

PAPER • OPEN ACCESS

## Real-Time Emotion Recognition System to Monitor Student's Mood in a Classroom

To cite this article: W B Putra and F Arifin 2019 *J. Phys.: Conf. Ser.* **1413** 012021

View the [article online](#) for updates and enhancements.

### You may also like

- [A fuzzy adaptive extended Kalman filter exploiting the Student's t distribution for mobile robot tracking](#)  
Xin Lai, Guorui Zhu and Jonathon Chambers
- [Investigating EEG-based functional connectivity patterns for multimodal emotion recognition](#)  
Xun Wu, Wei-Long Zheng, Ziyi Li et al.
- [Deep Neural Network for Electroencephalogram based Emotion Recognition](#)  
Shruti Garg, Soumyajit Behera, K Rahul Patro et al.

### ECS Toyota Young Investigator Fellowship

For young professionals and scholars pursuing research in batteries, fuel cells and hydrogen, and future sustainable technologies.

At least one \$50,000 fellowship is available annually.  
More than \$1.4 million awarded since 2015!



Application deadline: January 31, 2023



TOYOTA

**Learn more. Apply today!**

# Real-Time Emotion Recognition System to Monitor Student's Mood in a Classroom

**W B Putra and F Arifin**

Education Electronics and Information Technology,  
Universitas Negeri Yogyakarta, Yogyakarta, Indonesia

Email: [werda.buana2016@student.uny.ac.id](mailto:werda.buana2016@student.uny.ac.id), [fatchul@uny.ac.id](mailto:fatchul@uny.ac.id)

**Abstract.** Depression is a commonly unattended health problem, unbounded by age border, and greatly affecting student's performance in their study. To prevent it, the writer built a real-time facial emotion recognition system, so the teacher can monitor students' mood through class activity. The system should be reliable enough when running on mid-end computer specification. The student will be given questionnaire to measure their stress. The questionnaire result will be used to analyze whether the use of the system able to reduce student's stress or not. The result from the questionnaire shows that the use of the system able to detect student's mood early so the teacher may minimize student's stress.

## 1. Introduction

Human emotion recognition has been the spotlight for many researches to test various theories, algorithms, and new technologies. Emotion recognition discipline of Human-Computer Interaction relies on the algorithmic robustness and hardware sensitivity.

Social anxiety disorder (SAD) is one of the most common mental health disorders, with approximately 13% of the population meeting diagnostic criteria for SAD during their life. Hence, the need to quickly recognize such anomaly surfaced. If left untreated, SAD typically runs a chronic course and total remission is rare. Though depression is classified as an adult disorder, and depression in very young children is rare, middle to late adolescence is the most common age when depression symptoms first appear or a first major depressive chapter happens. Lewinsohn, Hops, Roberts, Seeley, and Andrews sampled random adolescents within a U.S. community and figured out that the average age of onset of major depression was 14 years old [1].

Though facial expressions obviously does not necessarily convey emotions, in computer vision community, the term of facial expression recognition often refers to the classification of human facial features into one of the six basic emotions: happiness, sadness, fear, disgust, surprise and anger [2] [3]. According to National Institute of Mental Health data in 2014, about 2,8 million American teenagers ranged 12-17 years old would at least suffered by one episode of major depression and only 30% of them get a proper treatment [4].

Depression is categorized as mental-emotional disorder in Indonesia and it's prevalence as high as 6% of total population [5]. A number of researches investigate the role some factors to depressive symptom, such as emotion regulation [6]. Illegal drugs usage is proven to be the byproduct of depression and because the bigger prevalence on men, depression and comorbidity affected by gender. Data from particular newspaper in Indonesia shows that 90% of drug users is on large, unattended [7]. Psychological



researchers have proved that there are negative effects of depression on student's academic performance [8]

In response to that condition, the writer proposed a way to monitor student's mood through class activity. The teacher should be able to constantly aware on the fluctuation of their student's mood by using facial emotion recognition system. At this point, the problem only lies on the willingness of the system developer and the educator to collaborate. But here is where the problems lies, Indonesia, where the research will take place upon, is a developing country. As sad as it sounds, 55% of under-15 Indonesian is functionally unable to read [9]. Implementing artificial intelligence to recognize human facial expression can be a complex and difficult task. High quality datasets are hard to found, and there are various pitfalls to avoid when designing such system. It is proved that facial expression could help lecturers to identify the involvement and comprehension of the student [10].

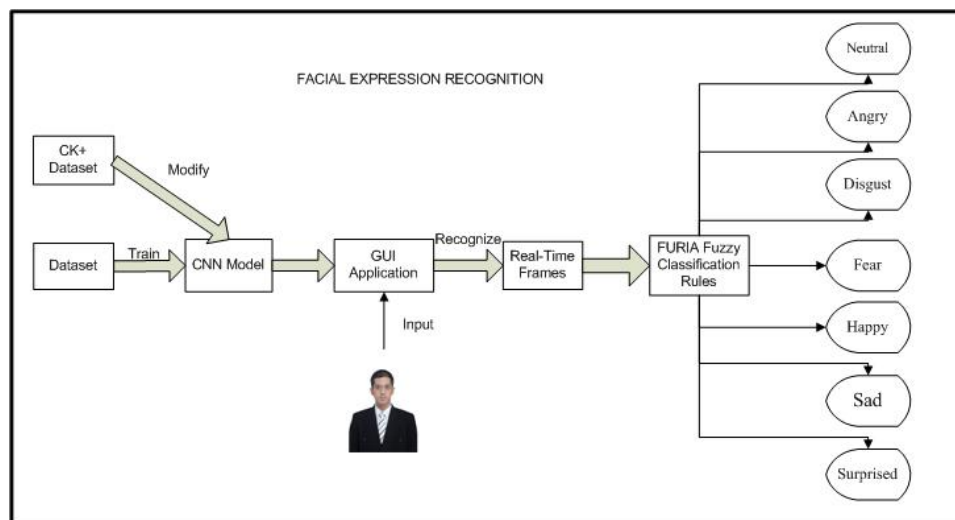
## 2. System Design

Real-time facial emotion recognition, in general, is divided by several phases. The first phases would be detecting general area of human face, this process include tracking system which require the hardware to monitor the general movement of facial layout. Second phase would be Facial landmarking, which pointing out more accurate facial point to be extracted. Of course, the writers have to choose the most efficient way to compute and doing facial analysis to prevent computer's memory overload. On the other hand the system still need a feature which is able to extract the subject's facial expression and then further processed it into the emotion as the result.

The writer needs to be sure the computer able to properly do image acquisition, facial extraction, facial landmarking, and logic computation to get the emotion from the subject. In short, a working webcam, and sufficient computation power on the subjected computer is crucial. The location of facial features can be represented as landmarks on the face. The image acquisition of human face can be represented as coordinated vector landmark and Action Units (AU). Hence, facial feature offered geometric information of each and overall shape of particular object [11]. Facial expression extraction and localization has become pivotal point among image analyst [12] [13].

## 3. Research Method

Prior to research conduct in school, a research proposal was requested to university board. After 3 days, the proposal granted and ready to be forwarded to the school board, which is *SDN Percobaan 2* Yogyakarta. The research proposal was submitted to Institutional Board and approved by the headmaster. Participant's selection began by consulting with supervisor and the school's teacher. The writers looking for students, who capable of basic computer operation and can be left by themselves to do their work. Another criterion is the subject shouldn't be burdened by external variables, such as the approaching National Exam, which could be applied to all of the sixth grade students. This narrowed the choice to fourth grader and fifth grader. After further discussion, the writer choose to pick students from fifth grade A class as test subjects due to their relatively docile nature compared to other classes on their grade or fourth grader.



**Figure 1.** System Design.

To gather the research data, a particular device to capture images of each student is needed, in this case it would be USB webcam. The writer needs to be sure the computer able to properly do image acquisition, facial extraction, facial landmarking, and logic computation to get the emotion from the subject. In short, the research needs a working webcam, and sufficient computation power on the subjected computer. After consulting with the headmaster and IT teacher, it is confirmed that even though they have 20 computers in their lab, only 4 computers for student use that were installed with working webcam (installed by default on monitor display). Like that, there are four client computer and one server computer as a working system (see figure 11 for reference) prepared by the school.

The system produced systematical processes which allowed the teacher to monitor their student's mood through the class activity by using computer's webcam to acquire the image real-time. The system will use a pre-made library to acquire image(s) from webcam. The process to determine facial feature from the acquired image will utilize pre-trained dataset [14] [15] [16], which is trained using convolutional neural network model. While trying some method to recognize facial feature, the writers stumbled at some problems such as:

1. Difficulty to combine the metric with powerful image recognition technique such as SVM, LDA, PPA, etc is particularly high.
2. The measurement computation is very complicated.
3. The distance does not obey triangle inequality. So sometime, two similar images can be both similar to a completely unknown object.

With these points all covered, the system simplify the Euclidean Distance computation. The system uses Standardizing Transform to do domain smoothing [17]. This directly relates image to smoothing and indicates that smoothing noiseless images can still increase image recognition rate [18].

#### 4. System Implementation

Table 1 through table 3 shows confusion matrix for pre-trained dataset that being used to build the system. The score that written in the table informs the reader how many images that represented an emotion that fit with the comparison. Table 1 is the comparison from the original database with the normalized shape feature benchmark. In table 2, the dataset will tested on Canonical Appearances benchmark. In table 3, the dataset will tested with the combination of both benchmarking system. This test is usually the most reliable of the three and the result often be made for the validity of the dataset.

**Table 1.** Confusion Matrix of Emotion Detection for the Similarity Normalized Shape [18]

	An	Di	Fe	Ha	Sa	Su	Co
An	<b>35</b>	40	0	5	5	15	0
Di	7.9	<b>68.4</b>	0	15.8	5.3	0	2.6
Fe	8.7	0	<b>21.7</b>	21.7	8.7	26.1	13
Ha	0	0	0	<b>98.4</b>	1.6	0	0
Sa	28	4	12	0	4	<b>28</b>	24
Su	0	0	0	0	0	<b>100</b>	0
Co	15.6	3.1	6.3	0	15.6	34.4	<b>25</b>

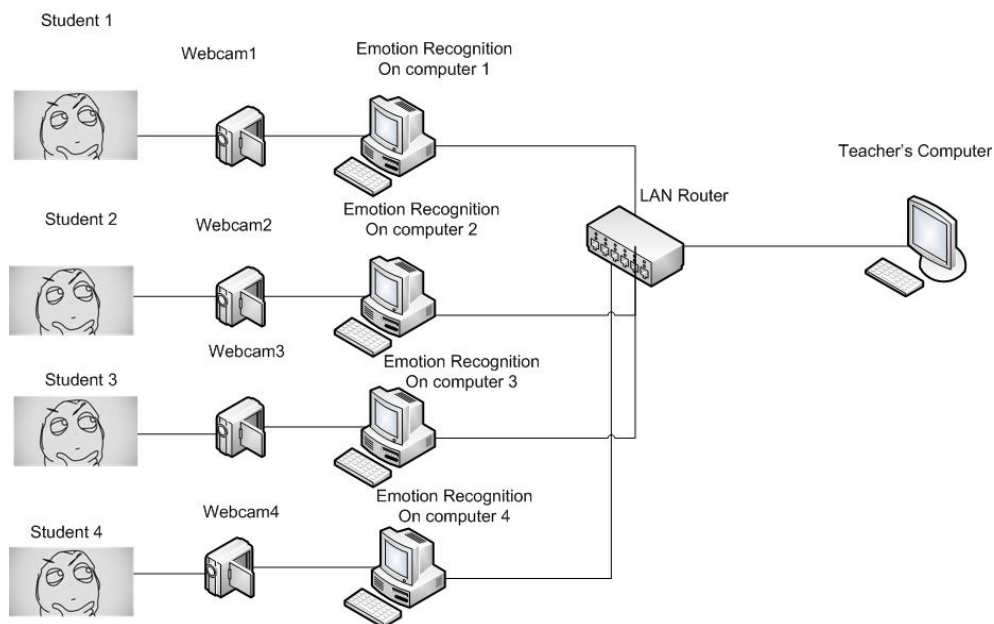
**Table 2.** Confusion Matrix of Emotion Detection for the Canonical Appearance

	An	Di	Fe	Ha	Sa	Su	Co
An	<b>70</b>	5	5	0	10	5	5
Di	5.3	<b>94.7</b>	0	0	0	0	0
Fe	8.7	0	<b>21.7</b>	21.7	8.7	26.1	13
Ha	0	0	0	<b>100</b>	0	0	0
Sa	16	4	8	0	<b>60</b>	4	8
Su	0	0	1.3	0	0	<b>98.7</b>	0
Co	12.5	12.5	3.1	0	28.1	21.9	<b>21.9</b>

**Table 3.** Confusion Matrix Of Emotion Detection For The Combination Of Features (SPTS+CAPP). The Fusion Of Both Systems Were Performed By Summing Up The Probabilities From Output Of The Multi-Class SVM

	An	Di	Fe	Ha	Sa	Su	Co
An	<b>75</b>	7.5	5	0	5	2.5	5
Di	5.3	<b>94.7</b>	0	0	0	0	0
Fe	4.4	0	<b>65.2</b>	8.7	0	13	8.7
Ha	0	0	0	<b>100</b>	0	0	0
Sa	12	4	4	0	<b>68</b>	4	8
Su	0	0	0	0	4	<b>96</b>	0
Co	3.1	3.1	0	6.3	3.1	0	<b>84.4</b>

JAFFE image set consisted of 213 images of Japanese female frontal facial pictures, classified by 7 basic emotions just like the classification used by the system. The writers will compare the result gotten from analyzing each of these 213 images using the system, with the default classification from JAFFE image set. In figure 2, the writers run the system to see how the emotion recognition system could detect facial features. The system proved to be able distinct emotion features of human expressions; this will be further tested using a specific expert-arranged image set which shown on table 4.



**Figure 2.** System Testing on Real Time Feed via Webcam

This suggests that an automated system can do just as a good job, if not better as a naive human observer and suffer from the same confusions due to the perceived ambiguity between subtle emotions. However, human observer ratings need to be performed on the CK+ database and automated results need to be conducted on the JAFFE database to test out the validity of these claims. Our tools also provide desirable result on the scanning and image analysis. Which could detect, locate, tracking, and analyze human expression that contained in captured image. This contributes greatly on this research mainly because this research is quite focused on the accuracy of the system. The writers have to measure the system's accuracy. The system uses JAFFE image set, which already classified based on specific emotions by expert. The writers will compare the measurement brought by our system, with the classification from the image set in a posed expression database; the database will display different basic emotional expressions simultaneously, while in spontaneous expression database, the expressions are natural. Spontaneous expressions are differed from posed ones remarkably in term of intensities, configurations, and durations. Apart from this, syntheses of some AUs are just barely achievable without the need of undergoing on associated emotional state. Therefore, at most cases, the posed expressions by human participants are exaggerated, while the spontaneous ones are quite subtle and differed in appearances.

The emotion annotation can be simply done in discrete emotion labels or on a continuous scale. Most of databases are usually based on the basic emotions theory proposed by Paul Ekman and Armindo Freitas-Magalhaes which assumes the existence of six discrete basic emotions (anger, fear, disgust, surprise, joy, sadness) [14] plus neutral emotion. JAFFE database was planned and assembled by Michael Lyons, Miyuki Kamachi, and Jiro Gyoba [19].

The result of teacher's questionnaire will be processed with a simple formula, mainly because this questionnaire is only has two respondents, which is the IT teachers. The formula is simply required the total score collected, which ranged from 0 to 20, from the questionnaire with maximum value possible for the questionnaire on the question. This method of data analysis is called SUS (System Usability Scale).

Originally created by John Brooke, it allows an evaluation of vast variety of products, be it hardware, software, operating system, mobile devices, and websites. SUS has been used in more than 1300 science

article and publications, or product quality assurance. There are 15 questions for each teacher, the score for each question will ranged from -2 to 2. The result is as follows:

Number of Questions	: 15
Number of Respondent	: 2
Score	: 50
Maximum Score	: 60

According to the acquired data which can be seen in Appendix I, the system achieved score of 50 out of 60. In percentage it means the system achieved 83,33% of total functionality. It proved the system can fulfilled its role as a tool to collect emotional information and monitor student's mood in a classroom activity.

To estimate the confidence, calculate t-value, and then lookup the inverse of CDF of Student's t-distribution with  $(N_a - 1) + (N_b - 1)$  degrees of freedom.  $N_a$  is the size of sample A and  $N_b$  is the size of sample B. To find t-value you start from calculating the mean  $M_x$  and sum of squared deviations, or sum of squares  $SS = \sum (X_i - M_x)^2$  for each sample. Then, estimate the variance of the source population as  $\{s_p^2\} = \frac{SS_a + SS_b}{(N_a - 1) + (N_b - 1)}$ . This estimation is called pooled variance, a method for estimating variance of several different populations when the mean of each population may be different, but one may assume that the variance of each population is the same.

The writer estimate the standard error of  $M_{xa} - M_{xb}$  as  $est. \sigma_{M-M} = \sqrt{\frac{\{s_p^2\}}{N_a} + \frac{\{s_p^2\}}{N_b}}$ . Finally,  $t = \frac{M_{xa} - M_{xb}}{est. \sigma_{M-M}}$  calculate t as.

Index	:
$t_x$	: Size of the difference relative to the variation in sample data x
$SS_x$	: Sum of Squares of x
$N_x$	: Size of sample x
$M_x$	: Sum of squared deviations from sample x

Here is the result from the two sample that run by testing the average of pre-test results and the average of post-test results as follows:

- Mean of samples a : 0.5
- Mean of samples b : 0.2
- t-value : 10
- Level of Confidence on one and two tall's test of significance : 100%

## 5. Conclusion

From the result collected in Data result and analysis, the research can be concluded to be able to answer the research purpose and research question while conducted in the area of problem limitation and research limitation. The conclusion can be summed as follows:

- (1) The Facial Expression recognition tools provided desirable results on the scanning and image analysis. The system also successfully implemented Euclidian Distance formula and FURIA fuzzy rules to extract necessary information from the acquired image or image sequence into one of 7 basic human emotions.
- (2) By comparing the result when the system detect the emotion from JAFFE image set, with the one that already classified by experts, the system able to reaches the accuracy of 90%.

It can be seen that when used in a classroom activity, the system could make a significant difference to student's mood. This result could greatly vary from school to school, not to mention when used on different grades. Because there will be many other external variable that could affected the state of the students. However, from this research, the writer concluded that it is possible to have a scientific measurement toward student's stress level by using a system which could detect human emotions.

## 6. References

- [1] Peter Lewinsohn, Hyman Hops, Robert Roberts, and John Seeley, "Adolescent Psychopathology: I. Prevalence and Incidence of Depression and Other DSM-III-R Disorders in High School Students," *Journal of Abnormal Psychology*, vol. 102, no. 1, pp. 133-144, March 1993.
- [2] Shivakumar Gundappa and P.A.Vijaya, "Facial Expression Based Human Emotion Recognition with Live Computer Response," *IJCSIT International Journal of Computer Science and Information Technology*, vol. 4, no. 2, pp. 81-84, December 2011.
- [3] Paul Ekman and Wallace Friesen, *Pictures of Facial Affect*. Palo Alto: CA: Consulting Psychologist, 1976.
- [4] NIH. (2014, January) The National Institute of Mental Health (NIMH). [Online]. <https://www.nimh.nih.gov/about/index.shtml>
- [5] Ministry of Health Republic of Indonesia. (2016, October) Ministry of Health Republic of Indonesia official website. [Online]. <http://www.depkes.go.id/article/print/16100700005/peran-keluarga-dukung-kesehatan-jiwa-masyarakat.html>
- [6] Brett Marroquin, Chloe C. Boyle, and Susan Nolen-Hoeksema, "Using Emotion as Information in Future-Oriented Cognition: Individual Differences in the Context of State Negative Affect," *Pers Individ Dif*, pp. 121-126, 2016.
- [7] Abu Hanifah and Nunung Unayah, "MENCEGAH DAN MENANGGULANGI PENYALAHGUNAAN NAPZA," *Informasi*, pp. 33-46, 2011.
- [8] Shumaila Khurshid, Qaisara Parveen, M. Imran Yousuf, and Abid Ghafoor Chaudhry, "EFFECTS OF DEPRESSION ON STUDENTS' ACADEMIC PERFORMANCE," *Sci.int.(Lahore)*, p. 27(2), 2015.
- [9] Yuli Yanna Fauzie, "Bank Dunia: Kualitas Pendidikan Indonesia Masih Rendah," 2018.
- [10] Mohamed Sathik and G.Sofia Jonathan, "Effect of facial expressions on student's comprehension recognition in virtual," *SpringerPlus*, pp. 1-9, 2013.
- [11] Staff Writer. (2017) The Mountain Ear. [Online]. <https://themtnear.com/2017/12/tech-trends-facial-recognition-history-and-privacy/>
- [12] Hamit Soyel and Hasan Demirel, "Facial Expression Recognition Using 3D Facial Feature Distances," *Affective Computing, Focus on Emotion Expression, Synthesis and Recognition*, 2008.
- [13] Seyedehsamaneh Shojaeilangari, Wei-Yun Yau, Karthik Nandakumar, Jun Li, and Eam Khwang Teoh, "Robust Representation and Recognition of Facial Emotions Using Extreme Sparse Learning," *IEEE Transactions on Image Processing*, vol. 24, no. 7, pp. 2140 - 2152, July 2015.
- [14] Paul Ekman and Wallace Friesen. (1993) PaulEkmanGroup Website. [Online]. <https://www.paulekman.com/product/pictures-of-facial-affect-pofa/>
- [15] Braathen, B. Bartlett, M. Littlewort, G. Smith, and E. Movellan, "An approach to automatic recognition of spontaneous facial actions.," in *International Conference on FGR*, 2002, pp. 345-350.
- [16] Patrick Lucey et al., "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression," *IEEE*, 2010.
- [17] Liwei Wangy, Yan Zhangy, and Jufu Fengy, "On the Euclidean Distance of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1334 - 1339, August 2005.
- [18] Moon Hwan Kim, Young Hoon Joo, and Jin Bae Park, "Emotion Detection Algorithm Using Frontal Face Image," *ICCAS2005*, pp. 2373-2378, 2015.
- [19] Michael Lyons, Shigeru Akamatsu, Miyuki Kamachi, and Jiro Gyoba, "Coding Facial Expressions with GaborWavelets," in *IEEE International Conference on Automatic Face and Gesture Recognition*, Nara, 1998, pp. 200-205.
- [20] Douglas S. Shafer and Zhiyi Zhang, *Beginning Statistics*. Charlotte: Creative Commons-licensed, 2012.



