Spoof Detection with RGB Cameras in Facial Recognition Systems

CS 230 Final Project Winter 2019

Ye Li (liye5), Kairen Ye (kairenye), Jiyao Yuan (yuan999)

Introduction

- ***Motivation:*** With the increasing application of biometric authentication, a robust 2D face anti-spoof algorithm can tremendously boost people's confidence in massively deploying facial recognition as the primary mode of authentication.
- ***Approach:*** We trained a deep neural network with CNN-RNN architecture on two auxiliary supervisions, the pseudo-depth map and the remote photoplethysmography (rPPG) signal, to discern live versus spoof images [1].

SiW Dataset

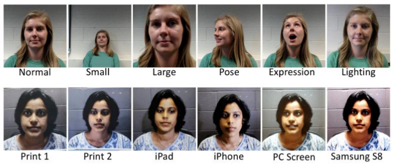

Figure 1. Example live (top) and spoof (bottom) videos in SiW

- The Spoof in the Wild (SiW) dataset contains 165 subjects. Each subject has approximately 8 live video clips and 20 spoof video clips [1].
- All videos are about 15 seconds in length with HD 1080P resolution recorded at 30 FPS.
- Live videos are taken with variations of illumination, distance, pose, and facial expression.
- Spoof videos are taken with different forms of attacks, such as paper with various textures (presentation attacks) and tablet devices of different brands (replay attacks).

SensorID	1	Canon EOS T6
	2	Logitech C920 webcam

			MediumID				**SessionID**	
			1	2	3	4	1	2
TypeID	1	Live	No lighting variation	Extra Lighting variation			Move backward and forward	Yaw-angle rotation & facial expression change
	2	Print Attack	High resolution image (5184 × 3456)	Low resolution image (1920 × 1080)			Glossy Paper	Matt Paper
	3	Print Attack	iPad Pro (2017)	iPhone 7 Plus	Asus MB168B	Samsung Galaxy S8	Randomly select a live video	

Table 1. Variety of dataset samples

Generating Supervision Data

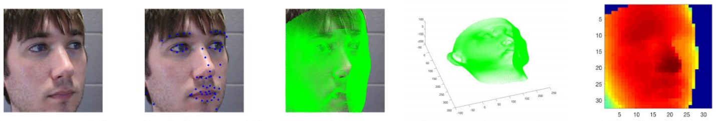

Figure 2. From left to right, 1) the original face cropped by given coordinates , 2) fitting the sparse 68 landmarks onto the face image, 3) fitting the dense 53,215 vertices onto the face image using DeFA, 4) the fitted 53,215 vertices from DeFA in 3D, 5) the 32x32 depth map by applying Z-Buffer to the 53,215 vertices

- Detecting and cropping faces according to the coordinates in *.face* file given in the SiW dataset
- Estimating 3D face shape with Dense Face Alignment (DeFA) [3]
- Generating 2D depth map from 3D face shape with Z-buffer
- Calculating chrominance-based rPPG signal from a tracked region on the subject's forehead in consecutive video frames [6]

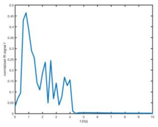

Figure 3. An example of rPPG signal

CNN-RNN Model Architecture

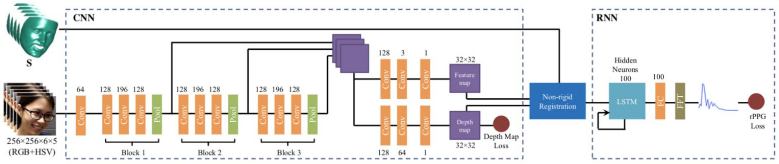

Figure 4. Overall CNN-RNN network architecture of the anti-spoof model

CNN-RNN Architecture:

- The convolutional layers help the network capture cues for generating the pseudo-depth from a *single* frame of image.
- The recurrent (LSTM) layers help the network capture the temporal information necessary for learning the rPPG signal across *multiple* image frames over time.
- The *non-rigid registration* layer applies face frontalization to align the feature maps so that the RNN can learn from more consistent inputs.

Two Loss Functions·

- Depth Map Loss: $\Theta_D = \arg\min_{\Theta_D} \sum_{i=1}^{N_d} ||\text{CNN}_D(\mathbf{I}_i;\Theta_D) - \mathbf{D}_i||_1^2$
- rPPG Loss: $\Theta_R = \arg\min_{\Theta_R} \sum_{i=1}^{N_s} ||\text{RNN}_R([\{\mathbf{F}_j\}_{j=1}^{N_f}]_i;\Theta_R) - \mathbf{f}_i||_1^2$

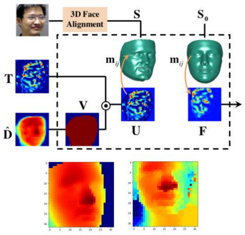

Figure 5. Non-rigid registration illustration(top) and an example of our implementation(bottom)

Results & Analysis

Evaluation Metric:

A score was computed on each test sample using the formula below:

$$score = ||\mathbf{f}||_2^2 + \lambda||\mathbf{D}||_2^2,$$

where **D** is the output depth map of the last frame and **f** is the output rPPG signal from the network. The weight λ is set to 0.015 [1]. The threshold on the score for classifying live vs. spoof is 6.3, with which the metrics are calculated and shown in Table 2.

	Validation Accuracy	Validation Precision	Validation Recall
Baseline (SVM)	91.80%	95.93%	86.62%
CNN-RNN	92.15%	90.37%	99.52%

Table 2. Evaluation and comparison

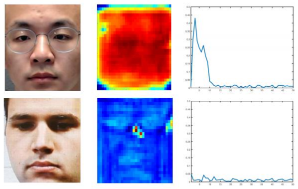

Figure 6. Examples of output depth maps and rPPG signals from the network for live (top) and spoof (bottom) images

Reference

[1] Y. Liu, A. Jourabloo, and X. Liu, "Learning deep models for face anti-spoofing: Binary or auxiliary supervision," in CVPR, 2018.

[2] D. E. King. Dlib-ml: A machine learning toolkit. JMLR, 10(Jul):1755–1758, 2009.

[3] Y. Liu, A. Jourabloo, W. Ren, and X. Liu. Dense face alignment. In ICCVW, pages 1619–1628, 2017.

[4] P. Paysan, R. Knothe, B. Amberg, S. Romdhani, and T. Vetter. A 3D face model for pose and illumination invariant face recognition. In AVSS, pages 296–301, 2009.

[5] C. Cao, Y. Weng, S. Zhou, Y. Tong, and K. Zhou. Facewarehouse: A 3D facial expression database for visual computing. IEEE Trans. Vis. Comput. Graphics, 20(3):413–425, 2014.

[6] G. de Haan and V. Jeanne. Robust pulse rate from chrominance-based rPPG. IEEE Trans. Biomedical Engineering, 60(10):2878–2886, 2013.