FINDING FACIAL EXPRESSION PATTERNS ON VIDEOS BASED ON SMILE AND EYES-OPEN CONFIDENCE VALUES

Setiawan Hadi[1], Asep K Supriatna[1], Faishal Wahiduddin[1], Wilis Srisayekti[2], Achmad Djunaidi[2], Efi Fitriana[2], Aceng Abdullah[3] and Dian Ekawati[3]

[1]Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Indonesia
[2]Faculty of Psychology, Universitas Padjadjaran, Indonesia
[3]Faculty of Communication Science, Universitas Padjadjaran, Indonesia

ABSTRACT

Facial expression recognition is one of the types of non-verbal communication that is not only commons for human but also plays an essential role in everyday lives. The development of science and technology allows the machine to automatically detect human facial expressions based on images and videos. Numerous facial expression detection methods have been proposed in the literature. This paper presents a method to find three basic facial expressions (neutral, happy, and angry) from two parameter values: smile and eyes-open. The analysis involves a preprocessing step using a combination of pre-designed proprietary algorithm and Luxand library. Firstly, the parameters were mapped into two-dimensional space and then grouped into three clusters using K-means, a popular heuristic clustering method. Secondly, more than 50,000 frames for each video were experimented using the proprietary research data. The result shows that the proposed method successfully performed a simple video analysis of facial expressions.

KEYWORDS

Facial expression, Image processing, Clustering analysis, Image Analysis, Video Processing

1. INTRODUCTION

Automatic human Facial Expression Recognition (FER) has attracted the attention of many researchers, not only from computer science but also from human behavior with a psychology background. However, the fundamental theory of facial expressions generally refers to Ekman [1], a pioneer in the emotion study and their association with facial expressions.

With the advance of science and technology development, FER has opened the possibilities of implementation in human life activities, including social-media-related applications. However, detecting human expression automatically in images or videos is not trivial due to many constraints, although some methods have been proposed and found in the literature [12-16]. Several previous studies [10-11] have shown that the FER method achieved 68.4% and 73.28 % accuracy using CNN on the FER2013 dataset.

This paper proposes a method to find facial expression patterns on video based on the smile and eyes-open confidence values obtained from Luxand SDK [2]. Three of six basic facial expressions have been classified, happy, angry, and neutral/normal, using the two parameters: smile and confidence. These parameters are quantitative numbers from 0 to 1 that show if the subject smiles and if the eyes are open or closed. For example, if the result of expression analysis is Smile=0.94768; EyesOpen=0.9998, it indicates that the subject smiles with a confidence level

DOI: 10.5121/ijaia.2021.12503

of 94.7%, and the eyes are open with a confidence level of 99.9%. Figure 1 shows the illustration of the expression analysis of images. The experiment was carried out using the predesigned proprietary facial expression database called DEWI, an Indonesian facial expression dataset collection.

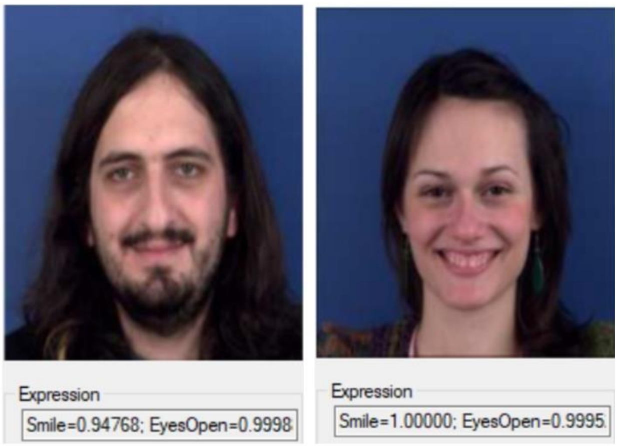

Figure 1. Smile and Eyes-open Values [3]

2. METHODS

Our proposed method is composed of five process blocks, as displayed in Figure 2. To start with, the face detection method was implemented and then followed by the feature selection method using the proprietary algorithm combined with the Luxand framework. This step resulted in two values, smile confidence values, and eyes-open confidence values. Based on these values, the targetted facial expression can be estimated.

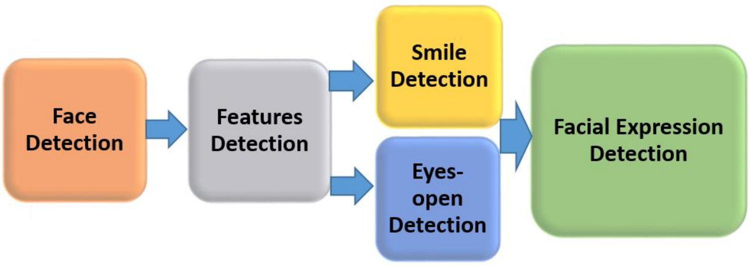

Figure 2. Process Blocks of the Expression Recognition Research

2.1. Face Detection Methods

Face detection is a process to localize faces in a digital image. An algorithm called DeWa is developed and applied to detect a facial object in a digital image. The detection step consists of three stages: segmentation, filtering, and localization. The algorithm was successfully tested using a expression database called DWI. The complete explanation of the method can be read in [4]. Specifically, Figure 3 shows the workflow of DeWa.

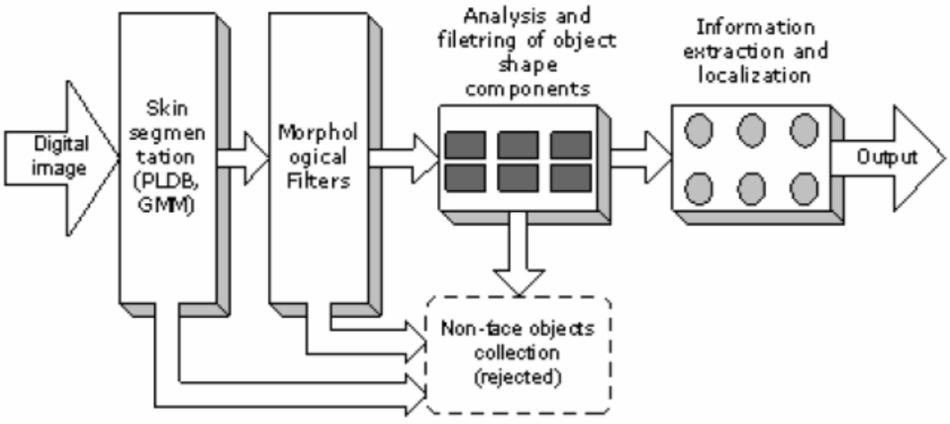

Figure 3. Face Detection Framework [4]

2.2. Facial Features

Facial features play an important role in facial expression recognition. In this research, 66 facial features adopted from Luxand were applied. These features show local coordinates of specific areas of a face that are categorized into (i) eyes, (ii) eyebrows, (iii) nose, (iv) mouth, (v) nasolabial, (vi) chin, (vii) contours. Figure 4 shows the local features of a face and an example of facial features marker number.

Seq	Facial Feature Name	Value
1	FSDKP_LEFT_EYE	0
2	FSDKP_RIGHT_EYE	1
3	FSDKP_LEFT_EYE_INNER_CORNER	24
4	FSDKP_LEFT_EYE_OUTER_CORNER	23
5	FSDKP_LEFT_EYE_LOWER_LINE1	38
6	FSDKP_LEFT_EYE_LOWER_LINE2	27
7	FSDKP_LEFT_EYE_LOWER_LINE3	37
8	FSDKP_LEFT_EYE_UPPER_LINE1	35
9	FSDKP_LEFT_EYE_UPPER_LINE2	28
10	FSDKP_LEFT_EYE_UPPER_LINE3	36
11	FSDKP_LEFT_EYE_LEFT_IRIS_CORNER	29
12	FSDKP_LEFT_EYE_RIGHT_IRIS_CORNER	30
13	FSDKP_RIGHT_EYE_INNER_CORNER	25
14	FSDKP_RIGHT_EYE_OUTER_CORNER	26
15	FSDKP_RIGHT_EYE_LOWER_LINE1	41
16	FSDKP_RIGHT_EYE_LOWER_LINE2	31
17	FSDKP_RIGHT_EYE_LOWER_LINE3	42

Figure 4. Example of Facial Features and Their Coordinate [2]

All features were processed based on the comparison between neutral or normal faces with other facial expressions. The distance between the two images was calculated and features were displayed the features in three classes: up, down, and regular (no changes). Figure 5 illustrates the process.

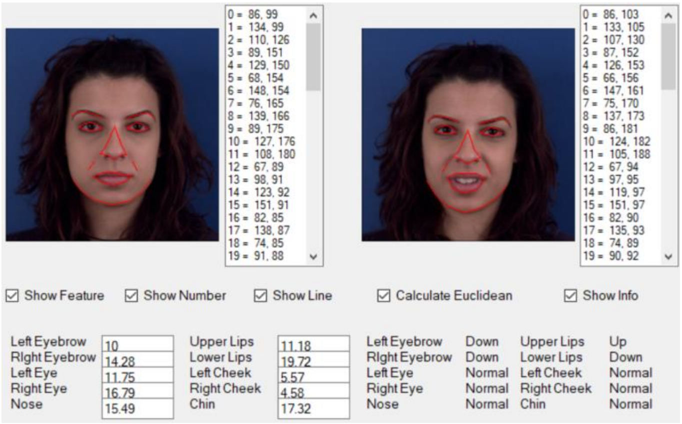

Figure 5. Facial Features Processing

2.3. Smile Detection and Eyes-open Detection

A smile can be considered as a form of micro-expression: brief facial cues that reveal the natural intention of a person. When a person smiles, his facial muscles will contract. A natural smile will include wrinkles, pushed-up cheeks, and a movement from the muscle that orbits the eye.

A Swedish anatomist, Carl-Herman Hjortsjo [5], has classified human facial movements on the face. Then, Ekman and Friesen [6] have improved this method and created the Facial Action Coding System (FACS). In 2002, Ekman, Friesen, and Joseph C. Hager published a significant update to FACS [7]. FACS is a general system to systematically categorize human facial expressions and is used in various disciplines, including by psychologists, neuro-analysts, and animators. Combinations of action units can build a facial expression in FACS [8]. Figure 6 shows the sample FACS coding of a facial expression.

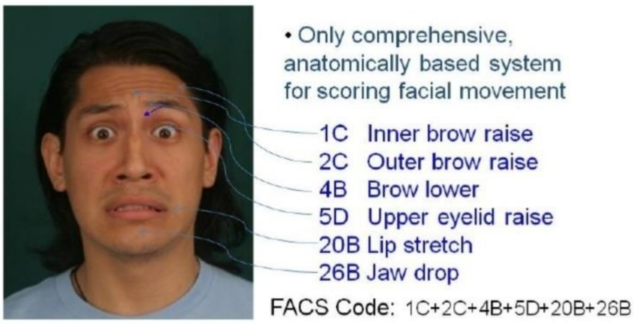

Figure 6. Sample FACS Coding of a Fear Expression

2.4. Modified K-means Clustering

K-means is an unsupervised method for clustering and is mainly used by researchers for various data mining implementations. However, traditional K-means have limitations. One of them is that to process the data, the number of a cluster has to be stated. This condition does not meet the requirement if the number of clusters to be processed is unknown. For this purpose, the K-means algorithm is was modified to solve this condition so K-means can run without a predefined number of clusters [9].

Mathematically, K-means minimizes a function as follows:

$$K = f(x) = \sum_{i=1}^{m} \sum_{x_j \in C_i} D(x_j, \mu_i)^2 \quad (1)$$

$D(x_j, \mu_i)$ is the distance between its closest centroid. There are numerous distance functions that can be used in measuring the distance between objects. However, the Euclidean distance is the most straightforward distance measurement used for calculation. The Euclidean formula for calculating the distance between two objects is presented below:

$$D_{i,j} = \sqrt{\sum_{k=1}^{n} (x_{ik} - x_{jk})^2} \quad (2)$$

As previously stated, in K-means, the number of clusters to classify has to be defined. In modified K-means, this limitation was eliminated by optimizing the value of k. The values were defined sequentially and processed to evaluate the highest-minimal value. The optimization function is defined as:

$$V = \frac{J'}{N \cdot \delta_{\min}^2} + \sum_{i=1}^{k'} \frac{1}{1 + |C_i|} \quad (3)$$

$\delta_{\min} = \min_{\neq j} \{\delta(\mu_i, \mu_j)\}$ is the centroids' shortest distance, k points to a number of the selected cluster provided with enough data, J' is used partly for clusters chosen, and $|C_i|$ is the cluster's cardinality. Thus, the first term of the above equation represents the goal function: finding clusters that were well-separated for better solutions. The second term was used for penalty factors for clusters with small members.

2.5. Generating Basic Facial Expression

The facial expression parameters were generated based on the ground-truth data (see Figure 7) that were processed using K-means clustering with k=3. The result of clustering is presented in Table 1 below.

Table 1. Result of the Clustering Process

Expressions	Smile	Eyes-open
Neutral	0.07048509	0.9998214
Happy	0.93574368	0.9618265
Angry	0.14346801	0.9572292
Neutral	0.07048509	0.9998214

The values show the centroids of each facial expression based on the parameters of smile and eyes-open. The centroid will be used for the distance calculation of tested data to approximate facial expression results.

3. RESULTS AND ANALYSIS

3.1. Data for Experiment

The experimental data provided by the Psychology Department of Universitas Padjadjaran are in the form of images and videos. The data were gathered through another research on the social anxiety of West Java people. Image data are considered as training data to generate the ground truth of facial expression clusters (refer to Table 1). The data characteristic of these images is shown in Table 2.

Table 2. Image Characteristics of Experimental Data

Expressions	Image Type and Resolution			
	# Grayscale	Resolution	# RGB	Resolution
Neutral	247	720 x 960	32	2085 x 3139
Happy	625	720 x 960	29	2164 x 3257
Angry	671	720 x 960	18	1865 x 2808

These image data were processed to obtain smile and eyes-open confidence values. The values were then mapped into a 2D-smile-eyes-open chart which shows the characteristics distribution of facial expression data as presented in Figure 7.

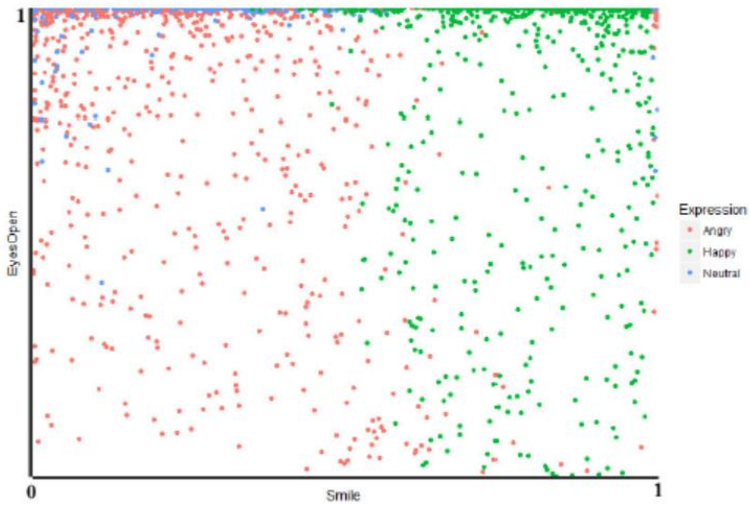

Figure 7. Ground-truth Data Mapping Space

The neutral, happy, and angry distribution of facial expressions are represented by point-shape colors grey, green, and red. As seen in the figure, the red dots and the green dots are more dominant, while the grey dots are almost invisible. This condition can happen due to fewer neutral data images (257 compared with 625 and 671, refer to Table 2).

The method was then tested using the video data. The characteristic of the videos is presented in Table 3.

Table 3. Video Characteristics of Experimental Data

Video	Size (MB)	Time	Resolution	Rates	# Frames
A	512	36:55:00	800 x 600	25 fps	55,380
B	739	53:19:00	800 x 600	25 fps	79,976
C	625	49:30:00	800 x 600	25 fps	74,255
D	403	31:42:00	800 x 600	25 fps	47,561

3.2. Discussion

To validate the approach, the ground-truth data were evaluated using clustering parameters. The result is presented in Table 4 below.

Table 4. Evaluation of Facial Expression Images

Expressions	#GT	#Neutral	#Happy	#Angry	Performance (%)
Neutral	**279**	**237**	35	7	84.95%
Happy	**654**	4	**618**	32	94.50%
Angry	**689**	505	72	**112**	16.26%

The evaluation result shows the excellent performance of the clustering method for neutral and happy expressions. The angry facial expression can be considered as a false result because it gave a minimal performance. Further analyzing this error, it was found that most of the images in the angry group have higher eyes-open confidence values, but lower smile confidence values. Thus, images with angry expressions were considered similar to images in a neutral expression. Other parameters were considered needed to differentiate between those two expressions.

The testing video was processed frame-by-frame sequentially. Initially, each frame was analyzed to obtain smile and eyes-open confidence values. These values were then calculated with the result of clustering (refer to Table 4) using the Euclidean distance formula (see Equation 1). Finally, the decision of facial expression was taken based on the smallest distance value. The result of video processing is presented in Table 5.

Table 5. Analysis of Facial Expression Videos

Video	# Frames	#Neutral	#Happy	#Angry
A	55,380	14,791	8,338	32,244
B	79,976	6,351	26,645	46,979
C	74,255	42,071	5,856	26,327
D	47,561	5,874	19,367	22,319

The numbers of happy facial expressions that are greater than the numbers of neutral facial expressions are found in Video B and Video D. On the other hand, Video A and Video C show that a neutral expression is more often found than a happy expression, which is presented graphically in Figure 8. The angry expression numbers shown in Table 5 were not analyzed and could be considered false detection errors. It is displayed for documentation only.

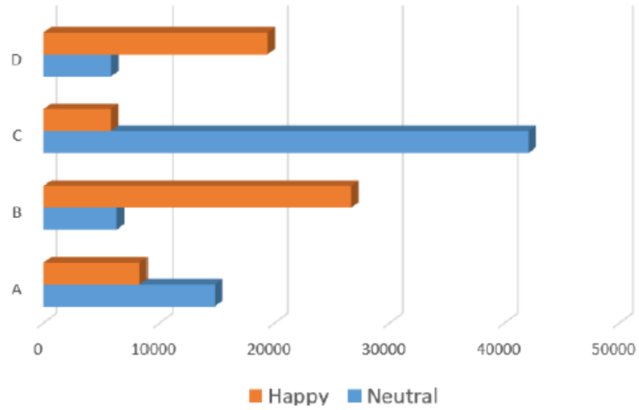

Figure 8. The result of Video Analytic in Graphics

3.3. Ekspreso: Testbed Software for Evaluation

For the purpose of this research, a testbed software for video and image expression evaluation called Ekspreso was developed. The software is built on .NET and runs on Microsoft Windows operating system. It can be used to evaluate the performance of our algorithm using images and videos. Figure 9 shows the expression evaluation for images, and Figure 10 depicts the expression evaluation for videos. The face areas are covered due to human-right privacy.

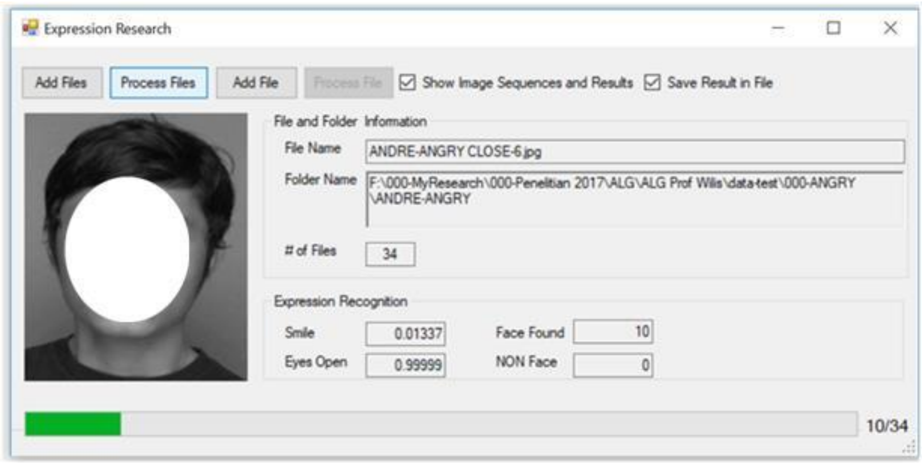

Figure 9. Image Evaluation using Ekspreso

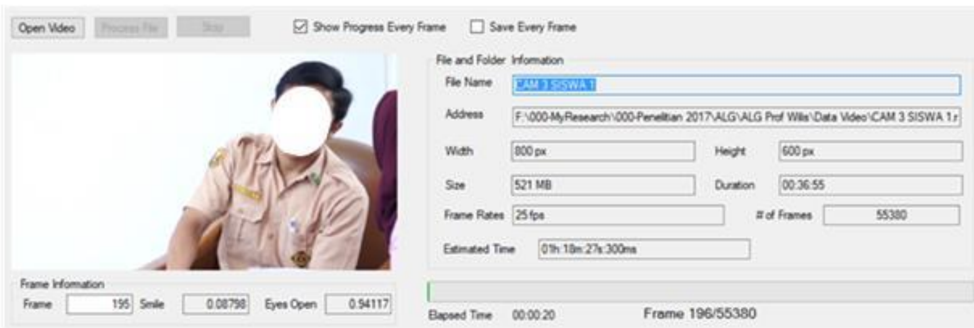

Figure 10. Video Evaluation using Ekspreso

4. CONCLUSIONS

In this paper, a method to find patterns of facial expressions is presented based on two parameters: smile and eyes-open confidence level values. A non-supervised method, K-means, was succesfully applied for clustering the data based on the proprietary ground truth images. The values were calculated to obtain information about three facial expressions: neutral, happy, and angry. It is shown that this method has successfully presented the accurate result for the two expressions: neutral and happy. The work is still in progress to improve the accuracy of neutral expression detection to reach the maximum accuracy as happy expression detection. However, for angry expressions, the performance was not remarkable. In this case, further research is recommended to explore and implement other techniques to solve this problem and increase the detection accuracy for angry expression.

ACKNOWLEDGMENTS

The authors would like to thank the Indonesian Government thru the Rector of Universitas Padjadjaran for supporting the research through Academic Leadership Grants (ALG) 2015-2021.

REFERENCES

[1] Paul Ekman, "Facial Expression and Emotion" American Psychology, 1993.
[2] Nikolay Ageev, Luxand SDK: Face Detection and Recognition Library Developer's Guide, Luxand Inc., 2017.
[3] Niki Aifanti, Christos Papachristou, and Anastasios Delopoulos, "The MUG Facial Expression Database" Image Analysis for Multimedia Interactive Service, pp. 1–4, 2016.
[4] Setiawan Hadi, Adang Suwandi Ahmad, Iping Supriana Suwardi, and Farid Wazdi, "DeWa: A Generic Multi-Aspect Framework for Multiple Face Detection in Complex Image" Journal of Information and Communication Technology, 1(1):281–286, 2007.
[5] Carl-Herman Hjortso, Man's face and mimic language, Lund: Studentlitteratur, 2016.
[6] Paul Ekman and Wallace Friessen, Facial Action Coding System: A Technique for the Measurement of Facial Movement, Consulting Psychologists Press, Palo Alto, 2016.
[7] Paul Ekman, Wallace V. Friesen, and Joseph C. Hager, Facial Action Coding System: The Manual on CD ROM, A Human Face. Salt Lake City, 2012.
[8] Takeo Kanade and Jeffrey F. Cohn, "Comprehensive database for facial expression analysis" In proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition, p.46–53, 2017.
[9] Setiawan Hadi, "Clustering Techniques Implementation for Detecting Faces in Digital Images" In proceedings of International Conference on Information and Communication Technology (ICTS), 2016.
[10] Zhou Yue; Feng Yanyan; Zeng Shangyou; and Pan Bing, "Facial Expression Recognition Based on Convolutional Neural Network," IEEE 10th International Conference on Software Engineering and Service Science 2019
[11] Yousif Khaireddin and Zhuo Chen, " Facial Emotion Recognition: State of the Art Performance on FER2013", ArXiv abs/2105.03588, 2021.
[12] Illiana Azizan and Fatimah Khalid, Facial Emotion Recognition: A Brief Review, International Conference on Sustainable Engineering, Technology and Management, 2018
[13] Fanzhi Kong, Facial expression recognition method based on deep convolutional neural network combined with improved LBP features. Pers Ubiquit Comput 23, 531–539, 2019.
[14] Jiannan Yang, Fan Zhang, Bike Chen, Samee U. Khan, Facial Expression Recognition Based on Facial Action Unit, 10th International Green and Sustainable Computing Conference, 2019
[15] Yifeng Zhao and Deyun Chen, A Facial Expression Recognition Method Using Improved Capsule Network Model, Journal of Scientific Programming Hindawi, 2020
[16] Humayra Binte Ali and David Martin Ward Powers, Facial Expression Recognition Based on WAPA and OEPA Fastica, International Journal of Artificial Intelligence & Applications 5(3), 2014

AUTHORS

Setiawan Hadi is an associate professor in the Computer Science Department Faculty of Mathematics and Natural Sciences Universitas Padjadjaran. His research interest is in computer vision, image processing, and computer graphics.

Asep K. Supriatna is a professor of Mathematics Department Faculty of Mathematics and Natural Sciences Universitas Padjadjaran. His research interest is in applied mathematics, biological and environmental modeling.

Faishal Wahiduddin graduated from the Computer Science Department Faculty of Mathematics and Natural Sciences Universitas Padjadjaran and now working in the information technology industry for innovative and intelligent systems.

Wilis Srisayekti is professor of Pyschology Departement Universitas Padjadjaran. Her research interest is stress and anxiety in society, post-traumatic stress disorder, diagnosis, and intervention, including psychometrics such as facial expression.

Achmad Djunaidi is a clinical psychologist focused on psychodiagnostic, directed, and non-directed intervention and counseling, including personal development. He is with Psychology Department Universitas Padjadjaran.

Efi Fitriana graduated from Psychology Department Universitas Padjadjaran. Her interest is in psychometrics, counseling psychology, psychiatry. In addition, she has skills and expertise in validation and methods, factor analysis, and reliability.

Aceng Abdullah is an associate professor in Communication Science Faculty Universitas Padjadjaran. His research is in mass communication, photographic journalism, and visual communication such as facial expression analysis.

Dian Ekawati is working in Germanistic Department Faculty of Cultural Sciences Universitas Padjadjaran. Her research is in intercultural communication, linguistics, conversational analysis, and language aspects of hypnosis therapy.