Exposing Deepfake Videos by Tracking Eye Movements

Meng Li[1], Beibei Liu[1,2], Yongjian Hu[1,2], Yufei Wang[2]

1. South China University of Technology, Guangzhou, China;
2. Sino-Singapore International Joint Research Institute, Guangzhou, China

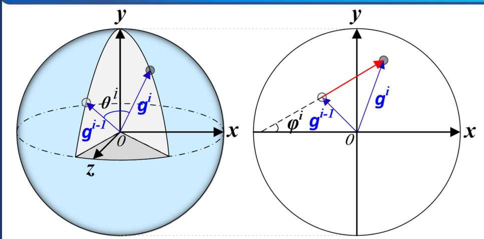
ICPR 2020
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021

華南理工大學
South China University of Technology

INTRODUCTION

This paper reports a novel approach to expose deepfake videos. We found that most fake videos are markedly different from the real ones in the way the eyes move. We are thus motivated to define four features that could well capture such differences. The features are then fed to SVM for classification. It is shown to be a promising approach that without high dimensional features we are able to achieve competitive results on several public datasets.

DEFINITION OF THE EYEMOVEMENT STATUS



$$\theta^i = \arccos\left(\frac{\boldsymbol{g}^i \cdot \boldsymbol{g}^{i-1}}{\|\boldsymbol{g}^i\| \cdot \|\boldsymbol{g}^{i-1}\|}\right)$$

$$\varphi^i = \begin{cases} \arccos\dfrac{g_x^i - g_x^{i-1}}{\sqrt{\left(g_x^i - g_x^{i-1}\right)^2 + \left(g_y^i - g_y^{i-1}\right)^2}}, & g_y^i - g_y^{i-1} > 0 \\ -\arccos\dfrac{g_x^i - g_x^{i-1}}{\sqrt{\left(g_x^i - g_x^{i-1}\right)^2 + \left(g_y^i - g_y^{i-1}\right)^2}}, & g_y^i - g_y^{i-1} < 0 \\ \pi/2 - \pi/2 \times \operatorname{sgn}\left(g_x^i - g_x^{i-1}\right), & g_y^i - g_y^{i-1} = 0 \end{cases}$$

θ^i indicates how rapid the eye moves; φ^i indicates the eye moving direction on the X-Y plane.

FEATURE EXTRACTION

1) Left/Right eye activity

$$A_l = \sqrt{\sum_{i=1}^{N-1} \frac{\left(\theta_l^i - \overline{\theta_l}\right)^2}{N-1}} \quad A_r = \sqrt{\sum_{i=1}^{N-1} \frac{\left(\theta_r^i - \overline{\theta_r}\right)^2}{N-1}}$$

2) Binocular angular shift correlation

$$C_{BAS} = \frac{\sum_{i=0}^{N-1}\left(\theta_l^i - \overline{\theta_l}\right)\left(\theta_r^i - \overline{\theta_r}\right)}{\sqrt{\sum_{i=0}^{N-1}\left(\theta_l^i - \overline{\theta_l}\right)^2}\sqrt{\sum_{i=0}^{N-1}\left(\theta_r^i - \overline{\theta_r}\right)^2}}$$

3) Binocular direction correlation

$$C_{BD} = \frac{\sum_{i=0}^{N-1}\left(\varphi_l^i - \overline{\varphi_l}\right)\left(\varphi_r^i - \overline{\varphi_r}\right)}{\sqrt{\sum_{i=0}^{N-1}\left(\varphi_l^i - \overline{\varphi_l}\right)^2}\sqrt{\sum_{i=0}^{N-1}\left(\varphi_r^i - \overline{\varphi_r}\right)^2}}$$

REFERENCES

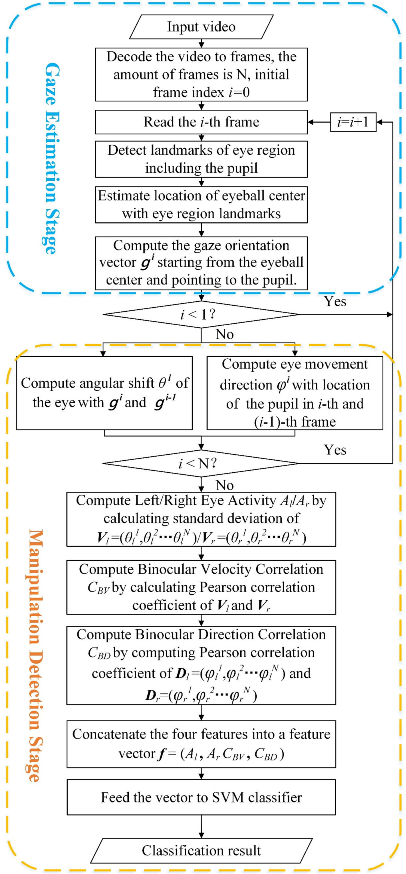
[1] Zhou P, X Han, et al. Two-Stream Neural Networks for Tampered Face Detection[C]. In Proc. Conference on Computer Vision and Pattern Recognition Workshops, 2017.

[2] Yang X, Li Y, Lyu S. Exposing Deep Fakes Using Inconsistent Head Poses[C]. In Proc. International Conference on Acoustics, Speech and Signal Processing, 2019.

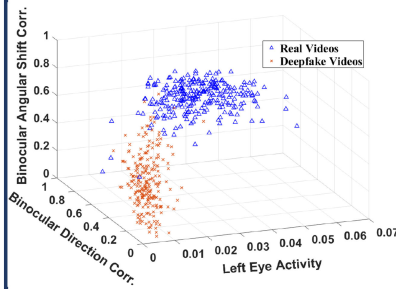
[3] F. Matern, C. Riess, and M. Stamminger, "Exploiting visual artifacts to expose deepfakes and face manipulations," in 2019 IEEE Winter Applications of Computer Vision Workshops (WACVW). IEEE, 2019.

[4] Matern F, Riess C , et al. Exploiting Visual Artifacts to Expose DeepFakes and Face Manipulations [C]. In Proc. IEEE Winter Applications of Computer Vision Workshops, 2019.

FLOWCHART



EXPERIMENT RESULTS



In order to prove the effectiveness of our method. we plot the scatter diagram of 500 randomly selected videos from the Deepfake-TIMIT dataset in feature space consisting of A_l , C_{BAS} and C_{BD} . It shows that the real and fake videos are well separated apart with the proposed features.

Table 1 AUC values compared to other methods

Method	FaceForensics ++-Deepfake	Deepfake-TIMIT HQ	Deepfake-TIMIT LQ
[1]	0.701	0.735	0.835
[2]	0.473	0.532	0.551
[3]	0.780	0.773	0.770
[4]	0.792	0.932	0.999
proposed	0.918	0.955	0.996

In Table 1, we list a few relative results compared with our method. It is obviously that our detection approach provides a competitive result.

Table 2 Cross-database test result

Training database	Testing database	AUC
Deepfake-TIMIT LQ	Deepfake-TIMIT HQ	0.951
	FaceForensics++-Deepfake	0.788
Deepfake-TIMIT HQ	Deepfake-TIMIT LQ	0.994
	FaceForensics++-Deepfake	0.825
FaceForensics++-Deepfake	Deepfake-TIMIT LQ	0.986
	Deepfake-TIMIT HQ	0.945

To further demonstrate the generality of our method, we also did some cross-database test, which is shown in Table 2.

CONCLUSION

Taking both temporal and spatial information into consideration, we have defined four features derived from eye movements to reveal the activity and coordination of eyes in the videos, which can be exploited to tell apart fake and genuine videos. Our approach works with a very small number of features and does not involve complicated neural networks.In addition, the proposed features could readily be incorporated into a comprehensive detection scheme that assembles various types of features, which we believe is a promising future work.