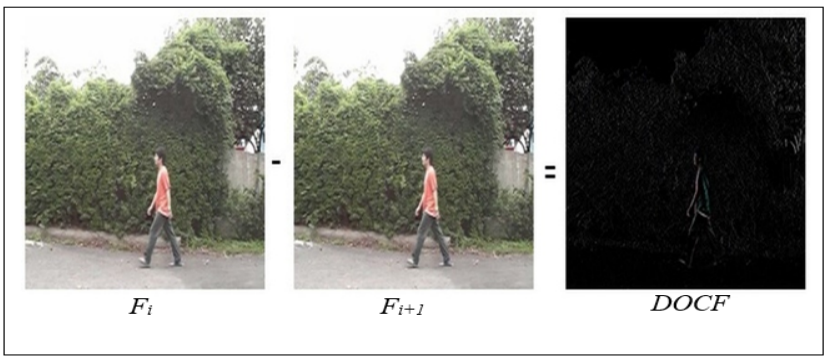

Figure 4. Difference of consecutive frames (DOCFs) computation

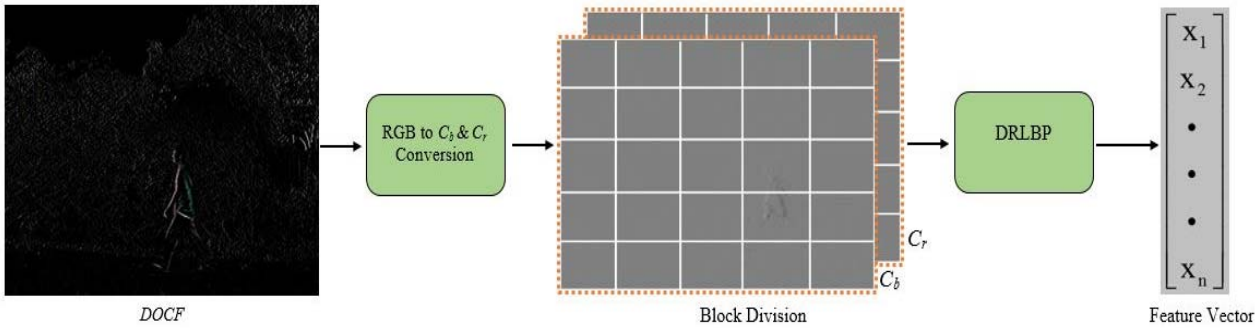

Figure 5. Visual description of Chrominance value of Consecutive Frame Difference (CCD) approach for feature extraction

1) Computation of Chrominance value of Consecutive Frame Difference (CCD)

In this process, a video V is divided into VSs and each VS consists of N=30 frames:

$$VS = \{F_1, F_2, ..., F_{30}\} \quad (1)$$

The overlapping difference of consecutive frames (DOCFs) of F_i and F_{i+1} is calculated which is denoted by D_i and its visual description is presented in Fig. 4:

$$D_i = F_i - F_{i+1} \text{ where } i = 1,2,..., N-1. \quad (2)$$

Each D_i is transformed from RGB space to YC_bC_r space using (3):

$$\begin{pmatrix} Y \\ C_b \\ C_r \end{pmatrix} = \begin{pmatrix} 0.299 & 0.587 & 0.117 \\ -0.299 & -0.587 & 0.886 \\ 0.701 & -0.587 & -0.114 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} + \begin{pmatrix} 16 \\ 128 \\ 128 \end{pmatrix}, \quad (3)$$

where Y is the luminance component, and C_b, C_r are the chrominance components.

During the process of tampering, efforts are made to hide the tampering traces so that they are not visible to human eyes. As human eyes are more sensitive to luminance, therefore it is assumed that tampering traces may be embedded in the chrominance components. The reason is that, luminance component mostly shows the contents of objects, while chrominance components encode the weak signal contents (edges and lines) which are not visible to human eyes. Edge irregularities of the objects caused by tampering form weak signals are embedded in video frames and are captured effectively using chrominance components [31-32]. Based on above facts C_b and C_r components are selected.

2) Computation of CCD-DRLBP descriptor

Each channel of CCD is divide into 5×5 blocks. These channels are divided into blocks due to following reasons. (i) Spatial locations have important role to detect the spatial (object-based) forgery in the videos. The local spatial information may get lost if features are extracted from the holistic frame. (ii) The frames are divided into number of blocks to keep the length of feature vector same on different resolution of videos. Following steps are followed to calculate the CCD-DRLBP descriptor of each block.

(i) Gradient $\omega_{x,y}$ of every pixel of each block of CCD is calculated using [33].

(ii) LBP image is calculated.

(iii) Weighted histogram H_{lbp} of LBP codes is also calculated.

(iv) The robust weighted histogram H_{rlbp} using [30] is computed to remove the reversal effects from the background and foreground.

(v) Discriminant weighted histogram H_{drlbp} is calculated to enhance the effects of pattern using [30].

(vi) DRLBP descriptor is computed by concatenating the H_{rlbp} and H_{drlbp} of each block of each channel. This descriptor is called CCD-DRLBP and its procedure is also explained in Fig. 5 and Algorithm 1.

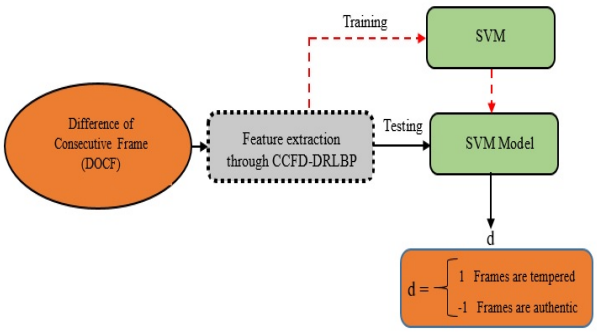

Figure 6. SVM based decision model

B. SVM Based Decision Model

Classification is a general process related to the categorization of predefined classes. In this study, classification model is trained with the help of SVM classifier [34] for detection of authentic and forged video segments because it is computationally efficient and robust. This model takes the difference of consecutive frames (DOCFs) as an input and gives the decision whether the frames are forged (1) or not (-1). The graphical representation of SVM based decision model is shown in Fig. 6. The idea behind SVM is to choose the hyper plane with the maximum margin. Let D be a training dataset with N points in a d-dimensional space:

$$D = \{x_i, y\}, \quad (4)$$

where i = 1, 2, ..., N and $y \in \{+1, -1\}$. A linear separating function $h(x)$ to split the original space into two half-spaces in d-dimensions is defined as:

$$h(x) = w^t x + b = w_1 x_1 + w_1 x_1 + ...w_d x_d + b, \quad (5)$$

where w is a d-dimensional weight vector and b is a scalar

bias. Points on the hyper plane have *h(x)* = 0, in other words, the hyper plane is defined by all points for which $w^t x = -b$. If the dataset is linearly separable *h(x)* will be less than 0 for all points having label -1 and for all points labeled +1 *h(x)* will be greater than 0 or a given separating hyper plane *h(x)* = 0, the distance between each point x_i and the hyper plane is calculated as:

$$\delta_i = \frac{yh(x)}{\| w \|}, \tag{6}$$

The margin is defined as the minimum distance of all *N* points to the separating hyper plane.

$$\delta = \min\left(\frac{yh(x_i)}{\| w \|} \right), \tag{7}$$

The set of points that satisfy this minimum distance are the support vectors for the linear classifier, but unfortunately, features are nonlinear for higher dimensions. In this case, kernel trick is used to make hyper planes linearly separable. Different types of kernel have been proposed like linear, polynomial, radial basis function (RBF), sigmoid etc. In this research, RBF kernel is selected empirically and we have achieved good accuracy by optimizing its parameters gamma γ and penalty *C*. Finding the best value of *C* and γ is very important and not easy because prediction of classes with good accuracy depends on these parFiameters. The optimized values of *C* and γ are $C=2^{-1}$, $\gamma=2^{1}$ which are tunned using grid-search algorithm [35] from ($C=2^{-2}$, 2^{-3}, …,2^{-15}) and ($\gamma=2^{-2}$, 2^{-3}, …,2^{-15}) and shown in Fig 8(f). The learning of SVM model is presented in Algorithm 2.

Algorithm 1: The computation of feature vector based on CCD-DRLBP from DOCFs

---

Input: Difference of Consecutive frames (DOCFs)

Output: Feature Vector (F.V) based on CCD-DRLBP descriptor of DOCFs

Procedure:

1. Take DOCFs as input which denoted by D_i where $i = 1,2,..., N-1$
2. Convert D_i from RGB to YCbCr
3. Select C_b and C_r space and divide into $K = L \times L$ blocks
4. For each channel $i \in \{C_b, C_r\}$ do Steps A~B

Step A: For each block $k = 1,2,..., K$ do

Step 1: For each pixel at location (x, y), compute weight $\omega_{x,y}$ [33]

$$w_{x,y} = \sqrt{I_x^2 + I_y^2}$$

where I_x and I_y are the first-order derivatives in the x and y direction.

Step 2: Compute image of LBP codes

Step 3: Compute weighted histogram H_{lbp} of LBP codes

$$H_{lbp}(j) = \sum_{x=0}^{L-1} \sum_{y=0}^{M-1} w_{x,y} \delta(LBP_{x,y}, j) \text{, where } j = 1,2,...,n$$

$$\delta(l,m) = \begin{cases} 1 & l = m \\ 0 & Otherwise \end{cases},$$

where n is the number of LBP codes.

Step 4: Compute robust LBP (RLBP) histogram H_{rlbp}

Step 5: Calculate difference LBP (DLBP) histogram H_{drlbp}

Step 6: Compute DRLBP descriptor by concatenating RLBP and DLBP histograms and named this descriptor as CCD-DRLBP

Step B: Concatenate CCD-DRLBP descriptors from K blocks

$CH_i = \{B1, B2,..., B_k\}$

5. Concatenate CH_i, $i \in$ {Cb , Cr }
6. $F.V_{D_i}$ = concatenate $\{CH_{Cb}, CH_{Cr}\}$

III. Evaluation Methodology

In this section, datasets, evaluation protocol, evaluation metrics, statistical analysis of CCD-DRLBP descriptor, parameters and parameter tuning are explained in detail.

A. Datasets Description

To evaluate the performance of the proposed method, the datasets utilized in [13], [36-38] are employed and annotated as D1, D2, D3, D4, D5, and D6. All these datasets collectively contain 130 authentic and 137 forged videos,

detail of which are given in Table I. In dataset D1 100 authentic videos are taken from [37], having frame rate 30fps and resolution 320x240. The videos are captured by different types of cameras like Canon, Fuji and Nikon with AVI and MOV formats. D2 [36] dataset contained a total of 20 videos, out of which 10 videos are authentic and 10 are forged. Each video in D2 varies in number of frames ranging from 210 to 583. In the forged videos, a total of 880 frames are forged using copy-move forgery [39]. The dataset D3 employed in [13], composed of total 20 videos, out of which 14 videos are authentic and 6 are forged. The forged videos have different number of frames ranging from 79 to 500, and a total of 1286 frames are forged in this dataset using copy-move, forgery. In dataset D4, authentic videos are taken from D1 and forged videos set is the combination of forged videos from datasets D2 and D3. In dataset D5 [38], each authentic video is forged with different geometric transformations (flipping, rotation, scaling and shearing), post-processing operations (luminance, RGB) and None (without any geometric transformation and post-processing operations. Moreover, this dataset has a total of 3800 forged frames. Dataset D6 is the combination of the authentic videos taken from D1, D2, D3, D5, and forged from D2, D3 and D5.

B. Evaluation protocol

A total of 12463 frames (6500 authentic and 5966 forged frames) are used for training and testing. Although the datsets collectively have a large number of authentic frames, but to avoid imbalance, 6500 authentic frames are collected by randomly selecting 50 frames from each authentic video. DOCFs of all the forged and authentic frames are calculated, then using CCD-DRLBP descriptor, we obtain the feature vectors and finaly label them as 1 and -1 against forged and authentic respectively.

For evaluation, 10-fold cross-validation is used. The feature vectors of authentic and forged DOCFs are randomly divided into ten-folds of equal size. Corresponding to each fold, a system is trained and tested. Nine-folds of authentic and forged frames are used for training and the leftover fold are used for testing. The average results of 10-folds are taken. SVM parameters are optimized with the training set and finally trained model is used for computing the predictions of test dataset. 50 authentic and 50 forged videos are divided into video segments for calculating the video accuracy.

.

Algorithm 2: Training of Video Forgery Detection Decision Model

Input: X is the set of forged/tampered DOCFs, *Y* is the set of authentic DOCFs, *c* and *g* (gamma) are the parameters of SVM with RBF kernel

Output: Trained classification model SVM

Procedure:

1 *for* each DOCFs in *X*,
  Create features vector (*f.v*) of each forged DOCFs using Algorithm 1
  $(f.v)_T = f.v$
  Assigned label 1 to $(f.v)_T$
end for

2 *for* each DOCFs in *Y*,
  Create features vector (*f.v*) of each authentic DOCFs using Algorithm 1
  $(f.v)_A = f.v$
  Assigned label -1 to $(f.v)_A$
end for

3 $D \leftarrow (f.v)_T \cup (f.v)_A$

4 *AC*=0

5 *for c= min to max do*

6 *for g= min to max do*
  Divide *D* into equally *k* folds (*k*=10)
  for i=1 *to k do*
  Train SVM(*c,g*) on *D*-F_i to get SVM_{Model} (*c, g*) % all training data *D* except i^{th} fold F_i
  Test SVM_{Model} (*c, g*) on fold F_i
  Record the *AC(i)* on fold F_i
  end for
  $AvgAC = \frac{1}{k}\sum_{i=1}^{k} AC(i)$ % compute average accuracy on *k* folds
  if (*AvgAC* > *AC*)
  AC=*AvgACC*, c_{fiinal}=c, g_{final}=*g*
  end if
  end for
  end for

7 Fit the SVM (c_{final}, g_{final}) model on training data

TABLE I. DETAIL OF ALL DATASETS

Datasets	`Authentic	Forged	Frame Rate	Resolution	Format				Camera			Length
					AVI	MOV	MP4	WMV	Cannon	Fuji	Nikon	
D1 [37]	100	-	30	320x240	60	40	-	-	41	30	29	Variable
D2 [36]	10	10	30	320x240	20	-	-	-	-	-	-	Variable
D3 [13]	14	6	30	720x480	15	-	-	5	-	-	-	Variable
D4	100 (D1)	16 (D2+D3)	30	720x480	35	-	-	5	-	-	-	Variable
D5 [38]	6	121	Variable	768x576	101	-	36	-	-	-	-	Variable
D6	130 (D1+D2+ D3+D5)	137 (D2+D3+ D5)	Variable		196	40	36	5	41	30	29	Variable

C. Metrics Used for Evaluation

The evaluation measures, namely detection accuracy (DA), true positive rate (TPR), true negative rate (TNR) and video accuracy (VAC) are commonly used to evaluate video forgery detection techniques [1], [12], [19]. Therefore, these measures are used to evaluate the proposed technique. DA accuracy is the proportion of total number of predictions (authentic or forged) that are correctly predicted and is defined as follows:

$$DA = \frac{TP + FP}{TP + FP + FN + TN}, \tag{8}$$

TPR is the proportion of positive cases (i.e., forged frames) that are correctly classified and is calculated using the following equation:

$$TPR = \frac{TP}{TP + FN} \tag{9}$$

TNR or specificity is the proportion of negative cases (i.e., authentic frames) that are correctly classified and is computed by:

$$TNR = \frac{TN}{TN + FP}, \tag{10}$$

where TP, FP, FN, and TN are the numbers of true positive, false positive, false negative and true negative cases, respectively.

The VAC is the proportion of correctly classified video segments (CCFVS) to the total number of video segments (*N*) in a video and is calculated as:

$$VAC = \frac{CCFVS}{N}, \tag{11}$$

D. Statistical Analysis of CCD-DRLBP Descriptor

The discrimination of CCD-DRLBP descriptors have been analyzed by statistical methods. Firstly, the pairwise distance within authentic frames, within forged frames and between the authentic and forged frames is calculated. The analysis of the histograms of these pairwise distances presented in Fig. 7. The pairwise distance within authentic and within forged class is shown in Fig. 7(a) and 7(b) which is approximately between 0.5 and 2 while Fig. 7(c) represents the distance between authentic and forged classes. It is depicted from Fig. 7(c) that the distance is overlapped from 1.2 to 2 with the pairwise distance within authentic and within forged classes. In this way, 6% of the pairwise distance values between authentic and forged classes are overlapped, which can produce wrong classification results. Since, we are using SVM, which does not work, based on distances, therefore, our accuracy is good and this concludes that the CCD-DRLBP based features are statistically significant.

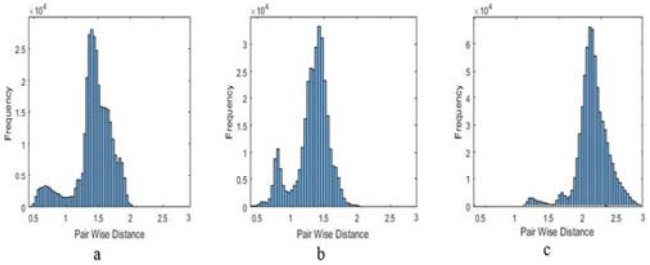

Figure 7. (a) Pairwise distance within authentic class. (b) Within forged class and (c) between authentic and forged class

Furthermore, features are analyzed by calculating the scatter matrix S_w (within authentic frames, within forged frames) and S_B between the authentic and forged frames. Let n samples of m-dimensional data represented by $X = [x_1, x_2, ..., x_n]$ and C is the class labels.

$$S_w = \sum_{i=1}^{C} \sum_{x \in D_i} (x - m_i)(x - m_i)^T, \tag{12}$$

$$S_B = \sum_{i=1}^{C} n_i (m - m_i)(m - m_i)^T, \tag{13}$$

where D_i is the i^{th} class, m_i is the class mean, m the overall mean, C the number of classes, n_i number of items in class D_i, T is matrix transpose and D is the set of all classes. The trace values S_w (within authentic and forged frames) and S_B (between the authentic and forged frames) are computed (see Table II). The trace values S_B is less while S_B are higher of all the datasets which is an evidence that the features extracted based on CCD-DRLBP are most discriminant in nature. The detail of datasets D2, D3, D4, D5, and D6 is given in section III (A).

TABLE II. TRACES OF S_w AND S_B OF CCD-DRLBP BASED FEATURES ON DIFFERENT DATASETS

Datasets	S_w	S_B
D2	1.35	2.81
D3	1.44	2.37
D4	1.24	2.34
D5	1.33	2.22

E. Parameter Tuning

The method contains different parameters like components of YC_bC_r system, the number of neighboring pixels (P), radius (R), type of mapping, number of blocks and the parameters values of RBF kernel (C & γ). Tuning of these parameters is essential for the better performance of the method. The results shown in Fig. 8 indicate that the best accuracy achieved when number of blocks =5x5, P = 16, R =2, uniform mapping=$u2$, $C = 2^{-1}$, $\gamma = 2^{1}$, and C_b & C_r components are fused. The optimal values of the parameters are shown in Table III.

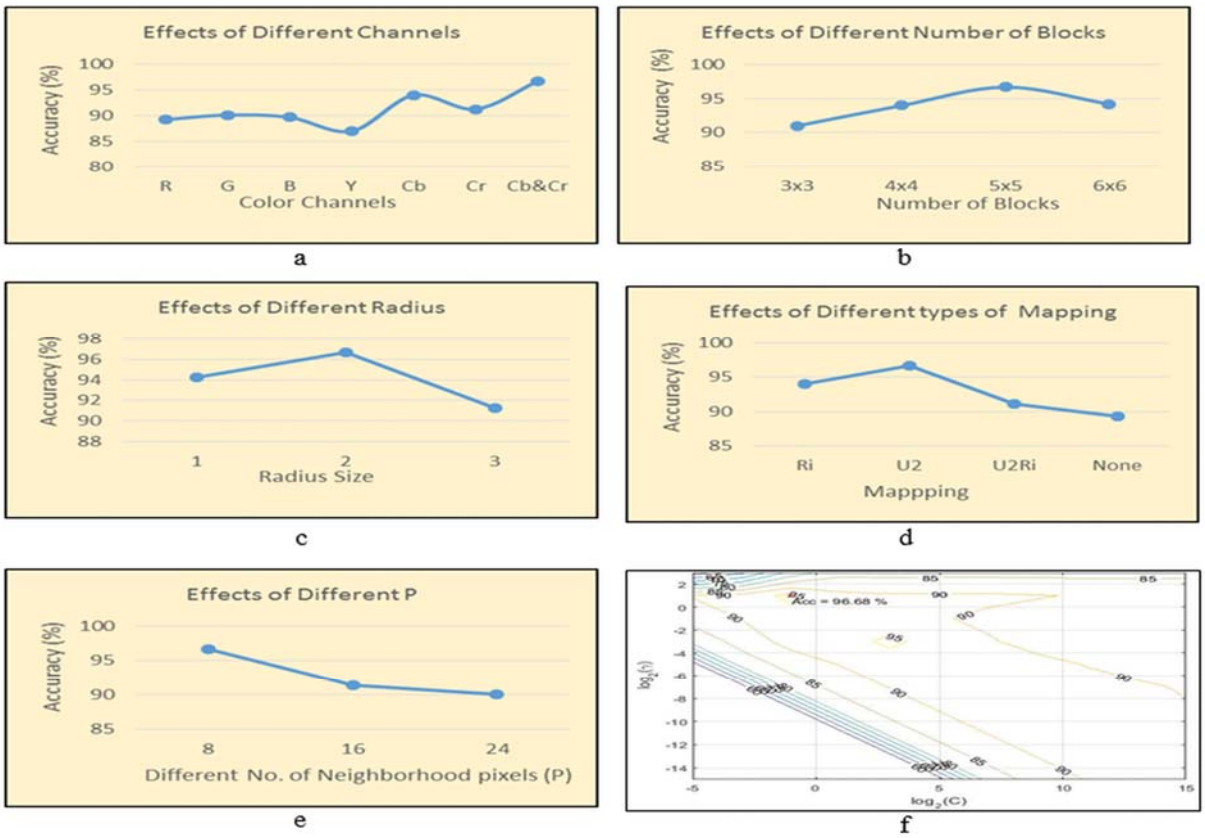

Figure 8. Effect of different parameters on Accuracy (%). (a) Effect of different channels. (b) Effects of different block numbers. (c) Effects of different radius . (d) Effects of different types of mapping. (e) Effects of different number of neighborhood pixels . (f) RBF kernel parameters (&) optimization process

TABLE III. THE OPTIMAL VALUES OF DIFFERENT PARAMETERS

Pre-Processing		**Feature Extraction**			**Classification using SVM**
Color Channel/s	**# of Blocks**	**P**	**R**	**Mapping Type**	
C_b , C_r	5x5	16	2	u2	RBF kernel with $C = 2^{-1}$, $\gamma = 2^{1}$

IV. RESULTS AND DISCUSSION

To exhibit the usefulness of the method, detailed discussion on its effectiveness on available datasets, various geometric transformations, different resolutions, comparison with state-of-the-art methods, and cross data validation is presented in the following sub-sections.

A. Effectiveness on Different Datasets

The DA achieved on different datasets is 93.4%, 87.7%, 91.6%, 95.7% and 96.68% while VAC obtained 94.36%, 90.48%, 93.23%, 96.37% and 98.32% on D2, D3, D4, D5 and D6 respectivly. The method obtained less DA on dataset D3 due to less number of forged videos available for training. Similarly, method achieved 95.7% accuracy on the dataset D5 which are forged by splicing. Table IV and Fig. 9 depicted that the method gives better DA 96.68% and VAC 98.32% on the dataset D6, due to larger number of authentic and forged frames of the videos. By comparing the results of different datasets given in Table IV, it is found that the accuracy for copy-move forgery detection is less than splicing detection. This is due to the reason that in copy-move forgery, the pasted object belongs to the same frame and the change in texture is not so prominent, while in splicing the pasted object comes from different frames or videos.

B. Effectiveness on Different Geometric Transformations and Post-Processing Operations

It is very difficult to detect a tampered object in a frame of a video, if it is manipulated using geometric transformations such as shearing, scaling, rotation, flipping and post-processing operations like luminance and RGB. The proposed method was evaluated by an individual or a combination of geometric transformations and gave 95.7% accuracy on the dataset D5 as described in Table IV. The results of individual and combination of geometric

transformations are shown in Fig. 10. The said figure reveals that the least accuracy of 87.59% is achieved against flipping while an accuracy of 93.97% is achieved against shearing. The best accuracy is achieved which is 95.7% when all the transformations and post-processing operations are grouped. The method has almost the same accuracy on each individual transformation and has achieved significant accuracy on their combination. The fluctuation in the trend might be due to the different number of forged frames in each transformation. When forgery is made with different types of geometric transformations and post-processing operations, then, to catch such forgery more discriminative features are required. CCD-DRLBP yields more discriminant features using edge and texture information in a single representation. Due to this reason, the best accuracy is achieved.

C. Effectiveness on Different Video Frame Resolution and Format

The method was evaluated on different video resolutions (320×240, 720×480, and 768×576) and formats (AVI, MOV, MP4, and WMV). The results presented in Table IV depict that the method has significant results on different resolutions and formats. The DA on high resolution is 95.7% but on low resolution, it is 87.7%. These results reveal that resolution also impacts on accuracy. However, as the model is trained on different type of resolutions, the proposed architecture yields the best DA 96.68% and VAC 98.32%.

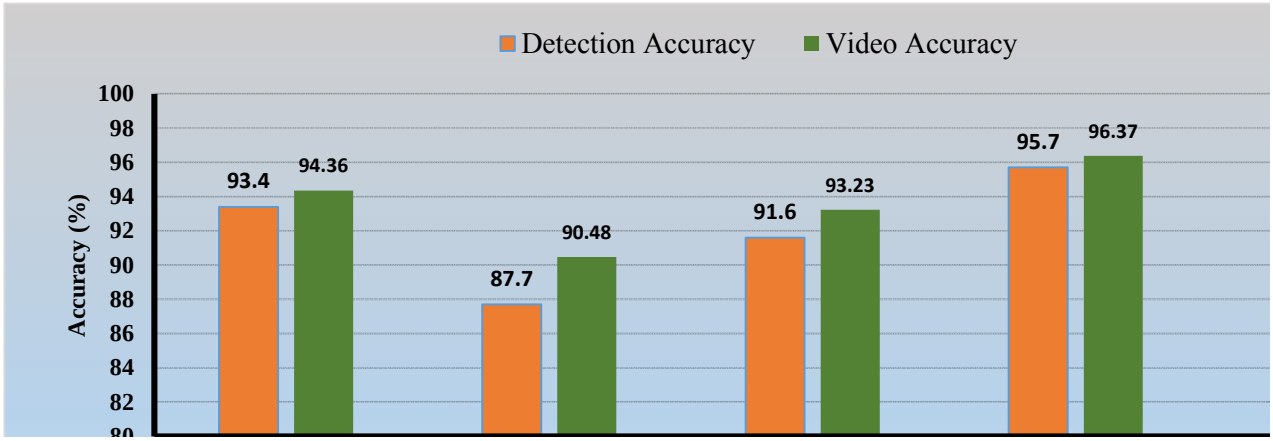

Figure 9. Accuracy (%) of the proposed method on four datasets (D2, D3, D4 and D5) and their combination (D6)

TABLE IV. PERFORMANCE OF THE PROPOSED METHOD ON FOUR DATASETS A ND THEIR COMBINATION

Datasets	TPR (%)	TNR (%)	DA (%)	VAC (%)
D2	92.2	92.6	93.4	94.36
D3	91.65	90.89	87.7	90.48
D4	91.34	89.22	91.6	93.23
D5	95.5	91.6	95.7	96.37
D6	96.5	93.6	96.68	98.32

D. Comparison with State-Of-The-Art Methods

In this section, the results of the proposed technique and the state-of-the-art methods are compared. To the best of our knowledge, only a few methods are available in the literature for object-based forgery detection. The proposed method is compared with the recent schemes [1] and [16] dealing with object-based forgery on dataset D6. The results are shown in Table V. Similarly the methods developed in [1], [9], [16], [25], [40], and [41] are tested on dataset D2 and compared with the proposed technique. The results are shown in Table VI with respect to DA. The proposed technique outperforms the methods developed in [1] and [16]. The method works well against different types of geometric transformations and post-processing operations as shown in Table V. The success of a forgery detection system depends on how accurately it models the structural changes occurring in video frames due to tampering. Lines, edges, and corners are introduced during the process of tampering, which are considered artifacts of forgery. Results are improved because features based on CCD-DRLBP represents these artifacts together with texture information. The method in [1] and [16] claimed accuracy 97.36% and 99.90% on a limited number of videos but when tested on larger videos the results produced less accuracy as shown in Table V.

TABLE V. COMPARISON OF THE PROPOSED ADN THE STATE-OF-THE-ART METHODS ON COMBINED D6 DATASET

Methods	TPR (%)	TNR (%)	DA (%)	VAC (%)	Testing Time (Sec.)
Proposed	96.5	93.6	96.68	98.32	0.1058
Method in [1]	59.42	61.23	63.6	67.36	0.1482
Method in [16]	68.31	70.34	72.67	81.87	0.2898

The accuracy decreases for both methods because the features used in these methods do not provide enough discrimination for all types of forged objects when applied on larger datasets with a variety of forged objects. Since the videos used in [1] and [13] are not publically available, the proposed method cannot be tested on these videos. The testing time of the proposed technique is also less than the other techniques which show that our technique is efficient in terms of time. The detection accuracy of the proposed method is also better than the other methods which are trained and tested on D2 dataset, which is shown in Table VI. The developed technique is invariant to all type of geometric transformations (scaling, rotation, shearing and mirroring) but the other methods are not invariant to

different transformations. The method presented in [25] is only mirroring invariant.

TABLE VI. COMPARISON OD DETECTION ACCURACY (DA) WITH OTHER ALGORITHMS ON D2 DATASET

Methods	DA (%)	Invariant to scaling, rotation, shearing and mirror
Proposed	93.4	All
Method in [1]	59.2	No
Method in [9]	89.7	No
Method in [16]	68.32	No
Method in [25]	92.6	Mirror
Method in [40]	70	No
Method in [41]	90.8	No

E. Cross Dataset Validation

For the success of real applications, the method should also perform well when tested on a completely different dataset, i.e., development of training model on one dataset and testing on another dataset. For cross-dataset validation, the proposed and the state-of-the-art methods have been trained on the D5 dataset and tested on the D4 dataset. Description of the datasets D4 and D5 is presented in Table I and the results are reported in Table VII. The DA and VAC of the proposed system through cross dataset validation is 70.6% and 78.23% respectively, which is better than the state-of-the-art methods. The reason for high accuracy on cross data is that the proposed system trained on multiple resolutions, formats and a large number of videos. Furthermore, the features used in proposed architecture are of more discriminant nature, which is proved by statistical analysis discussed in section III (D) as well.

TABLE VII. CROSS DATASET PERFORMANCE OF THE PROPOSED AND THE STATE-OF-THE-ART METHODS

Methods	TPR (%)	TNR (%)	DA (%)	VAC (%)
Proposed	69.32	67.67	70.6	78.23
Method in [1]	46.54	49.23	52.24	59.45
Method in [16]	55.45	57.46	58.57	62.72

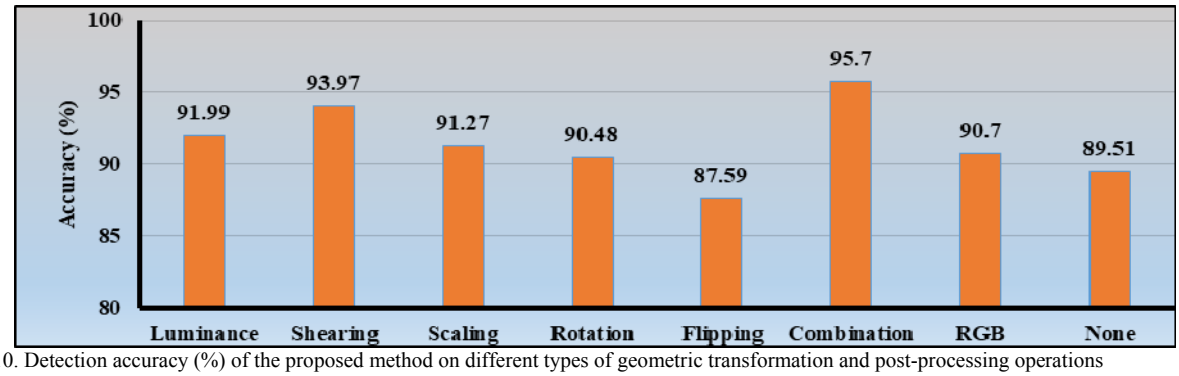

Figure. 10. Detection accuracy (%) of the proposed method on different types of geometric transformation and post-processing operations

V. CONCLUSION AND FUTURE WORK

Detection of tampered video is a challenging task. The state-of-the-art methods face limitations ranging from evaluation of the method on the small number of videos, single video formats, and resolution. Moreover, texture descriptors such as LBP and HOG have been used to represent structural changes occuring due to spatial forgery in video frames. These descriptors are not robust due to their weak representation of local patterns. In this study, discriminant features are extracted through a newly proposed CCD-DRLBP descriptor and SVM is employed to detect the segments of the videos as authentic or forged. The proposed method is also used for localization of forged frames. The system is achieving best DA of 96.68% and VAC of 98.32%. The proposed method works better to detect the spatial (object-based) forgery done by splicing method as compared to copy-move method. The DA and VAC obtained through cross-dataset validation is 70.6% and 78.23% respectively, which is not ideal but much better than existing methods.

The focus of this research is to identify forged videos and locate the forged frames. After the identification of the forged frames, it is desirable to find out which regions of the frames are tampered. Localization of forged regions in the frames is a part of our future work. Moreover, the future investigation will also be on the development of more robust representation of tampering using deep learning [42], extreme learning [43] and transfer learning [44] approaches.

Further research will also be required to enhance the accuracy through cross dataset validation, which is important for reliable and realtime applications. Another direction is the preparation of a benchmark dataset of forged videos for research community, which will be helpful for them to compare their methods with existing methods.

REFERENCES

[1] C. Richao, Y. Gaobo, and Z. Ningbo, "Detection of object-based manipulation by the statistical features of object contour", Forensic science international, vol. 236, pp. 164-169, 2014, doi:10.1016/j.forsciint.2013.12.022

[2] K. Asghar, Z. Habib, and M. Hussain, "Copy-move and splicing image forgery detection and localization techniques: a review", Australian Journal of Forensic Sciences, pp. 1-27, 2016, doi:10.1080/00450618.2016.1153711

[3] R. C. Pandey, S. K. Singh, and K. K. Shukla, "Passive forensics in image and video using noise features: A review", Digital Investigation, vol. 19, pp. 1-28, 2016, doi:10.1016/j.diin.2016.08.002

[4] R. D. Singh and N. Aggarwal, "Video content authentication techniques: a comprehensive survey", Multimedia Systems, pp. 1-30, 2017, doi:10.1007/s00530-017-0538-9

[5] K. Sitara and B. Mehtre, "Digital video tampering detection: An overview of passive techniques", Digital Investigation, vol. 18, pp. 8-22, 2016, doi:10.1016/j.diin.2016.06.003

[6] A. Alahmadi, M. Hussain, H. Aboalsamh, G. Muhammad, G. Bebis, and H. Mathkour, "Passive detection of image forgery using DCT and local binary pattern", Signal, Image and Video Processing, vol. 11, pp. 81-88, 2017, doi:10.1007/s11760-016-0899-0

[7] B. Yang, X. Sun, H. Guo, Z. Xia, and X. Chen, "A copy-move forgery detection method based on CMFD-SIFT", Multimedia Tools and Applications, vol. 77, pp. 837-855, 2018, doi:10.1007/s11042-016-4289-y

[8] P. Aflaki, M. M. Hannuksela, and M. Gabbouj, "Subjective quality assessment of asymmetric stereoscopic 3D video", Signal, Image and Video Processing, vol. 9, pp. 331-345, 2015, doi:10.1007/s11760-013-0439-0

[9] A. Subramanyam and S. Emmanuel, "Video forgery detection using HOG features and compression properties", in Multimedia Signal Processing (MMSP), 2012 IEEE 14th International Workshop on, Banff Center Banff, AB, Canada, 2012, pp. 89-94, doi:10.1109/mmsp.2013.6659337

[10] Z. Guo, L. Zhang, and D. Zhang, "A completed modeling of local binary pattern operator for texture classification", IEEE Transactions on Image Processing, vol. 19, pp. 1657-1663, 2010, doi:10.1109/tip.2010.2044957

[11] J. Ren, X. Jiang, and J. Yuan, "Noise-resistant local binary pattern with an embedded error-correction mechanism", IEEE Transactions on Image Processing, vol. 22, pp. 4049-4060, 2013, doi:10.1109/tip.2013.2268976

[12] J. Zhang, K. Huang, Y. Yu, and T. Tan, "Boosted local structured hog-lbp for object localization", in Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, 2011, pp. 1393-1400, doi:10.1109/cvpr.2011.5995678

[13] C.-C. Hsu, T.-Y. Hung, C.-W. Lin, and C.-T. Hsu, "Video forgery detection using correlation of noise residue", in Multimedia Signal Processing, 2008 IEEE 10th Workshop on, Queensland, Australia, 2008, pp. 170-174, doi:10.1109/mmsp.2008.4665069

[14] G. Singh and K. Singh, "Video frame and region duplication forgery detection based on correlation coefficient and coefficient of variation", Multimedia Tools and Applications, pp. 1-36, 2018, doi:10.1007/s11042-018-6585-1

[15] L. Su, T. Huang, and J. Yang, "A video forgery detection algorithm based on compressive sensing", Multimedia Tools and Applications, vol. 74, pp. 1-16, 2014, doi:10.1007/s11042-014-1915-4.

[16] S. Chen, S. Tan, B. Li, and J. Huang, "Automatic detection of object-based forgery in advanced video", IEEE Transactions on Circuits and Systems for Video Technology, vol. 26, pp. 2138-2151, 2016, doi:10.1109/tcsvt.2015.2473436

[17] M. Kobayashi, T. Okabe, and Y. Sato, "Detecting video forgeries based on noise characteristics," in Advances in Image and Video Technology. vol. 5414, ed: Springer, 2009, pp. 306-317, doi:10.1007/978-3-540-92957-4_27

[18] D.-K. Hyun, S.-J. Ryu, H.-Y. Lee, and H.-K. Lee, "Detection of upscale-crop and partial manipulation in surveillance video based on sensor pattern noise", Sensors, vol. 13, pp. 12605-12631, 2013, doi:10.3390/s130912605

[19] R. D. Singh and N. Aggarwal, "Detection of upscale-crop and splicing for digital video authentication", Digital Investigation, vol. 21, pp. 31-52, 2017, doi:10.1016/j.diin.2017.01.001

[20] R. C. Pandey, S. K. Singh, and K. Shukla, "Passive copy-move forgery detection in videos", in Computer and Communication Technology (ICCCT), 2014 International Conference on, 2014, pp. 301-306, doi:10.1109/iccct.2014.7001509

[21] J. Goodwin and G. Chetty, "Blind video tamper detection based on fusion of source features", in Digital Image Computing Techniques and Applications (DICTA), 2011 International Conference on, 2011, pp. 608-613, doi:10.1109/dicta.2011.108

[22] A. Bidokhti and S. Ghaemmaghami, "Detection of regional copy/move forgery in MPEG videos using optical flow", in Artificial intelligence and signal processing (AISP), 2015 International symposium on, 2015, pp. 13-17, doi:10.1109/aisp.2015.7123529

[23] C. Guo, G. Luo, and Y. Zhu, "A detection method for facial expression reenacted forgery in videos", in Tenth International Conference on Digital Image Processing (ICDIP 2018), 2018, p. 108061J, doi:10.1117/12.2502817

[24] O. I. Al-Sanjary, A. A. Ahmed, A. A. B. Jaharadak, M. A. Ali, and H. M. Zangana, "Detection clone an object movement using an optical flow approach", in 2018 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE), 2018, pp. 388-394, doi:10.1109/iscaie.2018.8405504

[25] L. Su and C. Li, "A novel passive forgery detection algorithm for video region duplication", Multidimensional Systems and Signal Processing, vol. 29, pp. 1173-1190, 2018, doi:10.1007/s11045-017-0496-6

[26] L. Li, X. Wang, W. Zhang, G. Yang, and G. Hu, "Detecting removed object from video with stationary background", in The International Workshop on Digital Forensics and Watermarking 2012, 2013, pp. 242-252, doi:10.1007/978-3-642-40099-5_20

[27] V. Conotter, J. F. O'Brien, and H. Farid, "Exposing digital forgeries in ballistic motion", ieee transactions on information forensics and security, vol. 7, pp. 283-296, 2012, doi:10.1109/tifs.2011.2165843

[28] M. Zampoglou, F. Markatopoulou, G. Mercier, D. Touska, E. Apostolidis, S. Papadopoulos, R. Cozien, I. Patras, V. Mezaris, and I. Kompatsiaris, "Detecting Tampered Videos with Multimedia Forensics and Deep Learning", in International Conference on Multimedia Modeling, 2019, pp. 374-386, doi:10.1007/978-3-030-05710-7_31

[29] Y. Yao, Y. Shi, S. Weng, and B. Guan, "Deep learning for detection of object-based forgery in advanced video", Symmetry, vol. 10, p. 3, 2017, doi:10.3390/sym10010003

[30] A. Satpathy, X. Jiang, and H.-L. Eng, "LBP-based edge-texture features for object recognition", Image Processing, IEEE Transactions on, vol. 23, pp. 1953-1964, 2014, doi:10.1109/tip.2014.2310123

[31] G. Muhammad, M. H. Al-Hammadi, M. Hussain, and G. Bebis, "Image forgery detection using steerable pyramid transform and local binary pattern", Machine Vision and Applications, vol. 25, pp. 985-995, 2014, doi:10.1007/s00138-013-0547-4

[32] X. Zhao, S. Li, S. Wang, J. Li, and K. Yang, "Optimal chroma-like channel design for passive color image splicing detection", EURASIP Journal on Advances in Signal Processing, vol. 2012, pp. 1-11, 2012, doi:10.1186/1687-6180-2012-240

[33] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection", in Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, 2005, pp. 886-893, doi:10.1109/cvpr.2005.177

[34] C. Cortes and V. Vapnik, "Support-vector networks", Machine learning, vol. 20, pp. 273-297, 1995, doi:10.1007/BF00994018

[35] J. Y. Hesterman, L. Caucci, M. A. Kupinski, H. H. Barrett, and L. R. Furenlid, "Maximum-likelihood estimation with a contracting-grid search algorithm", IEEE transactions on nuclear science, vol. 57, pp. 1077-1084, 2010, doi:10.1109/tns.2010.2045898

[36] P. Bestagini, S. Milani, M. Tagliasacchi, and S. Tubaro, "Local tampering detection in video sequences", in Multimedia Signal Processing (MMSP), 2013 IEEE 15th International Workshop on, Pula (CA), Italy, 2013, pp. 488-493, doi:10.1109/MMSP.2013.6659337

[37] G. Qadir, S. Yahaya, and A. T. Ho, "Surrey university library for forensic analysis (SULFA) of video content", in Image Processing (IPR 2012), IET Conference on, London, UK, 2012, pp. 1-6, doi:10.1049/cp.2012.0422.

[38] E. Ardizzone and G. Mazzola, "A Tool to Support the Creation of Datasets of Tampered Videos," in Image Analysis and Processing—ICIAP 2015, ed: Springer, 2015, pp. 665-675., doi:10.1007/978-3-319-23234-8_61

[39] G. Jin and X. Wan, "An improved method for SIFT-based copy-move forgery detection using non-maximum value suppression and optimized J-Linkage", Signal Processing: Image Communication, 2017, doi:10.1016/j.image.2017.05.010

[40] C.-M. Pun, X.-C. Yuan, and X.-L. Bi, "Image forgery detection using adaptive oversegmentation and feature point matching", ieee transactions on information forensics and security, vol. 10, pp. 1705-1716, 2015, doi:10.1109/tifs.2015.2423261

[41] W. Wang and H. Farid, "Exposing digital forgeries in video by detecting duplication", in Proceedings of the 9th workshop on Multimedia & security, New York, USA, 2007, pp. 35-42, doi:10.1145/1288869.1288876

[42] L. Shiqi, T. Shengwei, Y. Long, Y. Jiong, and S. Hua, "Android malicious code Classification using Deep Belief Network", KSII Transactions on Internet & Information Systems, vol. 12, 2018, doi:10.3837/tiis.2018.01.022

[43] X. Liu, S. Lin, J. Fang, and Z. Xu, "Is extreme learning machine feasible? A theoretical assessment (Part I)", Neural Networks and Learning Systems, IEEE Transactions on, vol. 26, pp. 7-20, 2015, doi:10.1109/tnnls.2014.2335212

[44] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning", Journal of Big Data, vol. 3, pp. 1-40, 2016, doi:10.1186/s40537-016-0043-6