JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 37, 793-808 (2021)
DOI: 10.6688/JISE.202107_37(4).0004

Contactless Deception Detection System with Hybrid Facial Features

JING-MING GUO[1], CHIH-HSIEN HSIA[2,+], LI-WEI HSIAO[1] AND CHEN-CHIEH YAO[1]
[1]*Department of Electrical Engineering*
National Taiwan University of Science and Technology
Taipei, 106 Taiwan
[2]*Department of Computer Science and Information Engineering*
National Ilan University
Yilan, 260 Taiwan
[+]*E-mail: chhsia625@gmail.com*

Facial deception detection has become a popular and challenging problem. In this study, an effective system is proposed to address this issue based on visual clues. The Parametric-Oriented Histogram Equalization (POHE) is presented to enhance image contrast and reduce the noise effect. A random forest classifier is applied to track the facial landmark points, and they are subsequently utilized to analyze the facial action unit based on the movement of the facial feature points. In addition, the geometrical features are also considered, and then the Sequential Forward Floating Selection (SFFS) is integrated to select the best feature combinations. To verify the extracted features for deception and truth identification, the Support Vector Machine (SVM) is applied. Experimental results demonstrate that even under uncontrolled factors, *e.g.*, illumination, head pose, and facial sheltering, the proposed method is consistent in achieving an effective recognition results and provides superior performance than that of the state-of-the-art methods.

Keywords: facial deception detection, support vector machine, parametric-oriented histogram equalization, sequential forward floating selection, biometrics image

1. INTRODUCTION

Though human cognition has evolved dramatically, the ability to detect deception is no more accurate than chance or flipping a coin. The applications are applicable to various community: Students, psychologists, judges, job interviewers, and law enforcement personnel [1]. Particularly, when investigating crime, the ability to detect deception accurately is critical for the police who must get criminals off the streets instead of detaining innocent suspects. Typically, deception detection assumes liars can exhibit some implicit cues, caused by their guilt and pressure about deception. Thus, it assists researchers to look for reliable behavioral evidences of deception. Some existing lie detection approaches are based on posture shifts, and foot and hand movements, yet they yield poor detection rate. Practically, it is difficult to train police personnel with a large number of case studies, and it is impossible for them to judge unconditionally. In this study, a new approach is proposed which can perform automated computer vision-based deception identification using facial clues. As machine decisions are rather consistent, this can avoid bios on decision making. Deception detection can be separated into contact and contactless approaches. Contact approaches normally utilize the polygraph and functional Magnetic Resonance Imaging (fMRI) which detects autonomic reactions [2]. These changes in body functions cannot be

Received September 26, 2020; revised November 22, 2020; accepted February 25, 2021.
Communicated by Ching-Chun Huang.

controlled easily by the conscious mind, including bodily reactions such as skin conductivity and heart rate. Having said that, it is not considered totally reliable. If the respondent already knows that it is to test lying, the physiological index will inevitably be affected by the environment and become less natural. In addition, it is inconvenient and restricted since one needs to wear the devices.

This study leans to explore contactless approaches to detect deception. In [4], it developed eye-tracking technology based on an emotional reaction similar to that of the polygraph but rather on a cognitive reaction. Subsequently, voice risk analysis or voice stress [5-8] analysis uses computers to compare pitch, frequency, intensity and micro tremors and can detect minute variations in the voice corresponds to signal lying or deception. Pérez-Rosas *et al.* [9] proposed a method extracting features on the linguistic and gesture modalities. Yet, only verbal and non-verbal features are involved, and it omits the most discriminative visual clues such as face for description. Jaiswal *et al.* [10] proposed a method considering the facial behavior and a lexical analysis on the spoken words to extract features. Besides, vision, audio, and text in detect micro-expression. Wu *et al.* [28] considers both human facial changes and audio transmitted by speaker in sequential input to evaluate the accuracy of micro-express detection. Ding *et al.* [29], use gestures and facial emotions can be strong features to determine lying reactions while communication. During facial alignment in dynamic motions from videos, explicitly detect micro-expressions as well. In [30] gave combinations of extracting video, audio, text for micro-expression in detection. However, the voice risk analysis method fails when the volume in the clips is too small or the background noise is too loud. The primary drawbacks of the aforementioned methods are that they are formulated based on verbal, non-verbal and voice analysis. Yet, with the introduction of vision devices, the most discriminative feature such as facial and visual clues or behavior can be accurately captured. With properly designed algorithms, a significant improvement can be achieved in terms of the recognition capability.

The two key factors in a biometric identification system are its high identification rate and convenience of device usage [27, 31]. This study focuses on the facial analysis for deception detection to compensate for the above-mentioned issues. In the proposed system, it first enhances the image contrast with the proposed POHE [27] and detects the facial landmark points [11]. Subsequently, this work extracts the geometrical features of the face which analyze the facial action unit, and then produces temporal profiles of each facial movement. The facial action unit measures the movement based on the facial landmark points that correspond to a displayed emotion. This work also extracts the geometrical features to represent the physiological response of the participants. Moreover, the SFFS [12] is utilized to select the best feature combinations. Finally, the SVM [13] is applied for recognition purpose. Experimental results show that the proposed deception detection method achieves excellent performance on the test dataset, which in turn suggests that the proposed method is an attractive candidate for practical deception or lying applications.

2. ALGORITHM DESCRIPTION

Fig. 1 illustrates the flow of the proposed method in training phase. First, the face localization is employed for the face detection [11] as the red rectangle shown in Fig. 2. Subsequently, the facial landmark extraction [11] is adopted for the required 68 facial

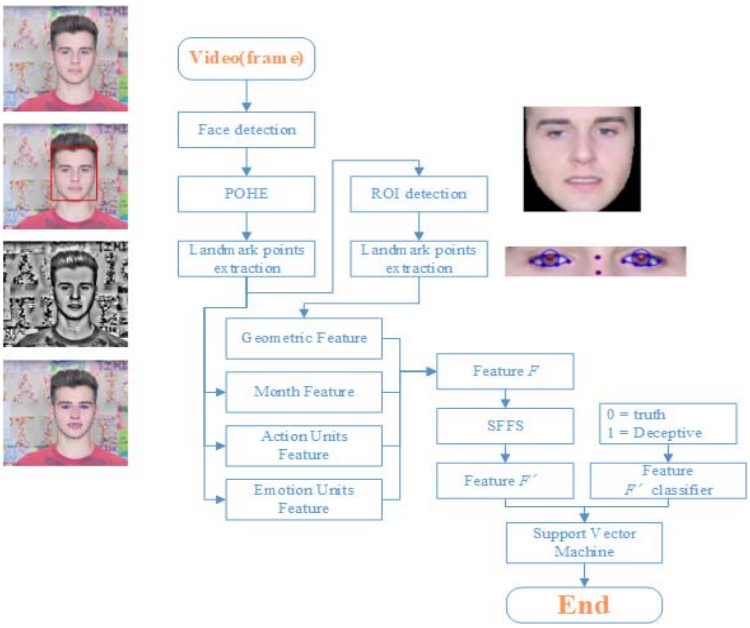


Fig. 1. The flow of the proposed method in training phase.

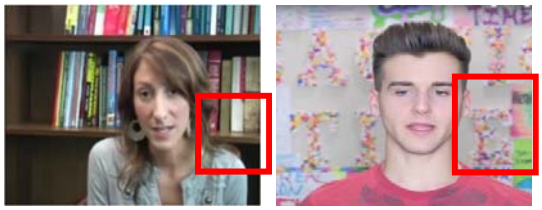


Fig. 2. Facial region.

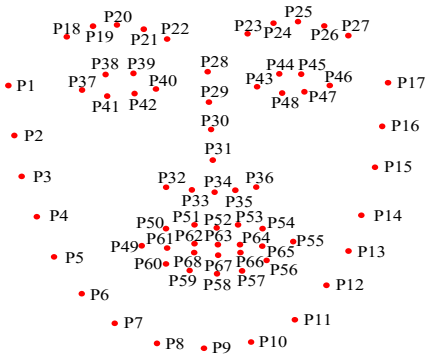


Fig. 3. 68 Facial landmarks.

landmark points with notations {P1, P2, ..., P67, P68} as indicated in Fig. 3, which is able to assist in extracting the following features expeditiously. Some samples of the extracted landmark points of facial images are reported in Fig. 4. To extract accurate eye features,

the captured facial images is further processed by extracting 28 feature points of the eye according to eye-specific Conditional Local Neural Fields (eye-specific CLNF) [14], as shown in Fig. 5. After that, a vector (F) is extracted for two kinds of characteristics are considered as features in this work as follows. (1) The movement of the irises and mouth can be captured by facial landmark extraction. It is able to distinguish whether a person is lying or not according to the movement of the irises, since the liars usually look around to cover their nervousness and anxious when they are telling lies. The movement of the mouth can be used to identify a person's speaking interval, which becomes an argument to the SVM algorithm in deception classification; (2) Owing to the guilty suspects make fake sadness or other emotions to cover their embarrassment when they are lying, the movement of the facial landmark points can observe the facial behavior and the emotions of the liars and truth-tellers. The description of the above-extracted features (F) is detailed in Section 3. Subsequently, the SVM is applied to detect the deception and truth with feature vector F' which is refined by the SFFS.

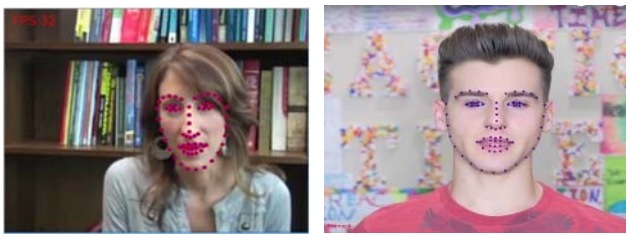
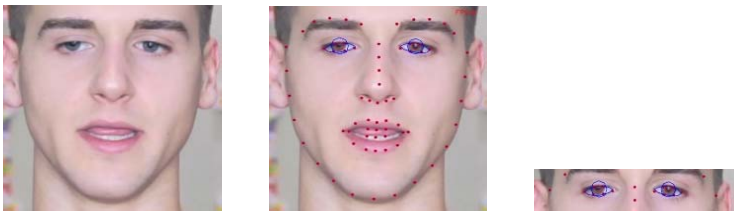


Fig. 4. Detected samples of landmark points of facial images.



(a) Face detection. (b) Landmark extraction. (c) Eye-specific CLNF.

Fig. 5. 28 feature points of the eye according to eye-specific CLNF.

3. FEATURE EXTRACTION METHODS

This section elaborates the components of the feature vector F which is previously defined in Section 2. The feature vector F can be categorized into two different properties: (1) Geometric features (G); (2) Action (A); and (3) Emotion (E) Units. The description of these features is further detailed as follows. Finally, the SVM is employed for deception classification.

3.1 Geometric Features

The feature G represents two geometrical properties: the movement of the eyes (G_{EY}) and the movement of the mouth (G_{MT}). To obtain the movement of the eyes as mentioned

in Section 2, the recently proposed eye-specific CLNF [14] is employed for iris detection, eye detection, and tracking. The CLNF [15] is a novel instance of the Constrained Local Model (CLM) [16] that was employed for optimization function and advanced patch experts. Moreover, eye-specific CLNF is a state-of-the-art shape-based method which can robustly adapt the ellipses to the boundary of iris by using image-aware RANSAC [17]. As a result, eye-specific CLNF is utilized in this study to detect and track iris as well. Some samples for detecting iris are shown in Fig. 6. The blue circles in each figure are the boundary of left and right iris.

Fig. 6. Samples of the detecting eye in the Real-Life dataset.

The eye-specific CLNF is adopted to locate the required 16 landmark points of the iris with notations $\{I_1^L, \ldots, I_8^L, I_1^R, \ldots, I_8^R\}$ as indicated in Fig. 7. Moreover, it locates the required 20 landmark points of the eye with notations $\{E_1^L, \ldots, E_8^L, E_1^R, \ldots, E_8^R\}$ as illustrated in Fig. 8. The landmark points of the iris (I) and eye (E) are employed for subsequent feature extraction. Specifically, the notations L and R denote the left-eye and right-eye, respectively. Notably, since the feature extraction process for the left and right iris as well as the left and right eye are identical, only the left iris and the left eye are considered to ease the presentation. Features extraction for the right iris and the right eye can be carried out by following the same procedure. As a result, the iris and occur landmark points can be utilized to extract features, which is the variance between the center point of iris and the center point of ocular region. Subsequently, the center points of the left iris (x_i^L, y_i^L) are defined as follows:

$$(x_1^L, y_1^L) = ((I_{1x}^L + I_{5x}^L)/2, (I_{1y}^L + I_{5y}^L)/2) \tag{1}$$

$$(x_2^L, y_2^L) = ((I_{3x}^L + I_{7x}^L)/2, (I_{3y}^L + I_{7y}^L)/2) \tag{2}$$

$$(x_i^L, y_i^L) = ((x_1^L, y_1^L)/2, (x_2^L, y_2^L)/2) \tag{3}$$

where points denote the $I_1^L = (I_{1x}^L + I_{1y}^L)$, $I_3^L = (I_{3x}^L + I_{3y}^L)$, $I_5^L = (I_{5x}^L + I_{5y}^L)$, and $I_7^L = (I_{7x}^L + I_{7y}^L)$ as shown in Fig. 7. The red dots are the center points of the left iris (x_i^L, y_i^L) and the right iris.

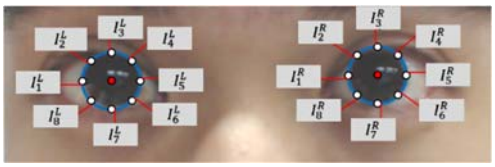


Fig. 7. Landmark points of the iris region.

Moreover, the center points of the left occur region (x_e^L, y_e^L) is defined as follows,

$$(x_1^L, y_1^L) = ((E_{1x}^L + E_{7x}^L)/2, (E_{1y}^L + E_{7y}^L)/2) \tag{4}$$

$$(x_2^L, y_2^L) = ((E_{4x}^L + E_{10x}^L)/2, (E_{4y}^L + E_{10y}^L)/2) \tag{5}$$

$$(x_e^L, y_e^L) = ((x_1^L, y_1^L)/2, (x_2^L, y_2^L)/2) \tag{6}$$

where points denote the $E_1^L = (E_{1x}^L, E_{1y}^L)$, $E_4^L = (E_{4x}^L, E_{4y}^L)$, $E_7^L = (E_{7x}^L, E_{7y}^L)$, and $E_{10}^L = (E_{10x}^L, E_{10y}^L)$ as shown in Fig. 8. The red dots are the center points of the left eye (x_e^L, y_e^L) and the right eye.

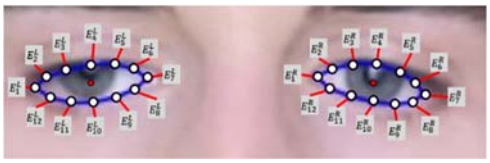


Fig. 8. Landmark points of the ocular region.

From our observation, the liars tend to look around to cover their embarrassment when they are lying. Consequently, the movement of the iris can be treated as a reliable feature. For this, the movements of the irises (I_m) are regarded as the features $I_m = \{m^L, m^R\}$ for detection. To derive the movement of the left iris, (x_i^L, y_i^L) as labeled in Fig. 6 is additionally located to describe the center point of the left iris; (x_e^L, y_e^L) as illustrated in Fig. 7 is additionally located to describe the center point of the left eye. Therefore, the movement m^L is derived from the $|(x_i^L, y_i^L)(x_e^L, y_e^L)|$, denotes its distance, which is not affected by the facial movement and head pose. Since this lie detection system is reading a video, a series of data can be obtained from each frame. Subsequently, the iris displacement can be extracted frame by frame. The energy represents the accumulation of the movement of iris, *i.e.*, m_t^L, within each frame, where t stands for the current frame. The variable eyemoving is to count the number of times for iris displacement. Subsequently, the average number of times for iris displacement over t frames is computed, denoted as I_{Avg}. The variable t is less than 100.

$$\begin{cases} energy = energy + m_t^L \\ energy = energy - m_{t-10}^L, \text{ if } (t\%100 > 9) \\ eyemoving + 1, \begin{cases} \text{if } (m_t^L \geq energy * 0.1 - 1) \\ \text{if } (m_t^L \leq energy * 0.1 - 1) \end{cases} \end{cases} \tag{7}$$

$$I_{Avg} = eyemoving/t, t \leq 100 \tag{8}$$

The obtained eye movement average displacement value is subtracted from the current eye movement displacement value. If the subtraction value reaches a certain level, it can be assumed that the iris is drifting. Subsequently, binary sequence events can be received per frame depend on whether it drifts. Finally, the SVM is utilized for deception classification. The 68 face feature points are obtained by using the face and feature point detection to obtain the detailed information of the mouth opening and closing, and the 20 feature points of the mouth part as shown in Fig. 9 are used for operation as detailed as follows.

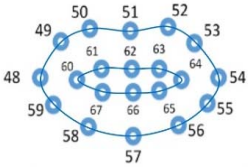


Fig. 9. Landmark points of the mouth region.

Subsequently, we can obtain the displacement of the mouth according to the corresponding feature points above and below. The mouth opening distance can be obtained by the outer lips D_1^o, D_2^o, D_3^o and distance by the inner lips D_1^i, D_2^i, D_3^i, as shown in Fig. 10.

$$D_1^o = |(x_{50}, y_{50})(x_{58}, y_{58})|,\ D_1^o = |(x_{51}, y_{51})(x_{57}, y_{57})|,\ D_3^o = |(x_{52}, y_{52})(x_{56}, y_{56})|,$$
$$D_1^i = |(x_{61}, y_{61})(x_{67}, y_{67})|,\ D_2^i = |(x_{62}, y_{62})(x_{66}, y_{66})|,\ D_3^i = |(x_{63}, y_{63})(x_{65}, y_{65})|$$

where point denotes the $N = (x_n, y_n)$. Two distance values are included on the mouth, and the mouth opening ratio (P^M) is defined as follows to help us determine if a person is talking.

$$P_n^M = D_n^i / D_n^o \tag{9}$$

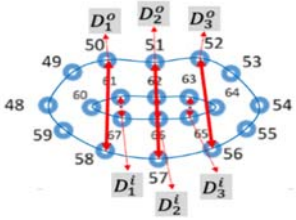


Fig. 10. Distance of opening mouth at feature points.

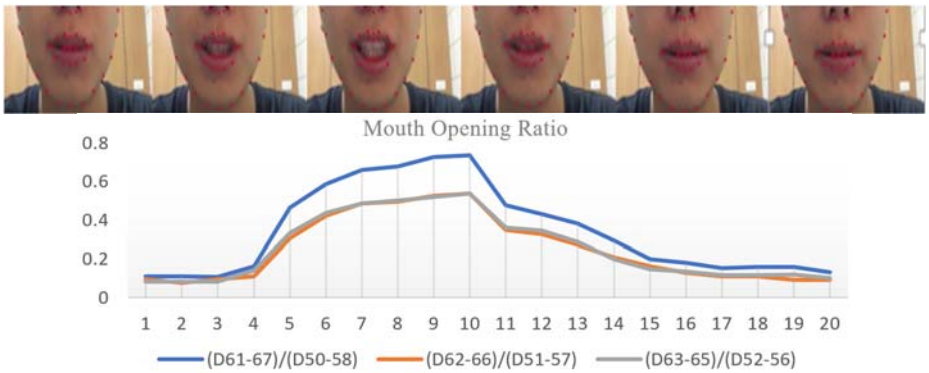


Fig. 11. Mouth opening ratio of each frame.

In Fig. 11, the curve from left to right is the process of speaking, and it shows the ratio of the mouth opening we measured. When the mouth opening ratio is greater than a certain threshold, then it can be determined that the person is opening the mouth and talk. After the testing, the threshold is preferably about 0.09, as shown in Fig. 11. The algorithm

defines a person's speaking interval according to the continuous movement of the mouth and distinguish the lines between his words by only reading the image.

The algorithm for judging speech is based on continuous images. Since a person speaks, the mouth is opened and closed repeatedly. Thus, a count value is set to determine the state of speech. When a person's mouth is in an open state, increase the value; otherwise, multiply this value by a buffer value, which is located in between 0 and 1. When the count reaches a certain value, it can be assumed that this is a paragraph. According to this algorithm, as long as a person keeps talking, it can be considered as a passage, and thus achieves a lie test for each sentence, not just a video.

3.2 Facial Action Unit

The Facial Action Coding System (FACS) [18] refers to a set of facial movements that correspond to a displayed emotion. Some examples of the Action Units are shown in Fig. 12. We can determine the displayed emotion of a participant using FACS. This is currently the only available technique for assessing emotions in real-time. Action Units have been employed to be potential observations for distinguishing the liars or truth-tellers in recent years. For instance, Porter *et al.* [19, 20] and Owayjan *et al.* [21] have shown that guilty suspects or liars make fake sadness or other emotions to cover their embarrassment when they are telling lies. Su *et al.* [22] have found that the potential indicators, *e.g.*, eye blinking, eyebrow motion and mouth motion, can also distinguish the liars or truth-tellers. Consequently, this study utilized FACS to measure the psychometric or deceptive tests as the true feeling in direct response of a participant.

Fig. 12. Examples of some action units extracted from Cohn and Kanades dataset [26].

The FACS can be categorized into two different properties: (1) Main Action Units (the feature A); (2) Emotions Units (the feature E). Each AU is associated with the facial movement and can affect in a motion of a part of the face or appearance changes in a facial region. In addition, multiple AUs can occur at the same time. To conclude, the proposed method aims to detect the Action Units and Emotions Units as summarized in Tables 1 and 2. Emotion Units are introduced when multiple Action Units show simultaneously. Subsequently, these potential deception indicators (AUs and EUs) are used to distinguish the deceptive and truthful suspects. These Action Units presence detection module is based on a recent state-of-the-art AU recognition framework [23, 24]. A more detailed description of the detection system can be found in Baltrusaitis *et al.* [24, 25]. The description of the method on feature extraction is detailed as follows.

First, the presence of each AU is extracted frame by frame, then subsequently calculate the presence of each EU using AU simultaneously as shown in Table 2.

Table 1. Potential indicators of deception.

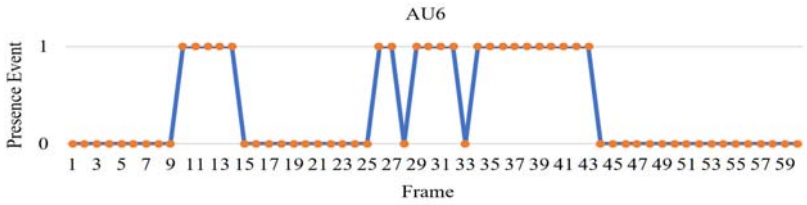
Action Unit	Description	Facial Region
AU1	Inner Brow Raiser	Eyebrows
AU2	Outer Brow Raiser	Eyebrows
AU4	Brow Lowerer	Eyebrows
AU5	Upper Lid Raiser	Eyes
AU6	Cheek Raiser	Eyes
AU7	Lid Tightener	Eyebrows + Eyes
AU9	Nose Wrinkler	Eyebrows + Nose
AU10	Upper Lip Raiser	Mouth
AU12	Lip Corner Puller	Mouth
AU14	Dimpler	Mouth
AU15	Lip Corner Depressor	Mouth
AU16	Lower Lip Depressor	Mouth
AU17	Chin Raiser	Mouth
AU20	Lip stretcher	Mouth
AU23	Lip Tightener	Mouth
AU26	Jaw Drop	Mouth
AU28	Lip Suck	Mouth
AU45	Blink	Eyes

Table 2. Potential indicators of deception (emotion units).

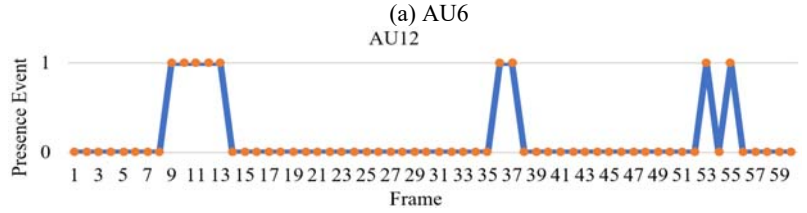
Emotion Unit	Description
AU6+12	Happiness / Joy
AU1+4+15	Sadness
AU1+2+5+26	Surprise
AU1+2+4+5+7+20+26	Fear
AU4+5+7+23	Anger
AU12+14	Contempt

Second, a binary sequence event is generated for each frame with each AU and EU, *e.g.*, AU6, AU12 and AU6+AU12 (EU) as shown in Fig. 13. A binary sequence event where one represents the frame involving the presence of the Action Unit or Emotion Unit and zero is not involving the presence. Finally, the AU and EU are extracted as be our features. Notably, we extracted the features with a number of frames (α) in the Real-Life dataset, where α is later discussed. The features are separated into two parts which are the sum of the present event (the binary sequence shows one) and the sum of the change of the AU and EU event (the binary sequence shows one to zero or zero to one). Fig. 13 shows the sequence of the present event of the AU6, AU12, and AU6+12. The AU6+12 means that the Emotion Unit is happiness or joy which is involving the presence of AU6 and AU12 simultaneously as shown in Fig. 13. Each orange dot on the curve corresponds to a frame in Fig. 12. The sum of the presence of the AU6, AU12, and AU6+12 extracted features are 9, 21, and 6 with 60 frames, respectively. Moreover, the sum of the change of the AU6, AU12, and AU6+12 extracted features are 8, 8, and 4, respectively. In this paper, the facial action unit and emotion unit are used to generate visual vector features (A+E) for analysis of 46 dimensions.

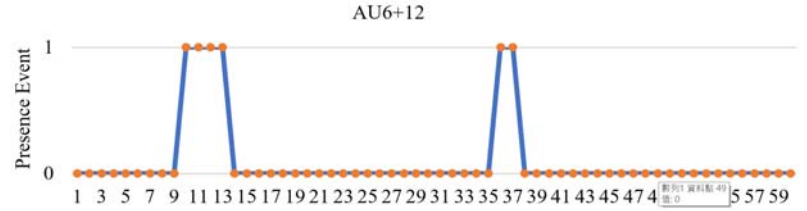
Subsequently, all of the extracted Action Unit and Emotion Unit features are fed to the SVM for deception classification.



(a) AU6



(b) AU12



(c) AU6+12 (EU)

Fig. 13. A binary sequence event where one represents the frame involving the presence of the AU or EU and zero is not involving the presence.

4. EXPERIMENTAL RESULTS AND DISCUSSION

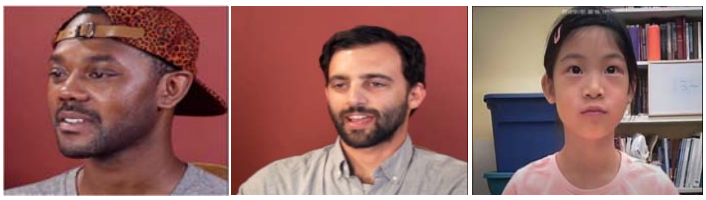
In the following experiments, two datasets, the Real-Life Trail Data [9] and the MSP-YTD, are adopted to evaluate the performance of deception detection using the proposed algorithm. The Real-Life dataset comprises of 28 deceptive and 38 truthful videos are from identified public multimedia sources, where some sample screenshots are shown in Fig. 14.

(a) Deceptive clips.

(b) Truthful clips.

Fig. 14. Sample screenshots showing facial displays from Real-Life Trail clips.

The average deceptive and truthful video lengths are 24.96 seconds and 27 seconds, respectively. The dataset composes of 21 female and 35 male speakers and are aged between 16 to 60 years. As mentioned in [9], three different trial results were utilized to correct label video clip as deceptive or truthful: guilty verdict, non-guilty verdict, and exoneration. For guilty verdicts, deceptive and truthful videos were collected from a defendant and witnesses in a trial, respectively. In some cases, deceptive videos are collected from a suspect denying a crime he committed while truthful clips are taken from the same suspect when answering questions concerning some facts that were verified by the police as truthful. Exoneration testimonies are assembled as truthful statements. On the other hand, the MSP-YTD dataset consists of 145 videos including 62 deceptive and 83 truthful videos sourced from various YouTube channels. The average video lengths of the deceptive and truthful are 9.9 sec. and 5.1 sec., respectively. The database consists of 15 female and 20 male participants. The video includes a clip of celebrity called a press conference but was verified to be a fraud later on by the police, a clip of polygraph testing the participants were lying or not and a clip of children to lie cause by some incidents. Some sample screenshots of the MSP-YTD dataset are shown in Fig. 15.



(a) Deceptive videos.

(b) Truthful videos.

Fig. 15. Sample screenshots showing facial displays from MSP-YTD dataset.

In our simulation, the 3-fold cross-validation is applied to the datasets for performance comparison. The detection accuracies discussed in this section highly depend on the detected landmark points, which was introduced in Fig. 4. It suggests that a failure in detecting the landmark points can lead to inability to detection. To this end, Table. shows the successful detection probabilities of the landmark points under these datasets. As it can be seen, a reliable result is achieved by using facial landmarks, and it ensures a consistent cognition between the values shown in the following experiments as well as the actual performance of the proposed method. The length of the videos is ranged from just a few seconds to about thirty seconds. In Section 3.1, it is mentioned the use of biometrics about the mouth to cut out the passages in the video. A paragraph is divided into K-segments and divide it once every 100 frames. The proposed method extracts the feature F every fragment; consequently, the feature at kth fragment of nth video F_n^k is extracted in all videos. Notably, the size of K is different in each video. Finally, the majority decision of the SVM

classifier is applied to classify each feature F_n^1, …, F_n^K. The feature F_n^1, …, F_n^k are then utilized to classify the n video by the majority decision. The accuracy (ACC) is adopted as the metric for the evaluation as follows.

Table 3. Probability of landmark points detection.

Datasets	Probability of detection
Real-life trail data [9]	0.9755
MSP-YTD dataset	0.9843

$$ACC = N_{corr}/N_{all} \tag{10}$$

where N_{corr} denotes the number of the correct classification video, N_{all} is the number of videos.

In this work, the 3 cross-validation method is utilized to randomly divide the film into 3 groups of data for training and testing of SVM. Because the lengths of the videos are not unified, the features of honesty or lying in each film are different. The best accuracy can be derived from these random samples that were generated by the random combinations of two datasets, Real-Life Trail Data [9] and MSP-YTD. The overall accuracy is computed through 3 folds of databases. The more balanced distribution of the positive and negative samples in each fold of data, the more reliable and robust training model for testing we can obtain from SVM classification. The original frame sizes are 640×480 and the format of a color image frame is 24-bit in an RGB system. All gray level frames are used, by transferring the RGB system to the YCbCr system. It is used for the proposed system for the deception detection of feature in real-time. The experimental environment is established using a CPU i7-8750H, 8 GB RAM, Microsoft Windows 10 and Open CV, OpenFace, Visual studio 2015. There are chosen as the software development platform. The frame rate for the proposed system is 50 FPS (Frame per Second). For the comparison, the corresponding performances of the former methods and the proposed method for the Real-Life dataset are shown in Table XX. In addition, the performance of the proposed algorithm for the MSP-YTD dataset and the combination of the MSP-YTD dataset and the Real-Life dataset are shown in Table XX. It also contains the performance ablation test of various feature configurations with the proposed method. To have an in-depth exploration of the gain from the combination of features in the proposed method, the Parametric-Oriented Histogram Equalization (POHE) [27] and the SFFS [12] feature selection method are used.

The application of POHE [27] can deepen the contour and enhance the extraction accuracy of the Facial Action Coding System. Among the multiple tested samples, the accuracy decreases on a small number of samples with POHE. These samples have factors that cause errors such as illumination, head posture, facial blur, and facial obscuration. Yet, the impact of this error is ignored, because POHE can introduce a positive effect on the overall results. The SFFS [12] feature selection method is a bottom-up search procedure improved by the basic Sequential Forward Selection (SFS) method, and it can resolve the drawback of the over fitting problem by excluding the worst features to guarantee the high performance. For simplicity, the label "(SFFS)" in Table 4 indicates that features employed are selected by the SFFS within a given pool of features. The proposed method is significantly superior to the former methods. Notably, the involved features in our feature pool greatly affect the performances. For instance, results fundamentally demonstrate great performances if it includes the facial action unit information feature into the feature pool.

Table 4. Accuracy of various deception detection methods (Circled denotes the best performance).

Method		Real-Life Trail [9]	MSP-YTD
Pérez-Rosas *et al.* [9]		0.752	
Jaiswal *et al.* [10]		0.7895	
Proposed Method	Features		
	$A + E + G_{MT}$ (SFFS)	0.8425	0.8404
	$A + E + G_{EY}$ (SFFS)	0.8301	0.8306
	$A + E + G_{MT}$ (SFFS+POHE)	0.8711	0.8176
	$A + E + G_{EY} + G_{MT}$ (SFFS)	0.8416	0.8401
	$A + E + G_{EY} + G_{MT}$ (SFFS+POHE)	0.8587	0.8245

Table 5. Accuracy of proposed method (Circled denotes the best performance).

Classification	Features	MSP-YTD + Real-Life Trail [9]
SVM	$A + E + G_{MT}$ (SFFS)	0.8011
SVM	$A + E + G_{EY}$ (SFFS)	0.7943
SVM	$A + E + G_{MT}$ (SFFS+POHE)	0.7887
SVM	$A + E + G_{EY} + G_{MT}$ (SFFS)	0.8071
SVM	$A + E + G_{EY} + G_{MT}$ (SFFS+POHE)	0.7916

5. CONCLUSION

This study presents a deception detection system with well-designed parameters, and it achieves optimal performance on two datasets, Real-Life Trail Data and MSP-YTD. The experimental results show the join of the classifications SVM with the feature sets "$A + E + G$" and SFFS can achieve the best performance compared with the state-of-the-art methods on deception or lie detection. In addition, the precise mouth information in the captured geometric features is utilized. This can cut every paragraph that a person says in a video, and let the polygraph system be for every word, not for a video. According to the experiments, both features '*A*' and '*E*' positively contribute to the system accuracy, and the join of the geometrical features can yield an additional improvement. As documented in the experimental results, the proposed method can be a very promising candidate for the practical application of the deception detection. Future possible improvements can be put to explore more robust features for further enhancing the performance on the excessive head movement or considering speech as well.

REFERENCES

1. F. Charles, Jr. Bond, and M. B. DePaulo, "Accuracy of deception judgments," *Personality and Social Psychology Review*, Vol. 10, 2006, pp. 214-234.
2. R. Adelson, "Detecting deception," *Monitor on Psychology*, Vol. 37, 2004, p. 70.
3. Office of Technology Assessment, United States Congress, *Scientific Validity of Polygraph Testing: A Research Review and Evaluation*, University Press of the Pacific, Washington, 1983.
4. "Education psychologists use eye-tracking method for detecting lies," psychologialscience.org, 2012.

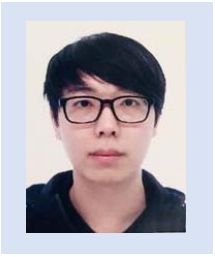
5. F. Horvath, J. McCloughan, D. Weatherman, and S. Slowik, "The accuracy of auditors' and layered voice analysis (LVA) operators' judgments of truth and deception during police questioning," *Journal of Forensic Sciences*, Vol. 58, 2013, pp. 385-392.
6. K. R. Damphousse, "Voice stress analysis: Only 15 percent of lies about drug use detected in field test," *NIJ Journal*, Vol. 259, 2008, pp. 8-12.
7. J. D. Harnsberger, H. Hollien, C. A. Martin, and K. A. Hollien, "Stress and deception in speech: evaluating layered voice analysis," *Journal of Forensic Sciences*, Vol. 54, 2009, pp. 642-650.
8. H. Hollien, J. D. Harnsberger, C. A. Martin, and K. A. Hollien, "Evaluation of the NITV CVSA," *Journal of Forensic Sciences*, Vol. 53, 2008, pp. 183-193.
9. V. Perez-Rosas, M. Abouelenien, R. Mihalcea, and M. Burzo, "Deception detection using real-life trial data," in *Proceedings of ACM International Conference on Multimodal Interaction*, 2015, pp. 59-66.
10. M. Jaiswal, S. Tabibu, and R. Bajpai, "The truth and nothing but the truth: multimodal analysis for deception detection," in *Proceedings of IEEE International Conference on Data Mining Workshops*, 2017, pp. 938-943.
11. T. Baltrusaitis, P. Robinson, and L.-P. Morency, "OpenFace: An open source facial behavior analysis toolkit," in *Proceedings of IEEE Winter Conference on Applications of Computer Vision*, 2016, pp. 1-10.
12. P. Pudil, J. Novovicova, and J. Kittler, "Floating search methods in feature selection," *Pattern Recognition Letters*, Vol. 15, 1994, pp. 1119-1125.
13. R. E. Fan, K. W. Chang, C. J. Hsieh, X. R. Wang, and C. J. Lin, "LIBLINEAR: a library for large linear classification," *Journal of Machine Learning Research*, 2008, pp. 1871-1874.
14. E. Wood, T. Baltrusaitis, X. Zhang, Y. Sugano, P. Robinson, and A. Bulling. "Rendering of eyes for eye-shape registration and gaze estimation," in *Proceedings of IEEE International Conference on Computer Vision*, 2015, pp. 3756-3764.
15. T. Baltrusaitis, L. P. Morency, and P. Robinson. "Constrained local neural fields for robust facial landmark detection in the wild," in *Proceedings of IEEE International Conference on Computer Vision Workshops*, 2013, pp. 354-361.
16. D. Cristinacce and T. F. Cootes, "Feature detection and tracking with constrained local models," in *Proceedings of British Machine Vision Conference*, 2006, Vol. 3, pp. 929-938.
17. L. Swirski, A. Bulling, and N. Dodgson, "Robust real-time pupil tracking in highly off-axis images," in *Proceedings of Symposium on Eye Tracking Research and Applications*, 2012, pp. 173-176.
18. P. Ekman and W. V. Friesen, "Measuring facial movement," *Environmental Psychology and Nonverbal Behavior*, Vol. 1, 1976, pp. 56-75.
19. S. Porter, L. Brinke, and B. Wallace, "Secrets and lies: involuntary leakage in deceptive facial expressions as a function of emotional intensity," *Journal of Nonverbal Behavior*, Vol. 36, 2012, pp. 23-27.
20. L. Brinke, S. Porter, and A. Baker, "Darwin the detective: observable facial muscle contractions reveal emotional high-stakes lies," *Evolution and Human Behavior*, Vol. 33, 2012, pp. 411-416.
21. M. Owayjan, A. Kashour, N. A. Haddad, M. Fadel, and G. A. Souki, "The design and development of a lie detection system using facial micro-expressions," in *Proceedings*

of IEEE International Conference on Advances in Computational Tools for Engineering Applications, 2012, pp. 33-38.
22. L. Su and D. L. Martin, "High-stakes deception detection based on facial expressions," in *Proceedings of IEEE International Conference on Pattern Recognition*, 2014, pp. 2519-2524.
23. M. F. Valstar, T. Almaev, J. M. Girard, G. McKeown, M. Mehu, L. Yin, M. Pantic, and J. F. Cohn, "Fera 2015-second facial expression recognition and analysis challenge," in *Proceedings of IEEE International Conference and Workshops on Automatic Face and Gesture Recognition*, Vol. 6, 2015, pp. 1-8.
24. T. Baltrušaitis, M. Mahmoud, and P. Robinson, "Cross-dataset learning and person-specific normalisation for automatic action unit detection," in *Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition*, Vol. 6, 2015, pp. 1-6.
25. Y. I. Tian, T. Kanade, and J. F. Cohn, "Recognizing action units for facial expression analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, 2001, pp. 97-115.
26. T. Kanade, J. F. Cohn, and Y. Tian, "Comprehensive database for facial expression analysis," in *Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition*, 2000, pp. 46-53.
27. C.-H. Hsia, J.-M. Guo, and C.-S. Wu, "Finger-vein recognition based on parametric-oriented corrections," *Multimedia Tools and Applications*, Vol. 76, 2017, pp. 25179-25196.
28. Z. Wu, B. Singh, L. S. Davis, and V. S. Subrahmanian, "Deception detection in videos," *AAAI Conference on Artificial Intelligence*, 2018, pp. 1695-1702.
29. M. Ding, A. Zhao, Z. Lu, T. Xiang, and J.-R. Wen, "Face-focused cross-stream network for deception detection in videos," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 7802-781.
30. G. Krishnamurthy, N. Majumder, S. Poria, and E. Cambria. "A deep learning approach for multimodal deception detection," *arXiv Preprint*, 2018, arXiv:1803.00344.
31. E. Turki, R. Alabboodi, and M. Mahmood. "A proposed hybrid biometric technique for patterns distinguishing," *Journal of Information Science and Engineering*, Vol. 36, 2020, pp. 337-345.

Jing-Ming Guo (郭景明) received the Ph.D. degree from the Institute of Communication Engineering, National Taiwan University, Taipei, Taiwan, in 2004. He is currently a full Professor with the Department of Electrical Engineering, and Director of Advanced Intelligent Image and Vision Technology Research Center. He was the former Vice Dean of the College of Electrical Engineering and Computer Science, National Taiwan University of Science and Technology, Taipei, Taiwan. He was also Director of the Innovative Business Incubation Center, Office of Research and Development. He was Visiting Scholar at the Digital Video and Multimedia Lab, Department of Electrical Engineering, Columbia University, USA from June to August, 2015, and the Signal Processing Lab, Department of Electrical and Computer Engineering, University of Cali-

fornia, Santa Barbara, USA from July 2002 to June 2003 and June-November, 2014. His research interests include multimedia signal processing, biometrics, computer vision, and digital halftoning.

Chih-Hsien Hsia (夏至賢) received the Ph.D. degree from Tamkang University, New Taipei, Taiwan in 2010. In 2007, he was a Visiting Scholar with Iowa State University, Ames, IA, USA. From 2010 to 2013, he was a Postdoctoral Research Fellow with the Department of Electrical Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan. From 2013 to 2015, he was an Assistant Professor with the Department of Electrical Engineering at Chinese Culture University, Taiwan. He was an Associate Professor with the Chinese Culture University and National Ilan University from 2015 to 2017. He currently is a Professor with the Department of Computer Science and Information Engineering, National Ilan University, Taiwan. His research interests include DSP IC design, multimedia signal processing, and cognitive learning.



Li-Wei Hsiao (蕭力瑋) received the B.S. degree from the Department of Electronic Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan, in 2019. He is currently pursuing the M.S. degree from the Department of Electronic Engineering, National Taiwan University of Science and Technology, Taipei. His current research focuses on machine learning, deep learning, biometrics, and computer vision.

Chen-Chieh Yao (姚振傑) received the B.S. degree from the Department of Electronic Engineering, Tamkang University, Taipei, Taiwan, in 2013. He is currently pursuing the M.S. degree from the Department of Electronic Engineering, National Taiwan University of Science and Technology, Taipei. His current research interests include machine learning, biometrics, and computer vision.