Human Profiling based Facial Micro Expressions and Speech Input

by Kalaivani Mam

Submission date: 11-Jan-2022 03:28PM (UTC+0530)
Submission ID: 1740030847
File name: Batch_29_ISE_Prakriti_1.docx (124.67K)
Word count: 3278
Character count: 18533

Human Profiling based Facial Micro Expressions and Speech Input

***Abstract*— Lie detection techniques include facial micro-expressions, body language, and voice analysis. Facial micro-expression is an uncontrollable reflex that cannot be faked. Stress affects liars, and it shows up in their mannerisms and vocal clues. Based on Paul Ekman's research, visual micro-expressions are acknowledged using Principal Component Analysis. Speech analysis is performed on the subject's input, and the findings are compared to the results of the face micro-expression detector. Stress analysis is performed on the human speech. This study also goes through the numerous bodily cues used by liars and presents a way for identifying an individual's mannerisms using the robust features approach to check if the subject has liar-like body language.**

***Keywords*—Micro-expressions, Deep Learning, Speech Input, Artificial Intelligence, Human psychology, Human Behaviour.**

I. INTRODUCTION

Micro-expression is a quick and fleeting facial expression that arises when people try to hide their true feelings, particularly in high-stakes circumstances. Micro-expression (micromomentary expression) was initially identified by Haggard and Isaacs, who saw it as suppressed emotions. Ekman discovered micro-expressions in an interview tape of a patient suffering from depression who attempted suicide. Since then, various studies in the subject of micro-expression have been done, although few results have been published.

Micro-expression has lately acquired prominence as a result of its possible uses in the diagnostic and national security processes. It is regarded as one of the most reliable indicators of deception and harmful behavior [2][3]. Screening Passengers by Observation Techniques has previously been used by the Transportation Security Administration in the United States, which was primarily based on the findings of micro-expression research.

In the therapeutic setting micro-expression might be utilised to decipher patients' true feelings and aid in the development of more effective treatments. Micro-expression, on the other hand, is regarded so ephemeral that it is practically invisible, making it impossible for humans to notice. Micro-expressions, which are far quicker than normal facial expressions and frequently overlooked [12], are described by Matsumoto as any facial expression that lasts less than 500 milliseconds.

An effective micro-expression detection system should be implemented to considerably minimise the amount of labour and time required to better apply micro-expressions in identifying falsehoods and risky behaviours. As a result, numerous academics have attempted to create an artificial micro-expression detection system that may assist individuals in detecting such transient facial emotions [5].

Micro expressions carry considerable and useful quantity of data regarding actual emotions that might be valuable in practical applications like security and interrogation. Due to the rapid movements of microexpressions, it is difficult to extract this information, and the characteristics must be more detailed.

One among the key features of micro expressions is their short time period along with the common norm being maximum of 500 milliseconds. This adds to the complexity. Micro expressions last less than 250 milliseconds, 330 milliseconds and less than half a second, according to several speed definitions evaluated [9][14]. A typical period considered is less than 200 milliseconds, following E and F as the first to identify a micro expression.

The key trait that distinguishes micro expressions from macro facial expressions is duration, which makes it harder in the following aspects:

• Humans have a hard time recognising microexpressions. This is because macro-expressions are larger and more distinct, whereas micro-expressions are smaller and more subtle muscle movements.

• Datasets Creation: Micro-expressions are harder to induce than macro-expressions. Micro-expression datasets that are currently accessible were induced in a laboratory curated setting.

• Macro-expressions may be captured with a standard camera. The speed and subtlety of micro-expressions, on the other hand, necessitate a high speed camera, which provides more noisy data than a standard camera.

• The creation of algorithms has a long and illustrious history. When compared to facial expression recognition, automated micro-expression recognition is very new (discovered work in 2009).

Despite the fact that both micro and macro expressions are linked by the element of face expression, these issues should be considered separate study concerns. Our goal is to give a complete assessment of micro-expressions as well as new discoveries.

II. LITERATURE SURVEY

a. Review on Facial Micro-Expressions Analysis: Datasets, Features and Metrics

In recent years, facial micro-expression (henceforth, micro-expression) analysis has been an important study topic. When a person tries to hide their genuine feeling, they exhibit micro-expressions. When a person becomes aware that they are making a facial expression, they may attempt to repress it since exhibiting the emotion may be inappropriate or violate a cultural display guideline. After the suppression, the person may cover over the original facial emotion, resulting in a micro-expression. In a high-stakes situation, these emotions are more common since there is more danger in displaying the emotion. Micro-expressions carry a considerable and effective quantity of information regarding actual emotions, which might be valuable in real world applications such as security and interrogation. It is hard to retrieve this detail because of the quick movements in micro-expressions, where the characteristics should be more comprehensive.

The challenge stems from one of the primary features of micro-expressions, which is their short duration, with the common standard being no more than 500 ms. Other speed definitions explored suggest that micro-expressions stay less than 250 ms, less than 330 ms, and less than half a second. Following in the footsteps of Ekman and Friesen, a typical time considered is less than 200 ms. The key trait that distinguishes micro-expressions from macro-facial expressions is longevity, which makes it more difficult than micro-expressions.

b. 3D Histograms of Oriented Gradients (3DHOG)

Polikovsky et al. demonstrated a method for recognizing face micro-expressions. They manually annotated areas on the face to divide it into 12 portions, then centred a rectangle on such points. To recognize motion in each location, 3Dhistograms of oriented gradients (3DHOG) were employed. A posed dataset of micro-expressions captured with a high-speed camera was used to evaluate this strategy (200 fps). In this study, 13 distinct micro-expressions were identified. micro-expression stages: muscle restrict, muscle growth (In-Action), and muscle release.

3DHOG was used again by Polikovsky and Kameda [30], however this time by using k-mean classifier and voting mechanism. It demonstrated an approach for detecting and tracking the timing of micro-expressions. Frame-by-frame classifications were performed to locate AUs in eight video cube sections. The Onset and Offset frames are more accurate than the Apex frame, implying that the concept is more suitable towards identification than classification in a static frame. To quantify AU temporal characteristics, the variations in bin values inside the 3D gradient orientation histogram were used to indicate the changes and motion accelerations of face movement. This temporal profile, they proposed, may be used to distinguish between posed and spontaneous micro-expressions.

b. Local Binary Pattern-Three Orthogonal Planes (LBP-TOP) and Variations

Pfister and colleagues suggested a system for distinguishing spontaneous facial microexpressions. To retrieve dynamic features, LBP-TOP was employed as a spatiotemporal local texture descriptor. Multiple Kernel Learning, Support Vector Machines and Random First were employed in the classification phase. This system was tested on an older version of SMIC, with data obtained from only six subjects and 77 microexpression specimens. To generate statistically reliable histograms, the Temporal Interpolation Model (TIM) was utilised to increase the number of frames. SMIC results were then compared to York Deception Detection Test (YorkDDT) results, which were captured at 25 frames per second with a resolution of 320240.

The approach was tested on two corpora using leave-one-subject-out (LOSO) and SMIC was down-sampled to 25 fps. They have two categories to choose from: emotional vs. non-emotional, and lying vs. true. YorkDDT's highest performance for classifying between first set is an accuracy of 76.2 percent utilising MKL and 10 frames. The best result for the second batch is 71.5 percent using MKL 10 frames and the same outcome using SVM. They assess SMIC as negative or positive, with the greatest result being 71.4 percent using MKL 10 frames and 64.9 percent using MKL and 15 frames for down-sampled SMIC.

Then, Pfister et al. devised a technique for distinguishing between spontaneous and posed facial expressions (Spontaneous Vs Posed (SVP)). They improved Guo et al. 's Local Bi nary Patterns (CLBP) to operate with dynamic texture descriptors and named it CLBP from Three Orthogonal Planes (CLBP-TOP). They assessed their suggested strategy by doing a leave-one-subject-out test on a corpus created by them, Spontaneous vs POSed (SPOS). In the session, this SPOS delivers both spontaneous and staged expression for the same subject. It has 7 subjects, 84 staged emotions, and 147 spontaneous expressions. The corpus was captured using two cameras, one for visible (VIS) data and the other for near-infrared data (NIR). Both cameras used 640480 resolution and 25 frames per second. As classifiers, SVM, LINEAR classifier (LIN), Multiple Kernel Learning (MKL), and fusion of SVM, LIN, and Random Forest via majority vote (FUS) were utilised. CLBP-TOP outperformed LBP-TOP with accuracy of 78.2 percent, 72 percent, and 80 percent on NIR, VIS, and combination, respectively.

Liu et al. [46] presented Main Directional Mean Optical-flow (MDMO) characteristics for micro-expression recognition. Their MDMO is made up of Regions of Interest (ROIs) that are partially based on AUs. One of the most notable benefits of MDMO is the modest features dimension, with the features vector length equal to 72, which is two features taken from each of the 36 ROI regions. To decrease noise caused by head motion, all frames have been aligned to the first frame. For detection, the SVM classifier was used. To test their approach, they employed the SMIC, CASME, and CASME II datasets. The result was better than the reference, which employed LBP-TOP and histogram of oriented optical flow (HOOF) features and achieved 68.86 percent, 67.37 percent, and 80 percent on CASME, CASME II, and SMIC, respectively.

Human spotting accuracy is estimated to be approximately 40%. Only with established measures can computer algorithms including machine learning and computer vision be assessed objectively. This section goes through the metrics that have been utilised in the literature.

We summarised and explained the assessment measures based on the full examination. Because the metrics for micro-expressions analysis are routinely employed for binary classification, they are appropriate for assessing True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) detections. More documentation on these measures may be found here. The majority of the outcomes in micro-expressions analysis in previous work are based on Accuracy, as stated in equation 1.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

At a later phase, performance was measured using the F1-Score (or F-Measure). Other measures, including as recall, precision, and the Matthews Correlation Coefficient (MCC), are being utilised to present the findings. Precision is a measure of exactness that determines a proportion of relevant replies from outcomes. Recall, also known as sensitivity, is a percentage of the findings that are relevant to the experiment and effectively retrieved.

$$Precision = \frac{TP}{TP+FP}$$

$$Recall = \frac{TP}{TP+FN}$$

Both of these indicators are commonly employed together to build an understanding of the significance of the results provided from experimental classification, therefore using them separately is improbable. The F-Measure is useful in calculating the harmonic mean between Precision and Recall, and it is utilised instead of accuracy since it gives a more complete examination of the data. The equation is as follows:

$$F\text{-}Measure = \frac{2TP}{2TP+FP+FN}$$

In terms of pre-processing approaches, hand-curated features and when necessary, machine learning (ML) classifications, the pipeline of a traditional micro-expression recognition strategy is essentially similar to that of a macro-expression recognition approach. Geometric feature-based approaches, on the other hand, are rarely employed since monitoring feature points on the face that scarcely changes would not yield effective output. Consequently, appearance-based attributes are mostly employed to attempt and try to characterise the micro-movement or to teach the algorithm to categorise micro-expressions.

The primary components for traditional techniques include spatial temporal configurations during information gathering, preparation state of data-set comprising face alignment and face regions split, feature extraction methods, and the kind of classifiers. Moving ahead, an end-to-end system capable of dealing with these difficulties is necessary. Deep learning approaches have yet to make a significant influence on micro-expression analysis; nonetheless, in order to offer a comprehensive evaluation of existing methodologies, we will conduct early research on deep learning (DL) and the relevance of it to micro-expressions.

We advise researchers to standardise the performance indicators they utilised in their review. Because the bulk of datasets are unbalanced, it appears that publishing the output of F Measure (or F1-Score) is actually a better decision. Using the traditional Accuracy metric may result in a bias or partial treatment towards classes with a large sample size, overestimating the capabilities of the examined technique.

The F-Measure micro-average is computed throughout the whole dataset and is based on the total true positives, false negatives, and false positives across 10-fold cross validation and Leave-one-subject-out (LOSO).[22] Researchers are encouraged to employ additional datasets for their research since each dataset contains little micro-expression samples. Unweighted average recall (UAR) and weighted average recall (WAR) are suggested for cross-dataset examination since they have been demonstrated to be effective in speech emotion identification.

WAR is described as the amount of properly categorised samples divided by the total number of samples, whereas UAR is defined as the sum of each class's accuracy divided by the number of classes without taking into account the number of samples per class. The outcomes of all folds are averaged to give the total scores. These measures were proposed during the First Micro Expressions Grand Challenge Workshop, which was held in connection with the Face and Gesture 2018 Conference.

III. PROPOSED MODEL

The proposed model consists of the below components:

1) Facial Micro Expression detector
2) Body Language Analyzer
3) Speech analysis during interrogation

Facial Micro expression Detection:

According to Dr Paul Ekman's research on facial micro expression, Facial Action Coding System (FAAC) which are-surprise, disgust, sadness, contempt, fear, happiness and anger have been termed as 'facial micro expressions'. A few specific characteristics of micro-expressions are shown in Fig 1.

The facial micro expressions usually always occur at a speed or rate of about 1/16 to 1/25th of a second and it is essential to use a high-speed camera that will be able to capture at least a minimum of 30 frames per second. After this the video will be broken down to a set/number of frames after which 'Facial Micro Expression' will be detected using-'Principal Component Analysis' technique. The speech input is converted to text using a speech-to-text convertor and is then sent for speech analysis. The result of the speech analysis will then be compared with the results obtained facial micro expression detector. Now, it will be checked for consistency.

Pre Processing:-

The visual appearance of any image will be subjected to noise removal, enhancement and also grey-scale conversion.

Feature Extraction and Face Detection: -

First facial region will be detected and then feature extraction will be done to reduce the dimensionality of the given input space. Principal Component Analysis can be used to recognize a particular statistical-pattern for data reduction and these extracted features play a big role in distinguishing of the input patterns.

Covariance matrix is calculated by subtracting the mean. After this the eigen vectors and eigen values are calculated. The components will be chosen and feature vector is then formatted. After this the image will be compared with the training image. The Euclidean distance between these is calculated and then feature vectors and the facial expression will be classified. After this 'Speech Analysis' is done.

Body Language Analysis:

From research, it has been understood that the stress experienced by a dishonest person usually manifests itself through several body language cues. Some of the most common examples of peculiar/exquisite body language of a liar are listed below:

1) When the hands touch the face or the mouth, these are some gestures that increase a lot when a person lies.
2) Touching of the nose is a very obvious cue.
3) Placing their hands over their mouth.
4) Biting their lips, fidgeting.
5) Clenched fists, crossed arms.

The system is pre-trained with many exquisite types of body language cues depicted by a liar and those shown by an honest person. The images that depict the body language of a liar and an honest person are subjected to the 'Limb Action Model Convertor'.

'Speeded Up Robust Features'(SURF) approach will then be used to classify the testing images that are provided after testing.

Analysis Of Speech:

Speech of the testing subject will be recorded during the interrogation and is then converted to text via a speech-to-text convertor. The converted text will be subjected to speech analysis after this. The speech analyzer uses Java in BlueJ environment and the speech recorded during interrogation will then be made to text and this text is checked for qualifiers, gulping, clearing, fillers of the throat, stammering etc. If such suspicious verbal cues are found to be used for multiple times after which the user will be notified.

The speech analyzer can also find out the emotional state that the user depicts during the interrogation by using the dictionary approach and is checked for consistency with the type of micro-facial expression. If the results do not match then the user is notified that the subject has a possibility of lying.

Fear Microexpression Raised Eyebrows that are drawn together. Wrinkles in the center. Lips retracted	**Anger Microexpression** Flaring nostrils Furrowed Brow Mouth Compressed	**Happiness Microexpression** Skin under eyes wrinked Corners of the lips are drawn back and up. Crows feet near the outside of the eyes.	**Surprise Microexpression** Raised Eyebrows Horizontal wrinkles across the forehead. Mouth Open; Eyes wide Open Jaw drops
Disgust Microexpression Raised Upper lip Wrinkled Nose Lower lip turned down	**Contempt Microexpression** One side of the mouth raises. Partial Closure of eyelids Eyes are turned away	**Sadness Microexpression** Lower lip pouts Inner corners of the eyebrows are raised. Corner of the lips are drawn down.	

Fig 1

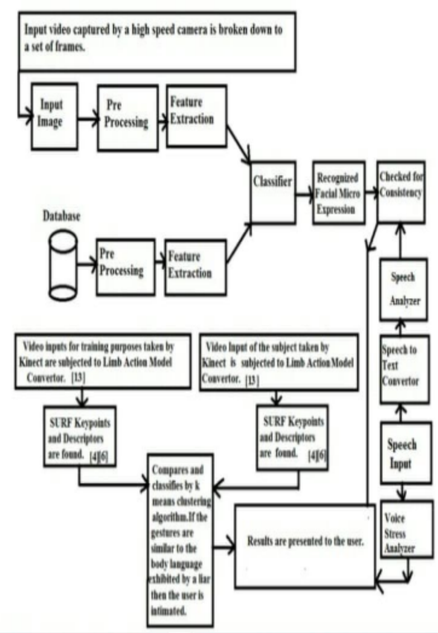

Fig 2

IV. CONCLUSION AND FUTURE WORKS

A technique for detecting a liar based on face micro-expression, body language, and interrogation voice analysis is addressed in this work. The speech analyzer's output is compared to the face micro-expression detector's output to ensure that they are in sync. If the findings do not match, the user is told. The system has shown to be reliable, efficient, and cost-effective. It's useful in a variety of situations, including security, investigations, counter-terrorism, and interpersonal interactions.

A body language analyzer will be built that analyses the subject's motions during interrogation with the usual body language of a liar and alerts the user if they are comparable. Images of a liar's and an honest person's usual body language clues will be utilized as training images.

The Kinect video clipping of the subject will be converted to a sequence of frames, and the photos will be run through Limb Model Convertor. In both the training and testing pictures, key descriptors and characteristics must be determined. To identify the testing photos, the SURF (Speeded Up Robust Features) technique will be employed. The K-Means algorithm will be used to match the photos. The user will be told if the subject's motions match the body language indications of a liar. According to Carol Kinsey Goman's article 12 Ways to Spot a Liar at Work, inconsistency between what a person is saying and his or her gestures comes as a result of psychological stress experienced when lying.

When saying yes, you could shake your head from side to side, and when saying no, you might nod. This may be identified by comparing the displacement of the subject's head between multiple frames acquired by a high-speed camera and, if detected, noting the direction of displacement to determine whether the individual is nodding or shaking his or her head from side to side. The voice to text converter's verbal output must be evaluated, and if it is found to be contradictory with the subject's head motions, the subject is lying. After telling a falsehood, a person's blink rate increases up to eight times the normal rate. This may also be noticed by comparing the position of the subject's eyelids in different frames of the video.

By training the system to integrate new behavioural research findings, it may be made more efficient. Future studies will entail developing a cost-effective and efficient portable non-invasive lie detector. Head shakes while saying yes and nodding while saying no are two instances.

Human Profiling based Facial Micro Expressions and Speech Input

ORIGINALITY REPORT

39%	38%	6%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	deepai.org Internet Source	18%
2	www.ijert.org Internet Source	17%
3	arxiv.org Internet Source	2%
4	Submitted to TAR University College Student Paper	1%
5	repository.sustech.edu Internet Source	<1%
6	www.isca-speech.org Internet Source	<1%
7	Hemin Ibrahim, Chu Kiong Loo, Fady Alnajjar. "Speech Emotion Recognition by Late Fusion for Bidirectional Reservoir Computing With Random Projection", IEEE Access, 2021 Publication	<1%

8	Hua Lu, Kidiyo Kpalma, Joseph Ronsin. "Motion descriptors for micro-expression recognition", Signal Processing: Image Communication, 2018 Publication	<1 %
9	www.arxiv-vanity.com Internet Source	<1 %
10	Sai Prasanna Teja Reddy, Surya Teja Karri, Shiv Ram Dubey, Snehasis Mukherjee. "Spontaneous Facial Micro-Expression Recognition using 3D Spatiotemporal Convolutional Neural Networks", 2019 International Joint Conference on Neural Networks (IJCNN), 2019 Publication	<1 %
11	Walied Merghani, Adrian K. Davison, Moi Hoon Yap. "The implication of spatial temporal changes on facial micro-expression analysis", Multimedia Tools and Applications, 2019 Publication	<1 %

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		