Image Feature Detectors for Deepfake Video Detection

Faten F Kharbat
College of Engineering
Al Ain University
Abu Dhabi, UAE
faten.kharbat@aau.ac.ae

Tarik Elamsy
College of Engineering
Al Ain University
Abu Dhabi, UAE
tarik.elamsy@aau.ac.ae

Ahmed Mahmoud
College of Engineering
Al Ain University
Abu Dhabi, UAE
201810062@aau.ac.ae

Rami Abdullah
College of Engineering
Al Ain University
Abu Dhabi, UAE
201810105@aau.ac.ae

***Abstract* – Detecting DeepFake videos are one of the challenges in digital media forensics. This paper proposes a method to detect deepfake videos using Support Vector Machine (SVM) regression. The SVM classifier can be trained with feature points extracted using one of the different feature-point detectors such as HOG, ORB, BRISK, KAZE, SURF, and FAST algorithms. A comprehensive test of the proposed method is conducted using a dataset of original and fake videos from the literature. Different feature point detectors are tested. The result shows that the proposed method of using feature-detector-descriptors for training the SVM can be effectively used to detect false videos.**

Index Terms – DeepFake Video, HOG, ORB, BRISK, KAZE, SURF, FAST.

I. INTRODUCTION

Deepfake are fake videos created by artificial intelligence which are becoming more and more convincing. These fake videos are generated using Generative Adversary Networks (GANs) which replace human faces in a video with another face [1]. Most importantly, many modern tools which are available on the internet have made it easier than ever for anyone to produce realistic "deepfakes" with minimal effort and basic hardware including smartphones.

As a concequence, detecting "deepFake" videos is one of the new challenges in the digital media forensics. Different tools and algorithms have been proposed to detect a falsification in a video; such as [2] and [3].

In this paper, we propose an new algorithm to detect deepfake videos. Our method uses feature points extracted from the video to train Artificial Intelligence (AI) classifiers to detect false videos. Identifying and extracting feature points from images is a fundamental problem for computer vision and used effectively for object detection [6], 3D reconstruction, image registration [4], and frame to frame trajectory estimation[5].

In this paper, HOG, ORB, BRISK, KAZE, SURF, and FAST algorithms are tested and compared for detecting DeepFake videos. The performance of the selected feature detector descriptors are investigated using Support Vector Machine (SVM) regression [7].

The paper is organized as follows. Section II covers a basic background for the main concepts. In Section III, our deepfake video detection algorithm is described. Section IV describes the conducted experiments and results. finally, section V draws the conclusion and some future directions.

II. BACKGROUND

This section describes the basic background for the main concepts used in the research. The covered items are as follows:

A. Support Vector Machine

Support vector machine (SVM) is one of the powerful statistical techniques used as a classification and regression tool [7]. SVM uses linear, polynomial, radial basis function, and sigmoidal kernels, which provide the technique with the ability to solve different problems in different areas.

In fact, [8] has compared the performance of SVM and deep learning in remote sensing image classification. The result was interesting in a way that SVM was found to perform better than deep learning in some circumstances.

SVM has been heavily used in the image processing literature; for example, [9] proposed an SVM based algorithm for multi-label learning for image annotation with the problem of the missing labels. In [10], SVM was used in pattern recognition with the help of some image processing filters. Their enhanced method resulted in a promising effective tool to be used in removing the noise from Electrocardiographic signals.

B. HOG

Histogram of Oriented Gradient (HOG) [11] is an effective descriptor method that basically divides an image into blocks from which it uses the histogram gradient information to compute the edge direction. In general, the HOG process go into three phases for the divided blocks [12]: (i) conduct optional global image normalization, (ii) computes the image gradients for both directions x and y, and (iii) use the overlapping local contrast normalizations to collect the HOG descriptors for all blocks.

C. ORB

The Oriented FAST and Rotated BRIEF (ORB) was introduced in 2011 [13] to enhance previous methods. The main

978-1-7281-5052-9/19/$31.00 ©2019 IEEE

advantage of ORB is that it is rotation invariant and resistant to noise and small changes [4].

D. BRISK

In 2011, [14] proposed the Binary Robust Invariant Scalable Key-points (BRISK) method to detect corners and edges (keypoints). The method depends on "computing brightness comparisons to form a binary descriptor string" [14]. BRISK is also rotation invariant and resistant to noise and small changes [4].

E. KAZE

KAZE is the wind in Japanese from which it was analogized with "nonlinear diffusion processes in the image domain" [15]. The algorithm works on retaining the boundaries of an object in images and reduces the noise. Thus, it has "more distinctiveness at varying scales with the cost of moderate increase in computational time" [4].

F. SURF

Speeded-Up Robust Features (SURF) [16] is a method that depends on Gaussian scale space analysis of an image, based on sums of Haar wavelet components, and uses integral images to enhance feature-detection speed [4].

G. FAST

Features from Accelerated Segment Test (FAST) [17] was proposed to solve the problem of complexity for real-time applications. It was proven that FAST is applicable with different views of a 3D scenes as well.

III. Deep Fake Detection Methodology

We propose a simple approach for detecting deepfake video. Due to the nature of deepfake video generation where multiple camera views, differences in lightning conditions or simply the use of different video codecs makes it difficult for deepfake auto-encoders to generate perfectly realistic faces under all conditions. In addition, because the encoder used for generating the deepfake video is not aware of the skin or other scene information it is very common to have boundary effects due to a seamed fusion between the new face and the rest of the frame. These boundary effects are normally reduced by blurring the color intensity of these boundary pixels. This usually leads to swapped faces that are visually inconsistent with the rest of the scene.

Our deepfake detection methods exploits this inconsistency by extracting features points using traditional edge feature detectors. These features are normal corner points which can be detected using several feature extraction algorithms such as HOG, SURF, KAZE,…etc. The extracted feature points descriptor is aggregated across the different frames of the video to form the input data for SVM training in order to classify pristine and manipulated (Deepfake) videos. The extracted feature points descriptors from an authentic video will provide more correlated feature points attributes than those features extracted from a manipulated face swapped videos.

Figure 1 illustrates our deepfake video detection method. First, the input video is transformed into a sequence of frames. In order to speed up the process, only a few frames per second are extracted. In addition, for each extracted frame, the auto-face detection algorithm is used to identify and crop the face into a rectangular area. This normally reduces the image size heavily for faster processing. Then, the desired feature point detection method is used to extract point descriptors which will be aggregated across the different frames and fed into the SVM classification model for training or detection.

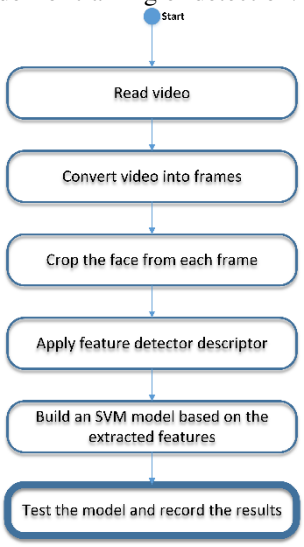

Figure 1. Detection Method

IV. EXPERIMENT AND RESULT

A. Dataset

Very few public datasets are used in literature to detect deepfake videos, however, [1] has generated a special dataset for deepfake videos contains 98 videos, half of them are fake videos and the other half are real videos. All the videos are in the format of mp4 and have approximately 30 seconds of duration.

C. Experimental Setup

MATLAB-2019a with computer Vision toolbox and image processing toolbox has been used for performing the experiments presented in this paper. Specifications of the computer system used are: Intel(R) Core(TM) i7-8750H CPU @ 3.70 GHz, 9M Cache and 16.00 GB RAM.

For each feature detector descriptor, the videos are divided into two folds. We used (85%) and (15%) of the data set for training and testing respectively. For the training set, each video is converted into a set of frames where 5 frames are extracted per second. For each frame, a face detection algorithm is used to detect and crop the face into a 200×200 pixels sub-image.

The different feature-detection algorithms are run on the extracted set of cropped faces to extract the feature points descriptors. The accumulative results for all the images will be used to build the SVM model. The testing set is used to test the accuracy of the produced model and to record the result in terms of the confusion matrix

All videos are divided into frames All parameters for BRISK, KAZE, FAST and SURF algorithms were set to default values. The cell size in the HOG algorithm was set to 4×4, and the scale factor and cell size in the ORB algorithm is set to 1.000001 and 100, respectively.

C. *Results*

Different tools are used to compare results between data mining algorithms [18], one of them is the confusion matrix which is used as "an indication of the properties of a classification (discriminant) rule". Confusion matrix has four cells which contains the correct and incorrect number of cases classified for each class. In this research, there are two classes: real and fake. True Real (TR) indicates the true classification for the real cases, and False Real (FR) indicates the false classification for the real cases. True Fake (TF) cell contains the true classification for the fake cases, and False Fake (FF) cell contains the false classification for the fake cases. The accuracy is computed by:

$$Accuracy = (TR+TF)/(TR+FR+TF+FF)$$

Although the accuracy is not the only indication used, it can be considered one of the important ones. Another indication is the sensitivity (how well the positive "Real" cases are recognized) and specificity (how well the negative "Fake" cases are recognized) [18].

$$sensetivity = \frac{TR}{TR + FF}$$

$$specificity = \frac{TF}{TF + FR}$$

TABLE 1
THE VALUES OF THE CONFUSION MATRIX

	TR	FR	TF	FF	Accurecy
HOG	95	5	94	6	**94.5%**
BRISK	75	25	99	1	**87%**
KAZE	64	36	89	11	76.5%
SURF	90	10	91	9	90.5%
FAST	74	26	99	1	86.5%
ORB	83	17	99	1	91%

It can be shown in TABLE I that the proposed algorithm provides acceptable accuracy for detecting original and fake videos using the different feature detection algorithms. Feature detection algorithm of HOG provides the best performance with accuracy of 94.5%. SURF and ORB also provides good accuracy exdeeding 90%. KAZE, on the other hand was the lest effective with an accuracy of 76.5%. BRISK and FAST scores above 86.5% accuracy.

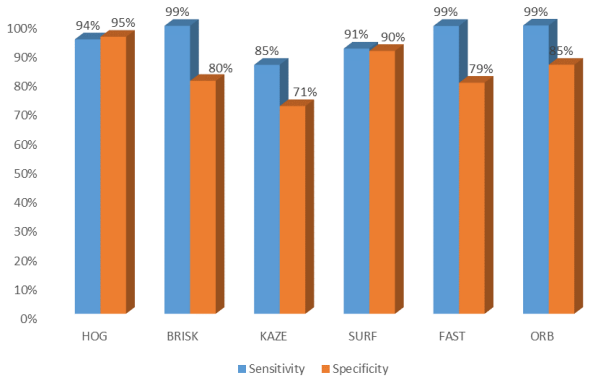

Fig. 2 Sensitivity vs Specificity.

IV. CONCLUSION

In this paper, we proposed a simple algorithm to automatically detect deepfake videos. The algorithm is based on Training SVM classifier with feature-point-descriptor of extracted points from the faces appearing in the videos. Around 95% accuracy has been achieved used HOG feature point extraction algorithm.

In our future work, we are planning to investigate aggregating the feature descriptors extracted using multiple feature point detectors as the input data set in training the SVM classifier for higher accuracy.

REFERENCES

[1] Li Y, Chang MC, Farid H, Lyu S. In ictu oculi: Exposing ai generated fake face videos by detecting eye blinking. arXiv preprint arXiv:1806.02877. 2018 Jun 7.

[2] Maras, M. H., & Alexandrou, A. (2018). Determining authenticity of video evidence in the age of artificial intelligence and in the wake of Deepfake videos. The International Journal of Evidence & Proof, 1365712718807226.

[3] Schwartz, Oscar (12 November 2018). "You thought fake news was bad? Deep fakes are where truth goes to die". The Guardian. Retrieved 14 November 2018

[4] Tareen SA, Saleem Z. A comparative analysis of SIFT, SURF, KAZE, AKAZE, ORB, and BRISK. In2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET) 2018 Mar 3 (pp. 1-10). IEEE.

[5] Kostusiak A. The comparison of keypoint detectors and descriptors for registration of RGB-D data. In International Conference on Automation 2016 Mar 2 (pp. 609-622). Springer, Cham.

[6] Andersson O, Reyna Marquez S. A comparison of object detection algorithms using unmanipulated testing images: Comparing SIFT, KAZE, AKAZE and ORB.

[7] Guenther N, Schonlau M. Support vector machines. The Stata Journal. 2016 Dec;16(4):917-37.

[8] Liu P, Choo KK, Wang L, Huang F. SVM or deep learning? A comparative study on remote sensing image classification. Soft Computing. 2017 Dec 1;21(23):7053-65.
[9] Liu Y, Wen K, Gao Q, Gao X, Nie F. SVM based multi-label learning with missing labels for image annotation. Pattern Recognition. 2018 Jun 1;78:307-17.
[10] Varatharajan R, Manogaran G, Priyan MK. A big data classification approach using LDA with an enhanced SVM method for ECG signals in cloud computing. Multimedia Tools and Applications. 2018 Apr 1;77(8):10195-215.
[11] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2005, pp. 886–893.
[12] Yao S, Pan S, Wang T, Zheng C, Shen W, Chong Y. A new pedestrian detection method based on combined HOG and LSS features. Neurocomputing. 2015 Mar 3;151:1006-14.
[13] ORB E. Rublee et al., "ORB: An efficient alternative to SIFT or SURF," in IEEE International Conference on Computer Vision, Barcelona, ICCV, 2011, pp. 2564-2571.
[14] BRISK S. Leutenegger et al., “BRISK: Binary robust invariant scalable keypoints,” in IEEE International Conference on Computer Vision, Barcelona, ICCV, 2011, pp. 2548-2555.
[15] KAZE P. F. Alcantarilla et al., "KAZE features," in European Conference on Computer Vision, Berlin, ECCV, 2012, pp. 214-227.
[16] SURF H. Bay et al., “Speeded-up robust features (SURF),” Computer Vision and Image Understanding, vol. 110, no. 3, pp. 346-359, 2008
[17] Fast E. Rosten and T. Drummond, “Machine learning for high-speed corner detection,” in Computer Vision–ECCV 2006, 2006, pp. 430–443.
[18] Chawla NV. Data mining for imbalanced datasets: An overview. InData mining and knowledge discovery handbook 2009. Springer, Boston, MA.