Robust Video Content Authentication using Video Binary Pattern and Extreme Learning Machine

Mubbashar Sadddique[1], Khurshid Asghar[2], Tariq Mehmood[3], Muhammad Hussain[4], Zulfiqar Habib[5]
Department of Computer Science, COMSATS, University Islamabad (Lahore Campus), Lahore, 54000, Pakistan[1, 5]
Department of Computer Science, University of Okara, Okara, 56300, Pakistan[2]
Department of Computer Science & Information Technology[3]
Superior University, Lahore, 54000, Pakistan
Department of Computer Science[4]
King Saud University, Riyadh
Saudi Arabia

***Abstract*—Recently, due to easy accessibility of smartphones, digital cameras and other video recording devices, a radical enhancement has been experienced in the field of digital video technology. Digital videos have become very vital in court of law and media (print, electronic and social). On the other hand, a widely-spread availability of Video Editing Tools (VETs) have made video tampering very easy. Detection of this tampering is very important, because it may affect the understanding and interpretation of video contents. Existing techniques used for detection of forgery in video contents can be broadly categorized into active and passive. In this research a passive technique for video tampering detection in spatial domain is proposed. The technique comprises of two phases: 1) Extraction of features with proposed Video Binary Pattern (VBP) descriptor, and 2) Extreme Learning Machine (ELM) based classification. Experimental results on different datasets reveal that the proposed technique achieved accuracy 98.47%.**

***Keywords*—Video forgery; spatial video forgery; passive forgery detection; Video Binary Pattern (VBP); feature extraction**

I. Introduction

In these days, digital video making has become very handy with the accessibility of video recording gadgets such as smartphones and digital cameras [2, 1]. These videos are an important part of our daily routine and also an important source of information. Digital videos present some of the most convincing documentary evidence to establish the truthfulness or falsehood of an issue under consideration, which is acceptable both inside and outside the court of law.

A few years back, digital videos were considered reliable proof, but a widespread availability as well as accessibility of easy-to-use video editing tools (VETs) such as (Pinnacle Studio 20 Ultimate, Adobe Premier Pro, Lightworks and Cinelerra, etc.) [21, 19], have negated this fact. Even a novice user can alter the contents of digital videos in such a manner that it is not possible to distinguish between the original and forged contents of a video with the naked eye. On one hand, video editing is a very useful and important tool for manipulating video scenes in film industry. On the other hand, it enables to forge video contents to distort the evidences for a court and propaganda on social, print and electronic media Therefore, authenticity of a video is a key issue when it is presented in a court as a prof of a crime [12].

Digital video forgery techniques are categorized into temporal, spatial, and spatio-temporal. In spatial category, digital videos are forged by changing contents within the frame(s) which modifies visual information. The object is taken from one location of a video frame and inserted on another place in the same frame or in other frame after some alteration [17]. This category consists of upscale-crop [15], copy-move [18] and splicing video forgery [5].

Temporal tampering (forgery) is done by removing, duplicating or inserting the number of frames from / in a digital video. Both object and frame level forgery is done in spatio-temporal category. Existing tampering detection techniques in digital videos are divided into active and passive. Active techniques need pre-embedded data such as watermarking, digital signatures, etc., whereas passive techniques do not depend on any pre-embedded information. Passive techniques are also called blind techniques. Fig. 1 shows categorization of video forgery techniques.

Various passive techniques are proposed to detect spatial video forgery which are not equally efficient for different datasets, static and moving objects. In this research, a robust video content authentication technique is presented. This paper is structured as follows: Section II explains related work; Section III describes a step-by-step research methodology used for development of proposed technique; datasets and performance evaluation parameters are described in Section IV; and experimental work is presented in Section V. Conclusion and future directions are presented in Section VI in the end of this paper.

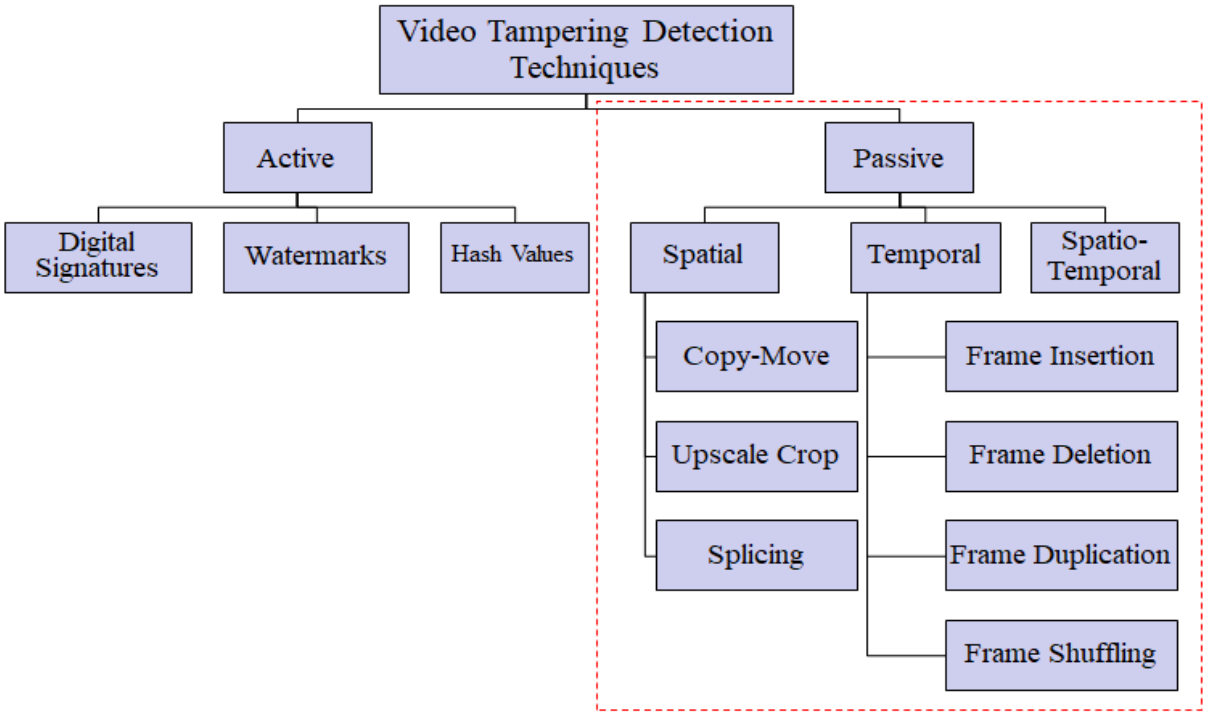


Fig. 1. Categorization of Video Tampering Techniques.

II. Related Work

Existing techniques for tampering detection in spatial domain can generally be classified as: 1) Statistical Based, 2) Compression Based, 3) Texture Based, and 4) Noise Based. Fig. 2 describes types of video forgery detection techniques. Each technique is briefly described in the following discussion.

A. Methods based on Statistical Features

Texture, tone, and context are always present in any frame (image). A texture is an important property, which is disturbed during the process of video forgery. The statistical features are used for representation of the texture [29]. Many researchers used statistical features for detection of object-based video forgery [10, 3, 2]. Singh et al. [2] exploited DCT and correlation coefficients to detect the duplicated regions. The method achieved accuracy 99.5% and 96.6% for detection of duplicated frames and regions respectively. This method cannot detect less number of duplicated frames and is not able to detect small duplicated regions. Richao et al. [10] used statistical features to detect forgery. Authors calculated the first four moments of the wavelet and average gradient of each color component. SVM classifier is applied for classification between original and forged video. Twenty videos with resolution 320 x 240 were used for an experiment. The accuracy and area under the curve (AUC) are 95% and 0.948 respectively. The receiver operating characteristic (ROC) curve showed a classification result of 85.45%. The results are tested on a limited dataset and not experimented on different compression ratios. Su et al. [3] proposed an algorithm based on exponential Fourier moments (EFMs) to detect the duplicate region and adaptive parameter based compression tracking algorithm is used to localize the tampered region. The detection accuracy 93.1% is achieved.

B. Methods based on Compression

Su et al. [9] proposed a new algorithm based on compressive sensing for video forgery detection. When the moving-objects deleted from a video frames, the traces (lines, edges, and corners) around the object are also altered. The compressive sensing is employed to perceive this tampering. The K- Singular Value Decomposition (K-SVD) was used to obtain and analyze difference between frames. The detection results of each frame were combined to obtain result. This technique has more compliance in problem solving and is more easy to use as compared to another similar technique proposed in [23]. However, the proposed system does not perform well for a very small deleted foreground.

A technique of forgery localizing in MPEG-2 videos was presented by Labartino et al. [14]. The proposed method first discovers twice intra-coded frames and then applies a double quantization analysis based on MPEG-2. The technique exploited the properties of MPEG-2 coding. This novel technique encoded by utilizing P-frames for analysis of video to apply double quantization. A well-known video dataset was used to carry out experiments having varied scenes having 720x576 resolution.

C. Methods based on Texture

Tamura texture features are exploited for detection of copy-move video forgery by Liao et al. [13], these features are utilized to localize the copy-move tampering. The method was verified on a dataset having 10 videos, which are captured with fixed and moving camera. The resolution of videos was 640 × 480 and a frame rate of 25 to 30 frames per second (fps). The results reveal 99.96% precision, which is comparatively higher than previous appropriate research. However, computation time of the method is much higher.

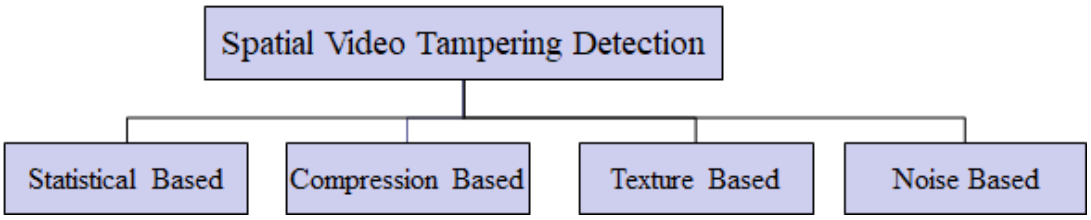


Fig. 2. Types of Techniques for Video Tampering Detection in Spatial Domain.

Subramanyam et al. [17] detected the spatial forgery by using compression and Histogram of Oriented Gradients (HOG) features. In this study, the authors used 6000 frames from 15 different videos with tampered regions of size 40x40, 60x60 and 80x80 in the same and different frames. Detection accuracy (DA) is 80%, 94% and 89% for 40x40, 60x60 and 80x80 blocks, respectively. This algorithm detected spatial forgery very well, but the training and testing are performed on a very limited dataset. The algorithm fails to detect tampering when different post-processing operations (rotation and scaling) are exercised to forge the regions and cannot localize the forged regions.

D. Methods based on Noise Characterisics

Noise variations between an authentic and tampered video sequence are exploited to detect forgery. The photon shot noise in digital camera was exploited by Kobayashi et al. [24]. The test was performed on gray scale videos recorded on 30 fps and 640×480 resolution, Huff-yuv, a lossless compression codec was used for compression. The experiments were performed on videos recorded from static scene only.

A bottom-up approach based on noise correlation was used by Hsu et. al. [25] to locate the forged / in-painted regions of a video. The noise residual is calculated by subtracting the noise free video frames and original frames. Wavelet de-noising filter is used to obtain noise free video frames. Then, every video frame was divided into non-overlapping $N \times N$ blocks. Then correlation of the noise residual was calculated between successive frames. Lastly, forged blocks are located by analysis of statistical features. Content dependency of the noise residual made correlation feature unstable for applications for moving cameras.

III. Proposed Methodology

In this section, a robust video content authentication technique in spatial domain is presented. The proposed methodology consists of two stages (see Fig. 3). (1) Feature extraction, and (2) classification based on ELM**, e**ach stage is described in the following discussion.

A. Feature Extraction

Feature extraction, a first step of proposed methodology, plays a very crucial role to distinguish whether a video is authentic or forged. These isolate important characteristics in the video. These unique characteristics of an authentic video do not exist in the forged video. Existing techniques for feature extraction have a high computational cost. Moreover, normally, features are extracted frame-wise from a video by taking these frames as static images. Therefore, no temporal correlation exists between these features, which are very important in video forgery detection. Importance of features demands the need of a descriptor that extracts features, which are, not only discriminate, but also has temporal correlation. These features are helpful in classification of original and forged videos. A descriptor is proposed to calculate the features named as Video Binary Pattern (VBP) and processes of a proposed descriptor are described below.

- As a first step, digital video is converted into '*N*' number of slices. The number '*N*' depends on '*Height (H)*' of video frames. For example, a digital video consisting of frames with image height H = 240 is converted into 240 slices. By converting video into slices, temporal correlation between the frames is accessed. Fig. 4 describes the process for slicing of digital video.
- Feature extraction from such a large number of slices is time consuming. To minimize the computation time, the number of slices are reduced in a manner that they do not affect classification, whatever the resolution of the video. We extract average of these slices using following formula:

$$X = \frac{N \text{ number of Slices}}{Y}, \quad (1)$$

where Y is 10 in our experiments.

$$\text{Average Slice, } W = \frac{\text{Sum of X Number of Slices}}{X}, \quad (2)$$

- The desired features are then extracted from these average slices using Local Binary Pattern (LBP) and are represented as:

  F_1 = LBP (S_1), F_2= LBP (S_2),….. F_n = LBP (S_N), where '*N*' and 'n' are indices of slice numbers and feature vectors, respectively.

- *n* Features of all slices are concatenated to make a final vector (V) which is represented as:

$$V = \text{Concatenate } (F_1, F_2, F_3, \ldots, F_n) \quad (3)$$

B. Classification based on ELM

Classification is a process of categorizing groups of data based on similarities. The classification models attempt to extract some decisions from observed data. When an input is given to classification model (trained model), it will predict and categorize the given value into one or more groups. For example, when a feature vector is given as input to the trained model, it will label as authentic or forged. Different models are available for classification, for example, support vector machine (SVM) [28], ELM [20], Decision Trees and Naïve Bayes [11], etc. The proposed descriptor was trained and tested using ELM. It has superiority in computational speed as

compared with other machine learning algorithms such as Decision Trees and Naïve Bayes etc. The proposed descriptor was trained and tested using ELM. It has a single hidden layer feedforward neural network and has superiority in computational speed as compared with other machine learning algorithms and, being a mathematical model, its implementation is easy. We have experimented different classifiers such as J48 which is used to generate decision trees, Naïve Bayes based on Bayes' theorem with strong independence assumptions between features, Multiclass classifier [27] which classifies instances into one of three or more classes, SVM a supervised learning model, simple ELM [26] feedforward neural networks based classifier and Kernel ELM [22] where the number of neurons is decided by the classifiers itself. The proposed technique finds best result with kernel ELM using RBF kernel.

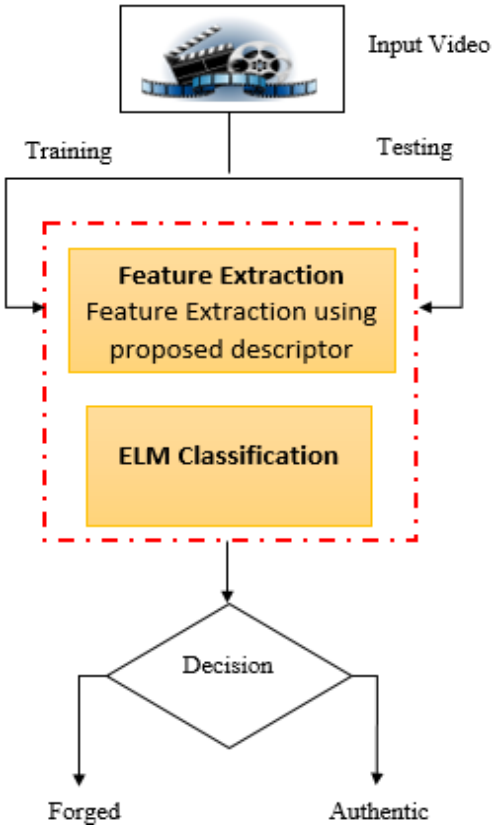


Fig. 3. Proposed Methodology.

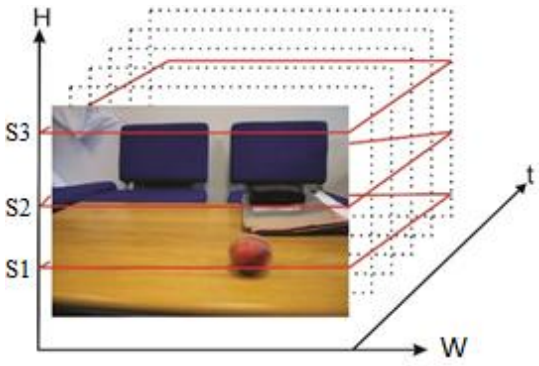


Fig. 4. Model for Slicing of Digital Video.

IV. Experimental Results

The performance of proposed technique was measured on different datasets taken from Hsu et al. [25], Bestaigini et al. [16], Ariddizone and Mazzola [8], and Sanjary et al. [6] which are summarized in Table I. Sample frames taken from authentic and forged video sequences of these datasets are shown in Fig. 5 and 6 which describe different objects removed, inserted or duplicated to forge digital videos. The elephant object is inserted at frame No. 250 in forged frame (see Fig. 5(f)) to forge the video content. Similarly the car object is copied and inserted in the same video. The forged frames (50 and 375) of forged video are shown in Fig. 6(d) and 6(f).

The performance of proposed technique is evaluated by different evaluation parameters same as discussed by Richao et al. [10]. A frame is labelled as true positive (TP), if a forged frame is also classified as forged. Otherwise, if the frame is classified as an original, then it is labelled as false positive (FP). A frame is labelled as true negative (TN), if an original frame is also classified as original. Otherwise, the frame is labelled as false negative (FN). Accuracy (AR) is the ratio of sum of TPs and TNs to the sum of TPs, TNs, FNs and FPs. True positive rate (TPR) performance parameters is a ratio of total TPs to the sum of total TPs and FPs.

False positive rate (FPR) is a ratio of total FPs to the sum of total TNs and FNs. AR, TPR and FPR are represented by the following equations:

$$AR = \frac{TP+TN}{TP+TN+FN+FP} \quad (4)$$

$$TPR = \frac{TP}{TP+FP} \quad (5)$$

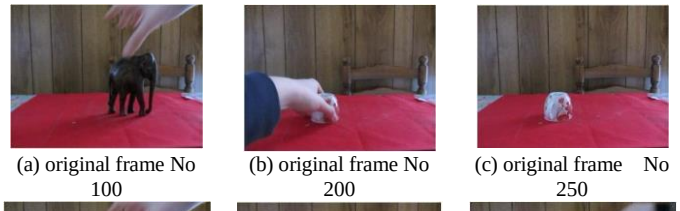
$$FPR = \frac{FP}{TN+FN} \quad (6)$$

Experiments were performed on Intel ® Core ™ i5-2400 CPU @ 3.10GHz, 64-bit Windows operating system with 4GB RAM using MATLAB version R2018a. Experimental results of proposed technique tested on DS1 to DS5 are shown in Table II. The results reveal that the proposed technique demonstrates higher accuracy rate 98.45% using ELM Kernel Classifier on data set DS5 whereas lowest on data set DS2. The accuracy rate is lowest on DS2 because this dataset has less video sequences (10 videos). The classifier, on less data learned specific pattern but generalization of this pattern was not possible. Table III demonstrates effects of different classifiers on accuracy rate using proposed techniques. The best accuracy is achieved using Kernel ELM classifier. Kernel ELM classifier decides by itself the number of neurons to be set.

The performance of our proposed technique is compared with two other existing state-of-the-art techniques proposed by Chen et al. [4] and Richao et al. [10] in spatial domain. Chen et al. claimed accuracy of 99.9% whereas Richao et al. claimed 97.36% accuracy on limited video dataset. We implemented these techniques and obtained results on dataset DS5, because their datasets are not publicly available. The comparison is presented in Table IV. It can be seen from Table IV that our proposed technique outperforms the other techniques with AR = 98.47%, TPR = 98.50% and FPR = 98.37%. Moreover, performance of both the techniques reduced significantly on the proposed datasets.

TABLE. I. DETAIL OF DATA SETS USED FOR EXPERIMENTS

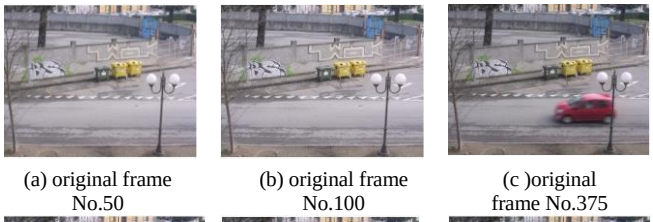
Datasets	Types of forgery	Authentic	Forged	Frame Rate	Static/Moving Camera	Resolution	Format			Length (Sec.)
							AVI	MP4	WMV	
DS1 [16]	Copy-move	10	10	30	Static	320x240	20	-	-	7-19
DS2 [25]	Splicing	14	6	30	Moving	720X480	15	-	5	3-17
DS3 [8]	Copy-move & Splicing	6	121	25-30	Static & Moving	768X576	101	26	-	2-16
DS4 [6]	Copy-move & Splicing	20	20	29	Static & Moving	720x1280	-	40	-	14-15
DS5 (DS1+DS2+DS3+DS4)	Variable	50	157	25-30		Variable	136	66	5	2-19



(a) original frame No 100 (b) original frame No 200 (c) original frame No 250

(d) forged frame No 100 (e) forged frame No 200 (f) forged frame No 250

Fig. 5. Sample of Frames (a)-(c) Taken from Original Videos of DS1 and (d)-(f) from Forged Videos of DS1 [16].



(a) original frame No.50 (b) original frame No.100 (c)original frame No.375

(d) forged frame No. 50 (e) forged frame No. 100 (f) forged frame No. 375

Fig. 6. Sample of Frames (a)-(c) Taken from Original Video and (d)-(f) Taken from Forged Video [16].

TABLE. II. PERFORMANCE OF PROPOSED TECHNIQUE USING ELM

Dataset	AR (%)	TPR (%)	FPR (%)
DS1	96.23	97.51	12.23
DS2	92.56	96.20	30.12
DS3	93.83	92.35	25.67
DS4	97.47	96.50	21.32
DS5	98.47	97.10	14.23

TABLE. III. EFFECTS OF DIFFERENT CLASSIFIERS ON ACCURACY

Classifier	AR (%)	TPR (%)	FPR %)
J48	82.35	82.4	43.34
Naïve Bayes	66.45	66.74	36.12
SVM	60.89	60.30	32.67
Multiclass classifier	83.08	83.34	40.56
Simple ELM	97.45	96.10	19.45
Kernel ELM	98.47	97.10	14.23

TABLE. IV. COMPARISON OF PROPOSED TECHNIQUE WITH STATE OF ART

Methods	AR (%)	TPR (%)	FPR (%)
Proposed Technique	98.47	97.10	14.23
Method proposed by Chen et al. [7]	72.67	68.31	29.66
Method proposed by Richao et al. [10]	63.6	59.42	38.77

V. CONCLUSION

Detection of forgery in a video is a challenging task, because it substantially affects content of the video. In this paper, we proposed a two-stage technique (feature extraction and classification) for forgery detection in spatial domain. For features extraction a descriptor Video Binary Pattern (VBP) is proposed to extract features from average slices of videos and ELM classifier is used for detection and classification of video forgery. Experimental results on different datasets reveal that the proposed technique achieved accuracy rate 98.47% using ELM classifier. The technique is also robust to different formats and variety of datasets.

Further research will also be required to enhance the accuracy through cross dataset validation, which is important for reliable and real time applications.

REFERENCES

[1] 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, 10.1007/978-3-030-05710-7_31.

[2] G. Singh and K. Singh, "Video frame and region duplication forgery detection based on correlation coefficient and coefficient of variation", Multimedia Tools and Applications, vol. 78, pp. 11527-11562, 2019, 10.1007/s11042-018-6585-1.

[3] L. Su, C. Li, Y. Lai, and J. Yang, "A Fast Forgery Detection Algorithm Based on Exponential-Fourier Moments for Video Region Duplication", IEEE Transactions on Multimedia, vol. 20, pp. 825-840, 2018

[4] 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.org/10.1109/tcsvt.2015.2473436

[5] 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.org/10.1080/00450618.2016.1153711

[6] O. I. Al-Sanjary, A. A. Ahmed, and G. Sulong, "Development of a video tampering dataset for forensic investigation", Forensic science international, vol. 266, pp. 565-572, 2016

[7] S. Chen, S. Tan, B. Li, and J. Huang, "Automatic Detection of Object-based Forgery in Advanced Video", 2015

[8] E. Ardizzone and G. Mazzola, "A tool to support the creation of datasets of tampered videos", in International Conference on Image Analysis and Processing, 2015, pp. 665-675, doi.org/10.1007/978-3-319-23234-8_61

[9] 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, 10.1007/s11042-014-1915-4.

[10] 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.org/10.1016/j.forsciint.2013.12.022

[11] T. R. Patil and S. Sherekar, "Performance analysis of Naive Bayes and J48 classification algorithm for data classification", International journal of computer science and applications, vol. 6, pp. 256-261, 2013

[12] Z. Parmar and S. Upadhyay, "A Review on Video/Image Authentication and Temper Detection Techniques", International Journal of Computer Applications, vol. 63, 2013

[13] S.-Y. Liao and T.-Q. Huang, "Video copy-move forgery detection and localization based on Tamura texture features", in Image and Signal Processing (CISP), 2013 6th International Congress on, Hangzhou, China, 2013, pp. 864-868

[14] D. Labartino, T. Bianchi, A. De Rosa, M. Fontani, D. Vazquez-Padin, A. Piva, and M. Barni, "Localization of forgeries in MPEG-2 video through GOP size and DQ analysis", in Multimedia Signal Processing (MMSP), 2013 IEEE 15th International Workshop on, 95 Pula (CA), Italy, 2013, pp. 494-499

[15] 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

[16] 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, 10.1109/MMSP.2013.6659337.

[17] 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.org/10.1109/mmsp.2013.6659337

[18] V. S. Pujari and M. Sohani, "A Comparative Analysis On Copy Move Forgery Detection Using Frequency Domain Techniques", International Journal of Global Technology Initiatives, vol. 1, pp. E104-E111, 2012

[19] S. Milani, M. Fontani, P. Bestagini, M. Barni, A. Piva, M. Tagliasacchi, and S. Tubaro, "An overview on video forensics", APSIPA Transactions on Signal and Information Processing, vol. 1, p. e2, 2012, doi.org/10.1017/atsip.2012.2

[20] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification", IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 42, pp. 513-529, 2012

[21] A. Rocha, W. Scheirer, T. Boult, and S. Goldenstein, "Vision of the unseen: Current trends and challenges in digital image and video forensics", ACM Comput. Surv., vol. 43, pp. 1-42, 2011, 10.1145/1978802.1978805.

[22] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification", IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 42, pp. 513-529, 2011

[23] J. Zhang, Y. Su, and M. Zhang, "Exposing digital video forgery by ghost shadow artifact", in Proceedings of the First ACM workshop on Multimedia in forensics, 2009, pp. 49-54

[24] 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.

[25] 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.org/10.1109/mmsp.2008.4665069

[26] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: theory and applications", Neurocomputing, vol. 70, pp. 489-501, 2006

[27] T. Li, C. Zhang, and M. Ogihara, "A comparative study of feature selection and multiclass classification methods for tissue classification based on gene expression", Bioinformatics, vol. 20, pp. 2429-2437, 2004

[28] J. A. Suykens and J. Vandewalle, "Least squares support vector machine classifiers", Neural processing letters, vol. 9, pp. 293-300, 1999

[29] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural features for image classification", Systems, Man and Cybernetics, IEEE Transactions on, pp. 610-621, 1973.