



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJIRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 11, Issue 6, June 2022

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.118



9940 572 462



6381 907 438



ijirset@gmail.com



www.ijirset.com

Facial Emotion Recognition using Combined ANN and SVM

Aswathy Anand ¹, Hari Sivan.AS ², Shahana Habeeb Mohammed ³

P.G Student, Department of Communication Engineering, Mount Zion Engineering College, Kadamanitta, Kerala, India¹

Associate Professor, Department of Electronics and Communication Engineering, Mount Zion Engineering College, Kadamanitta, Kerala, India²

Associate Professor, Department of Electronics and Communication Engineering, Mount Zion Engineering College, Kadamanitta, Kerala, India³

ABSTRACT: This paper presents combined ANN and SVM based emotion recognition system. The system is comprised of face recognition stage followed by an emotion recognition stage. Using HOG and Gabor filter face and facial are extracted. Next, the relevant identifying points are extracted from each facial feature. In the emotion recognition stage uses ANN and SVM, we compare both classifier result score and highest score classifier output will display. These strengths are then used to determine the subject's displayed emotion. Here we use 10 emotions The performance resulting in an overall successful detection rate of 84.9%..

KEYWORDS: Gabor Filter, Emotion Recognition system, ANN, SVM.

I. INTRODUCTION

Emotion recognition is a problem that is relevant in a wide variety of fields such as AI based human interaction, Robotics, AI based autism treatment. The recognition of emotion from facial expressions has been a main focus of many psychological studies over the past decade [1]. It is no wonder that many studies have arisen from this topic. Applications of this topic include automated security systems, interactive robotic aids assistants, and diagnosing medical patients [2]. In a security system, accurate emotion recognition could help identify potential threats. Similarly, it would allow robotic aids to more effectively interact with and help a human counterpart. In each case, accurate and timely recognition plays an integral role in improving the performance of the system.

Today, there has been a growing interest in improving interaction between humans and computers. Some say that to get the best response, computers need to interact with the user just like humans interact with each other. Emotional responses can be seen through gestures and speech. Even though speech is the main method, gestures can portray a lot of vital information about a current state or emotion [3, 4]. Psychological theory states that human emotions can be classified into six archetypal types: surprise, fear, disgust, anger, happiness, and sadness. To display these archetypal types, the muscles in a person's face can change, the tone of voice can be altered, and the energy of a person's voice can increase or decrease. The shape and formation of the lips can also greatly contribute to understanding of speech in a noisy environment. All of these play a role in the process of communicating different feelings. Even if these signals are subtly displayed, humans can recognize these signals by processing information from the ears and eyes [5, 6]. It has been said that individuals focus more attention on projecting their own facial expressions and perceiving others' facial expressions than they do other nonverbal channels and often more than verbal communication as well [7].

II. RELATED WORK

When approaching emotion recognition, a variety of techniques can be employed. Approaches include analysing body language, voice patterns, and facial features. In analysing the face, Ekman and Friesen, in 1978, presented a now well-known facial expression model called the Facial Action Coding System (FACS). FACS breaks the face into 46 different Action Units (AU's). Each AU is defined as the contraction or relaxation of one or a group of muscles. Some, but not all, of these AU's contribute to the identification of six basic emotions identified by Ekman

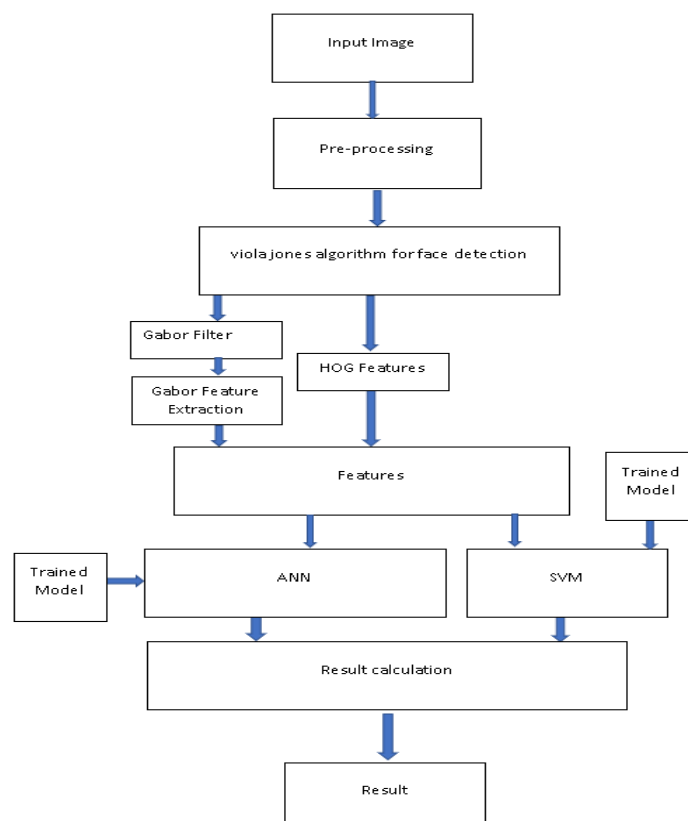
and Friesen: happiness, sadness, fear, anger, disgust, and surprise. The AU's pertaining to the eyes, eyebrows, and mouth are most associated with the expression of emotion [8-11].

Gajarla and al. proposes a system based on VGG Image Net [12] to perform sentiment analysis on images obtained from Twitter and Flickr APIs in accordance to the FACS system. The underlying architecture is a Support Vector Machine (SVM) classifier which yields results only for the happy emotion. The network uses a deep layer approach (50 layers) reaching an accuracy of 73%. Revina and al. proposes an approach which is based on the Gaussian Filter [14] to resize input images, Viola/Jones algorithm [15] to detect faces and Scale Invariant Feature Transform (SIFT) for face alignment. A combination of the Jaffe database [11] and Cohn Kanade database [17] have been used to test the classifier built using a Support Vector Machine alongside a Convolution Neural Network (CNN). Their results demonstrate high accuracy and identification of 7 expressions. The work of Tarnowski and al. is based on the Kinect Device [18] which measures the displacement and distance of the 121 datapoints identified by the FACS system. The classifier is based on a Mutli-Layer Perceptron (MLP) with two hidden layers trained and successfully tested using the KDEF [19] database.

III. METHODOLOGY

The research work is based on an experimental approach, which consists of incremental improvements on the developed prototype of the system. A traditional Facial Expression Recognition (FER) system constitutes of three key steps: facial detection, feature extraction, and classification of the facial expression. Face detection refers to

the pre-processing stage where face regions are located. Facial feature extraction aims to find the most fitting representation of facial images for recognition. As for the last step, facial expression classification, a probabilistic classifier is used to identify different expressions relative to the extracted facial features. The Block diagram of proposed system as shown in fig 3.1.





2.1 Input Image

We use full size and face only image to obtain 10 emotion databases. All the category of emotion set contain male and female images.

2.2 Pre-Processing

Pre-processing of input image enhance the image and reduce the noise. We use base image enhancement like contrast and brightness enhancement also apply basic low pass filter.

2.3 Face detection

We use the Viola-Jones algorithm Detect objects. Viola-Jones is quite powerful and its application has proven to be exceptionally notable in real-time face detection.

There are 2 stages in the Viola-Jones Algorithm, Training and Detection. Viola-Jones was designed for frontal faces, so it is able to detect frontal the best rather than faces looking sideways, upwards or downwards. Before detecting a face, the image is converted into grayscale, since it is easier to work with and there's lesser data to process. The Viola-Jones algorithm first detects the face on the grayscale image and then finds the location on the coloured image.

Viola-Jones outlines a box (as you can see on the right) and searches for a face within the box. It is essentially searching for these haar-like features, which will be explained later. The box moves a step to the right after going through every tile in the picture. In this case, I've used a large box size and taken large steps for demonstration, but in general, you can change the box size and step size according to your needs.

With smaller steps, a number of boxes detect face-like features (Haar-like features) and the data of all of those boxes put together, helps the algorithm determine where the face is.

Haar-like features are named after Alfred Haar, a Hungarian mathematician in the 19th century who developed the concept of Haar wavelets (kind of like the ancestor of haar-like features). The features below show a box with a light side and a dark side, which is how the machine determines what the feature is. Sometimes one side will be lighter than the other, as in an edge of an eyebrow. Sometimes the middle portion may be shinier than the surrounding boxes, which can be interpreted as a nose.

The integral image plays its part in allowing us to perform these intensive calculations quickly so we can understand whether a feature of a number of features fit the criteria. We're feeding it information, and subsequently training it to learn from the information to predict. So ultimately, the algorithm is setting a minimum threshold to determine whether something can be classified as a feature or not. The algorithm learns from the images we supply it and is able to determine the false positives and true negatives in the data, allowing it to be more accurate. We would get a highly accurate model once we have looked at all possible positions and combinations of those features. Training can be super extensive because of all the different possibilities and combinations you would have to check for every single frame or image.

Cascading is another sort of "hack" to boost the speed and accuracy of our model. So we start by taking a sub window and within this sub window, we take our most important or best feature and see if it is present in the image within the sub window. If it is not in the sub window, then we don't even look at the sub window, we just discard it. Then if it is present, we look at the second feature in the sub window. If it isn't present, then we reject the sub window. We go on for the number of features have, and reject the sub windows without the feature. Evaluations may take split seconds but since you have to do it for each feature, it could take a lot of time. Cascading speeds up this process a lot, and the machine is able to deliver results much faster.

2.4 Gabor Features

A gabor object represents a linear Gabor filter that is sensitive to textures with a specified wavelength and orientation. A filter bank is a set of filters that represent combinations of multiple wavelengths, orientations, and other optional parameters. For example, if you specify two wavelengths and three orientations, then the Gabor filter bank consists of six filters for each combination of wavelength and orientation.

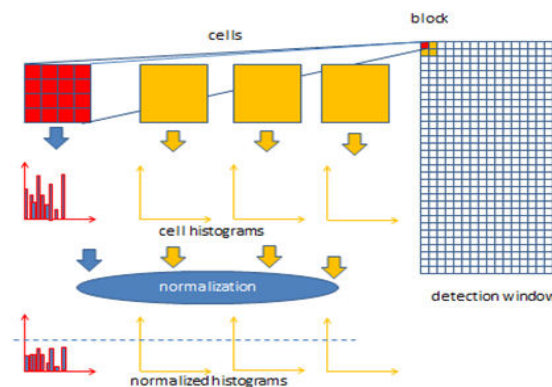
2.5 HOG

Histogram of oriented gradients (HOG) is a feature descriptor used to detect objects in computer vision and image processing. The HOG descriptor technique counts occurrences of gradient orientation in localized portions of an image - detection window, or region of interest (ROI).

Implementation of the HOG descriptor algorithm is as follows:

1. Divide the image into small connected regions called cells, and for each cell compute a histogram of gradient directions or edge orientations for the pixels within the cell.
2. Discretize each cell into angular bins according to the gradient orientation.
3. Each cell's pixel contributes weighted gradient to its corresponding angular bin.
4. Groups of adjacent cells are considered as spatial regions called blocks. The grouping of cells into a block is the basis for grouping and normalization of histograms.
5. Normalized group of histograms represents the block histogram. The set of these block histograms represents the descriptor.

The following fig 3.2 demonstrates the algorithm implementation scheme:



2.6 Artificial neural networks (ANN)

An ANN is based on a collection of connected units or nodes called artificial neurons, which loosely model the neurons in a biological brain. Each connection, like the synapses in a biological brain, can transmit a signal to other neurons. An artificial neuron receives signals then processes them and can signal neurons connected to it. The "signal" at a connection is a real number, and the output of each neuron is computed by some non-linear function of the sum of its inputs. The connections are called edges. Neurons and edges typically have a weight that adjusts as learning proceeds. The weight increases or decreases the strength of the signal at a connection. Neurons may have a threshold such that a signal is sent only if the aggregate signal crosses that threshold. Typically, neurons are aggregated into layers. Different layers may perform different transformations on their inputs. Signals travel from the first layer (the input layer), to the last layer (the output layer), possibly after traversing the layers multiple times.

ANNs are composed of artificial neurons which are conceptually derived from biological neurons. Each artificial neuron has inputs and produces a single output which can be sent to multiple other neurons.[47] The inputs can be the feature values of a sample of external data, such as images or documents, or they can be the outputs of other neurons. The outputs of the final output neurons of the neural net accomplish the task, such as recognizing an object in an image.

2.6 Support Vector Machine(SVM)

A support vector machine (SVM) is a supervised learning algorithm used for many classification and regression problems, including signal processing, medical applications, natural language processing, and speech and image recognition.

The objective of the SVM algorithm is to find a hyperplane that, to the best degree possible, separates data points of one class from those of another class. "Best" is defined as the hyperplane with the largest margin between the two classes, represented by plus versus minus in the fig 3.4 below. Margin means the maximal width of the slab parallel to the hyperplane that has no interior data points. Only for linearly separable problems can the algorithm find such a

hyperplane, for most practical problems the algorithm maximizes the soft margin allowing a small number of misclassifications.

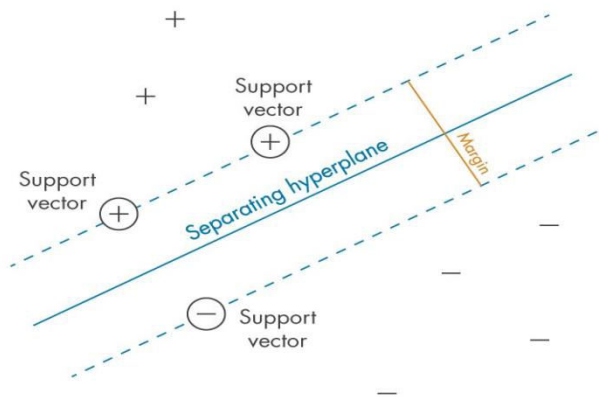


Fig 3.4. Defining the “margin” between classes – the criterion that SVMs seek to optimize

Defining the “margin” between classes – the criterion that SVMs seek to optimize. Support vectors refer to a subset of the training observations that identify the location of the separating hyperplane. The standard SVM algorithm is formulated for binary classification problems, and multiclass problems are typically reduced to a series of binary ones.

IV. EXPERIMENTAL RESULTS

We use Matlab 2020a software for coding

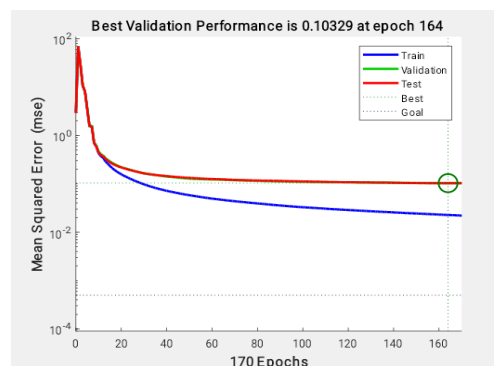


Fig 4.1 Training performance

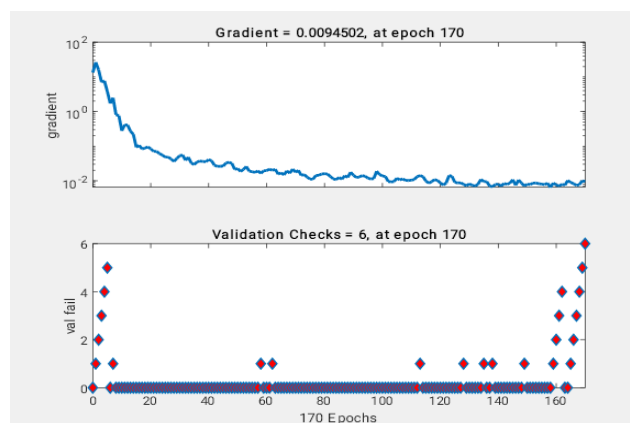


Fig 4.2 Training state graph

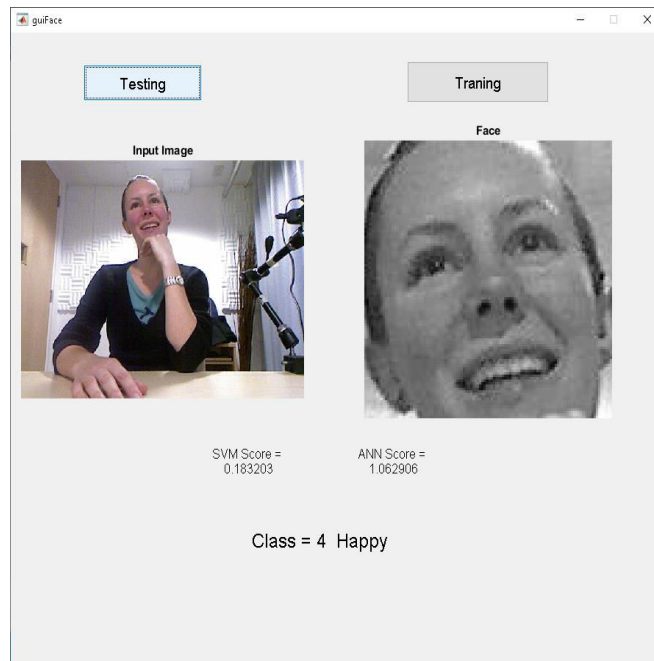


Fig 4.3 Test Result

		Confusion Matrix									
Output Class	1	2	3	4	5	6	7	8	9	10	
	274 7.2%	2 0.1%	7 0.2%	8 0.2%	22 0.6%	0 0.0%	7 0.2%	4 0.1%	2 0.1%	3 0.1%	83.3%
	0 0.0%	37 1.0%	4 0.1%	10 0.3%	13 0.3%	1 0.0%	13 0.3%	1 0.0%	3 0.1%	0 0.0%	45.1%
	3 0.1%	0 0.0%	82 2.2%	11 0.3%	15 0.4%	3 0.1%	14 0.4%	2 0.1%	0 0.0%	1 0.0%	62.6%
	10 0.3%	3 0.1%	6 0.2%	747 19.7%	36 1.0%	3 0.1%	13 0.3%	9 0.2%	2 0.1%	2 0.1%	89.9%
	12 0.3%	12 0.3%	18 0.5%	53 1.4%	832 22.0%	4 0.1%	37 1.0%	15 0.4%	6 0.2%	10 0.3%	83.3%
	2 0.1%	6 0.2%	8 0.2%	19 0.5%	37 1.0%	49 1.3%	12 0.3%	2 0.1%	2 0.1%	0 0.0%	35.8%
	6 0.2%	3 0.1%	6 0.2%	11 0.3%	39 1.0%	1 0.0%	237 6.3%	4 0.1%	4 0.1%	1 0.0%	76.0%
	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	231 6.1%	0 0.0%	0 0.0%	100%
	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	7 0.2%	82 2.2%	0 0.0%	92.1%
	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	644 17.0%	100%
		1	2	3	4	5	6	7	8	9	10
		89.3%	58.7%	62.6%	87.0%	83.7%	80.3%	71.2%	84.0%	81.2%	97.4%
		10.7%	41.3%	37.4%	13.0%	16.3%	19.7%	28.8%	16.0%	18.8%	2.6%
		1	2	3	4	5	6	7	8	9	10
		Target Class									

Fig 4.4 Confusion Matrix

V. CONCLUSION

In this paper, we can classify 10 emotions at efficiency of 84.9%. we propose a combined method classifiers of SVM and ANN to improve the efficiency. Due to the limitations of data base, we cannot able to training more images, in future increasing the training images to increasing the efficiency

REFERENCES

- 1] R. Adolphs, "Recognizing emotion from facial expressions: psychological and neurological mechanisms," Behavioral and Cognitive Neuroscience Reviews, vol. 1, pp. 21-62, March 2002.
- [2] C. Gagliardi, E. Frigerio, D. Buro, I. Cazzaniga, D. Pret and R. Borgatti, "Facial expression recognition in Williams syndrome," Neuropsychologia 41, 2003, pp. 733-738



- 3] A. Azcarate, F. Hageloh, K. Sande, and R. Valenti, "Automatic facial emotion recognition," Universiteit van Amsterdam, pp. 1-6, June 2005.
- [4] A. Metallinou, C. Busso, S. Lee, and S. Narayanan, "Visual emotion recognition using compact facial representations and viseme information," ICASSP 2010, pp. 2474-2477, 2010.
- [5] C. Busso, Z. Deng, S. Yildirim, M. Bulut, C. Lee, A. Kazemzadeh, S. Lee, U. Neumann, and S. Narayanan, "Analysis of emotion recognition using facial expressions, speech, and multimodal information," Proceedings of the 6th international conference on Multimodal interfaces, pp. 205-211, 2004.
- [6] S. Ioannou, A. Raouzaoui, V. Tzouvaras, T. Mailis, K. Karpouzis, and S. Kollias, "Emotion recognition through facial expression analysis based on a neurofuzzy network," Neural Networks, pp.423- 435, 2005
- 7] H. Elfenbein, A. Marsh, and N. Ambady, "Emotional intelligence and the recognition of emotion from facial expressions," The wisdom of Feelings: Processes Underlying Emotional Intelligence, pp. 1-19, 1998.
- [8] P. Ekman, "Emotions in the human face," New York, NY, Cambridge University Press, 1982.
- [9] P. Ekman and R. J. Davidson, "Nature of emotion," New York, NY, Oxford University Press, 1994.
- [10] P. Ekman and W. V. Friesen, "The facial action coding system," Palo Alto, CA, Consulting Psychologists Press, 1978.
- [11] P. Ekman and K. R. Scherer, "Approaches to emotion," Hillsdale, NJ, Erlbaum Associates, 1984. Gajarla, V and Gupta, A. 2016. Emotion Detection and Sentiment Analysis of Images. Georgia Institute of Technology
- [12] Gurney, K. 1997. An Introduction to Neural Networks. Taylor & Francis, Inc., Bristol, PA, USA.
- [13] Revina, M and Emmanuel, W. 2018. A Survey on Human Face Expression Recognition Techniques. Dept of Computer Science, Christian College, Sunadaranar University, Tirunelveli, Tamil Nadu, India
- [14] Biswas, S., 2015. An Efficient Expression Recognition Method using Contourlet Transform. Int. Conf. Percept. Mach. Intell. pp. 167-174.
- [15] Lyons, Michael, Kamachi, Miyuki, & Gyoba, Jiro. (1998). The Japanese Female Facial Expression (JAFPE) Database [Dataset].
- [16] Kanade, T., Cohn, J. F., & Tian, Y. (2000). Comprehensive database for facial expression analysis. Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition (FG'00), Grenoble, France, 46-53.
- [17] Tarnowski, P., Kołodziej, M., Majkowski, A. and Rak, J. (2017). Emotion Recognition using facial expressions. ICCS Zurich.
- [18] Lundqvist, D., Flykt, A., & Öhman, A. (1998). The Karolinska Directed Emotional Faces – KDEF, CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet, ISBN 91-630-7164-9.
- [19] Cootes, T., Edwards, G. and Taylor, C., "Active appearance models", IEEE Trans. Pattern Anal. Mach. Intell., vol. 23, no. 6, Jun. 2001, pp. 681- 685.



INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.118



ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 **9940 572 462**  **6381 907 438**  **ijirset@gmail.com**



www.ijirset.com

Scan to save the contact details