Camera Based Non-contact Lie Detector

Yu-Peng Hsieh, Zi-Yi Tai, Chien-Chi Hsu
BS in Electrical Engineering at National Taiwan University, Taiwan
b06901007@ntu.edu.tw, b06901184@ntu.edu.tw, b06901015@ntu.edu.tw
Advisor: Pai-Chi, Li

Abstract

Current existing lie detection methods are mostly based on physiological measurements through contact, which may be easily interfered by intended poor contact. This study proposed a non-contact lie detection system with camera-based physiological measurements. We acquire video data by conducting experiments on 15 participants, following the protocols in [1]. Heart rate and eye aspect ratio of each participant are extracted from the videos respectively with rPPG and OpenCV library. Machine learning models are then trained with the measurements. Provided with information about the participant's baseline features, our model achieved an AUC of 0.57 and 0.53 on training and testing set respectively.

1 Introduction

The definition of **lying** is yet to be universally accepted, while **deception** defines an act to hide the truth with a false statement, which requires an *intention* for lying [2]. Deception is pervasive, and some would argue necessary, in human communication. Yet this very action stirs up indignation. As a result, for as long as there have been lies, there have been methods of lie detection [3].

Human are very poor lie detectors when unaided. Therefore, interest in technology that can aid in lie detection has ensued. Since it was known that vital signs such as blood pressure (BP), heart rate (HR), and breathing could be affected by the stressful situation brought on by deception, quantifying and measuring those responses in an effort to detect lying became a goal [4] . The standard polygraph today measures respiration, HR, BP, and sudomotor function (sweating). For the polygraph measurement of cardiovascular activity, the standard tool is the sphygmomanometer arm cuff [5], which also comes in wrist cuff and finger cuff varieties. An alternative is the photoelectric plethysmograph, which is usually clipped to a finger or ear [6].

Although not painful, this procedure can be detrimental to subjects with sensitive skin, and it certainly detracts from the ideal psycho-physiological measurement, which should be unobtrusive [7]. Therefore, we propose a non-contact lie detector involving machine learning along with camera-based measurements of physiological parameters, including heart rates and eye aspect ratios. Also, this study is aimed to be robust among different subjects.

2 Materials and Methods

2.1 Camera Based Heart Rate Monitor: Remote Photoplethysmography (rPPG)

It is now established that, in a Concealed Information Test (CIT), the recognition of crime-related items results in larger skin conductance responses, respiratory suppression, heart rate deceleration when compared to neutral control items [8]. In this study, heart rate measurement (HRM) was achieved using low-cost, consumer-grade digital cameras and ambient light sources. These techniques are commonly referred to as remote photoplethysmography (rPPG) because of their similarity to traditional PPG, but instead with camera-based measurements [9]. Two main approaches have emerged from existing studies on rPPG: (i) HRM based on periodic variation of the subject's skin color, and (ii) HRM based on periodic head movement. Of the two approaches, that based on skin color variation has been discussed in many more studies and is utilized in our experiment.

During this procedure, pixel values (PV) for the red (R), green (G) and blue (B) channels are read for each video frame. To determine power spectral density of the signal, fast Fourier transform (FFT) algorithm is deployed. It has been shown that the signal for the green channel contains the strongest plethysmographic signal, consistent with the fact that hemoglobin, which plays a role in the transportation of oxygen in red blood cells, absorbs green light better than it does red and blue. A general rPPG algorithm can be divided into three key steps: (i) extraction of the raw signal from several video frames, (ii) estimation of the plethysmographic signal, and (iii) HR estimation. The framework is illustrated in Figure 1 [10].

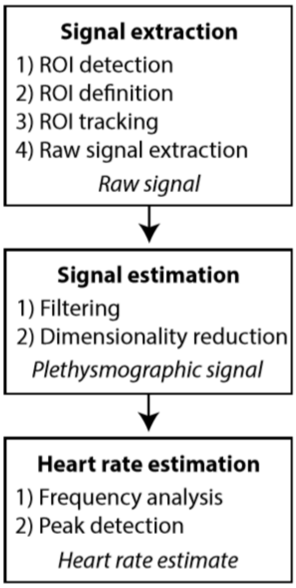

Figure 1: Generalized rPPG algorithm framework.

2.2 Eye Blink Measures

Another measurement acquired in this study is the eye aspect ratio (EAR). Studies show that variations in blink measures could differentiate between those with false intent and truthful intent individuals [11]. The rate of eye blinking is associated with concentration; the higher cognitive demand, the less eye blinks. Therefore, it is reasonable that lying is associated with a decrease in eye blinks, and is followed by a compensatory effect: an increase in eye blinks directly after the lie is told and cognitive demand has ceased [12]. This project utilizes OpenCV Library, a Library that is used to carry out image processing using programming languages like python, to implement real-time blink detection. For every video frame, the eye landmarks are detected, and the EAR is computed [13]. The EAR is computed as

$$EAR = \frac{\|p_2 - p_6\| + \|p_3 - p_5\|}{2\|p_1 - p_4\|} \tag{1}$$

where the sites of p_1 to p_6 are labeled in Figure 2. It is approximately constant while the eye is open, but rapidly falls to zero when a blink is taking place, as shown in Figure 3. The two featuring measurements are now obtained, the following content describes two models used in this study.

2.3 LSTM

Long short-term memory (LSTM) [14] encodes a latent vector for each input segment. Unidirectional LSTM preserves information seen from the past, since it only inputs from the front, while Bidirectional LSTM inputs in both ways: one from past to future and the other from future to past. This gives Bidirectional LSTM a major difference from Unidirectional LSTM that using the two hidden states combined, information from both past and future in any point in time is able to be well-preserved. Our detection model contains two Bidirectional LSTM and a LSTM layer.

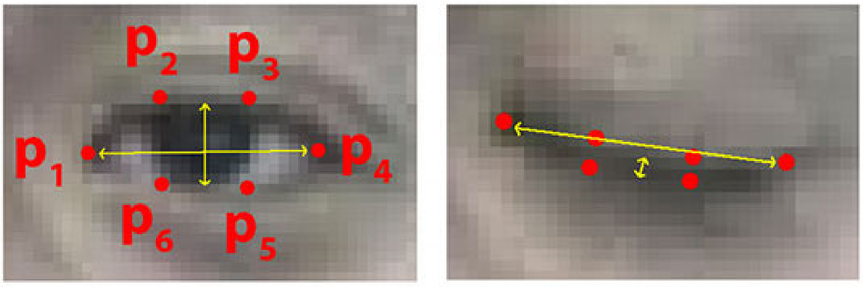

Figure 2: The sites of p_1 to p_6

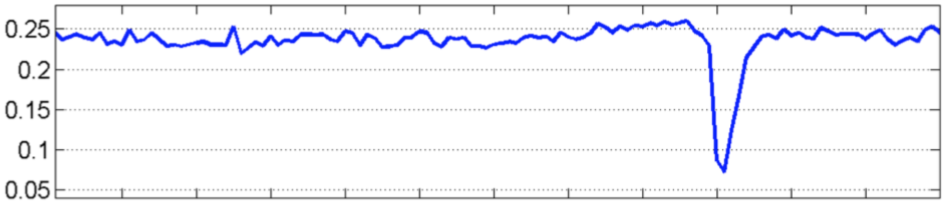

Figure 3: The EAR plotted for several frames of a video sequence. Open and closed eyes with landmarks. A single blink is present.

2.4 Deep Neural Network

The other model is deep neural network (DNN). A DNN model comprises multiple hidden layers to form a complex mapping function between the inputs and outputs. In DNN, the relationship between the input, x, and the output of the first hidden layer, h1, is described as

$$h_1 = f(W_1 x) + b \tag{2}$$

where W_1 and b_1 are the weight matrix and bias vector, respectively, and f (.) is the activation function. In this study, we use two Rectified Linear Unit (ReLU) function and one sigmoid function for the activation function. The relationship between the current hidden layer and the next hidden layer can be expressed as

$$h_{i+1} = f(W_{i+1} h_i + b_i + 1),\ i = 1, 2, ..., L1 \tag{3}$$

where L is the DNN hidden layer number.

3 Workflow

3.1 Data Acquisition

In order to collect the training and testing data for our lie detector, we followed a protocol proposed by Gonzalez-Billandon J et al [1]. The participants were asked to watch two crime scene videos as witnesses. During the interview, the participants were assigned to either testify or defend the suspects. If a participant was told to testify, he or she would have to answer all the questions honestly to help identify the criminals. Otherwise, he or she would lie about the crime so that the suspect would be exonerated. Each participant would have to answer 60 questions in total, which can be divided into three phases as follows:

1. General questions.

   This session contains simple questions irrelevant to the crime videos, such as ”Who is the president of the United States?” The participants were asked to answer each question honestly and without stress. The purpose of this phase is to record each participant’s usual heart rate and eye movements, allowing us to establish a baseline for the monitored physiological parameters.

2. First session of questions.

   After watching the first crime video, the participants would be randomly assigned to a role. The participants were asked to answer the questions according to their roles.

3. Second session of questions.

   The participants would also be assigned to the other role after watching the second crime video, and answer the questions about the crime.

We filmed all the investigations and pruned the videos so that each video contains one participant answering a question. Each video would be labeled as truth or lie. If a participant was assigned to testify, all 20 questions in that session would be labeled as true. Otherwise, the participant may choose to lie on some of the 20 questions. After the interrogation, we would verify with the participant about which answer was a lie.

In total, we collected data from 15 participants. Descriptions about the subjects and their data are shown in Table 1. After acquiring videos of participants answering questions, remote photoplethysmography (rPPG) is utilized to obtain heart rates. Eye aspect ratio of the participant was calculated with the OpenCV library. Both methods are described below.

	Lie about the first crime	Lie about the second crime
female	3	5
male	3	3
total	6	8

Table 1: Number of Participants that lie about each crime

	Truth	Lies
Number of answers	470	130
ratio	78.3%	21.6%

Table 2: Number of Truths and lies told by all participants in total

3.2 Physiological Measurements and Data Preprocess

For each video we acquired, heart rate(HR) was extracted with rPPG and eye aspect ratio(EAR) was calculated using the OpenCV library. So far, we would have series of these measurements of different length for each video. All of the series are padded with *0*s to 600 in length. Then, two types of concatenation are utilized in this study.

1. Omit the data from general questions and combine the remaining HR and EAR series in parallel. In other words, each data contains HR and EAR of the participant when answering the crime-related questions.

2. For each participant, pair the measurements from a general question and another crime-related question. The four sequences are concatenated into a (4, 600) dimension array. Figure 4 presents the data structure of this concatenation. With 20 general questions and 40 questions related to video I and II, there are 800 mixed data for each participant in total.

3.3 LSTM Classifier

3.3.1 Training Without Information From General Questions

For the first type of concatenation, we constructed a lie detector model (Figure 5). This model contains six layers. The first two layers are Bidirectional LSTM, followed by another layer of LSTM. The rest are three DNN layers. The output of this model is the probability that the participant is lying to a crime-related questions. No information from the general questions are included in this type of classifier.

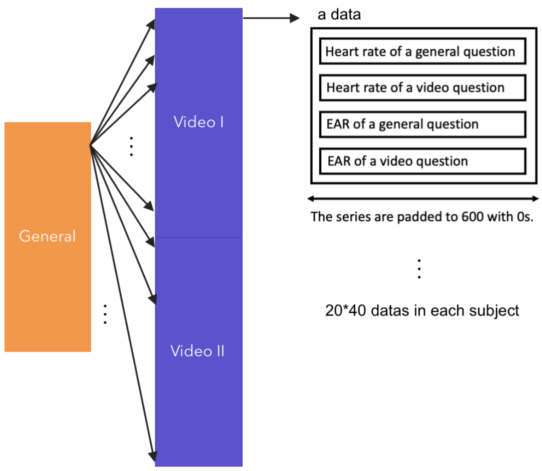

Figure 4: The data structure.

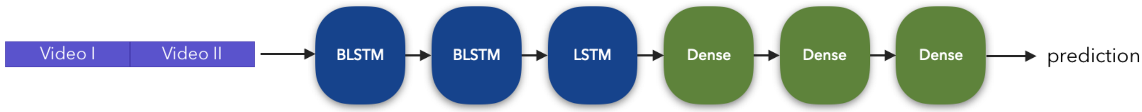

Figure 5: The workflow of the detector model without general questions.

3.3.2 Training With Information From General Questions

For the second type of concatenation, we also constructed a LSTM classifier as our lie detector model. Each crime-related data is evaluated along with one of the data from a general question. Hence, for every crime-related video, we would have 20 predicted probabilities (There are 20 general questions). We would then sort and evaluate the 20 predicted probabilities in order to obtain the final decision whether this video was a lie.

We performed two types of evaluation. The first one is to vote equally by calculating the average of all the probabilities. Another way is to vote with different weights. The weights of each data is based on if its general question is verified more easily. Higher weights are assigned to those questions, which are written in bold letters in [1]. The reason to this is the authenticity of those questions can be easily distinguished and may be more advantageous for the detector model than others. Figure 6 shows the workflow of this type of classifier.

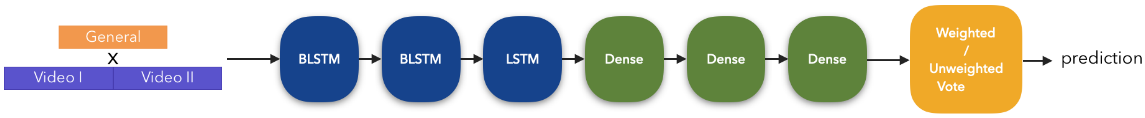

Figure 6: The workflow of the detector model with general questions.

4 Results

4.1 Dummy Classifier: Linear Regression

In order to evaluate how the LSTM classifier proposed above is performing, we constructed a dummy model to set a baseline. If the LSTM classfier does not outperform this dummy classifier, we can tell that it is not learning from out training data. The dummy model is simply a linear regression model with a sigmoid activation function at the output neuron. The comparison between the results of the dummy model along with the LSTM classifier are shown in Table 3.

We can tell from Table 3 that the dummy model can reach an accuracy of approximately 75% on the validation data. This may be a surprise at first sight, considering the fact that the dummy model is not reasonable. However, if we look at the ratio of lies and truths in our dataset, it turns out that the dummy model can reach an accuracy this high if it simply classfies every data as a truth. Hence, accuracy is not a robust metric in our senario. Instead, we calculate recall, f1 score, and AUC (Area under the ROC curve) to evaluate our models.

	Train	Test
Accuracy	79.3%	75.8%
Recall	0	0
f1 score	0	0
AUC	0.5	0.5

Table 3: Performance of the dummy model

4.2 Training Without Information From General Questions

The LSTM classifier was first trained without the data of general questions. To prevent the model from classifying every data as truth, we modified the **loss function** from *binary crossentropy*, the most common loss function for binary classification, to *weighted binary crossentropy* instead. The weight of classes were set to be inverse proportionally to their ratio in the dataset.

We randomly chose 11 participants as the training data and 3 participants as the test data. The model was trained for 20 epochs and save the best model with the highest f1 score. We repeatedly chose different sets of participants as test data and calculate AUC (Area under the ROC curve) each time. On average, the model reach an AUC of 51.0% on the training set and 49.0% on the testing set. Figure 7 shows the distribution of AUC on training and testing sets.

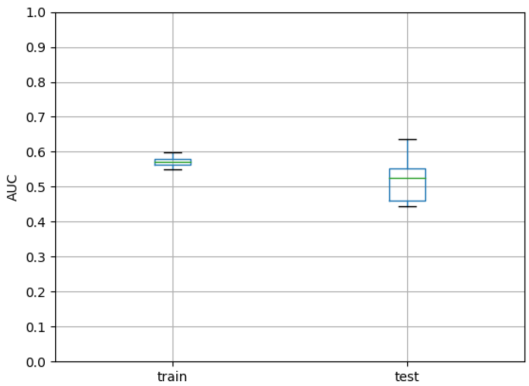

Figure 7: Distribution of AUC on training and testing sets

Mode1	Dummy Model		LSTM Classifier	
	Train	Test	Train	Test
Accuracy	79.3%	75.8%	74.3%	75.5%
Recall	0.0	0.0	0.06	0.06
f1 score	0.0	0.0	0.10	0.09
AUC	0.5	0.5	0.51	0.49

Table 4: Performance Comparison.

4.3 Training with Information From General Questions

We can tell from Table 4 that the LSTM classifier did not improve from the dummy model. This is because for each input to the model, the data only contains information about the participant answering the question. This means the model does not have information about the participant itself. In other words, the model does not know the usual HR and EAR for the participant. Since lie detection is based on slight changes in HR and EAR, we would have to provide a baseline so that our model could detect those slight changes.

The data from general questions are provided as a baseline. We randomly chose 11 participants as the training data and 3 participants as the testing data. The model was trained for 20 epochs and the best model with the highest f1 score was saved. We repeatedly chose different sets of participants as test data and calculated AUC (Area under the ROC curve) of each model. The average AUC improved from 51.0% to 55.0% on the training set and 49.0% to 52.0% on the testing set if the predictions are voted equally. The average AUC improved by 2% to 57.0% on the training set and 53.0% on the testing set if the prediction was based on different weights for the general questions. Figure 8 shows the distribution of AUC on training and testing sets. We can conclude that the LSTM Classifier improves when information from general questions is provided.

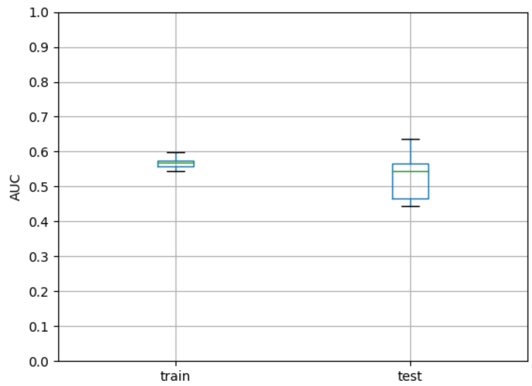

Figure 8: Distribution of AUC on training and testing sets

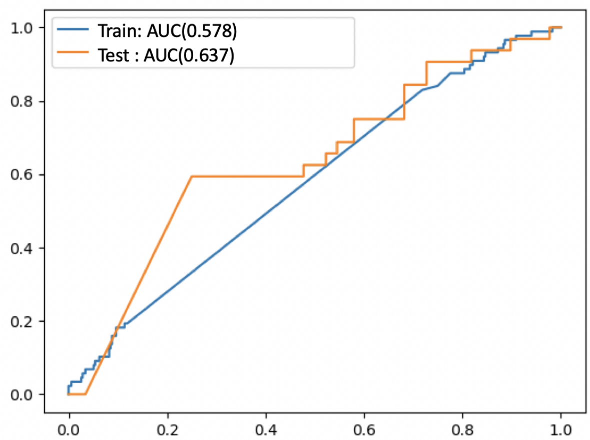

Figure 9: An example of the ROC curves.

5 Discussions

1. Since the definition of *lying* is yet to be universally agreed, to determine whether the participant was lying or not isn't always clear and even falls to be too assertive for some cases. In this report, we verify the true/lie labels with the participants to assure if they held the intentions to lie.

2. Remotely measuring heart rates and eye aspect ratio is often unstable, which may produce data with poor qualities, and further affect the results of machine learning. To eliminate data noises due to

Mode1	LSTM Classifier w.o general questions		LSTM Classifier with general questions	
	Train	Test	Train	Test
Accuracy	74.3%	75.5%	67.1%	65.1%
Recall	0.06	0.06	0.23	0.27
f1 score	0.10	0.09	0.19	0.18
AUC	0.51	0.49	0.57	0.53

Table 5: Performance Comparison with/w.o. general questions.

unstable measurements, data preprocessing methods such as scaling and smoothing are utilized, but no significant difference is shown in the results.

3. The effectiveness of machine learning based lie detection using contact measurements is yet to be discovered. As previously described, remote measurements may lead to noises in data. The results of detection using contact measurements can serve as a comparison to those using non-contact measurements.
4. There is still debate over the inherent problems associated with any attempt to infer psychological states based on peripheral nervous system activity. Heart rate, blood pressure and other cardiovascular processes are affected by a wide range of factors, including perceived threats, increased physical or mental activity, the anticipation of a threat or activity, and effectively any form of specific or general arousal. They can differ subtly or radically amongst individuals and even for the same individual under different circumstances. Thus current scientific opinions pertaining to the polygraph are still polarized.
5. The interview was not conducted well. The subjects may had little incentives to lie since there was neither reward nor punishment. Moreover, we asked no further questions when the subject told a flawed lie, which is unlikely to happen in real interrogation.

6 Conclusions and Future Work

Deception is one of the complicated human behaviors. This study proposed a lie detection system with camera-based non-contact physiological measurements. We followed the experiment protocols in [1]. Heart rate and eye aspect ratio of each participant are calculated respectively with rPPG and OpenCV library. The measurements are then trained on two types of LSTM classifier. Provided with information about the participant's baseline features, our model achieved an AUC of 0.57 and 0.53 on training and testing set. Due to the factors mentioned in the *Discussions* section, our attempt to develop a non-contact lie detector was not really successful. Our future work will focus on enhancing the reliability of remote measurements, designing a better protocol which gives the subjects a stronger incentive to lie, and introducing other physiological parameters to achieve higher accuracy of lie detection.

References

[1] Jonas Gonzalez-Billandon et al. "Can a Robot Catch You Lying? A Machine Learning System to Detect Lies During Interactions". In: *Frontiers in Robotics and AI* 6 (2019), p. 64. ISSN: 2296-9144.

[2] James Edwin Mahon. "The Definition of Lying and Deception". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Winter 2016. Metaphysics Research Lab, Stanford University, 2016.

[3] Martina Vicianova. "Historical Techniques of Lie Detection". In: *Europe's Journal of Psychology* 11.3 (Aug. 2015), pp. 522–534.

[4] Lcdr Glen Cook and Lt Charles Mitschow. "Beyond the Polygraph: Deception Detection and the Autonomic Nervous System". eng. In: *Federal practitioner : for the health care professionals of the VA, DoD, and PHS* 36.7 (July 2019). PMC6654171[pmcid], pp. 316–321. ISSN: 1078-4497.

[5] Martin J Turner and Johan M van Schalkwyk. “Blood pressure variability causes spurious identification of hypertension in clinical studies: a computer simulation study”. In: *American journal of hypertension* 21.1 (Jan. 2008), pp. 85–91. ISSN: 0895-7061.

[6] John Synnott, David Dietzel, and Maria Ioannou. “A review of the polygraph: history, methodology and current status”. In: *Crime Psychology Review* 1.1 (2015), pp. 59–83.

[7] John J. Furedy and Ronald J. Heslegrave. “Validity of the Lie Detector: A Psychophysiological Perspective”. In: *Criminal Justice and Behavior* 15.2 (1988), pp. 219–246.

[8] Matthias Gamer et al. “Combining physiological measures in the detection of concealed information”. In: *Physiology behavior* 95 (Oct. 2008), pp. 333–40.

[9] Wim Verkruysse, Lars O Svaasand, and J Stuart Nelson. “Remote plethysmographic imaging using ambient light.” In: *Opt. Express* 16.26 (Dec. ts , = http://www.opticsexpress.org/abstract.cfm?URI=oe-16-26-21434, = 10.1364/OE.16.021434,), pp. 21434–21445.

[10] Philipp Rouast et al. “Remote heart rate measurement using low-cost RGB face video: A technical literature review”. In: *Frontiers of Computer Science (electronic)* 12 (Sept. 2018), pp. 858–872.

[11] Frank Marchak. “Detecting false intent using eye blink measures”. In: *Frontiers in Psychology* 4 (2013), p. 736. ISSN: 1664-1078.

[12] Sharon Leal and Aldert Vrij. “Blinking During and After Lying”. In: *Journal of Nonverbal Behavior* 32 (Jan. 2008), pp. 187–194.

[13] Tereza Soukupová and Jan Cech. “Real-Time Eye Blink Detection using Facial Landmarks”. In: 2016.

[14] Sepp Hochreiter and Jürgen Schmidhuber. “Long Short-Term Memory”. In: *Neural Comput.* 9.8 (Nov. 1997), pp. 1735–1780. ISSN: 0899-7667. DOI: 10.1162/neco.1997.9.8.1735. URL: https://doi.org/10.1162/neco.1997.9.8.1735.