# A Novel CNN-Based Approach for Recognizing Facial Emotion

**Huu Hiep Nguyen†, Luong Anh Tuan Nguyen†, Anh Quan Tran† and Thi Ngoc Thanh Nguyen†**
Ho Chi Minh City University of Transport, Viet Nam†

**Summary**
Emotional recognition based on facial expressions is a very important research topic in the field of Human-Computer-Interaction. This paper proposed a new method to recognize emotions from facial image. First, based on Haar like features to detect left eye, right eye, left eyebrow, right eyebrow, nose and mouth. Then, using convolutional neural network (CNN) to recognize emotions with the above-mentioned detected images as input images. A software system based on the proposed method was designed to experiment. The experimental results with CK+ database and database show that the proposed solution achieved the best recognition rate of 98.1%.
***Key words:***
CNN, Facial Expressions, Haar Like

## 1. Introduction

Facial expressions recognition (FER) has been an active research in the computer vision field. in [1-4], authors presented facial expression recognition with hand-crafted feature extractors. In early 1970s, Ekman showed that there are six expressions across all cultures, namely disgust, anger, happiness, sadness, surprise and fear [2]. These expressions could be identified by observing the facial signals.
The recent success of convolutional neural networks (CNNs) in tasks like image classification [4] has been extended to the problem of facial expression recognition [5]. Unlike traditional machine learning and computer vision approaches where features are defined by hand, CNN learns to extract the features directly from the training database using iterative algorithms like gradient descent.
In computer vision, significant amount of researches on facial expression classification has led to many systems adopting different approaches. Survey descriptions can be found in Pantic et al. [6] and Fasel et al. [7]. Three main approaches exist: optical flow analysis from facial actions [8-12]; model based approaches [13-16]; and fiducial points based approaches [17-23].
This paper proposed a new method based on CNN to recognize facial emotions. The main contributions of this paper focuses on the following sections:

- Detecting left eye, right eye, left eyebrow, right eyebrow, nose and mouth based on Haar like features.
.
- Building CNN model to regconize facial emotions with left eye, right eye, left eyebrow, right eyebrow, nose and mouth images as input images.
- Designing a software system by C# to experiment with two databases.

The rest of the paper is structured as follows. The proposed method is described in detail of Section 2. Section 3 presents the experimental data, the experimental process and the experimental results. Section 4 gives conclusions and outlines future research directions.

## 2. Proposed Method

The system architecture including 2 modules is shown in Figure 1.

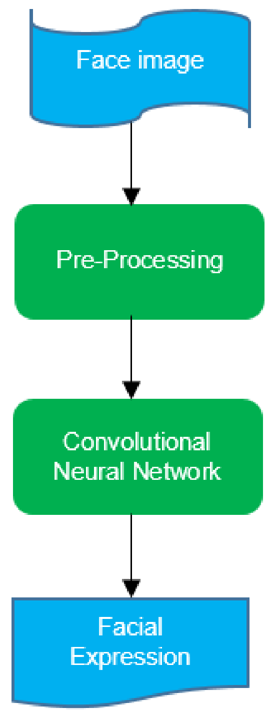


Fig. 1 Proposed System Architecture

- The first module is the "Pre-Processing" module which extracts features from facial image, and these features are the inputs of CNN.
- The second module is the CNN module which recognizes facial emotions from left eye image, right

Manuscript received May 5, 2019
Manuscript revised May 20, 2019

eye image, left eyebrow image, right eyebrow, nose image, mouth image.

## 2.1 “Pre-Processing” Module

This module is described in Figure 2.

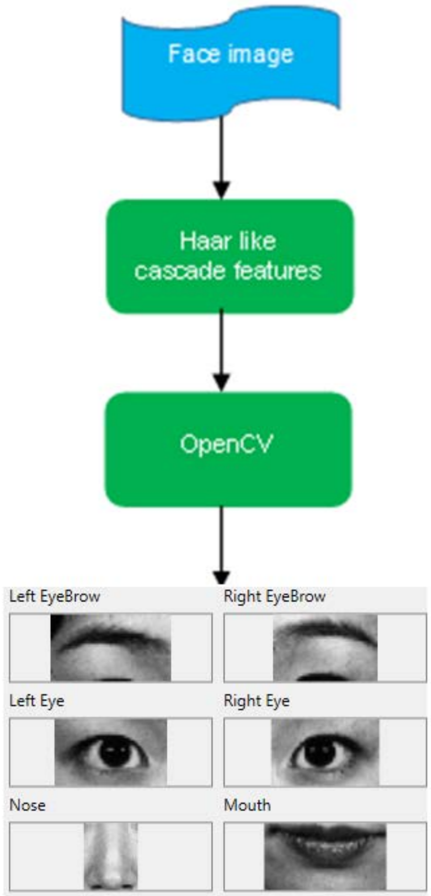


Fig. 2 Pre-Processing Module

This module extracts Haar features of facial image. Then using OpenCV to detect left eye, right eye, left eyebrow, right eyebrow, nose and mouth based on these Haar features.

## 2.2 CNN Module

This module is shown in Figure 3. This module includes the following layers.

### 2.2.1 Convolutional Layer

The convolutional layer extracts features from input images. Convolution preserves the relationship between pixels by learning image features using small squares of input data. In this layer, when input images are color images, three kernels corresponding to R, G, B shown in Figure 4 are used to generate convolved features. When input images are grayscale images, one kernel shown in Figure 5 is used to generate convolved feature. Stride=1 and padding=2 are used in proposed method.

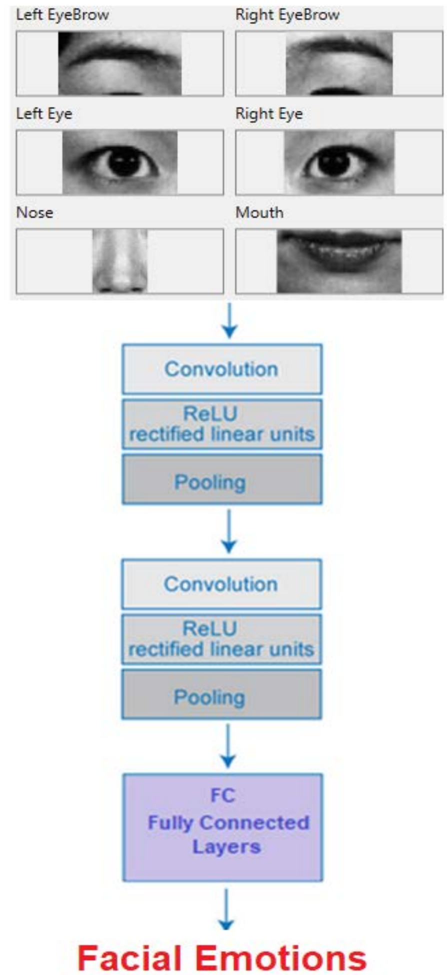


Fig. 3 CNN Module

| | | |
|---|---|---|
| -1 | -1 | 1 |
| 0 | 1 | -1 |
| 0 | 1 | 1 |

| | | |
|---|---|---|
| 1 | 0 | 0 |
| 1 | -1 | -1 |
| 1 | 0 | -1 |

| | | |
|---|---|---|
| 0 | 1 | 1 |
| 0 | 1 | 0 |
| 1 | -1 | 1 |

Fig. 4 Three Kernels for color image

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Fig. 5 One Kernel for grayscale image

2.2.2. ReLU Layer

ReLU stands for Rectified Linear Unit for a non-linear operation. The output is ***f(x) = max(0, x).*** Since, the real world data would be non-negative linear values. Figure 6 shows an example of ReLU.

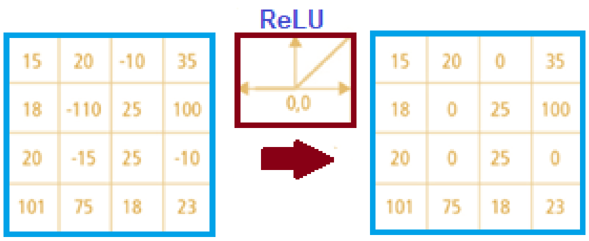


Fig. 6 Example of ReLU

2.2.3 Pooling layer

Pooling layers make reduce the number of parameters when the images are too large. The output of this pooling layer is a pooled featured map. Max Pooling and Average Pooling are two widely used pooling techniques. This paper uses Max Pooling with size=(2,2), stride=2 and padding=0. Figure 7 shows Max Pooling with 2x2 window and stride=2.

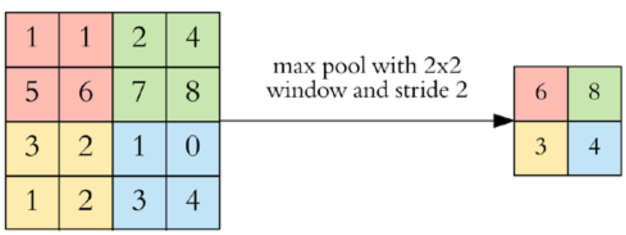


Fig. 7 Max Pooling

2.2.4 Fully Connected Layer

The flatten layer used to transfer pooled feature map into a one-dimensional vector. The flattened matrix goes through a fully connected layer to classify emotions. In proposed method, using the softmax activation function in the output layer to represent a categorical distribution over class labels, and obtaining the probabilities of each input element belonging to a label. Fully Connected Layer is described in Figure 8.

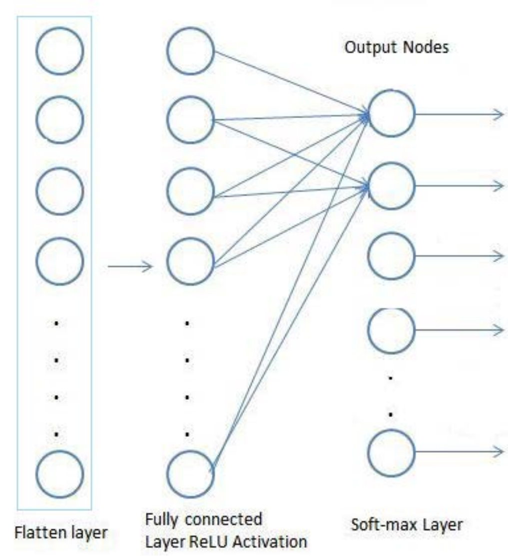


Fig. 8 Fully Connected Layer

# 3. Evaluation

## 3.1 Database Description

This paper utilized the JAFFE database and the extended Cohn–Kanade (CK+) database to evaluate the proposed method of facial expression recognition.
The JAFFE Database [24] contains 213 images in total. There are 10 subjects and 7 facial emotions for each subject. Each subject has about twenty images and each emotion includes two to three images. The seven emotions are anger, happiness, disgust, sadness, surprise, fear and neutral respectively. Figure 9 shows the seven emotions from one subject.


Fig. 9 JAFFE Database

The CK+ database [25] consists of 593 image sequences from 123 subjects, including eight basic emotion categories, which are anger, contempt, disgust, fear, happiness, sadness, surprise and neutral, as shown in Figure 10.


Fig. 10 CK+ database

## 3.2 Experimental Process

Experimental data was divided into 2 datasets as follows:

- The training dataset was gathered from 70% of the data.
- The testing dataset was gathered from 30% of the data.

Experimental process includes training phase and testing phase. A software is programmed by C# in Visual Studio 2017 to experiment.

### 3.2.1 Training Phase

The fully connected layers need to pass the training phase to learn. The flow of the training algorithm represents a backpropagation learning procedure [26]. This phase uses the training dataset to train with adaptive learning rate [27]. The parameters were set as follows:

- Mean error threshold value: $1\text{x}10^{-5}$
- Number of Epochs: 5,000
- The weights: initialize weights random values from -1 to 1.

### 3.2.2 Testing Phase

This phase uses weighting sets trained from training phase to test with the JAFFE testing dataset and the CK+ testing dataset.

## 3.3 Experimental Result

### 3.3.1 The JAFFE testing dataset

After performing the testing phase with the JAFFE testing dataset, the data of the confusion matrix [28] evaluating the performance of system was depicted in Figure 11.

| | | Actual Class | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | Neutral | Happiness | Sadness | Fear | Anger | Disgust | Surprise |
| Recognized Class | Neutral | 95.2% | 1.7% | 0.9% | 0.7% | 0.1% | 0.2% | 0.1% |
| | Happiness | 2.3% | 95.6% | 0.7% | 1.2% | 0.3% | 0.7% | 0.3% |
| | Sadness | 1.1% | 0.9% | 94.3% | 2.1% | 0.8% | 0.5% | 0.1% |
| | Fear | 0.4% | 0.2% | 2.1% | 94.7% | 0.4% | 0.9% | 0.2% |
| | Anger | 0.5% | 0.4% | 1.1% | 0.6% | 96.4% | 2.1% | 0.7% |
| | Disgust | 0.1% | 0.7% | 0.5% | 0.3% | 0.3% | 93.6% | 1.4% |
| | Surprise | 0.4% | 0.5% | 0.4% | 0.4% | 1.7% | 2.0% | 97.2% |

Fig. 11 Illustration of Confusion Matrix for The Result of Classification.

The success rates obtained by the testing with the JAFFE testing dataset are given in Figure 12. The average accuracy recognition rate is 94.9%.

| | TP | TN | FP | FN | AC | P |
|---|---|---|---|---|---|---|
| Neutral | 95.2% | 94.9% | 5.1% | 4.8% | 95.0% | 94.9% |
| Happiness | 95.6% | 94.8% | 5.2% | 4.4% | 95.2% | 94.9% |
| Sadness | 94.3% | 95.0% | 5.0% | 5.7% | 94.7% | 95.0% |
| Fear | 94.7% | 95.0% | 5.0% | 5.3% | 94.8% | 95.0% |
| Anger | 93.9% | 95.1% | 4.9% | 6.1% | 94.5% | 95.0% |
| Disgust | 93.6% | 95.2% | 4.9% | 6.4% | 94.4% | 95.1% |
| Surprise | 97.2% | 94.6% | 5.5% | 2.8% | 95.9% | 94.7% |

Fig. 12 The Experimental Result of The JAFFE Testing Dataset.

Here, True positives (TP) refers to the positive tuples that were correctly labeled by the classifier. True negatives (TN) refers to the negative tuples that were correctly labeled by the classifier. False positives (FP) refers to the negative tuples that were incorrectly labeled as positive. False negatives (FN) refers to the positive tuples that were mislabeled as negative. The *accuracy* (*AC*) is the proportion of the total number of predictions that were correct. The *precision* (*P*) is the proportion of the predicted positive cases that were correct.

The accuracy recognition rate of each emotion is shown in Figure 13.

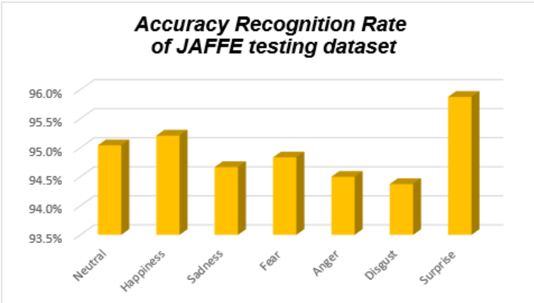


Fig. 13 The accuracy recognition rate of the JAFFE testing dataset

### 3.3.2 The CK+ testing dataset

After performing the testing phase with the CK+ testing dataset, the data of the confusion matrix [28] evaluating the performance of system was depicted in Figure 14.

| | | Actual Class | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Neutral | Happiness | Sadness | Fear | Anger | Disgust | Surprise | Contempt |
| Recognized Class | Neutral | 97.3% | 0.3% | 0.6% | 0.2% | 0.2% | 0.2% | 0.0% | 0.0% |
| | Happiness | 0.4% | 98.7% | 0.5% | 0.3% | 0.1% | 0.3% | 0.0% | 0.1% |
| | Sadness | 0.7% | 0.5% | 96.9% | 1.1% | 0.2% | 0.1% | 0.0% | 0.1% |
| | Fear | 0.5% | 0.1% | 1.2% | 97.7% | 0.1% | 0.2% | 0.0% | 0.2% |
| | Anger | 0.4% | 0.0% | 0.1% | 0.2% | 98.8% | 0.3% | 0.2% | 0.7% |
| | Disgust | 0.2% | 0.1% | 0.2% | 0.3% | 0.2% | 97.9% | 0.3% | 0.6% |
| | Surprise | 0.2% | 0.2% | 0.1% | 0.1% | 0.1% | 0.8% | 99.1% | 0.3% |
| | Contempt | 0.3% | 0.1% | 0.4% | 0.1% | 0.3% | 0.2% | 0.4% | 98.0% |

Fig. 14 Illustration of Confusion Matrix for The Result of Classification.

The success rates obtained by the testing with the CK+ testing dataset are given in Figure 15. The average accuracy recognition rate is 98.1%.

| | TP | TN | FP | FN | AC | P |
|---|---|---|---|---|---|---|
| Neutral | 97.3% | 98.2% | 1.8% | 2.7% | 97.7% | 98.2% |
| Happiness | 98.7% | 98.0% | 2.1% | 1.3% | 98.3% | 98.0% |
| Sadness | 96.9% | 98.3% | 1.8% | 3.1% | 97.6% | 98.2% |
| Fear | 97.7% | 98.1% | 1.9% | 2.3% | 97.9% | 98.1% |
| Anger | 98.8% | 97.9% | 2.1% | 1.2% | 98.4% | 98.0% |
| Disgust | 97.9% | 98.1% | 1.9% | 2.1% | 98.0% | 98.1% |
| Surprise | 99.1% | 97.9% | 2.1% | 0.9% | 98.5% | 97.9% |
| Contempt | 98.0% | 98.1% | 1.9% | 2.0% | 98.0% | 98.1% |

Fig. 15 The Experimental Result of The CK+ Testing Dataset.

The accuracy recognition rate of each emotion is shown in Figure 16.

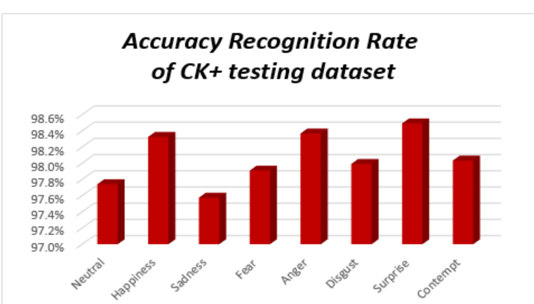


Fig. 16 The accuracy recognition rate of the CK+ testing dataset

### 3.3.3 Some demos running on proposed method

The following are demos running on software that are classified into 8 emotions.


Fig. 17 Anger Emotion

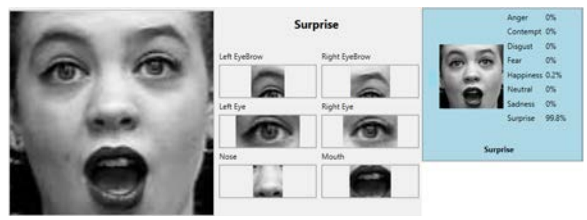


Fig. 18 Surprise Emotion

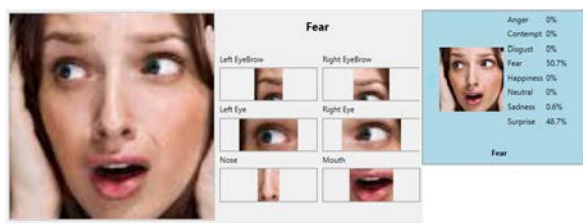


Fig. 19 Fear Emotion

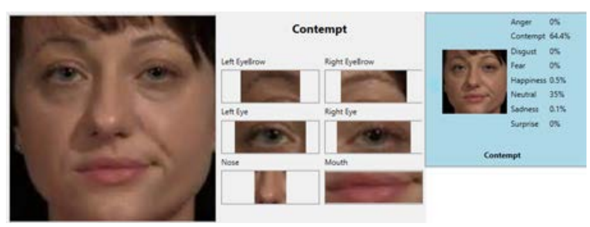


Fig. 20 Contemp Emotion

The following are demos running on software that are classified into 7 emotions.

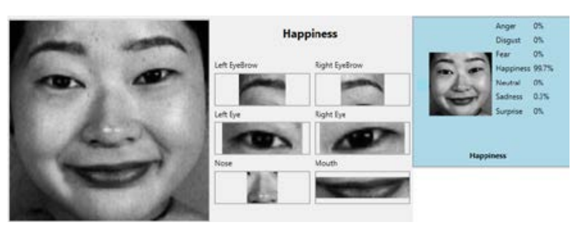


Fig. 21 Happiness Emotion


Fig. 22 Sadness Emotion

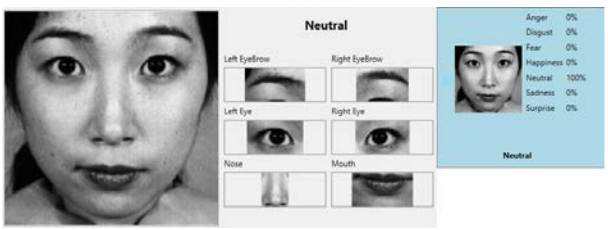


Fig. 23 Neutral Emotion

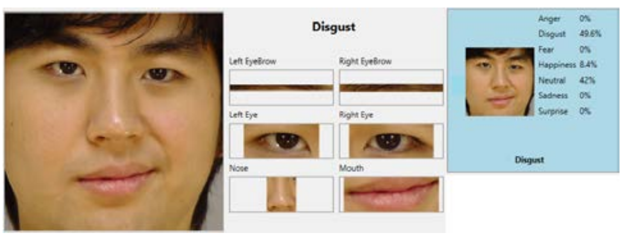


Fig. 24 Disgust Emotion

3.3.4 Comparing to other works

With CK+ database, Nianyin Zeng *et al.* [29] using DSAE achieved the accuracy rate of 96%; Patrick Lucey *et al.* [30] using SVM achieved the accuracy rate of 86%. However, the average accuracy rate of proposed method is 98.1%.
With JAFFE database, André Teixeira Lopes et al.[31] using a combination of CNN achieved the accuracy rate of 54%; Anisha Halder *et al.* [32] using custom methods (IT2FS, GT2FS) achieved the accuracy rate of 95%. However, the average accuracy rate of proposed method is 94.9%.

## 4. Conclusion

This paper proposed a novel approach to recognize facial emotions based on CNN. The proposed method was experimented with CK+ database and JAFFE database via a software system programmed by C#. The experimental results show that the proposed method achieved good results. However, the achieving results are not the best. Thus, in the future, our group will improve the system aim to get more and more better results and experiment in the real world.


## Acknowledgments

The authors thank reviewers for their reading of our manuscript and their insightful comments and suggestions.


## References


[1] Dosovitskiy, A. et al.: Discriminative Unsupervised Feature Learning with Exemplar Convolutional Neural Networks. ArXiv14066909 Cs. (2014).

[2] Ekman, P., Friesen, W.V.: Constants Across Cultures in the Face and Emotion. J. Pers. Soc. Psychol. 17, 2, 124–129 (1971).

[3] Goodfellow, I.J. et al.: Challenges in Representation Learning: A Report on Three Machine Learning Contests. ArXiv13070414 Cs Stat. (2013).

[4] He, K. et al.: Deep Residual Learning for Image Recognition. ArXiv151203385 Cs. (2015). Demo-Video. Neurophone (September, 2017). http://www.youtube.com/watch?v=tc82Z_yfEwc.

[5] Kim, B.-K. et al.: Hierarchical Committee of Deep Convolutional Neural Networks for Robust Facial Expression Recognition. J. Multimodal User Interfaces. 10, 2, 173–189 (2016).

[6] M. Pantic, M. Rothkrantz, Automatic analysis of facial expressions: the state of the art, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (12) (2000).

[7] B. Fasel, J. Luettin, Automatic facial expression analysis: a survey, Pattern Recognition 1 (30) (2003) 259– 275.

[8] Y. Yacoob, L.S. Davis, Recognizing human facial expressions from long image sequences using optical flow, IEEE Transactions on Pattern Analysis and Machine Intelligence 18 (6) (1996) 636–642.

[9] M. Rosenblum, Y. Yacoob, L. S Davis, Human expression recognition from motion using a radial basis function network architecture, IEEE Transactions on Neural Networks 7 (5) (1996) 1121–1137. Z. Hammal et al. / Internat. J. Approx. Reason. 46 (2007) 542–567 565.

[10] M.J. Black, Y. Yacoob, Recognizing facial expression in image sequences using local parameterized models of image motion, International Journal of Computer Vision 25 (1) (1997) 23–48.

[11] I.A. Essa, A.P. Pentland, Coding, analysis, interpretation, and recognition of facial expressions, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (7) (1997) 757–763.

[12] J.F. Cohn, A.J. Zlochower, J.J. Lien, T. Kanade. Feature-point tracking by optical flow discriminates subtles differences in facial expression, in: Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition, April, Nara, Japan, 1998, pp. 396–401.

[13] M. Lyons, S. Akamatsu, Coding facial expressions with gabor wavelets, in: Proceedings of the IEEE International

Conference on Automatic Face and Gesture Recognition, April, Nara, Japan, 1998, pp. 200– 205.

[14] Z. Zhang, L. Lyons, M. Schuster, S. Akamatsu, Comparison between geometry-based and Gabor waveletsbased facial expression recognition using multi-layer perceptron, in: Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition, 1998, pp. 454–459.

[15] Y. Gao, M.K.H. Leung, S.C. Hui, M.W. Tananda, Facial expression recognition from line-based caricatures, IEEE Transactions on System Man and Cybernetics – PART A: System and Humans 33 (3) (2003).

[16] N. Oliver, A. Pentland, F. Be´rard, LAFTER: a real-time face and tracker with facial expression recognition, Pattern Recognition 33 (2000) 1369–1382.

[17] B. Abboud, F. Davoine, M. Dang, Facial expression recognition and synthesis based on appearance model, Signal Processing: Image Communication 19 (8) (2004) 723–740.

[18] J.J. Lien, T. Kanade, J.F. Cohn, C. Li. Subtly different facial expression recognition and expression intensity estimation, in: Proceedings of the IEEE Computer Vision and Pattern Recognition, June 23–25, Santa Barbara, CA, 1998, pp. 853–859.

[19] M. Pantic, L.J.M. Rothkrantz, Expert system for automatic analysis of facial expressions, Image and Vision Computing Journal 18 (11) (2000) 881–905.

[20] Y. Tian, T. Kanade, J.F. Cohn, Recognizing action units for facial expression analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (2) (2001) 97–115.

[21] I. Cohen, F.G. Cozman, N. Sebe, M.C Cirelo, T.S. Huang. Learning Bayesian network classifiers for facial expression recognition using both labeled and unlabeled data, in: Proceedings of the IEEE Computer Vision and Pattern Recognition, 2003.

[22] M. Parda`s, Extraction and tracking of the eyelids, in: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 4, Istambul, Turkey, June 2000, pp. 2357–2360.

[23] N. Tsapatsoulis, K. Karpouzis, G. Stamou, F. Piat, S.A. Kollias, A fuzzy system for emotion classification based on the MPEG-4 facial definition parameter set, in: Proceedings of the 10th European Signal Processing Conference, Tampere, Finlande, September 2000.

[24] The Japanese Female Facial Expression (JAFFE) Database. http://www.kasrl.org/jaffe.html

[25] Patrick Lucey, Jeffrey F. Cohn, Takeo Kanade, Jason Saragih, Zara Ambadar, Iain Matthews. The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression. IEEE, 978-1-4244-7030-3/10, 2010, pp. 94-101.

[26] Jiawei Han, Micheline Kamber, Data Mining: Concepts and Techniques, Second Edition, Elsevier Inc. All rights reserved, 2006.

[27] Z. X. Bo and W. L. Li, "Handwritten digit recognition based on improced leaning rate BP algorithm," in *Proc. 2nd International Conference on Information Engineering and Computer Science*, pp. 1-4, 2010.

[28] Kohavi, R. and F. Provost (1998). Glossary of terms . Editorial for the Special Issue on Applications of Machine Learning and the Knowledge Discovery Process.

[29] Zeng, Nianyin, et al. "Facial expression recognition via learning deep sparse autoencoders." Neurocomputing 273 (2018): 643-649.

[30] Patrick Lucey, Jeffrey F. Cohn, Takeo Kanade, Jason Saragih, Zara Ambadar, Iain Matthews. The Extended Cohn-Kanade Dataset (CK+):A complete dataset for action unit and emotion-specified expression. IEEE, 978-1-4244-7030-3/10, 2010, pp. 94-101.

[31] Lopes, André Teixeira, et al. "Facial expression recognition with convolutional neural networks: coping with few data and the training sample order." Pattern Recognition 61 (2017): 610-628.

[32] Anisha Halder, Amit Konar, Rajshree Mandal, Aruna Chakraborty, Pavel Bhowmik, Nikhil R. Pal and Atulya K. Nagar, General and Interval Type-2 Fuzzy Face-Space Approach to Emotion Recognition, IEEE Transactions on Systems, Man, and Cybernetics: System, vol. 43, No. 3, May 2013, pp. 587-605.

**Huu Hiep Nguyen** received B.Sc. degree in Computer Science from Ho Chi Minh City University of Transport, Viet Nam. Currently, He is preparing to get a M.Sc degree in Computer Science. His current research interests include Computer Vision, Image Processing and Artificial Intelligence.

**Luong Anh Tuan Nguyen** is the head of Information System Depatment, Ho Chi Minh City University of Transport, Viet Nam. He received B.Sc. and M.Sc. in Computer Science from University of Science – Viet Nam National University Ho Chi Minh City. He also received Ph.D. degree in Control Engineering and Automation. His current research interests include Artificial Intelligence, Soft Computing, Control Engineering and Security.