Iraqi Journal of Information and Communications Technology(IJICT)
Conference Series: The 1st Conference of Applied Researches in Information Engineering(ARIE2021), 2021

ISSN:2222-758X
e-ISSN:2789-7362

OPTIMIZED ACTION UNITS FEATURES FOR EFFICIENT DESIGN OF DECEPTION DETECTION SYSTEM

Shaimaa H. Abd [1], Ivan A. Hashim [2], Ali S. Abdulhadi [3]
[1,3] College of Information Engineering, Al-Nahrain University, Baghdad, Iraq
[2] Department of Electrical Engineering, University of Technology, Baghdad, Iraq
{shaimaa.hameed,ali.sadeq}@coie-nahrain.edu.iq [1,3], 30095@uotechnology.edu.iq [2]
Received:2/5/2021, Accepted:29/5/2021

***Abstract*- Deception detection is becoming an interesting field in different areas related to security, criminal investigation, law enforcement and terrorism detection. Recently non-verbal features have become essential features for the deception detection process. One of the most important kinds of these features is facial expression. The importance of these expressions come from the idea that the Human face contains different expressions each of which is directly related to a certain state. In this research paper, facial expressions' data are collected for 102 participants (25 women and 77 men) as video clips. There are 504 clips for lie response and 384 for truth response (total 888 video clips). Facial expressions in a form of Action Units (AUs) are extracted for each frame with the video clip. The AUs are encoded based on Facial Action Coding System (FACS) which are 18 AUs. These are: AU 1, 2, 4, 5, 6, 7, 9, 10, 12, 14, 15, 17, 20, 23, 25, 26, 28 and 45. Based on the collected data, only six AUs are the most effective and have a direct impact on the discrimination process between liar and truth-teller. These AUs are AU 6, 7, 10, 12, 14 and 28, as discussed in detail in Table I.**

keywords: **Deception detection, Facial expressions, Appearance features, Geometry features, Face detection, Landmark detection, FACS.**

I. INTRODUCTION

Deception is defined as concealing the truth from individuals using face and body gestures [1] . People tend to use deception for many reasons. From a psychological perspective, there are two types of deception, which are related to low-stakes (face-saving) and high-stakes (malicious deception) [2] . Many research works and studies are conducted to detect the second type. Moreover, the person that tends to lie uses more cognitive load than an innocent person because deception requires thinking and imagine before answering any question [3] . In recent years, Deception Detection System (DDS) is widely used in different applications like security, hiring new employees for business, criminal investigation, law enforcement, terrorism detection ....etc.[4] . The early implementation of the deception detection system is the polygraph test or usually referred to as a lie detector. The polygraph detects guilty participants based on measuring different physiological cues like blood pressure, pulse rate, brain activity, respiration and skin changes [5] . The polygraph test is widely adapted for several years but it has several drawbacks. The first problem is related to low detection accuracy. The second, problem is that the device is considered an invasive technique (required physical contact) [6] . These problems led to the move to more reliable and non-invasive techniques for feature extraction from the suspect's body. This research paper contains the following sections: literature survey, the proposed deception detection system with explaining its stage and finally the optimization techniques for AUs selection process.

II. LITERATURE REVIEW

Recently, different studies are performed in the area of DDS. These studies depend on one of two kinds of features either verbal or non-verbal. Verbal features are related to voice analysis while non-verbal means are cues including full-body

ISSN:2222-758X
e-ISSN:2789-7362

motion, head movement, facial expressions, eye gaze, pupil dilation and eye blinking [7]. Bedoya-Echeverry et al. [8] designed a lie detection system by utilizing temperature change in a lacrimal puncta area. This study was performed on 27 subjects and utilized the simplest classification technique that depends on comparing the estimated temperatures in control questions and the remaining parts of the interrogation. The detection accuracy was 79.2%. In another study performed by Azar and Campisi [9] in this study, the designed system also depends on infrared imaging through detecting and measuring the temperature change in the nose area. Eleven participants were used in this study and two kinds of methods were used; first, in the time domain and second, in frequency domain analysis. The accuracy for the first method was 69% while the second method was about 84%. The third study was performed by Jain et al. [10] for lie detection using thermal imaging. This study was performed on 16 participants, and the achieved detection accuracy was about 83.5%. Thannoon et al. [11] designed DDS based on facial expressions, these expressions encoded based on FACS in a form of AUs. They collected a database for 43 participants and the detected AUs applied to Virtual Generalizing Random Access Memory Weightless Neural Network (VG-RAM WNN) classifier. Moreover, the measurement accuracy was 84%. Another study performed by Demyanov et al. [12] based on AUs detection for 270 participants that were taken from Mafia TV show and the archived accuracy was 70.26%. Su and Levine [13] proposed a DDS for detecting high-stakes; this study was also based on facial expressions. The expressions taken are (eye-blink, eyebrow motion, wrinkle occurrence and mouth motion). Video databases are collected from an open-source like YouTube. This study used AU1, AU2, AU4, AU12, AU15, and AU45 as an indication to determine deceptive or liar subjects. The designed system achieved an accuracy of about 76.92%. All the mentioned studies suffer from several drawbacks like: a limited number of participants, performed in a constrained environment and there is no optimization process on the selected features. But this work is performed on a real database that contains 102 subjects and is performed in unconstrained environments, moreover, the optimization process is performed on the extracted features to select only the effective ones.

III. Materials and Methods

The Deception Detection System (DDS) simply consists of three stages organized as shown in Fig. 1 [11] . The first stage is related to video recording and pre-processing and then perform editing process. The second stage is referred to as the features extraction stage in which features are extracted and used for discriminating between liar from truth-teller. These features are applied to the classification stage to be classified into one of two classes: either liar or truth-teller.

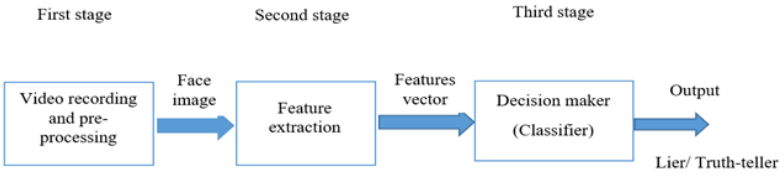


Figure 1: Block diagram of deception detection system [11]

Video recording and pre-processing: The first stage consists of three steps: video capturing and pre-processing, face detection, and landmarks detection. The first step is related to video capturing for participants when the interviewer asks

ISSN:2222-758X
e-ISSN:2789-7362

them several questions. Then put these videos in a single database. This database is used to test and validate the system and looks like a base for the system. Pre-processing step performs the editing of these videos and prepares them for subsequent stages. The second step is related to performing face detection. It means detecting a face in the input image that has an arbitrary size. The face detection algorithms do not only determine or detect face but also determine where it is located. So it is important to use these algorithms to determine face from non-face areas in the input image without taking into consideration facial expressions, orientation and lighting conditions. The third step is Facial Landmark Detection and Localization which deals with the process of placing salient points on the region of interest (ROI) in the human face. ROI in the human face represents eye, eyebrows, nose, mouth and face boundary.

Features Extraction: The second stage in DDS is the features extraction stage. Facial extractions are considered as a kind of non-verbal feature is represented in a form of AUs [14]. Facial features are described and analyzed based on a standard coding technique that is usually referred to as FACS. FACS is considered the most popular and comprehensive system for encoding facial expressions. FACS simply use Action Unit (AU) as a description of the facial muscles movement. Different studies presented on using AUs features for discriminating between the liar and truth-teller [11] - [13] . The detection of AUs depends on using two types of features these are: geometry and appearance [15]. Geometry based features are determined and measured on both landmark point locations (as each point have a specific location) and shape parameters. For appearance features that are extracted from utilizing Histograms of Oriented Gradients (HOGs) [16] . HOG is characterized by high sensitivity to any object deformation, and facial expression movements or namely facial AUs considered as a kind of these deformations, so for better description to facial features, HOG is the efficient and suitable choice [17] . HOG descriptor simply counts the occurrences of gradient orientations in a localized part or within the local patch in the input image. To capture and represent special information using the HOG technique, the input image must be divided into small cells that the HOG computed [18] . For face detection, Viola-Jones (VJ) is considered the most common face detector. The main advantages of the VJ algorithm are: the ability to detect multiple faces in a single image, process input images very rapidly, support real-time, detect faces that have different skin colors and the ability to detect faces even with eyeglass [19] . There are different landmark detection algorithms, but in this work, a Constrained Local Neural Fields (CLNF) algorithm is chosen due to the Robustness in an unconstrained environment that suffers from different problems like poor or high lighting conditions, presence of occlusions and extreme variation in pause [20]. Placing a landmark point in the 3D distribution model (PDM) is performed by applying Eq. 1, so each point is controlled by parameters [s, R, q, t] that are given by the equation [21] .

$$X_i = s \cdot R_{2D} \cdot (\bar{X}_i + \phi_i q) + t \tag{1}$$

Where:

ϕ_i is the principal component matrix,

$\bar{X}_i = [\bar{X}_i, \bar{Y}_i, \bar{Z}_i]$ is the mean value of the i_{th} feature.

q: m dimensional vector of parameters controlling the non-rigid shape.

s: scaling term that controls how close the face is to the camera.

ISSN:2222-758X
e-ISSN:2789-7362

$t = [t_x, t_y]^T$ is the translation term.

R_{2D} : 2×3 rotation matrix.

The decision maker (classifier): When the Feature extraction process is complete, it becomes necessary to apply decision classifiers. Features are combined and applied to the classification stage to distinguish innocent or truth-teller from liar subjects. Previous studies applied some kind of classification algorithms like Support Vector Machine (SVM)[22] and Virtual Generalizing Random Access Memory Weightless Neural Network (VG-RAM WNN) [11] .

IV. Proposed Action Units (AUs) Features Optimization

Before starting with the AUs detection and optimization process it becomes necessary to deal with the collected database.

A. Video Capturing and Pre-Processing

Video capturing means placing the participant in front of the camera and ask a set of questions. All recorded videos for participants are captured using a digital camera type Canon 2000 D. the main properties for this camera, it has a tiny LCD screen, The type of file format for all recorded video is MOV extension and the videos are recorded with a 1920×1080 pixel per frame (PPF). In this work, the video recording stage include videos for 102 participants, 25 of them are females and 77 are males. Their age range is from 18-55 years. The data are collected from The College of dentistry/AL-Mustansyria University and Ibn Sina University of Medical and Pharmaceutical Sciences. These videos are considered as the base of the proposed system, and these videos are used later to test the validity and reliability of the system. Fig. 2 shows some sample images for participants during the interview.

Figure 2: Sample images for participants during the interview

After video recording, it is necessary to perform the editing process. Editing means determining the necessary and effective frames. These frames represent those occurring during thinking time that proceeds while answering the question, which represents cognitive load. The resulting frames from this duration are called video clips which are approximately 1-second because this time is sufficient to capture AUs changes. Notice that some of the necessary frames may be discarded especially when participants are wearing hats or putting the hand on their faces that might lead to hiding of facial expressions. The resulting video clips obtained are 384 clips for truth and 504 clips for a lie (total is 888 clips).

ISSN:2222-758X
e-ISSN:2789-7362

B. Action Units Detection

As mentioned early, the AUs detection process requires capturing two kinds of features, these are: geometry and appearance features. These features are directly extracted from the participant's face. Geometry features depend on capturing both features (landmark) point location and non-rigid shape parameters. Before starting with the extraction of appearance-based features, it is necessary to remove any non-facial parts from the given image. After removing these parts, the masking operation is done. The masking process is performed using a common operation know as a convex hull on the region surrounded by the aligned landmark points. The output from this step is the face image that contains only the facial part and it is usually referred to as alignment and masking. Fig. 3 shows the essential steps in the process detection of AUs.

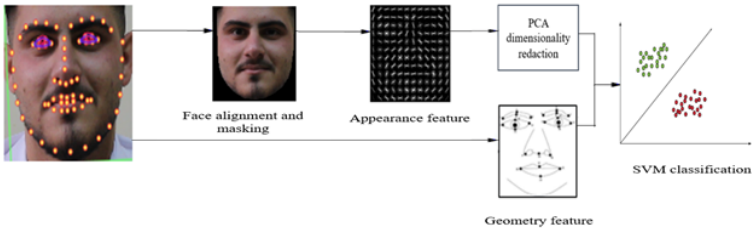


Figure 3: Facial AUs detection based on determining both appearance and geometry features

The CLNF is simply defined as the algorithm that is used for a landmark (feature) points detection and localization process. After the landmark detection process, it is necessary to calculate both appearance and geometry features. In the alignment and masking step that completely removes non-facial information from the face input image, the output image has 112×112 pixels and contain a 45-pixel interpupillary distance (IPD). IPD is defined as the distance between pupils of eyes. The aligned face image is applied to the HOG algorithm to calculate the appearance feature. The aligned image is divided into blocks and each block contain cells that also contain pixels so the final appearance information is an approximately 4464-dimensional vector. Then, Principal Component Analysis (PCA) is applied for dimensionality reduction so the resulting vector now contains 1379 dimensions. The geometry feature combines both the non-rigid shape parameter (q) and the location of the detected landmark. The q is represented with a 23-dimensional vector while the landmark location is represented with a 204-dimensional vector. The final geometric feature vector contains 23+204= 227-dimensional vectors. Finally combining both geometry and appearance features will result in a $227 + 1379 = 1606$ dimensional vector that represents combined features. Then apply the SVM classifier to detect the presence of AUs.

C. Action Unit (AU) Optimization Process

for each participant, eighteen AUs are extracted. These are AU 1, 2, 4, 5, 6, 7, 9, 10, 12, 14, 15, 17, 20, 23, 25, 26, 28, 45. It is worth mentioning that not all AUs have the same effect on the designed DDS. Table I shows the AUs with their effectiveness for the discriminating process.

ISSN:2222-758X
e-ISSN:2789-7362

TABLE I
Effectiveness of AUs in Lie and Truth Response

Action Units (AU)	Effectiveness in lie response	Effectiveness in truth response
AU1	0.2%	0
AU2	0	0
AU4	16.07%	18.49%
AU5	51.19%	58.07%
AU6	**34.72%**	**5.47**
AU7	**40.08%**	**21.61%**
AU9	0	0
AU10	**35.32%**	**21.88%**
AU12	**37.30%**	**14.58%**
AU14	**64.29%**	**47.92%**
AU15	0.60%	0
AU17	4.76%	0
AU20	0	0
AU23	66.59%	66.41%
AU25	0.79%	0
AU26	0.20%	0
AU28	**33.73%**	**25.78%**
AU45	0.79%	0

From Table I, AU6 (Cheek Raiser) is presented in lie video clips (response) with 34.72% while in truth 5.47%, which means that the value of AU6 has a great variance between lie and truth response. In other words, AU6 is raised when a person tends to lie while lowered when a person is telling the truth, so this AU can be considered as an effective AU that accurately discriminates between lie and truth response. The same thing for AU7 (Lid Tightener) that raised during deception with 40.08% while lowered in truth with a value of 21.16%. For AU10 (Upper Lip Raiser), lip raised during deception with 35.32% and lowered during truth with21.88%, which means that it is an efficient indication for deception. AU12 (Lip Corner Puller), 14(Dimple) and 28(Lip Suck) refer to the AUs that happen in the lip region, that are raised during deception and lowered when a person is telling the truth. Finally, depending on the above mentioned analytical result for the collected data set we conclude that only six AUs have a great effect on the discrimination process and are recognized as effective AUs. These are AU 6, 7, 10, 12, 14 and 28. The most effective AUs are listed in the Table II. These AUs are extracted for each frame in a given video clip. There is a different number of the frame for each video clip because the number of frames depends on the length of the video clip. In this work, the video clips contain approximately 9-44 frames. Each frame owns its associated AUs that are arranged in the form of a vector, so this leads to having 9-44 vectors for each clip (note that the number of frames is equal to the number of AUs vector). In the process of learning and testing the selected classifiers, it is necessary to perform normalization in which all video clips should have the same number of vectors even with having different frame numbers. To perform this step, this work uses the process of selection for the most repeated AUs vectors or in other words, each AU may occur in the frame within a video clip so the decision for the most repeated AUs takes over all frames within a single video clip. This lead to forming AUs vectors and these AUs represent the most accredited or most repeated. Table III shows an example of AUs extraction process. AU12 presented in

ISSN:2222-758X
e-ISSN:2789-7362

8 frames (AU12=1) and absent in 10 frames. So the final value of this AU based on the majority of AU12 values in the final AUs vector equals 0.

TABLE II
The Effective AUs in The Designed DDS with Determining Both AU Number and The Associated Region

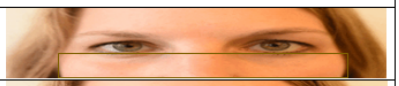
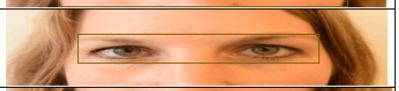
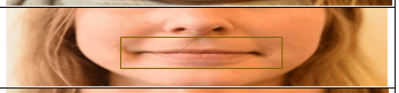
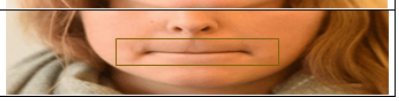
Action Unit (AU)	**Name-based on FACS**	**Associated facial region**
AU6	Cheek Riser	
AU7	Lid Tightener	
AU10	Upper Lip Raiser	
AU12	Lip Corner Puller	
AU14	Dimple	
AU28	Lip Suck	

TABLE III
Final AUs Vector Representation for A Video Clip That Consists of Eighteen Frames, The Value in The Final Vector for Each AU Based on Determining The Most Repeated Value

Frame number	**AU6**	**AU7**	**AU10**	**AU12**	**AU14**	**AU28**
1	1	1	1	1	0	1
2	1	1	1	1	1	1
3	1	1	1	1	1	1
4	1	1	1	1	1	1
5	1	1	1	1	1	1
6	1	1	1	1	1	1
7	1	1	1	1	1	1
8	1	1	1	1	0	1
9	1	1	1	0	0	1
10	1	1	1	0	0	1
11	1	1	1	0	0	1
12	1	1	1	0	0	1
13	1	1	1	0	0	1
14	1	1	1	0	0	1
15	1	1	1	0	0	1
16	1	1	1	0	0	1
17	1	1	1	0	0	1
18	1	1	1	0	1	1
Final AUs vector (most repeated)	**1**	**1**	**1**	**0**	**0**	**1**

ISSN:2222-758X
e-ISSN:2789-7362

V. CONCLUSION

Facial expressions are important clues for a deception detection system. In this paper. Facial expressions are extracted in an optimized way based on combining both geometry and appearance features. Geometry features provide sufficient information about face geometry while appearance features are used to subtle changes in appearance, so combining these features lead to identifying which AUs are presented or absent. Finally, the optimization technique is applied to AUs extraction process so that only six (AUs 6, 7, 10, 12, 14 and 28) out of eighteen features are selected as the most effective features. This decreases the complexity of the system and the detection time.

REFERENCES

[1] B. Diana, "A Cognitive Approach to Deception Detection: Multimodal Recognition of Prepared Lies" , Ph. D thesis, University of Milano Bicocca, 2014.
[2] G. An, "Literature Review for Deception detection" , 2015.
[3] J. Masip, "Deception Detection: State of The Art and Future Prospects" , Psicothema, Vol. 29, No. 2, pp. 149-159, 2017.
[4] M. Burzo, M. A. V. Perez-Rosas, and R. Mihalcea, "Multimodal Deception Detection" , in The Handbook of Multimodal-Multisensor Interfaces: Signal Processing, Architectures, and Detection of Emotion and Cognition, Vol. 2, No. October, pp. 419-453, 2018.
[5] S. Azhan, A. Zaman, and M. R. Bhuiyan, "Using Machine Learning for Lie Detection: Classification of Human Visual Morphology" , 2018.
[6] M. Jaiswal, S. Tabibu, and R. Bajpai, "The Truth and Nothing But The Truth: Multimodal Analysis for Deception Detection" , in 2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW), IEEE, pp. 938-943, 2019.
[7] H. Bouma et al. , "Measuring Cues for Stand-Off Deception Detection Based On Full-Body Nonverbal Features in Body-Worn Cameras" , in Society of Photo-Optical Instrumentation Engineers (SPIE) , Vol. 9995, p. 23, 2016.
[8] S. Bedoya Echeverry, H. Belalcazar Ramirez, H. Loaiza Correa, S. E. Nope Rodriguez, C. R. Pinedo Jaramillo, and A. D. Restrepo Giron, "Detection of Lies by Facial Thermal Imagery Analysis" , Rev. Fac. Ing. , Vol. 26, No. 44, pp. 47-59, 2017.
[9] Y. Azar and M. Campisi, "Detection of Falsification Using Infrared Imaging: Time and Frequency Domain Analysis" , in International Conference on Advances in Computing, Communications and Informatics (ICACCI), IEEE, pp. 1021-1026, 2014.
[10] U. Jain, B. Tan, and Q. Li, "Concealed Knowledge Identification Using Facial Thermal Imaging" , IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) , pp. 1677-1680, 2012.
[11] H. H. Thannoon, W. H. Ali, and I. A. Hashim, "Design and Implementation of Deception Detection System Based on Reliable Facial Expression" , J. Eng. Appl. Sci. , Vol. 14, No. 15, pp. 5002-5011, 2019.
[12] J. Bailey, S. Demyanov, K. Ramamohanarao, and C. Leckie, "Detection of Deception in The Mafia Party Game" , in Proceedings of the 2015 ACM International Conference on Multimodal Interaction, pp. 335-342, 2015.
[13] L. Su and M. D. Levine, "High Stakes Deception Detection Based on Facial Expressions" , in 2014 22nd International Conference on Pattern Recognition, IEEE, No. 1, pp. 2519-2524, 2014.
[14] M. S. Bartlett, G. C. Littlewort, M. G. Frank, C. Lainscsek, I. R. Fasel, and J. R. Movellan, "Automatic Recognition of Facial Actions in Spontaneous Expressions" , J. Multimed, Vol. 1, No. 6, pp. 22-35, 2006.
[15] J. Chen, Z. Chen, Z. Chi, and H. Fu, "Recognition of Facial Action Units with Action Unit Classifiers and An Association Network" , in Asian Conference on Computer Vision. Springer, pp. 672-683, 2014.
[16] D. Forsyth, "Object Detection with Discriminatively Trained Part-Based Models" , IEEE Trans. Pattern Anal. Mach. Intell. , Vol. 32, No. 9, pp. 1627-1645, 2009.
[17] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection" , 2005 IEEE Comput. Soc. Conf. Comput. Vis. pattern Recognit. , Vol. 1, pp. 886-893, 2005.
[18] J. Pao, "Emotion Detection through Facial Feature Recognition" , 2016.
[19] C. Zhang and Z. Zhang, "A Survey of Recent Advances in Face Detection" , 2010.
[20] T. Baltrusaitis, P. Robinson, and L. P. Morency, "Constrained Local Neural Fields for Robust Facial Landmark Detection in The Wild" , in Proceedings of the IEEE international conference on computer vision workshops, 2013, pp. 354-361.
[21] T. Baltrusaitis, "Automatic Facial Expression Analysis" , MSc thesis, University of Cambridge, 2014.
[22] A. I. Simbolon, A. Turnip, J. Hutahaean, Y. Siagian, and N. Irawati, "An Experiment of Lie Detection Based EEG-P300 Classified by SVM Algorithm" , in 2015 International Conference on Automation, Cognitive Science, Optics, Micro Electro-Mechanical System, and Information Technology, ICACOMIT. IEEE, 2016, pp. 68-71.