iJRASET
International Journal For Research in Applied Science and Engineering Technology



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 7 **Issue:** VI **Month of publication:** June 2019

**DOI:** http://doi.org/10.22214/ijraset.2019.6332

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177*
*Volume 7 Issue VI, June 2019- Available at www.ijraset.com*

# Using CNN and HOG Classifier to Improve Facial Expression Recognition

Joshua Gisemba Okemwa[1], Victor Mageto[2]
[1]*Research Scholar, Department Of Computer Science, Bhupal Nobles' University, Udaipur, Rajasthan*
[2]*Research Scholar, Department Of Computer Science, Suresh Gyan Vihar University, Jaipur, Rajasthan*

***Abstract:*** ***Facial expression recognition (FER) is growing on a large scope due to diversification of its field of application. FER is now applicable in crime prevention, smart city development, as well as other economic sectors like: transportation, advertising and health. There are a number of benefits accompanied through proper and correct analysis of emotions: Security is enhanced, proper event prediction, inter person communication channel, easy extraction of details and so on. There are various facial emotions identified from various previous studies, which make up the basis for effective and affective communication among people of different culture, race, ethnicity and gender. Many feature extraction methods and classification techniques have previously been developed to give better accuracy and performance in face recognition. A convolution neural network CNN is an unsupervised deep learning algorithms with ability to learn image characteristics and make differentiation of one aspect to another. It is a trending technique in this field due to its positive results, and fast computation. However, the still there are issues with accuracy and complexity challenges in face recognition. In this paper we perform experiments on FER to solve problems associated with orientation and different light conditions. We applied HOG classifier for feature extraction and CNN to detect and classify the expressions. Overall we achieved high accuracy and optimization results of 77.2%. This method achieved higher results than previous work done using SVM algorithm and HOG classifier with accuracy of 55%.***
***Keywords: CNN, FER, HOG, SVM***

## I. INTRODUCTION

Technology is rapidly growing and providing solutions to problems we encounter in our daily life. We depend on technology as part of our daily interactions with machines. One of them is Facial expression recognition (FER). Face plays a vital role in most social interactions; expressions are also very important in exposing a person's mental state as well as their views. [14] A lot can be analyzed from the face during communication, views of people and their intentions are unveiled. [5] There are seven basic facial emotions they include: happy, sad, anger, disgust, neutral, fear and surprise. These emotions make up the basis for effective communication among people of different culture, race, ethnicity and gender. [1] Two thirds of human communication are comprised of nonverbal communication; facial expression is the most common channel of communication which has currently gained a lot of attention within the computer vision environment. [6] The facial recognition system is vital because of its capabilities of understanding human coding skills and interpreting psychological state of individuals. [7] The gestures displayed form part of clues for enabling efficiency in correlation to a society; they complement what is missed out during verbal communication therefore conveying the intended message. Biometric machines are commonly used to detect human emotion response displayed on the face. It is described as a sentimental drive technique that enables automatic detection of universally recognized emotional expressions. The product of the recognition technique is unfiltered, unbiased emotional response that encourages more accuracy in extraction and interpretation of information. Algorithms are used in facial recognition process aimed at ascertaining the emotional state. Facial analysis using computer powered cameras include: face detection, locating face in the image scene, facial land mark detection by extracting information from features detected on the face, facial expression & emotional classification through depth analysis of appearance of facial features, changes and classification into interpretive categories. [8]

With consideration of 95% of communication being nonverbal, facial recognition and image processing domain have become one of the most active research areas. Research shows that deep learning has a high accuracy level in finding feature landmarks in data, different deep learning techniques that focus on appearance and geometric features when applied increase accuracy. In some cases, accuracy is achieved further by application of hybrid, ensemble and optimization techniques. [12] In our study we demonstrated how convolution Neural Network CNN one of the deep learning feature classification techniques works. We also illustrate how application of Histogram of orient Gradients, sliding window and face landmarks impacted our results positively in facial expression process.

©IJRASET: All Rights are Reserved

*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177*
*Volume 7 Issue VI, June 2019- Available at www.ijraset.com*

## II. PROSPECTIVE APPLICATION

Facial recognition is an open research area in machine learning domain. The facial expression systems are important in non-verbal gesture interpretation used during communication. Emotion recognition is attained from facial expression. This is done automatically in most cases using geometric and appearance approaches. Geometric techniques take into account the shape and location of facial components; ears, nose, cheeks, eyes and mouth. The geometric relationships between these components are used to form feature vectors which include position, distances and angles. Appearance approaches models the face images by applying an image filter or filter barks on the whole face or some specific regions of the face to extract change in facial appearance. Some of these model techniques include; PCA (principal component analysis), ICA (Independent component analysis), Gabor wavelet, LBP (local binary pattern) applied for the purpose of texture analysis and its variants, efficient to sensitive non-monotonic illumination and robust to poor performance in random noise.

Human beings are consistent in recognizing emotions but have variance from one person to another. Artificial intelligence has changed the entire operation with numerous benefits accompanied including automation of the process. This technology reduces the effort applied by man by making machines slaves able to carry out the activities that were performed by men previously. Machines speed up the work and give accurate results. Introduction of machines guided applications brings the idea of having an error free world.

Different methods have been developed by researchers working closely in improving accuracy in face recognition. Deep learning is an area in computer vision and machine learning which is promising in terms of improving accuracy. It adopts neural network architecture consisting of a number of multiple layer perceptions from a number of multiple hidden ones enabling easy finding of high level features in data that provides a hierarchical representation. [10]

Accuracy of development of deep learning approaches give a result that is more accurate when compared with most of the traditional methods. With expectation of face recognition being having a wider scale of application in different sectors it will prompt researcher to investigate and examine optimization of face recognition methods. [11, 9]

Research shows that deep learning can solve complex problems by using a number of multilayer architectures thus making the problem solving to be shorter and accuracy of result is boosted. Sub sampling techniques found in deep learning are applied which increases efficiency in giving solutions to complex problems. [12] It works well in prediction of data that is not well known and widely implemented in areas like face recognition and fingerprint recognition. [9]

One deep learning approach used in face recognition is convolution Neural Network CNN. A key feature of this method is the ability to note the characteristic of an image at entry point thus classifying it. CNN is a variant of distortion and geometric transformation. [13]

## III. CONVOLUTION NEURAL NETWORK

Convolution Neural Network CNN is a deep learning algorithm with ability to take an image and learn weight and biases given to it, and make differentiation of one aspect to another. It has less preprocessing requirements when compared to other methods even if they are primitive, with proper training. [2] Convolution neural networks are unsupervised learning algorithms, whereby each neuron receives some input, performs a dot product and optionally follows it with a non-linearity. The whole network expresses different score functions from the raw image pixels on the end to another. They compose loss function on the last fully connected layer and all the kernel trick developed is applied here. The conversion of raw pixels to multi-level perceptron image is shown below.

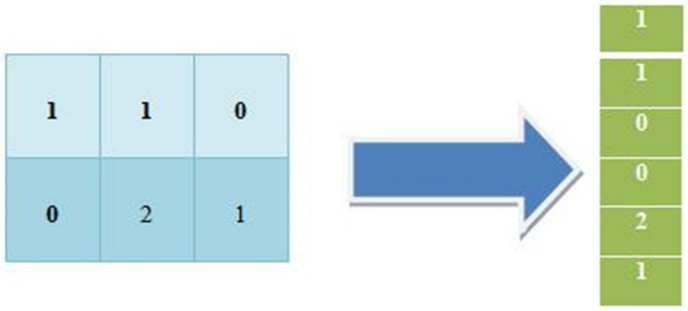



Fig1. conversion of raw pixel to multi-level perceptron

©IJRASET: All Rights are Reserved

For instance, an image is a matrix of pixel values. The image above with a matrix of 3 x 2 is flattened into 6x1 vectors then fed to a top a multi-level perceptron, this method might show precision but have zero accuracy when dealing with complex images having interdependence all along.

CNN enables effective capturing of spatial and temporal dependencies in an image. To create dependences between image pixels, filters are applied to reduce overfitting and outliers to allow reduction in parameters involved in reusability. Most importantly we note that through training a network can be able to understand sophistication of an image in a better manner.

### *A. Layer Architecture*

CNN has four layer architectures which are usually involved in checking different degrees of shift scales and distortions.

Convolution layer: it's the underlying layer for the process of applying a function in a repeated manner to the output of another function, with intention of extracting features of the image that has been added.

Pooling layer: The layer ensures reduction of spatial size of convolved feature thus enhancing reduction of computational power. Most importantly it reduces dominant figure in order to maintain effective model training. There are two types of pooling namely max which returns maximum value of image that is covered while on the other hand average pooling returns the average value of portion of image that is covered. It always noted that a max pool has better performance compared to average pooling.

Full Connection Layer: the role of this layer is to perform transformation on data dimensions so that data can be classified. The number of such layers is always flexible and can be adjusted in order to capture a number of details along with computation cost.

Output layer: Is a product of the last layer a product of convolution [13]

### *B. Impact of CNN Application*

Application of this method is accompanied with positive and negative outcome they include: its ability to implement various image resolutions, ability to solve problems that have a high complexity with computation involving very many parameters, detailed computation that reduces the level of errors and classification of both known and unknown data.

CNN process can appear quite long with complex computations involved, inability to do face description in a certain position and not suitable simple. Note the accuracy of CNN is quite high

## IV. IMPLEMENTATION OF CNN

In our study we illustrate the use of convolution neural network along with hog classifier to improve facial expression recognition.

1) *Dataset Selection:* The FER2013 dataset was selected for analysis which contained 35887 images. This open source dataset was created by Pierre Luc Carrier and Aaron Courville [3] for the Kaggle competition in 2013.The dataset is made up of 48x48 sized gray scale images with 7 different emotions. The expressions present include anger, sad, happy, surprise, neutral, disgust and fear.




Fig 2 FER 2013 dataset

The corresponding size of the images in each expression from the dataset were as shown in the table below

| Index of Emotion | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| Emotion | Angry | Disgust | Fear | Happy | Sad | Surprise | Neutral |
| Number of Images | 4593 | 547 | 5121 | 8989 | 6077 | 4002 | 6198 |

Fig 3. Size of images for each emotion

©IJRASET: All Rights are Reserved

2) *Training And Testing The Data:* These images were further classified into two groups; training phase and testing phase. It was done in order to verify the performance and accuracy of our results. The sample of images that were used for testing and training have been listed in the table below.

| Emotion | Angry | Disgust | Fear | Happy | Sad | Surprise | Neutral |
|---|---|---|---|---|---|---|---|
| Training | 3177 | 349 | 3253 | 5801 | 959 | 2577 | 3939 |
| Testing | 818 | 87 | 844 | 1414 | 1026 | 3871 | 594 |

Fig 4.0 Testing and training Input data

## V. RESULT AND DISSCUSION

Our research involved comparison of performance using CNN classifier with raw pixels of training images and addition of optimization parameters; (sliding window, HOG Face landmarks). The outcome of the experiment showed that when additional optimization parameters are added to the existing system the accuracy results are improved. Due to presence of overfitting and outliers among the network the accuracy levels were low in the first network.

In order to elevate the accuracy and remove the overfitting, a regularization technique called dropout [4] was applied to reduce the interdependency among the training data. This technique greatly improved the accuracy by +4.6%.

Face detection was done using Opencv3, HOG and face landmarks features were extracted using python's Dlib dependency into a convolutional neural network architecture.

| Technique | Model A | Model B | Difference |
|---|---|---|---|
| CNN(with raw pixels) | 74.4% | 75.5% | +1.1% |
| CNN + Face Landmarks | 75.5% | 76.4% | +0.9% |
| CNN + Face Landmarks + HOG | 70.7% | 75.3% | +4.6% |
| CNN + Face Landmarks + HOG + Sliding Window | 73.5% | 77.2% | +3.7% |

Fig 4.0 Comparison of different accuracy results

## VI. CONCLUSION

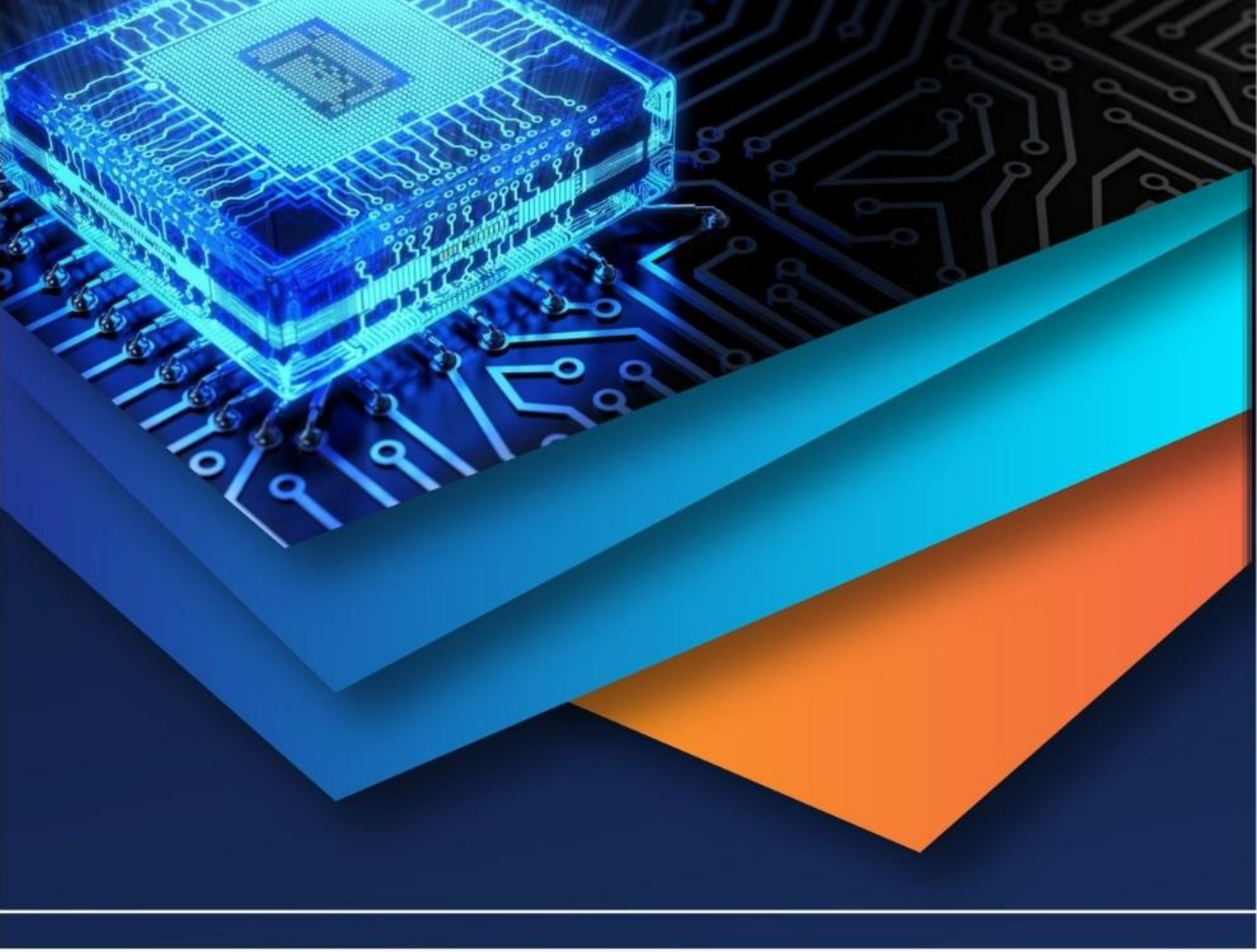
Convolution neural networks are unsupervised deep learning algorithms with ability to learn image characteristics and make differentiation of one aspect to another. Deep learning approaches give better result when compared to other traditional method based on accuracy. Emotional recognition can be achieved through facial expression processing using computer methodologies. Factually machines are more accurate than humans in predicting, the training and simulation. We demonstrated how convolution neural network works in facilitating emotion differentiation and use of hog classifier to achieve high accuracy and optimization results. The result shows that raw images produces low levels of performance. Histogram of orient Gradients, sliding window and face landmarks impacted our results positively. Through different testing phases of each feature it was evident that there is still more room for future research work.

©IJRASET: All Rights are Reserved

## REFERENCES

[1] Ekman, Paul. "An argument for basic emotions." Cognition & emotion 6.3-4 (1992): 169-200.
[2] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems. 2012.
[3] Carrier, Pierre-Luc, et al. "FER-2013 face database." Technical report (2013).
[4] Srivastava, Nitish, et al. "Dropout: a simple way to prevent neural networks from overfitting." The Journal of Machine Learning Research 15.1 (2014): 1929-1958.
[5] Byoung Chul Ko. A Brief Review of Facial Emotion Recognition Based on Visual Information. Sensors 2018, pp 1-20; doi:10.3390/s1802040
[6] Kaulard, K.; Cunningham, D.W.; Bülthoff, H.H.; Wallraven, C. The MPI facial expression database—A validated database of emotional and conversational facial expressions. PLoS ONE 2012, 7, e32321.
[7] Mehrabian, ³Silent Messages-A Wealth of Information About Nonverbal Communication (Body Language)´ 1981
[8] Roshan Jameel, Abhishek Singhal, Abhay Bansa. A Comprehensive Study On Facial Expressions Recognition Techniques. 2016 6th International Conference. Pp 478-483
[9] M. M. Kasar, D. Bhattacharyya, and T. Kim, "Face Recognition Using Neural Network : A Review," vol. 10, no. 3, pp. 81–100, 2016.
[10] M. Li, C. Yu, F. Nian, and X. Li, "A Face Detection Algorithm Based on Deep Learning," vol. 8, no. 11, pp. 285–296, 2015.
[11] A. Rikhtegar, M. Pooyan, and M. T. Manzuri-shalmani, "Genetic algorithm-optimised structure of convolutional neural network for face recognition applications," pp. 559–566, 2016
[12] M. Bianchini and F. Scarselli, "On the complexity of neural network classifiers: A comparison between shallow and deep architectures," IEEE Trans. Neural Networks Learn. Syst., vol. 25, no. 8, pp. 1553–1565, 2014.
[13] H. Khalajzadeh, M. Mansouri, and M. Teshnehlab, "Face Recognition Using Convolutional Neural Network and Simple Logistic Classifier," Soft Comput. Ind. Appl., pp. 197–207, 2014
[14] Dr. Sanjay Kumar, Ayushi Gupta. Facial Expression Recognition: A Review. Special Conference Issue: National Conference on Cloud Computing & Big Data. Pp 159- 162

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)