Emotion Recognition using Deep Neural Network with Vectorized Facial Features

Guojun Yang, Jordi Saumell y Ortoneda and Jafar Saniie

Department of Electrical and Computer Engineering
Illinois Institute of Technology
Chicago, Illinois, United States

***Abstract*—Emotion reveals valuable information regarding human communications. It is common to use facial expressions to express emotions during a conversation. Moreover, some interpersonal communication can be achieved using facial expressions only. Some facial expressions are universal, they express the same emotion across cultures. If a machine were able to interpret its user's facial expression correctly, it might be able to help its user more efficiently. In this paper, a novel vectorized facial feature for facial expression will be introduced. The vectorized facial feature can be used to build an DNN (Deep Neural Network) for emotion recognition. Using the proposed vectorized facial feature, the DNN can predict emotions with 84.33% accuracy. Nevertheless, compared with CNNs (Convolutional Neural Network) with similar performance, training such DNN requires less time and data.**

I. Introduction

Emotions are, reaction or a series of reactions of certain events [1] [2]. With different emotion, people may react differently on certain event. Without knowing others' emotion, one might misunderstand other's intensions. Therefore, if a machine were able to perceive its users' emotion, it might be able to help its users more efficiently. Nevertheless, allow machines to perceive emotions is challenging. Emotions are hard to be modeled. More specifically, some of these emotions can be observed or measured easily through some physical indicators, such as heart rate variation, sweating or even vomiting, other emotions cannot be observed. Yet, for those observable emotions, it is challenging to constantly monitoring one's physical indicators consistently. Hence observing facial expression is the most common way to perceive others emotion during human communication.

The ability of perceiving others' emotion correctly is called emotional intelligence. To be more specific, emotional intelligence can be divided into three categories, appraisal and expression of emotion; regulation of emotion; utilization of emotion [2]. The first category, regarding appraisal and of emotion, is the main topic of this paper. The importance of expression of emotion come from two different aspects, one is some of the emotions are reflected on facial expressions, the other is some emotions are universal.

Regarding the expression of emotions through facial expression. In many cases, people express their emotions through facial expression spontaneously. Once emotion arise, the subject will express their feeling accordingly. There are studies concludes, some expressions of emotions are emotions biologically and genetically related [3]. Thus, through facial expression, many emotions can be perceived correctly.

Moreover, some emotions are universal. Therefore, those emotions can be recognized by others regardless of their culture or religion. This hypothesis, that facial expression can be recognized across culture was first proposed by Darwin. Later in different res studies, researchers found evidence supporting such hypothesis [4]. The homogeneity across culture in the link between facial expression and emotion can be found in infants' reaction to their mothers, which imply such ability does not come from influence of culture [5]. The hegemony in facial expression suggested that, if a computer system were able to analyze the facial expression of its user, then it can anticipate its user better understand the emotion of that person.

Due to the natural of physical status of an individual, it is impractical to observe emotion through sensor attach to test subjects. Therefore, analyzing facial expression is a reasonable alternative. Facial expressions are formed by the contracting different combination of muscles on facial area. These changes on facial area create facial features. Facial features can be categorized in two different groups, one group of features are permanent features, the other are features that only appear when certain muscles been contracted. Permanent features include eyebrow lip cheek, these features will appear regardless of one's facial expression [6]. By combining these information, a coding for facial expression can be created for computer vision systems to analyze emotion.

In this paper, vectorized facial features will be introduced to represent facial expression. The proposed facial feature model can not only reflect facial expressions correctly, it can also be used for DNN with high efficiency. To test the efficiency of such method, a DNN is trained to recognize some universal expressions. Finally, the result will be presented and discussed.

978-1-5386-5398-2/18/$31.00 ©2018 IEEE

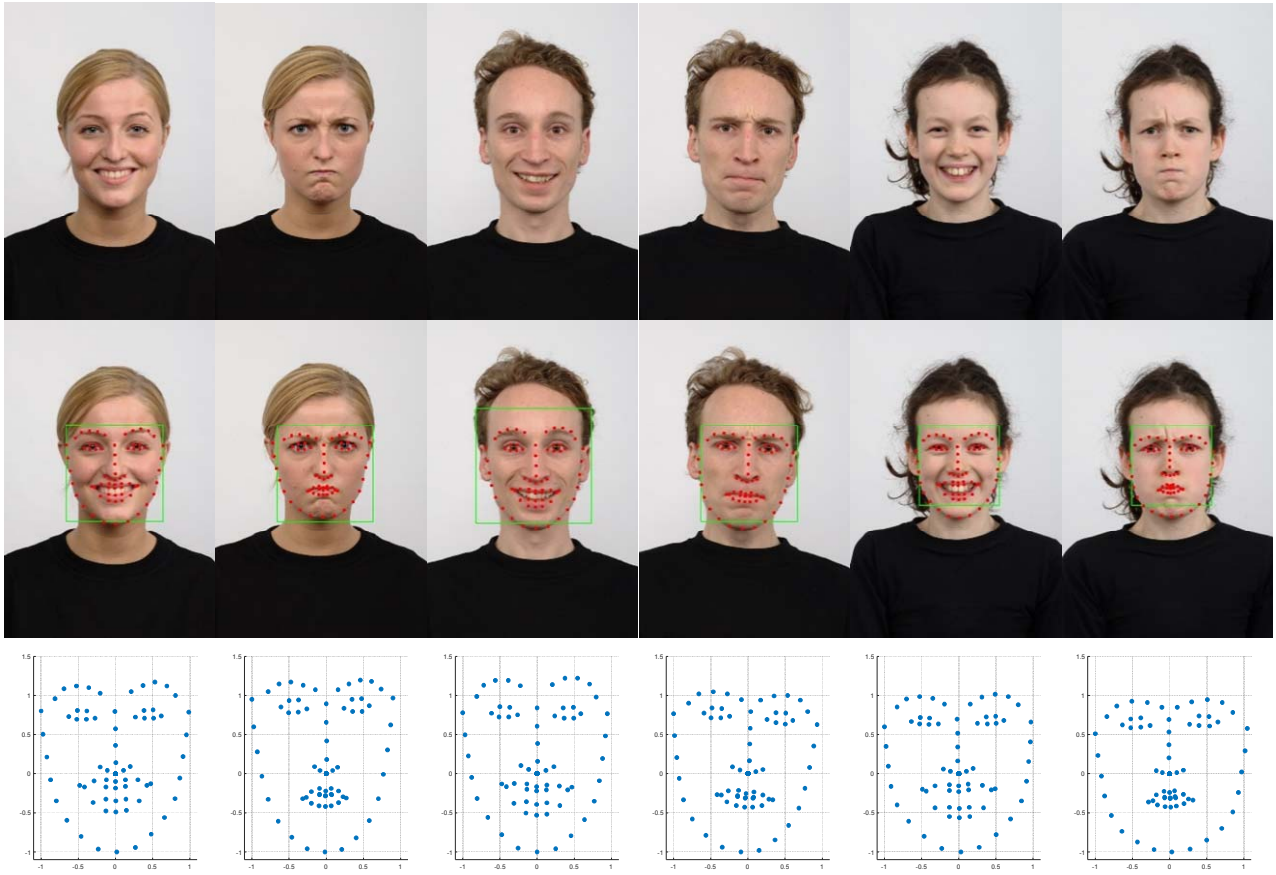

Figure 1. Example of Test Subjects and Respective Facial Landmarkers, In the first row, raw images collected from test subjects are shown. In the second row, faces and facial feature landmarkers are shown as green box and red dot respectively. In the third row, normalized facial feature landmarkers are shown in red dots.

II. Emotion Recognition

Facial expressions reflect the emotions of person, which reveal valuable information of one's feeling, reaction etc. To correctly recognize these emotions is challenging. Using machine recognizing emotions been explored by researchers extensively. There are varies emotion recognition methods. In this section, existing emotion recognition methods will be discussed. Then emotion recognition methods using computer vision techniques will be discussed and compared with the proposed method.

A. Traditional Emotion Analysis Methods

Existing solution towards emotion recognitions are usually recognize the emotion of the test subjects by analyzing their physical status. For example, heart rate, body temperature etc. ECG signal can be one indicator to emotion of human. Through ECG (Electrocardiography) signal, the system can even recognize the emotion of volunteers even when participants are trying to hide their emotions [7]. There are systems that is able to analyze the emotion of individual through other physiological signals, such as electromyogram (EMG) signal, electroencephalogram (EEG), galvanic skin response (GSR), blood volume pressure (BVP), heart rate (HR) or heart rate variability (HRV), temperature (T), respiration rate (RR) [7]. These methods require using designated sensor(s) to collect information from test subjects. For none scientific applications, relying on sensors attach to skin of users is impractical.

There are researchers claiming combining Wi-Fi signal and machine learning algorithms, it is possible to sense the emotion of the subjects in a room. By observing the disturbance on Wi-Fi signals created by breathing, the system is able to unveil ECG signals from test subjects, consequently the emotions of the test subjects can be recognized [8]. Nevertheless, such method can only differentiate the emotion that involving ECG change.

B. Computer Vision Based Method

Computer vision techniques appear to be a viable solution towards this emotion recognition. As human, with sufficient knowledges, computer vision powered systems are able to generate proper results photos are provide. Computer vision powered systems can help navigating indoor [9], recognize individua, and perform other tasks that involving visual perception. Modeling facial expressions for computer vision systems is challenging. Naturally, facial expressions are hard to be described and quantified. Some emotions will generate certain temporary facial features, such as winkles around eye may deepen when smiling etc. Temporary facial features were

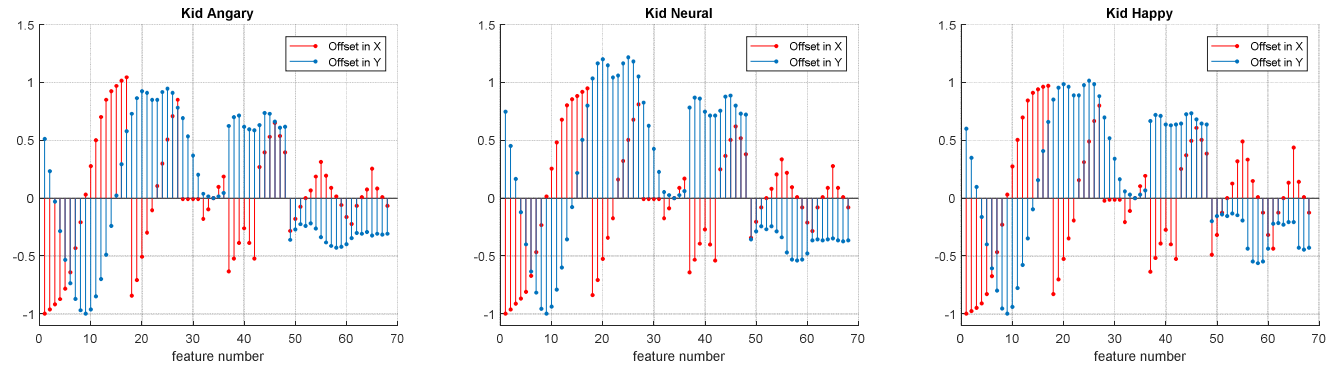

Figure 2. Visualization of Proposed Vectorized Facial Landmarkers.

used in some researches to classify emotions of test subjects. However, temporary facial expressions vary across different people. Even for the same person, his (her) temporary facial features my change through his (her) life due to aging. Therefore, using temporary facial expression will make the computer vision system perform inferior in terms of uniformity. Another more practical approach is using the statistical models of permanent facial feature to model facial expression.

Facial landmarkers are key points describing the location of permanent facial features. When test subjects express certain emotions through their faces, the stretch and compress of facial muscles will cause shifting of these facial landmarkers. These shifting of landmarkers are easy to be detected by human eyes, but hard using conventional methods to analyze, since linear classifiers are not optimal when solving problem. With proper preprocessing, DNNs are able to classify these emotions with good accuracy. In this project, an annotation tool implemented by Dlib library [10] [11].

III. Vectorized Facial Features

As discussed in the previous sections, it is challenging to differentiate different emotions using solely linear classifier and facial landmarkers. Nonetheless, through proper training, a DNN is able to build a model automatically to tackle this problem. In this section, the proposed vectorized emotion model will be introduced.

A. Facial Feature Normalization

Neural network is a powerful tool. Yet it needs to be fed with proper data. Normalization of training and input data has tremendous impact on the result of training, consequently hinge the performance of the system. Moreover, neural networks are good at finding differences that human always ignore, however it needs clues for such difference. With only unprocessed raw data, training a neural network will require a larger database and the neural network will be less likely to converge.

The relative position of all facial landmarkers are unique for each emotion, yet share certain degree of similarity across test subjects. Based on such assumption, a DNN was built and trained. To place emphasis on the shifting of facial muscles, these landmarkers are store as vector of coordinates. Therefore, by comparing the relative position of all facial landmarkers, the shifting can be detected and modeled.

The size of faces varies among people. Without any adjustment, such difference in sizes might affect the overall performance of the neural network. To elaborate, the distance between each landmarker will appear to be different even when same facial expression is presented. To eliminate the affect created by the size differences of faces, a normalization procedure was taken. Firstly, to best describe the relative position of facial landmarkers, instead of using absolute position of each landmarker, the facial landmarkers were all positioned relative to the landmarker associated with the tip of the nose. Secondly, the width face will be scaled to 1, all other landmarkers are scaled in horizontally according to this factor. Lastly, the height from the tip of the nose to bottom of the chin is always scaled to 1(actually -1 for the chin), and all other facial landmarkers are scaled in vertical direction accordingly. With these three steps, all facial landmarkers are normalized to a 1-by-1 rectangular box. To further explain the normalization process, as shown in Figure 1, each test subject, as shown in the first row of the figure, will be extract with facial landmarkers. These extracted landmarkers (as shown in the second row of the figure), were further been normalized (as shown in the third row of the figure). Pictures shown in Figure 1 are collected from 3 different test subjects from Radboud [12] dataset, these three test objects obviously have different size of faces, moreover, the position of those facial landmarks are different relative to the tip of nose (shown as the second row of Figure 1). After normalization, these variations of landmarkers caused by differences in appearances among people are reduced (shown as the third row of Figure 1). Such normalization is carried out by transformation in basis and shifting.

B. Vectorization of Facial Features

Compared with more complicated neural network structures, such as CNNs (Convolutional Neural Netweork) and RNNs (Recurrent Neural Network), a DNN can be trained with less data, therefore, to be built quicker. To be more specific, with simpler inputs, a neural network can have a simpler structure, hence, can be trained with less data, and converge quicker. To serve the simplicity purpose, vectorized facial landmarkers were used to train a DNN.

The vectorization of facial landmarkers is achieved by putting tensors of 2-dimensional coordinates into a vector. Since these coordinates are normalized, when the vectorized facial landmarkers being fed into the DNN, the network can be trained properly. As shown in Figure 2, three different data of the same test object been presented. In each subfigure of Figure 2, the X axis and Y axis represent the feature number and the magnitude of the variable respectively, stems in red represent the offset of each landmarker in horizontal (X) direction, while stems in blue represent the offset of each landmarker in vertical (Y) direction. Each subfigure in Figure 2, denotes the vectors of all landmarkers associated with happy, neutral and angary respectively.

By observing these stems in Figure 2, a few facts can be highlighted. First, when compare angry with neutral, feature 19 – feature 26 are increased in magnitude in horizontal (Y) direction. These features are associated with the eyebrows, while angary, eyebrows will drop, which is reflected in the vectorized feature chart. Another intuitive observation can be made by comparing stems in happy and stems in neutral, feature 60 – 68 has larger offset in horizontal (X) direction. These features are associated with mouth, which will be stretched outwards while smiling. These two examples intuitively illustrate how emotions can be reflected on vectorized facial landmarkers.

DNNs are known to be effective while performing classification tasks. To perform emotion recognition task, a 273-by-545-by-273 DNN was built. This DNN has three hidden layers, all layers are fully connected. As shown in Figure 3, with vectorized input layer, 136 nodes are in the input layer. Since there are only eight different emotions to be recognized, only 8 nodes present in the output layer.

IV. Implementation and Experiment Results

With proposed vectorized facial landmarker method, a series experiments were performed to evaluate the efficiency of the proposed method.

A. Training the Neural Network

To train the neural network, TensorFlow [13] was used to accelerate the process. The step taken for training a DNN has been explored extensively. The training of the DNN for emotion detection bare little difference. After extracting facial landmarkers from Radboud database, normalization and vectorization was performed. As introduced in vectorization process, all raw images were first extracted with facial landmarkers, then normalized and vectorized. Moreover, to avoid overtraining and biased network, 5% of the training data was reserved for evaluation.

The DNN was implemented in python with TensorFlow. With the help of GPU, TensorFlow was able to finish training on the network in an hour. Mean-squared-error was used to evaluate the cost in each iteration. As shown in Figure 4, the loss rate drops to almost zero after 10000 of iterations. After training, the DNN is able to predict the emotion of with 92% accuracy on evaluation dataset.

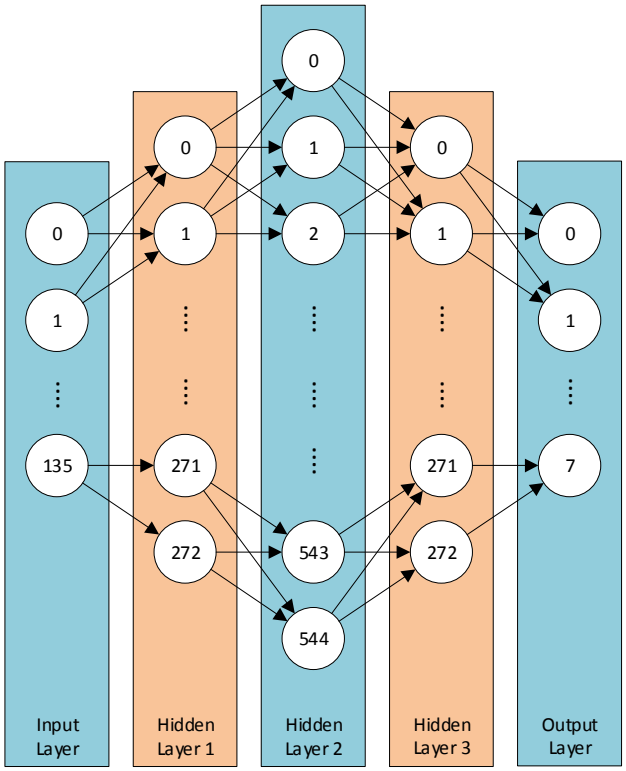

Figure 3. Structure of the DNN

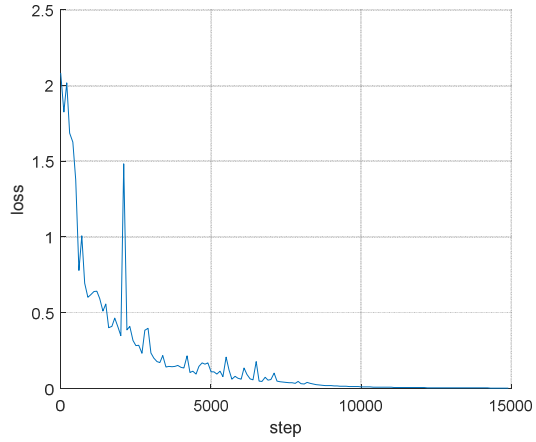

Figure 4. Loss Rate Through Iterations

B. Real World Testing

Finally, the DNN was used to test on volunteers from the lab. To collect data efficiently, an app was implemented. Using this app, all test subjects were asked to express emotion as instructed. A photo will be taken for each expression. With the trained DNN each of the photos will be evaluated.

The result is shown in Table , each row shows the one emotion test subjects were asked to express, each column in a row represents the output from the DNN trained using vectorized facial features. For example, in the second row, 39 photos were taken for contemptuous expression, the system was able to predict 37 of them correctly, one was recognized as

neutral and one was recognized as sad. With such statistical method the system is able to predict emotions from volunteers with 84.33% of accuracy.

TABLE I. CROSS CORRELATION OF ACCURACY EMOTIONS EXPRESSED(COLUMN) AND EMOTION DETECTED(ROW)

Angry	39	0	0	0	0	0	0	0
Contemptuous	0	37	0	0	0	1	1	0
Disgusted	0	0	39	0	0	0	0	0
Fearful	0	0	0	34	0	3	0	2
Happy	0	0	0	0	39	0	0	0
Neutral	0	1	0	1	0	34	3	0
Sad	0	0	0	3	0	0	35	1
Surprised	0	0	0	0	0	0	0	39

V. CONCLUSION

Emotions can be reflected on facial expressions. Thus, with proper modeling, emotions can be recognized using tools such as, computer vision and DNN. In this paper, the proposed vectorized facial feature can be used to train a DNN for emotion recognition. Compared with other computer vision powered system, vectorized facial features can achieve similar accuracy as other machine learning algorithms (CNN). Yet, it reduces the data as well as the time required for training. Such advantages can significantly increase the speed of building applications involving emotion recognition.

REFERENCE

[1] S. W. Porges, "Emotions and Interpersonal Communication," in *The polyvagal theory: Neurophysiological foundations of emotions, attachment, communication, and self-regulation (Norton Series on Interpersonal Neurobiology)*, WW Norton & Company, 2011.

[2] P. Salovey and J. D. Mayer, "Emotional Intelligence," *Imagination, Cognition and Personality ,* vol. 9, no. 3, pp. 185-211, 1990.

[3] D. Matsumoto and B. Willingham, "Spontaneous facial expressions of emotion of congenitally and noncongenitally blind individuals.," *Journal of personality and social psychology,* vol. 96, p. 1, 2009.

[4] C. E. Izard, "Innate and universal facial expressions: evidence from developmental and cross-cultural research.," *Psychological Bulletin,* vol. 115, no. 2, pp. 288-299, 1994.

[5] P. Ekman and H. Oster, "Facial expressions of emotion," *Annual review of psychology,* vol. 30, pp. 527--554, 1979.

[6] Y.-I. Tian, T. Kanade and J. F. Cohn, "Recognizing action units for facial expression analysis," *IEEE Transactions on pattern analysis and machine intelligence,* vol. 23, no. 2, pp. 97-115, 2001.

[7] F. Agrafioti, D. Hatzinakos and A. K. Anderson, "ECG Pattern Analysis for Emotion Detection," *IEEE Transactions on Affective Computing,* vol. 3, no. 1, pp. 102-115, 2012.

[8] M. Zhao, F. Adib and D. Katabi, "Emotion recognition using wireless signals," in *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*, New York City, 2016.

[9] G. Yang and J. Saniie, "Indoor Navigation for Visually Impaired Using AR Markers," in *Electro Information Technology (EIT), 2017 IEEE International Conference on* , Lincoln, 2017.

[10] C. Sagonas, G. Tzimiropoulos, S. Zafeiriou and M. Pantic, "A Semi-automatic Methodology for Facial Landmark Annotation," in *2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2013.

[11] D. E. King, "Dlib-ml: A Machine Learning Toolkit," *Journal of Machine Learning Research,* vol. 10, pp. 1755-1758, 2009.

[12] O. Langner, R. Dotsch, G. Bijlstra, D. Wigboldus, S. Hawk and A. Knippenberg, "Presentation and validation of the Radboud Faces Database.," in *Cognition & Emotion*, 2010, pp. 1377-1388.

[13] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mane, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viegas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu and X. Zheng, "TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems," 2015. [Online]. Available: https://www.tensorflow.org. [Accessed 1 1 2018].