International Journal of Computational and Electronic Aspects in Engineering
Volume 3, Issue 3, September 2022, pp 47 - 54
https://doi.org/10.26706/ijceae.3.3.2210551

Human Behavioral Analysis Based in Facial Emotion and Gesture (Survey)

Raghad Ghalib Abd[1]
Raghadghalib@uomustansiriyah.edu.iq

Ameen A. Noor[2]
a.ameen63@uomustansiriyah.edu.iq

Abdul-Wahab Sami Ibrahim[3]
dr.wahasb.sa@uomustansiriyah.edu.iq

[1,2,3]Computer Science Department, College of Education, University of Almustansirya Baghdad, Iraq.

***Abstract*—This survey paper describes a number of fields in which human behavior analysis through facial gestures is applied, the various method that is utilized with the most recent technologies. Cutting-edge camera technology was used to capture images and assess a person's emotions. Human behavior is studied through facial and body gestures, and in this study, we divide human emotions into universally acknowledged expressions such as "sad," "happy," "surprised," "worried," and "liar." The limitations and benefits of competing and complementary technologies are discussed in this study, as well as the diversity of research in the area of human behavioral analysis based on face and gestures.**

Keywords—Facial recognize, Human behavioral analysis, Human face and gesture analysis, Human facial recognition, Gesture analysis.

I. Introduction

Biometric technologies have exploded in popularity in the field of identity and identification as information technology has advanced in recent years. Recognition is achieved in these systems. To analyze human behavioral, this study utilizes a physiological or behavioral attribute for individuals such as the face, which will be used in this study, and facial gestures. We will work in three stages: - image capture and processing from face gesture and classifying to happy, sad, lie, surprised or worried. The purpose of this study is to clarify the use of human behavioral analysis based on face and gestures in clinics, games, and psychology, as well as acceptance of offers to promote goods and on schools and other fields. Based on Mehrabian etal [1][2] , for communications of the human feelings, 55% of the messages are conveyed through facial expressions, 38% through paralanguage like intonation (i.e. vocal), and 7% through linguistic language (i.e. verbal). Facial Expression Recognition (FER) is regarded an essential study issue in research community due to its vast range of applications.

The paper is organized as follows. Section 2 describes the human face and gestures analysis, and Section 3 represents the data analysis in related work.

II. Human Face and Gestures Analysis

In the field of marketing [13]. Customer interest quantification is considered as one of the most inventive, promising, and fascinating trends in marketing research. This study describes a deep learning-based system used to monitor the behavior of customers, with a focus on interest identification. The suggested approach determines customer attention by estimating head posture. The system monitors facial expressions and reports customer interest for the customers whose heads are turned towards the product of interest or advertisement. The suggested model begins through identifying the frontal poses of the face; the facial components that are essential for the facial

Technical Article
Available online on – 05 September 2022

Cite this article – Raghad Ghalib Abd, Ameen A. Noor, Abdul-Wahab Sami Ibrahim, "Human Behavioral Analysis Based in Facial Emotion and Gesture (Survey)", *International Journal of Computational and Electronic Aspects in Engineering*, RAME Publishers, vol. 3, issue 3, pp. 47-54, 2022.
https://doi.org/10.26706/ijceae.3.3.2210551

expression recognition are after that segmented and iconized face image will be created; and lastly, the analysis of the facial expressions is carried out with the use of the values of confidence of the acquired iconized facial image in combination with raw face images. For robust facial expression recognition, this approach has combined the holistic facial information with the local part-based characteristics. With the suggested processing pipeline, head posture estimation and face expression detection may be done with a simple imaging device like a webcam. The suggested pipeline may be utilized for tracking focus groups' emotional responses to diverse concepts, sounds, images, phrases, and other stimuli.

In clinical and healthcare [14]. The purpose of the Gesture and Face Analysis for Health Informatics workshop is to discuss and share challenges and successes in applying machine learning (ML) and computer vision for automatic human behavior analysis and modeling in healthcare applications and clinical research. The workshop's goal is to promote present research and support multidisciplinary cooperation in order to further this cutting-edge work. The conference brings together researchers interested in ML and computer vision, multi-modal signal processing and fusion, assistive technologies, behavioral sensing, human-centered computing, and medical tutoring systems for medicine and healthcare.

With a baseline-intervention-retest approach, Facial Imitation Enhances the Recognition of the Emotions in the Adults with various Subclinical Autistic Trait Levels [15] investigated effects of imitations on the recognition of the facial emotions using computer-based automatic expression analysis. The participants, who were 55 young adults with varied autistic trait degrees, were given face images showing 1 of 6 main emotional expressions to perform an emotion recognition task. After then, the task was repeated with the addition of instructions for imitating the expressions. A camera captured the faces of participants throughout the experiment so that their imitation performances could be automatically assessed. The instructions for imitation of enhanced the emotion recognition and imitation performance. In the imitation block, emotion recognition improvement was greater in those with higher autistic trait levels, while imitation improvement was unaffected by autistic qualities. The discovery that imitation instruction enhances the recognition of emotion and that the imitation is a positive within-participant recognition accuracy predictor in imitation block has supported the concept of the relation between the motor expressions and perception in the emotion processing that could be mediated through mirror neuron system. Yet, given there has not been any indication of the fact that those who have higher autistic features differed in the imitative conduct in particular, their disproportionate emotion recognition advantages might have resulted from the indirect imitation instruction impacts.

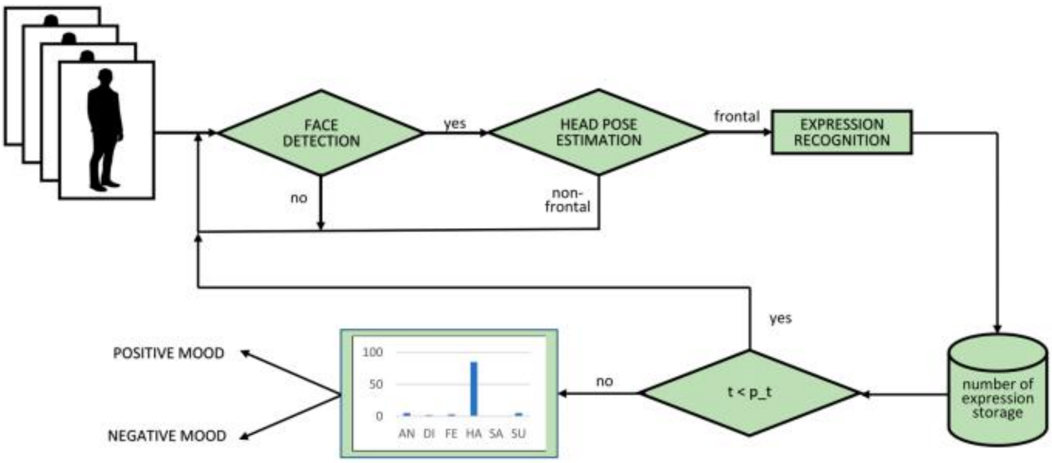


Figure 1. processing pipeline [13]

Toys [16]. In the future, interactive television will automatically give each viewer with user-personalized features. Interactive TV must identify the viewers, along with their preferences or emotions in order to provide automatic user-personalized services. This study presents a unique architecture for future interactive television, as well as a system with real-time face analysis which has the ability to recognize and detect human face images, including their expressions, and so understand their internal emotional states. Face recognition, facial expression recognition and face detection are the three image processing modules of suggested face analysis system. Face detection is used for tracking and detecting a large number of persons watching TV. Facial expression recognition and face recognition are used to distinguish individual TV viewers as well as their internal emotional states, required for customized user interfaces and services. This work describes the Ada-LDA learning technique, which is adequate for multi-class pattern classification and depends on efficient and simple MspLBP features, for a robust system of the real-time face analysis. Extensive testing shows that suggested system of face analysis performs well in real-time and has a high recognition rate. It has a frame rate of more than 15 frames per second.

III. DATA ANALYSIS IN RELATED WORK

The Affect-in-the-Wild (Aff-Wild) data-base has 248 videos totaling 30 hrs in length. Videos are recorded in this dataset under a variety of conditions. A total of 200 people were registered, 70 of whom were women and 130 of whom were men. Between 6 and 8 subjects manually annotate all videos. All of the recorded videos have arousal and valence annotations. This set includes videos that convey a range of emotions. It is the world's largest facial behavior video dataset [5][6]. Between 6 and 8 experts assigned ratings to such videos based on two continuous dimensions: arousal and valence. Arousal and valence are measures of how negative or positive a facial expression is, as well as the strength of the emotion's activation. Samples from the Aff-Wild database are depicted in Figure 3.

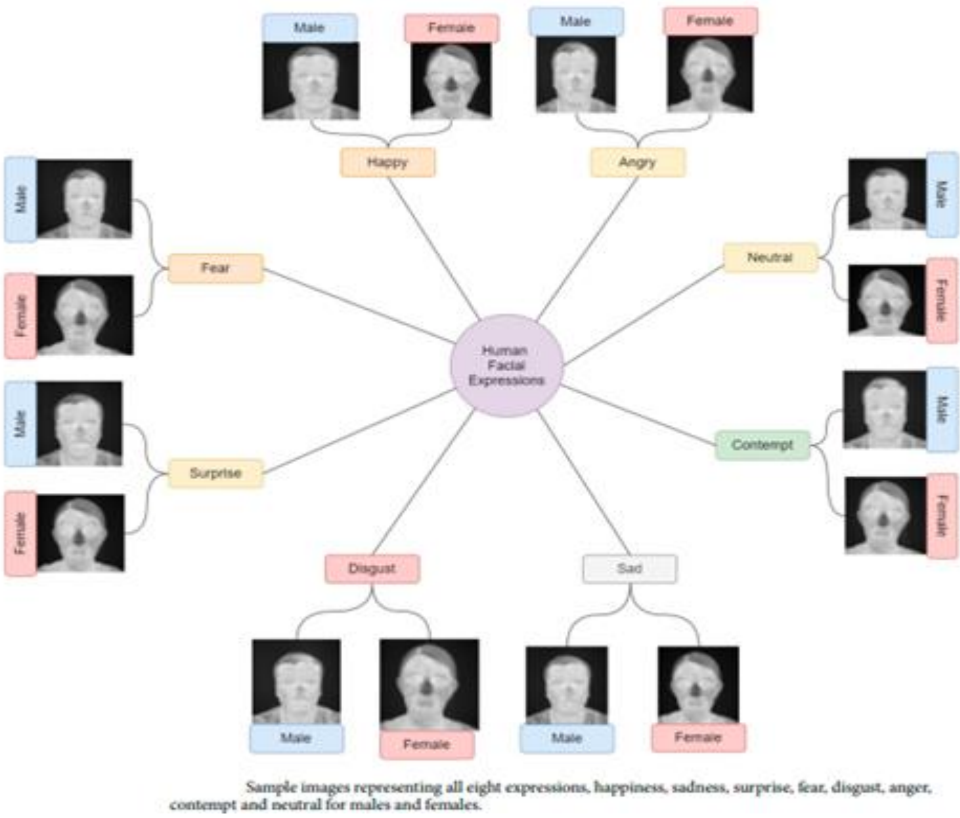


Figure 2. Sample images representing all eight expressions, happiness, sadness, surprise, fear, disgust, anger, contempt and neutral for males and females. Image is developed using Draw.io [4]

Human Behavior Understanding in the Big Multi-media Data Utilizing the CNN based FER [7]Dallas et al. [8] suggested the HOG feature extraction method. In comparison to other current features, it has demonstrated state-of-the-art performance [9]. The HOG represents standard global descriptor, similar to the SURF and SIFT, that is utilized to describe object representation.

Figure 3. Aff-Wild sample images

We used the database created by Kopaczka for the current work [4]. The images in the database were captured with an Infratech HD820 high-resolution thermal infrared camera that has been supplied with 30mm f/1.0 prime lens and 1024 768 pixel-sized micro-bolometer sensor that has a 0.03K thermal resolution at a temperature of 30oC. The subjects have been filmed at 0.9 m distance from camera, which results in a face spatial resolution of about 0.5mm for each pixel. To reduce background variation, the recordings were made against a thermally neutral backdrop. The database was created by manually screening video recordings of participants at frame rate of 30 frames per second and extracting images. Anger, Fear, Disgust, Contempt, Neutral, Happy, Surprise, and Sad are among the eight classes of expressions represented in the database.

Human Behavior Understanding in the Big Multi-media Data Utilizing the CNN based FER [7]Dallas et al. [8] suggested the HOG feature extraction method. In comparison to other current features, it has demonstrated state-of-the-art performance [9]. The HOG represents standard global descriptor, similar to the SURF and SIFT, that is utilized to describe object representation.

HOG extracts data from each pixel's face edges, directions and visibility. Vertical and horizontal gradients are calculated with the help of two filters (F2) and (F1). Second, the formulas in Eq. (2) and Eq. (1) are used to obtain magnitude (g) and gradient orientation (θ). M × N cells make up the HOG feature, which represents an image in grids. Bins in each one of the cells store information about edge orientations. A variety of gradient orientations are represented by the bin.

-1 0 1 (F1)

-1
1 (F2)
0

$g=\sqrt{(g^2 x+g^2 y)}$ (1)

$\theta=\arctan gy/gx$ (2)

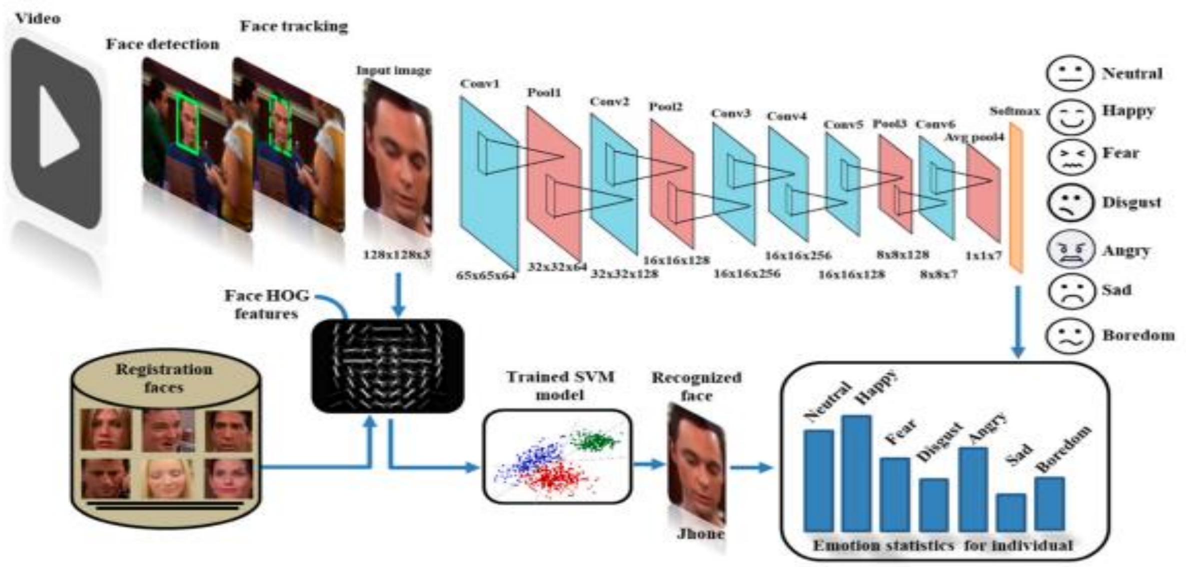


Figure 4. The suggested model for the analysis of the human behavior utilizing the FER of the persons from popular TV-series [7]

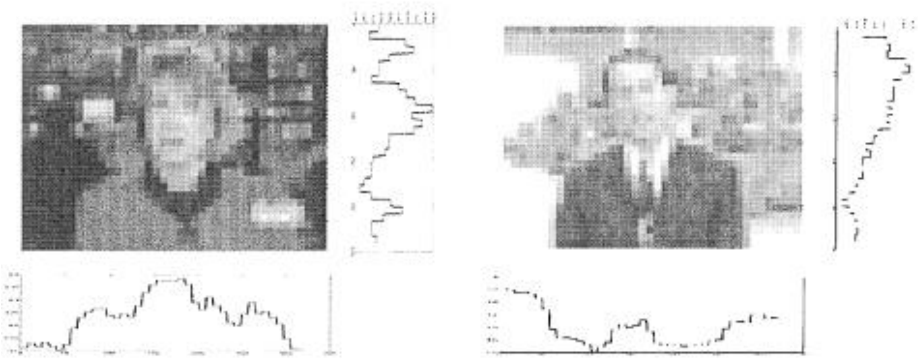
The smaller cell- and block- sizes extract more robust face representation and capture small eye, mouth, and nose shape patterns. As a result, in the suggested framework, we used a 2 x 2 block size, an 8 x 8 cell size, and 18 bins per cell to represent a 128 by 128 face image. This system extracts specific traits that are powerful enough to depict a face in a way that can be recognized.face feature extract in still image [12].

For face identification, Sirohey suggested a localization technique for segmenting a face from a cluttered background [10]. It removes and groups edges using an edge map (i.e. Canny detector) and heuristics, leaving only the ones on the contour of the face. The boundary between background and head region is after that fitted with an ellipse. On data-base that includes 48 images with cluttered backgrounds, the system has obtained an 80% rate of the accuracy. Instead of using edges, Lerch & Chetverikov [11] proposed a simple approach of the facial detection based on blobs and streaks (i.e. linear sequences of edges with similar orientations).

The cheekbones, eyes, and nose are represented by 2 dark blobs and 3 light ones in their facial model. The outlines of eyebrows, lips and face are represented by streaks in the model. The spatial relation between the blobs is encoded with the use of 2 triangular layouts. In order to aid blob detection, a Laplacian image with low-resolution is created. The image is then examined for certain triangular occurrences that might be used as candidates. A face would be identified in the case where streaks are detected around the candidate.

Face analysis systems are evaluated in a realistic virtual environment under dynamic conditions. [17] The purpose of this study is to present an innovative tool for evaluating systems of facial analyses under the dynamic experimental conditions. The technology is made up mostly of the virtual environment in which a virtual agent (i.e. simulated robot) performs facial analyses (facial recognition and detection). This virtual agent might be navigating in a virtual world containing at least one subject, and it could examine subjects' faces from various angles and distances (i.e. pitch, yaw, and roll), as well as under various lighting conditions (outdoor or indoor). The agent's present view, or the image which an agent sees, is created through combining the real face with the background images that have been obtained before they are used in virtual environment.

Various types of agents and their orientations could be simulated in the virtual environment, like agent that navigates in a scene with individuals looking in many orientations, (which mimic a home-like environment), an agent that performs circular scanning (like at security check-point), or camera-based surveillance system that observes an individual. Furthermore, the agent might actively adjust its view-point throughout the process of recognition in order to enhance the results of the recognition.



(a) n=8　　　　(b) n=8

Figure 5. Horizontal and vertical profiles. By looking for peaks in vertical and horizontal profiles, it is possible to identify a single face. Yet, this same approach has an issue in the detection of the faces in complex backgrounds or many faces as demonstrated in 2.3(b) (b) [12]

The suggested tool gives developers all the tools that he or she needs to create the scenario of the evaluation, including a group of the images of real faces that include real background information, a scenario configuration (position, pose and number of subjects that are to be observed), a virtual agent with navigation capabilities, generations of simulated agent's view-dependent images, an agent trajectory definition, some fundamental active vision mechanisms, as well as ground truth data (such as the face id and pos), which allow evaluating approaches of the face analysis within realistic conditions. The study of robustness regarding face recognition and detections approaches under position variations, as well as evaluating integrated system of face analysis for use by a service robot, are provided. Developers and researchers of face analysis tools, particularly in the biometrics and robotics sectors, might be interested in the suggested approach. The availability of the standard data-bases, evaluation methods, and benchmarks is also critical for adequate development and comparison regarding systems of face analysis in related works in this field.

There are many face databases and associated evaluation approaches which take into account various numbers of camera sensors, subjects, and conditions of the image acquisition, and are suitable for testing many aspects of facial recognition problems, like aging, illumination invariance, expression invariance, and so on (comparative studies and surveys [18–19,20]). [21, 22] provides a basic information and an overview regarding current face databases. Even though certain new databases (LFW data-base [23] and the Photoface data-base [24]) are developed for including real-world images, the majority of data-bases and evaluation protocols (which include the LFW database [23] and the Photoface data-base [24]) have been developed for testing approaches with the use of images captured via static cameras. Face recognition infrared images are also routinely employed using comparable approaches [25,26].

TABLE 1. CORE TECHNOLOGIES USED IN HUMAN BEHAVIOR ANALYSIS USING FACIAL IMAGE THAT ARE PRESENTED IN THE SURVEY PAPERS.

	Year	Methods	dataset	conclusion	Present result
F1	2013 A.J Ezhil, K. Adalarasu [27]	FPGA	JAFFE dataset	future an attempt may be made for developing hybrid method for the extraction of the facial features and recognition accuracy may be additionally enhanced with the use of the same NN method	95.33%
F2	2007 Hatice Gunes, Massimo Piccardi [28]	two separate cameras	Image	Early fusion appears to yield better recognition accuracy than late fusion, and the summation rule showed to be the optimal strategy to fuse such 2 modalities among the three late fusion approaches. The consistency of such findings on full-length expressive video sequences will be verified in future extensions of this research.	91.1%
F3	2021 Ankan Bhattacharyya, et al [2]	Deep learning with infraredcamera	JAFFE database	It may make it easier to make high-level decisions with confidence because the decisions are based on a set of data features that have been carefully chosen to accomplish the desired objectives. A few feature selection approaches might be used to acquire the most relevant features, which may reduce the computational overhead. Aside from that, the suggested model may be tested on other huge data-bases, which will help ensuring suggested mode's robustness.	82.836%
F4	2020 Muhammad Sajjad1&Sana Zahir1&Amin Ullah2&Zahid Akhtar4&Khan Muhammad [7]	Cnn (convolutional neural network)	Image	We conducted a subjective review to assess our technique's overall performance and were able to produce persuasive results for the problem at hand. Due to the high processing complexity regarding our suggested architecture, we only used DL technology for one module. In the future, we'll look into using light-weight DL approaches for all of the suggested framework's modules, such as tracking, face detection, recognition, and face expression analysis	93.39%

IV. Conclusion

Many researchers in various disciplines, including human interaction systems, mental disease detection, and affect recognition, have been interested in facial expression recognition (FER). Although few existing methods/techniques are helpful in the actual world, even with low recognition rates, most applications are applied to controlled laboratory settings. After examining several feature extraction, face detection, and expression classification methodologies and approaches, we conclude that a deep learning methodology such as a convolution neural network may successfully recognize facial expressions.

References

[1] Mehrabian, A. Communication without words. Psychol. Today 2, 53–56 (1968). https://doi.org/10.4324/9781315080918-15

[2] Bhattacharyya, A., Chatterjee, S., Sen, S., Sinitca, A., Kaplun, D., & Sarkar, R. (2021). A deep learning model for classifying human facial expressions from infrared thermal images. Scientific Reports, 11(1). https://doi.org/10.1038/s41598-021-99998-z

[3] Ekman, P. & Friesen, W. V. Facial Action Coding System (Consulting Psychology Press, 1978)

[4] Kopaczka, M., Kolk, R. & Merhof, D. A fully annotated thermal face database and its application for thermal facial expression recognition. In IEEE International Instrumentation and Measurement Technology Conference (I2MTC) 1–6 (2018). https://doi.org/10.1109/I2MTC.2018.8409768

[5] Zafeiriou S, Kollias D, Nicolaou MA, Papaioannou A, Zhao G, Kotsia I (2017) Aff-wild: valence and arousal in-the-wild challenge. In: IEEE CVPR workshop. https://openaccess.thecvf.com/content_cvpr_2017_workshops/w33/html/Zafeiriou_Aff-Wild_Valence_and_CVPR_2017_paper.html

[6] Khan, G., Samyan, S., Khan, M. U. G., Shahid, M., & Wahla, S. Q. (2020). A survey on analysis of human faces and facial expressions datasets. International Journal of Machine Learning and Cybernetics, 11(3), 553–571. https://doi.org/10.1007/s13042-019-00995-6

[7] Sajjad, M., Zahir, S., Ullah, A., Akhtar, Z., & Muhammad, K. (2020). Human Behavior Understanding in Big Multimedia Data Using CNN based Facial Expression Recognition. Mobile Networks and Applications, 25(4), 1611–1621. https://doi.org/10.1007/s11036-019-01366-9

[8] Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: international Conference on computer vision & Pattern Recognition (CVPR'05). IEEE Computer Society. https://doi.org/10.1109/CVPR.2005.177

[9] Shokrani S, Moallem P, Habibi M (2014) Facial emotion recognition method based on Pyramid Histogram of Oriented Gradient over three direction of head. In: Computer and Knowledge Engineering (ICCKE), 2014 4th International eConference on. IEEE. http://dx.doi.org/10.1109/ICCKE.2014.6993346

[10] S. A. Sirohey. Human face segmentation and identification. Technical Report CS-TR-3176, University of Maryland, 1993. http://hdl.handle.net/1903/400

[11] D. Chetverikov and A. Lerch. Multiresolution face detection. In Theoretical Foundations of Computer Vision, volume 69 of Mathematical Research, pages 131-140. Akademie Verlg, 1993.

[12] Yang, M.-H., & Ahuja, N. (2001). Face Detection and Gesture Recognition for Human-Computer Interaction (Vol. 1). Springer US. https://doi.org/10.1007/978-1-4615-1423-7

[13] Yolcu, G., Oztel, I., Kazan, S., Oz, C., & Bunyak, F. (2020). Deep learning-based face analysis system for monitoring customer interest. Journal of Ambient Intelligence and Humanized Computing, 11(1). https://doi.org/10.1007/s12652-019-01310-5

[14] Hammal, Z., Huang, D., Bailly, K., Chen, L., & Daoudi, M. (2020). Face and Gesture Analysis for Health Informatics. ICMI 2020 - Proceedings of the 2020 International Conference on Multimodal Interaction, 874–875. https://doi.org/10.1145/3382507.3419747

[15] Kowallik, A. E., Pohl, M., & Schweinberger, S. R. (2021). Facial imitation improves emotion recognition in adults with different levels of sub-clinical autistic traits. Journal of Intelligence, 9(1), 1–14. https://doi.org/10.3390/jintelligence9010004

[16] Ho An, K., Jin Chung, M., & Member, S. (2009). Cognitive Face Analysis System for Future Interactive TV.

[17] Correa, M., Ruiz-Del-Solar, J., & Verschae, R. (2016). A realistic virtual environment for evaluating face analysis systems under dynamic conditions. Pattern Recognition, 52, 160–173. https://doi.org/10.1016/j.patcog.2015.11.008

[18] J. Ruiz-del-Solar, R. Verschae, M. Correa, Recognition of faces in unconstrained environments: a comparative study, EURASIP J. Adv. Signal Process. (2009) 19184617 (Recent Advances in Biometric Systems: A Signal Processing Perspective). https://doi.org/10.1155/2009/184617

[19] Samal, A., & Iylngar, P. A. (1992). automatic recognition and analysis of human faces and facial expressions: a survey (Vol. 25, Issue 1). https://doi.org/10.1016/0031-3203(92)90007-6

[20] MichałBeretaabPawełKarczmarekacWitoldPedryczadMarek Reformata, “Local descriptors in application to the aging problem in face recognition”, Pattern Recognition, Volume 46, Issue 10, October 2013, Pages 2634-2646. https://doi.org/10.1016/j.patcog.2013.03.010

[21] Face Recognition Home Page (Available on June 4th, 2012): http://www.facerec.org/databases/

[22] R. Gross, Face databases, in: S. Li, A.K. Jain (Eds.), Handbook of Face Recognition, Springer-Verlag, 2005, pp. 301–327.

[23] G.B. Huang, M. Ramesh, T. Berg, E. Learned-Miller, Labeled faces in the wild: a database for studying face recognition in unconstrained environments (Technical Report 07-49), University of Massachusetts, Amherst, 2007. https://hal.inria.fr/inria-00321923/

[24] S. Zafeiriou, M. Hansen, G. Atkinson, V. Argyriou, M. Petrou, M. Smith, L. Smith, The photoface database, in: Proceedings of Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2011, pp. 132–139. https://doi.org/10.1109/CVPRW.2011.5981840

[25] G. Hermosilla, J. Ruiz-del-Solar, R. Verschae, M. Correa, A comparative study of thermal face recognition methods in unconstrained environments, Pattern Recognit. 45 (7) (2012) 2445–2459. https://doi.org/10.1016/j.patcog.2012.01.001

[26] R.S. Ghiass, O. Arandjelović, A. Bendada, X. Maldague, Infrared face recognition: a comprehensive review of

methodologies and databases, Pattern Recognit. 47 (9) (2014) 2807–2824. http://dx.doi.org/10.1016/j.patcog.2014.03.015

[27] Ezhil, A. J., & Adalarasu, K. (2013). FPGA Implementation of Human Behavior Analysis Using Facial Image. In International Journal of Engineering Trends and Applications (IJETA) (Vol. 2). www.ijetajournal.org

[28] Gunes, H., & Piccardi, M. (2007). Bi-modal emotion recognition from expressive face and body gestures. Journal of Network and Computer Applications, 30(4), 1334–1345. https://doi.org/10.1016/j.jnca.2006.09.00