iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 Issue: VI Month of publication: June 2022

DOI: https://doi.org/10.22214/ijraset.2022.45141

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: 7.538
Volume 10 Issue VI June 2022- Available at www.ijraset.com

Facial Expression Recognition System Using Haar Cascades Classifier

Varshita Usem[1], Neha Dammannagari[2]
[1]Graduate, Dept. of Computer Science and Engineering, Chaitanya Bharathi Institute of Technology, Telangana, India
[2]Graduate, Dept. of Information Technology, Chaitanya Bharathi Institute of Technology, Telangana, India

Abstract: Facial expression conveys non-verbal cues, which play a crucial role in social relations. Facial Expression Recognition is a significant yet challenging task, as we can use it to identify the emotions and the mental state of an individual. In this system, using image processing and machine learning, we compare the captured image with the trained dataset and then display the emotional state of the image. To design a robust facial feature recognition system, Local Binary Pattern(LBP) is used. We then assess the performance of the suggested system by using a database that is trained, with the help of Neural Networks. The results show the competitive classification accuracy of various emotions.

Keywords: Facial expression recognition, Local binary Pattern(LBP), Convolutional Neural Networks(CNN), Feature extraction, Haar Cascade Classifier.

I. INTRODUCTION

A person's facial expressions serve as a medium of communication in verbal and nonverbal communication because they are the outward expression of their affective response, line of thinking, intent, and persona. Generally, gestures or facial expressions or even involuntary languages are used by humans to express their emotional state, instead of verbal communication. It has been studied for a long time and has made significant progress in recent decades. Despite the progress, identifying and understanding facial expressions accurately is still challenging since the complexity is very high and the variety of facial expressions are too many. It has the ability to become a very useful, nonverbal way for people to communicate with one another in the near future. What matters is how well the system detects or extracts facial expressions from images. It is gaining popularity because it has the potential to be widely used in a variety of fields such as lie detection, medical assessment, and human-computer interface.

A. Facial Expression Recognition

Facial expression recognition is the process of classifying the expressions on face images into various categories, such as anger, disgust, sorrow, happiness, surprise and so on. Human relations rely heavily on facial expressions. Recognition of these expressions can prove to be useful to understand the individual's emotions and their way of communication.

B. Haar Cascade Classifier

Viola-Jones Face Detection Technique, popularly known as Haar Cascades, is an object detection algorithm used to identify faces in an image or a real time video. It uses edge or line detection features which Viola and Jones described in their research paper, named Rapid Object Detection using a Boosted Cascade of Simple Features. The algorithm is given a lot of positive images containing faces, and a lot of negative images not consisting of any face for it to train. Haar Cascades is described in detail in section 3.

C. Local Binary Pattern (LBP)

Local Binary Pattern is a technique used for feature extraction. The original LBP operator initially points all the pixels in an image with decimal numbers, which are called LBP codes. These LBP codes of each pixel then encode the structures that are local to or around every pixel. The value of each pixel is then compared with its eight neighboring pixels in a 3x3 neighborhood by removing the center pixel value, which is the pixel itself. As a result, the non-negative values are encoded as 1 and the negative values are given 0. For a particular pixel, we get a binary number by combining all the neighboring binary values in a clockwise direction, starting from any one of its top-left neighbors [1]. Thus, the decimal values for every binary number generated is obtained and used for labeling that particular pixel. The binary numbers obtained are known as LBPs or LBP codes.

D. Convolutional Neural Networks

A convolutional neural network is an architecture in deep learning that usually directly learns from the data, thus reducing or eliminating the manual effort needed for feature extraction. Convolutional neural networks perform better than other neural networks with image, speech, or audio signal inputs. There are 3 important types of layers of ConvNets: Convolutional layer, Pooling layer and Fully-connected layer. These layers eliminate the need for manual feature extraction, produce highly accurate recognition results and can be retrained for new recognition tasks, enabling us to build on pre-existing networks.

II. LITERATURE SURVEY

Facial expressions not only represent the emotions of an individual but also their mental state, the way they think and communicate. So, due to its significance in our lives, detecting them and analyzing them has become quite important and active in terms of research [2]. It can be applied in various fields which require analysis of facial expressions. The most commonly used analysis methods and the comparison between them are presented in [3].

Approaches for extracting facial movement and deformation and classification techniques are addressed not only with regard to issues including face normalization, the dynamics of facial expression, and the intensity of facial expression, but also their robustness to environmental changes [4].

In paper [5], a brand-new technique for detecting facial signals that relies on alterations to the position of the facial points captured in a video showing a close-up of the face was introduced. The existing systems use techniques to detect the entire face of the user. The paper [6] suggests the use of Local Binary Pattern to develop a robust facial expression detection system that can give more accuracy by locating only certain landmarks of the face.

III. METHODOLOGY

The proposed system offers a solution which minimizes the resources required for identification of emotion using facial features. The trained face detector scans an image by a sub-window at various scales. A cascade classifier made of several stage classifiers tests each sub-window. If the sub window clearly does not consist of a part of the face, one of the initial stages in the cascade will reject it. If it is difficult to identify, a more specific one will further classify it.

The proposed system uses the Haar Cascades or Viola-Jones Technique to identify if the images consist of a face or not. If a face is present, the areas containing eyes, mouth are determined and cropped out from the image. Sobel edge detection method is used to detect filters and edges, followed by feature extraction. We train the feature extraction model using the neural networks and then use it to classify the emotions.

A. Phases in Facial Expression Recognition

After detecting the face and extracting features from the images, we classify them into six classes belonging to six basic expressions. The system includes the training and testing phase followed by image acquisition, face detection, image preprocessing, feature extraction and classification.

1) *Image Acquisition:* Images used for facial expression recognition are static images or image sequences. Images of faces can be captured using web camera.
2) *Face Detection:* Face Detection is carried out on a training dataset using a Haar classifier called the Viola-Jones face detector. It uses the intensity in different parts of the image to get the difference in the average intensities. It consists of black and white connected rectangles. The difference of the sum of pixel values in these regions is the value of the feature.
3) *Image Pre-processing*

The major part of any image pre-processing is noise removal and normalization against variation of the pixel position. It includes:

a) Color Normalization

b) Histogram Normalization

4) *Feature Extraction:* We use the image of the face after preprocessing for extracting the significant features from it. Such features are identified and extracted using the Linear Binary Pattern algorithm, which is described below.
5) *Local Binary Pattern:* As described in section 1.2, LBP codes are obtained for each pixel. Further extension of LBP is to use uniform patterns. We consider the binary string as circular and check if there are any transitions from 0 to 1 or 1 to 0. If any such transitions are present, the LBP is said to be uniform.

E.g. 00000000, 001110000 and 11100001 can be called as uniform patterns.
A histogram of a labeled image, say f1(x, y) is defined as

$$H_i = \sum_{x,y} I\,(fl(x, y) = i),\; i = 0, \ldots, n-1$$

Where n is the number of distinct labels given as the result by the LBP operator and

$$I(A) = 1 \text{ if } A \text{ is true}$$

$$0 \text{ if } A \text{ is false}$$

This histogram consists of information about the distribution of edges, spots and flat areas all over the image. For efficient face representation, features extracted should store the previous spatial information. Hence, the image is divided into m small regions and we define a spatially enhanced histogram as

$$H_i = \sum_{x,y} I\,(fl(x, y) = i)\,I((x, y) \in R_j)$$

6) *Classification:* The dimensionality of data obtained from the feature extraction method is very high so it is reduced during classification. Features take different values for objects belonging to different classes so classification will be done using CNN. A typical architecture of a convolutional neural network contains an input layer, some convolutional layers, some fully connected layers, and an output layer at the end. The CNN is designed with some modifications on LeNet Architecture. The architecture of the Convolution Neural Network used in the project is shown in the following Figure 3.1.1.

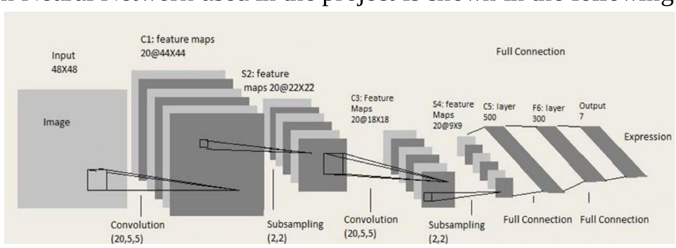


Figure 3.1.1 Architecture of Convolutional Neural Network

Output layer consists of seven distinct classes. Using the Softmax activation function, the probabilities for all the seven classes are calculated individually, and the output is obtained. The class with the highest probability is the resultant predicted class.

B. Architecture of the Proposed System

The below Figure 3.2.1 depicts the system design for the proposed system. It shows the steps that are taken to train the model to detect the face, extract and classify distinct features from the images and test the model.

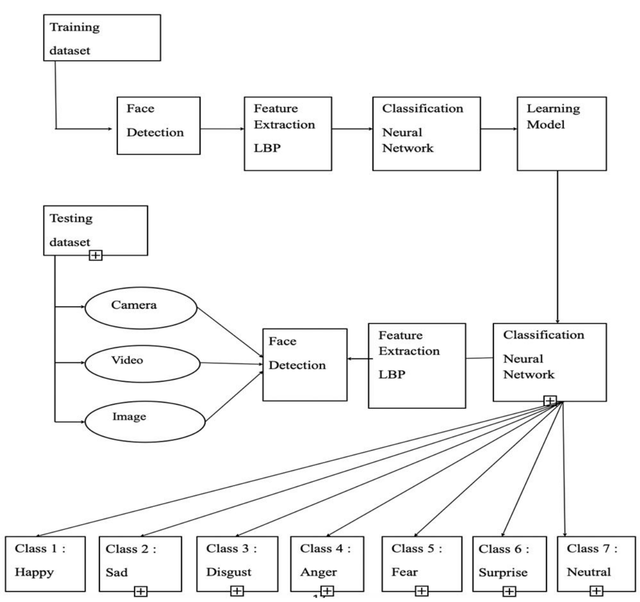


Figure 3.2.1 Architecture of the proposed system

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VI June 2022- Available at www.ijraset.com

IV. PERFORMANCE METRICS

To evaluate and analyze the performance of the proposed system, we used few of the commonly utilized metrics: Precision, Recall and F1 score. A detailed description of each of these metrics is given below.

A. *Precision*

Precision can be a positive value or a negative value based on the class for which it is being calculated for. It evaluates the predictive power of the algorithm. Precision is the ratio of true positives to the total number of positives as predicted by the model.

$$Precision = tp / (tp+fp)$$

where,
tp = number of correctly predicted positives,
fp = number of negatives that are incorrectly predicted as positives.

B. *Recall*

Recall is a function of the model's examples that are classified correctly and its misclassified examples (true positives and false negatives). Recall is the percentage of correctly assigned expressions in relation to the total number of expressions.

$$Recall = tp / (tp+fn)$$

where,
tp = number of correctly predicted positives
fn = number of incorrectly predicted negative classes.

C. *F1 score*

F1 Score is the harmonic mean of Precision and Recall. It takes both false positives and negatives into account. It is more difficult to understand than accuracy, but it is more useful when the class distribution is uneven.

$$F_1 = 2 * \frac{precision * recall}{precision + recall}$$

V. RESULT ANALYSIS

The confusion matrix and the normalized confusion matrix for all seven facial expression classes are shown below in figures 5.1 and 5.2 respectively :

Predicted Actual	Anger	Disgust	Fear	Happy	Sad	Surprise	Neutral
Anger	686	16	8	5	10	6	9
Disgust	35	700	0	0	0	0	0
Fear	43	7	698	4	11	8	12
Happy	85	5	54	980	60	29	75
Sad	37	6	64	25	720	17	15
Surprise	7	4	5	6	8	583	10
Neutral	37	4	31	80	32	13	750

Figure 5.1 Confusion matrix

Predicted Actual	Anger	Disgust	Fear	Happy	Sad	Surprise	Neutral
Anger	0.9270	0.02162	0.0108	0.00675	0.0135	0.00810	0.01216
Disgust	0.04761	0.95238	0	0	0	0	0
Fear	0.05491	0.00893	0.89144	0.00510	0.014048	0.01021	0.01532
Happy	0.06599	0.00388	0.041925	0.76086	0.046583	0.02251	0.05822
Sad	0.04185	0.00678	0.07329	0.02828	0.81447	0.01923	0.01696
Surprise	0.01123	0.00642	0.00802	0.00963	0.01284	0.93579	0.16051
Neutral	0.03907	0.00422	0.03273	0.08447	0.033790	0.01372	0.79197

Figure 5.2 Confusion matrix

The performance metrics for each emotion is calculated and shown in the below table.

	Precision	Recall	F1-score
Anger	0.91	0.93	0.92
Disgust	0.95	0.99	0.97
Fear	0.91	0.89	0.90
Happy	0.75	0.76	0.75
Sad	0.83	0.81	0.82
Surprise	0.90	0.94	0.92
Neutral	0.81	0.79	0.80
Average	0.87	0.87	0.87

The below Figure 5.3 shows the successful detection of the user's face and classification of the expression as happy when the user smiles.

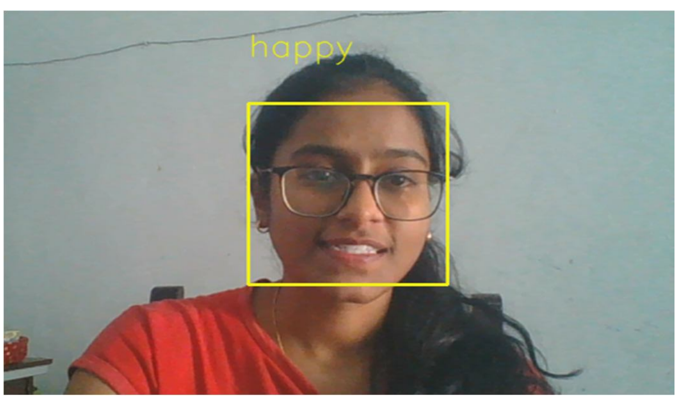


Figure 5.3 System recognised the expression as happy

The below Figure 5.3 shows the successful recognition of the expression as fear when the user is scared.

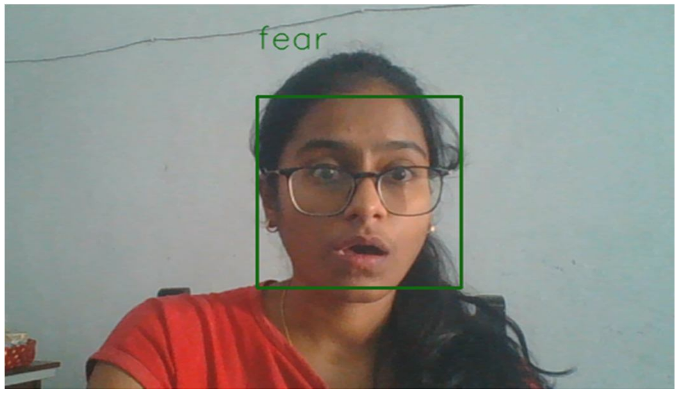


Figure 5.4 System recognised the expression as fear

The below Figure 5.5 shows the successful recognition of the expression as surprise.

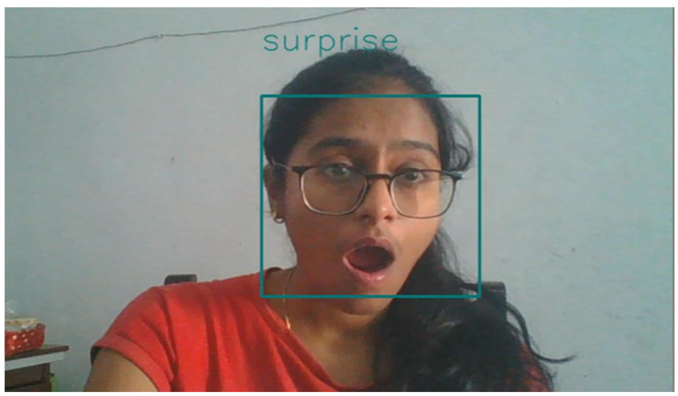


Figure 5.5 System recognised the expression as surprise

The below Figure 5.6 shows the successful recognition of the expression as neutral when there is no visible expression on the user's face.

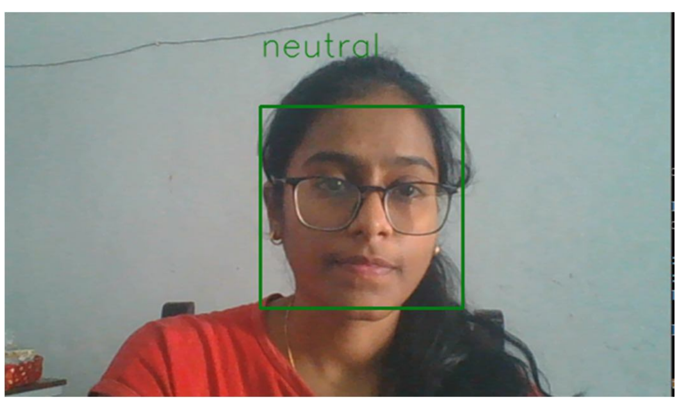


Figure 5.6 System recognised the expression as neutral

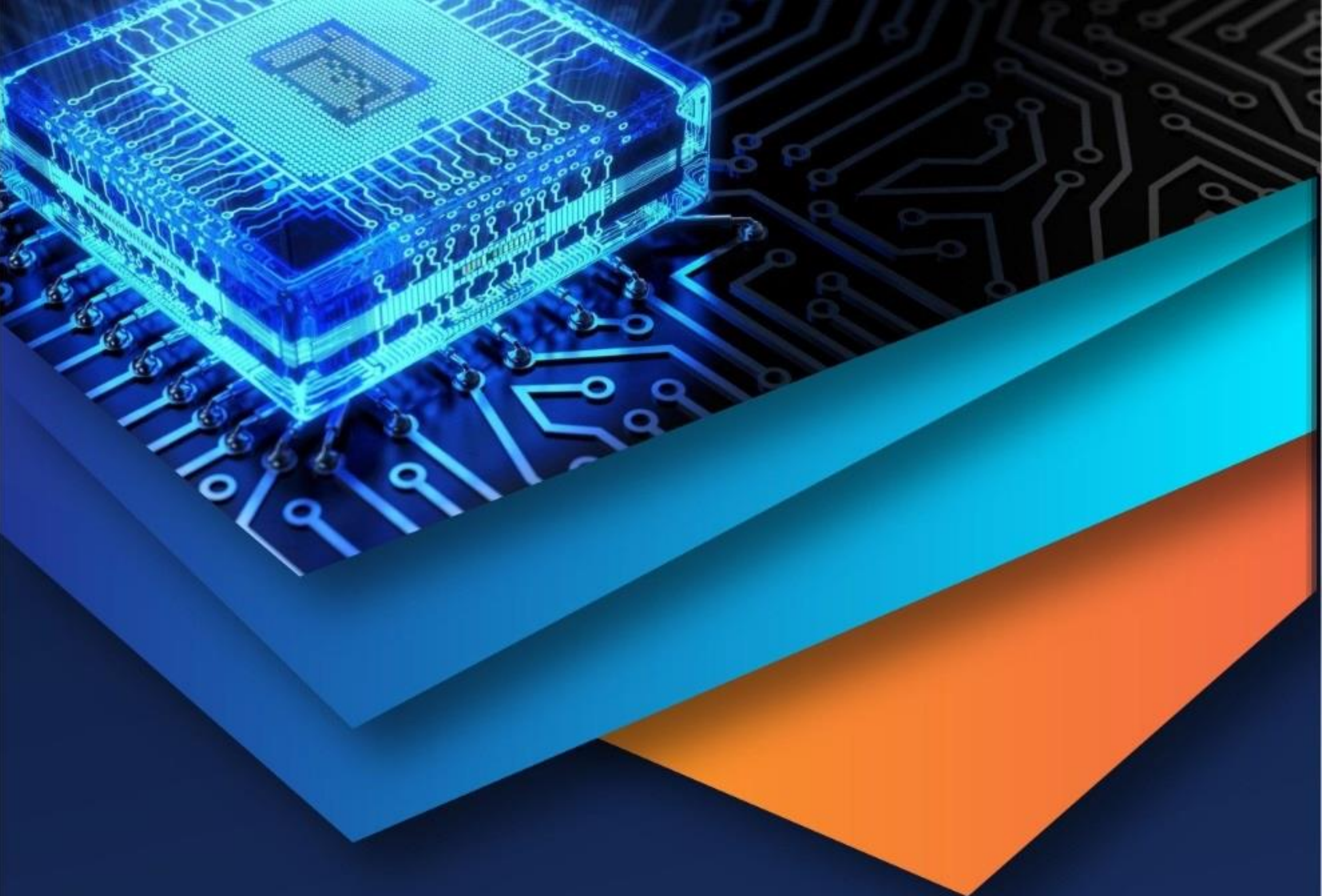
To measure the performance of proposed algorithms and methods and check the results accuracy, we have evaluated the system with its resultant values of precision, recall and F1 score. The same dataset was used for both training and testing by splitting the datasets into training samples and testing samples. The Accuracy obtained from Kaggle dataset was 86.7%, precision was 0.87, recall was 0.87 and F1-score was 0.87.

VI. CONCLUSION

We can say that the model was successfully able to detect and classify various human emotions. There is a difference in the accuracy obtained from the model as the landmark positions of the face are known and fixed, i.e. the mouth and both the eyes. So it can be concluded that this method can be used instead of a traditional face detection, which requires locating more landmarks. A good combination of efficiency and accuracy can be achieved using Haar Cascades for detection and identifying various types of emotions through neural networks.

REFERENCES

[1] Dornaika, Fadi & Bosaghzadeh, Alireza & Salmane, Houssam & Ruichek, Yassine. (2014). Graph-based semi-supervised learning with Local Binary Patterns for holistic object categorization. Expert Systems with Applications. 41. 7744–7753. 10.1016/j.eswa.2014.06.025.
[2] Y. Tian, L. Brown, A. Hampapur, S. Pankanti, A. Senior, and R. Bolle, “Real world real-time automatic recognition of facial expression,” in IEEE PETS, Australia, March 2003.
[3] B. Fasel and J. Luettin, “Automatic facial expression analysis: a survey,” Pattern Recognition, vol. 36, pp. 259–275, 2003.
[4] Fasel, Beat & Luettin, Juergen. (1999). Automatic Facial Expression Analysis: A Survey.
[5] Srivastava, Saumil. (2012). Real Time Facial Expression Recognition Using a Novel Method. The International Journal of Multimedia & Its Applications. 4. 10.5121/ijma.2012.4204.
[6] Shan, C., Gong, S., & McOwan, P. W. (2005, September). Robust facial expression recognition using local binary patterns. In Image Processing, 2005. ICIP 2005. IEEE International Conference on (Vol. 2, pp. II-370). IEEE.
[7] Chen, J., Chen, Z., Chi, Z., & Fu, H. (2014, August). Facial expression recognition based on facial components detection and hog features. In International Workshops on Electrical and Computer Engineering Subfields (pp. 884-888).
[8] Happy, S. L., George, A., & Routray, A. (2012, December). A real time facial expression classification system using Local Binary Patterns. In Intelligent Human Computer Interaction (IHCI), 2012 4th International Conference on (pp. 1-5). IEEE.
[9] Zhang, S., Zhao, X., & Lei, B. (2012). Facial expression recognition based on local binary patterns and local fisher discriminant analysis. WSEAS Trans. Signal Process, 8(1), 21-31.



10.22214/IJRASET

INDEX COPERNICUS
45.98

IMPACT FACTOR:
7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR:
7.429

ISI

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY