Feasibility of Using Phone and Web Cameras to Detect Micro-Expressions for Lie Detection

Lanz Kendall Lim, Eden Paige Ong, Carl Arjan Ramos,
Mico Dellosa, and Fritz Kevin Flores
St. Edward School, General Trias City, Cavite

Abstract: This study explores the feasibility of using low-resolution cameras as a means of detecting facial movements for lie detection. Micro-expressions, however, are difficult to detect by the human eye due to their short duration and low intensity, thus the research explores the possibility of extracting micro-expressions from phone or web cameras that have low resolution and framerate. The collected videos are the processed using time series processing, to obtain both facial data points extracted from facial landmark detection models, as well as image generation from the obtained datapoints to produce a face structure. The classification mainly focuses on the use of common machine learning algorithms, to detect facial movement patterns, in the hopes of classifying people telling truths or lies. The tests ultimately proved to have a low accuracy in classification, but the results show that the methodology may contribute to other domains, such as in person identification, as well as possible recommendations for future works.

Key Words: Facial Landmarks, Image Transform, Machine Learning, Time Series.

1. INTRODUCTION

People have the innate ability to make facial expressions. Feelings of joy, sadness, anger, fear, etc., are expressed on faces via what is commonly interpreted as macro-expressions; however, there are also unseen movements in the face, known as micro-expressions. According to Ekman and Friesen (1969), micro-expressions are involuntary emotional responses that can be used to reveal one's true emotions and claimed that micro-expressions might be the most promising approach to detecting lies; however, micro-expressions are imperceptible to the untrained eye due to their short duration, with an upper limit of 0.5s, and low intensity. Zhang et al. (2014) described that micro-expressions are linked to the flow of expressions and occur unconsciously when emotions are concealed or repressed, expressing certain movement patterns which may aid in identifying individuals telling the truth or lies. Experts would classify micro-expressions according to the emotion being manifested, which is both tedious and time consuming.

Face detection programs have since then been investigated as a means of perceiving micro-expressions. Researchers use high-resolution cameras with framerates of 120 fps or above, to see subtle movements; however, these cameras tend to be expensive and are not always accessible. The study aims to investigate the feasibility of using low-resolution and low-framerate cameras as a means of capturing micro-expressions and facial patterns to classifying truth or lies.

2. Methodology

The methodology is comprised of the Preparation and Data Collection, Facial Landmark Detection and Preprocessing, and Image Generation and Time Series Processing phases.

2.1. Preparation and Data Collection

During this phase, participants undergo an Emotional Intelligence (also known as EQ) Test, to determine if the participant is capable of understanding and managing ones emotions, which may relate to a degree of emotional expressiveness and is used to determine eligibility for participation. This allows selected participants who are generally able to better express emotions, to participate in the study, in the hopes of being able to obtain more descriptive data for processing.

The fifteen selected participants are then oriented regarding the study and requested their consent for participation. They then proceeded to a recorded online interview, comprised of twenty (20) yes-or-no questions, of which they have the option to tell the truth or lie. At the end of the video, the participants are requested to label their answers, to serve as the ground truth.

Each video recording of the online interviews uses 60 fps and is processed by splitting the video per question, and further trimming each one by only retaining the segments where the participants answered. These video segments, containing the participants providing yes-or-no answers, are then labelled indicating if they are truth or lie, then are sent to the next phase.

2.1. Facial Landmark Detection and Preprocessing

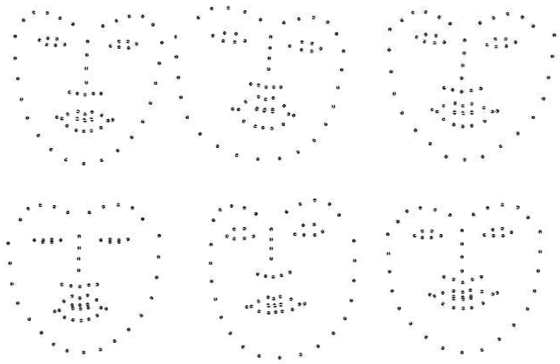

Figure 1. Samples of Facial Landmark Annotation from Video Segment Frames

The labelled video segments are run through a feature extraction algorithm using computer vision and image processing. The process begins by splicing the video segments into individual frames, allowing each frame to be subjected to the facial landmark extraction algorithm. The algorithm enables detection and extraction of sixty-eight (68) facial landmarks from each frame in as presented in Figure 1. This also determines the location of the whole face in the entire frame.

Since different participants are positioned differently in the frame and do not have a constant position across different video segments, even within the same interview, additional preprocessing was implemented. For each video segment, the size of the face relative to the chin and eyebrows, as well as the position of the face relative to the nose, were used as guidelines to perform affine transformation methods of scaling and translation respectively, and in addition, the images were also scaled to an arbitrary 1000px by 1000px image size. This is to normalize the size and position of the faces, across each video segment, for use of processing in the study.

2.2 Image Generation and Time Series Processing

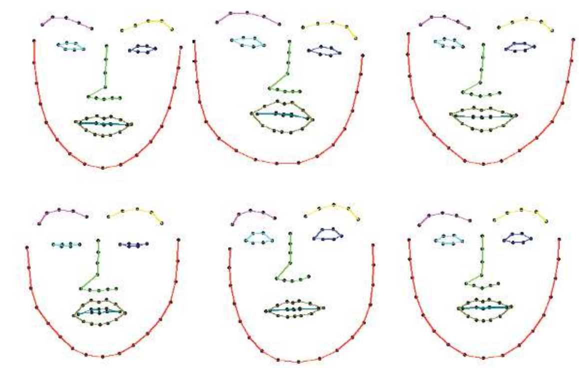

Figure 2. Samples of Facial Landmark Annotation with Facial Feature Annotation

The transformed facial landmark points form the video segments are then processed further to generate two datasets: the image dataset and time series dataset. The image dataset begins by annotating the facial landmark points by using colored lines to connect certain facial features, as seen in Figure 2. The colors and lines are assumed to be useful in processing, by guiding the machine learning algorithms of the different facial features, identified by a unique color scheme.

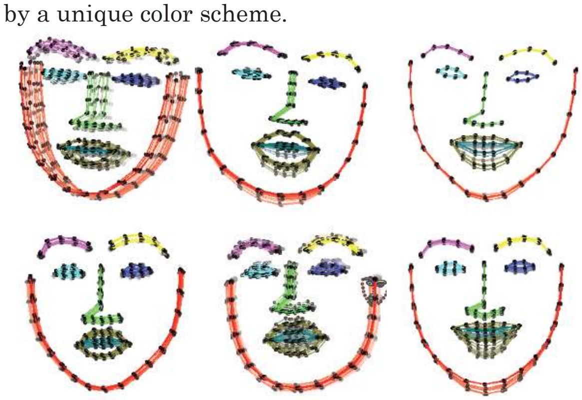

Figure 3. Samples of Facial Movement Annotation with Alpha Adjustment with Respect to Time

Since each video segment is a series of frames, an additional temporal aspect needs to be included, expressing information from the start frame to the end frame of each video segment. Therefore, the solution was to annotate all frames, but have a gradual increase of transparency from 0% to 100%, to represent a gradual transition from transparent to opaque, representing time.

Table 1. Subset from the Initial Facial Landmark Dataset from Feature Extraction

label	participantID	testNum	frameNum	jaw0-x	jaw0-y	jaw1-x	jaw1-y	jaw2-x	jaw2-y
L	0	1	1	714	452	715	503	721	552
L	0	1	2	742	427	742	475	747	521
L	0	1	3	773	439	773	487	776	532

The time series dataset on the other hand, was processed from the initial facial landmark points, containing both x-axis and y-axis for each of the 68 points, with a total of 136 features. Unfortunately, each video segment contains varying number of frames (rows), due to the participants having various durations in providing answers, thus requiring time series processing, to represent the entire sample in only a single row. The algorithm implemented to address the concern was to implement a MinMaxAve Scaling for time intervals with overlap.

The time interval used in the study was to split the video segment into three (3) equal parts, indicating the start, middle, and end of an answer, by dividing the number of frames into three (3) groups. An overlap between the start-middle and middle-end were also obtained using the same number of frames used on the split, creating a total of five (5) subsegments. Each subsegment contained the 136 features from the facial landmark points, and is further processed to get the minimum, maximum, and average of each subsegment, reaching a total of 2040 feature columns per sample. The entire dataset is then normalized using MinMax Scaler for machine learning.

3. TEST RESULTS AND DISCUSSIONS

The time series dataset was investigated using Support Vector Classifier and Linear Regression algorithms to determine its ability to classify the data between truth or lie.

Table 2. Support Vector Classifier Result

Label	Precision	Recall	F1-Score	Support
0 – Lie	50%	25%	33%	12
1 – True	62%	83%	71%	18
Accuracy			60%	30

Using different Train-Test proportions, the best result was obtained using a proportion of 90% train and 10% test. Hyperparameter tuning includes a grid search of the following values: C of 0.01, 0.1, 1, or 10; gamma of 0.1, 0.01, or 0.001; and kernel of rbf, poly, sigmoid, or linear; having the best hyperparameters with C=10, gamma=0.1, and kernel=linear. Unfortunately, with all tests implemented, the accuracy ranged from 47% to 60%, with the highest shown in Table 2 for Support Vector Classifier. Logistic Regression results ranges from 49% to 56%. This means that the classifiers have difficulty in determining the difference of true and lie from the dataset.

Another investigation was done by performing algorithms in dimensionality reduction via Principal Component Analysis to simplify the dataset features. Additionally, K-Means clustering was performed in order to determine if the clusters are able to distinguish some differences.

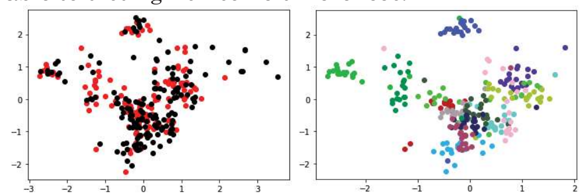

Figure 4. PCA Visualization for Truth and Lies (left) and by Participant (right)

Figure 4 present the results obtained from the PCA Visualization. The left figure uses Red points to denote lie samples and Black points for truth samples. It can be seen that there is not much differences with the truth and lie samples, hence the low accuracy obtained by the Support Vector Classifier and Logistic Regression algorithms. However, even though the images were already normalized through transformations in translation and scaling, figure on the right of Figure 4, presenting a unique point color per participant, still indicated visually observable grouping and clusters for each individual, indicating a similarity in the individuals data despite normalization.

This indicates the possibility that the methodology implemented in the study can, in a way, have the capability for other implementations, such as to identify a participant or individual from the rest, only through the use of the facial landmarks and facial movement of the individual.

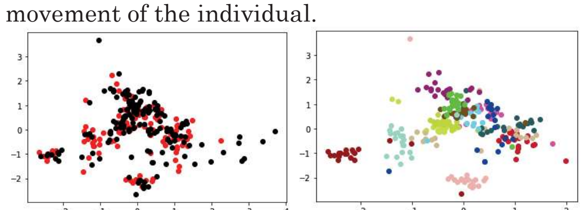

Figure 5. PCA with StandardScaler Visualization for Truth and Lies (left) and by Participant (right)

Additional testing of the hypothesis that the methodology can be used for studies in identification was implemented by implementing PCA one again, but this time, processing the data using Standard Scaling. The result in Figure 5, reveals a result that is somewhat similar to that of Figure 4, where the left image has some difficulty in identifying truth from lie samples, but the image on the right, still shows some form of clustering per participant.

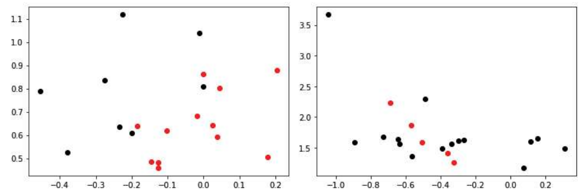

Figure 6. Sample PCA Visualization per Participant showing Visual Differences

Further observing the visualizations with regards to the individual participant data can be seen in Figure 6. The left image, obtained from one participant, seems show that there is a visually observable separation from the truth (black points) and lie (red points) samples of the participant. This could indicate quantitative evidence that the micro-expressions of an individual, when lying, is specific to that individual. This also merits further investigation on user-dependent and user-fold validation techniques as recommendation for future researchers.

4. CONCLUSION

The study performed various techniques from extracting facial landmark points from video frames, performing affine transformation methods of scaling and translation to provide uniformity to the data, implementation of time series processing via MinMaxAve scaling for time intervals with overlap due to the temporal aspect of the data, generate facial feature visualizations representing time using changes in transparency, as well as implementation of basic machine learning algorithms for classifications. Although the accuracy obtained in the study did not exceed 60%, even with simple binary classification, other results show some opportunity for improvement as well as use of the methodology in different applications. With limited dataset and resources, the study is still at its early stages, and more improvement can be made, such as normalizing the samples since the number of truth and lie samples in the data collection is uneven; further transformation of rotation and scaling for the images, since some participants tend to rotate and angle their faces; implementation of Convolutional Neural Network and other Deep Learning algorithms instead of simply using visual observation, and others.

5. REFERENCES

Agrawal , S., & Khatri, P. (2015). Facial Expression Detection Techniques: Based on Viola and Jones algorithm and Principal Component Analysis. 2015 Fifth International Conference on Advanced Computing & Communication Technologies, (pp. 108-112). Haryana. doi:10.1109/ACCT.2015.32

Ekman, P., & Friesen, W. V. (1969). Nonverbal Leakage and Clues to Deception. Psychiatry: Journal for the Study of Interpersonal Processes, 32(1), 88-106.

Yan, W.-J., Wang, S.-J., Liu, Y.-J., Wu, Q., & Fu, X. (2014). For micro-expression recognition: Database and suggestions. Neurocomputing, 136, 82-87. doi:https://doi.org/10.1016/j.neucom.2014.01.029

Zhang, M., Fu, Q., Chen, Y.-H., & Fu, X. (2014). Emotional Context Influences Micro-Expression Recognition. Plos One, 9(4). doi:10.1371/journal.pone.0095018